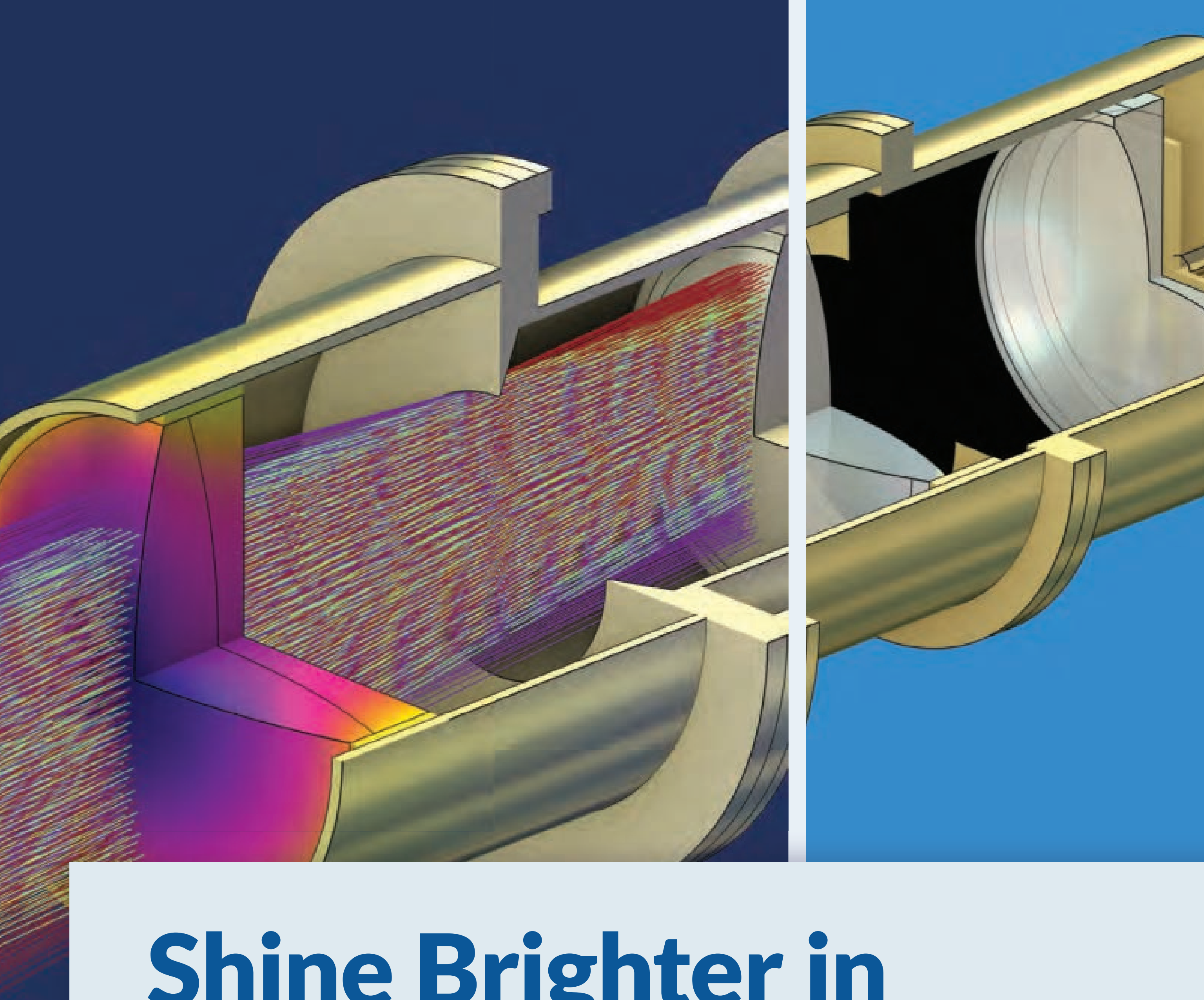


PHYSICS TODAY

June 2024 • volume 77, number 6

A publication of the American Institute of Physics

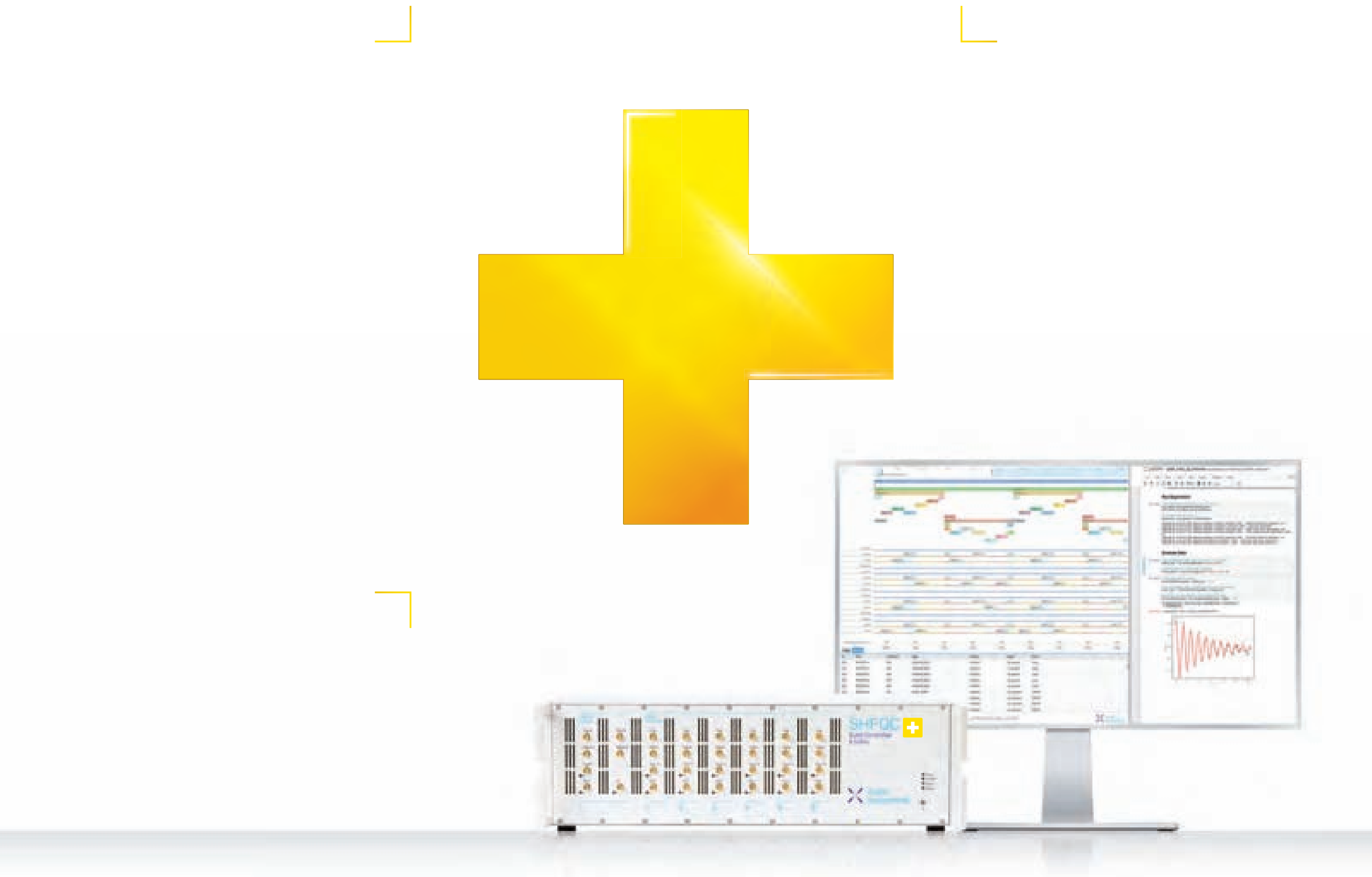
MicroLEDs

Ernest Moniz on nuclear-threat reduction

Firearms forensics gets more quantitative

A nuclear transition in a haystack


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

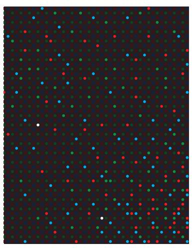

SHFQC

PHYSICS TODAY

June 2024 | volume 77 number 6

FEATURES

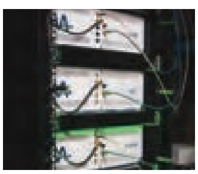
SPECIAL FOCUS ON OPTICAL ENGINEERING



ON THE COVER: MicroLEDs produce light by the same process that their larger LED cousins use: When electrons and holes in a semiconductor material recombine, photons are emitted. But compared with conventional LEDs and other light emitters, microLEDs are much brighter—an advantage that has drawn the interest of makers of augmented- and virtual-reality headsets, whose displays need to compete with the Sun in outdoor environments. On **page 30**, Vikrant Kumar, Keith Behrman, and Ioannis Kymissis discuss the manufacturing challenges of microLEDs and highlight display-industry applications. (Image by J. Keisling.)

Recently on PHYSICS TODAY ONLINE

www.physicstoday.org

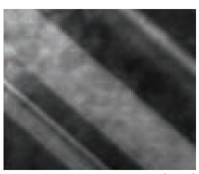


WILL LUNDEN

Seaworthy optical clock

Optical atomic clocks are exquisitely accurate and precise, but their bulkiness and sensitivity limit them largely to lab settings. Now researchers have debuted a more mobile optical clock, based on molecular iodine, that has the performance and robustness to stably keep time on a ship navigating rough seas.

physicstoday.org/Jun2024a



WU ET AL., *NATURE* **626**, 779 (2024)

Flexible ceramic

For many industrial components, ceramics are essential, but they tend to be brittle and intolerant to strain. A newly designed ceramic material is unusually strong and flexible. The key to the advance was fabricating stacks of twisted-layer 2D structures that were then integrated into an interlocked 3D matrix.

physicstoday.org/Jun2024b

CF INDUSTRIES

Ammonia catalysts

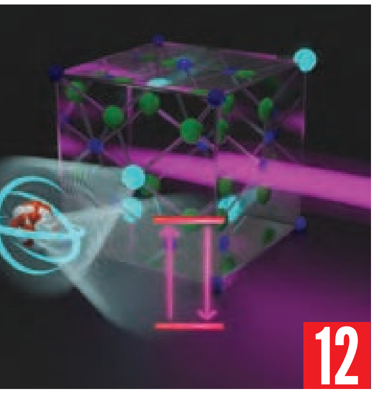
The Haber–Bosch process requires high temperatures and pressures to synthesize ammonia. But new candidates emerge for potential low-energy catalysts—if their magnetism is suppressed. Researchers demonstrated the approach by placing lanthanum near a cobalt catalyst to enhance ammonia production.

physicstoday.org/Jun2024c

PHYSICS TODAY (ISSN 0031-9228, coden PHTOAD) volume 77, number 6. Published monthly by the American Institute of Physics, 1305 Walt Whitman Rd, Suite 110, Melville, NY 11747-4300. Periodicals postage paid at Huntington Station, NY, and at additional mailing offices. POSTMASTER: Send address changes to PHYSICS TODAY, American Institute of Physics, 1305 Walt Whitman Rd, Suite 110, Melville, NY 11747-4300. Views expressed in PHYSICS TODAY and on its website are those of the authors and not necessarily those of AIP or any of its member societies.

PHYSICS TODAY
www.physicstoday.org

DEPARTMENTS



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

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

PhysicsToday **@physicstoday**

AIP

Ultra-stable *DC Voltage Source*

DC205 ... $2995 (U.S. list)

- ±100 VDC range
- True 6-digit resolution
- 1 ppm/°C stability
- 0.0025 % accuracy (1 yr)
- Triggerable voltage scans
- Low-noise design
- Linear power supply

When you need a quiet, stable, high-resolution bias voltage, the DC205 is the right tool. Its bipolar, four-quadrant output delivers up to 100 volts with microvolt resolution and up to 50 mA of current. In 4-wire mode (remote sense), the instrument corrects for lead resistance delivering accurate potential to your load. The DC205's output stability is a remarkable ±1 ppm over 24 hours. With its linear power supply, there is no need to worry about high-frequency noise.

The DC205 can generate triggerable scans when voltage ramping is required. It is also fully programmable over RS-232 and USB, and there's a fiber optic interface for use with the SX199 Optical Interface Controller.

thinkSRS.com/DC205

Supporting emerging astronomers across Africa

We appreciated reading "More Africans are pursuing STEM graduate studies in the US" by Toni Feder in the January 2024 issue of Physics Today (page 19). The report describes initiatives, such as the African School of Physics and the African Institute for Mathematical Sciences, that support African university students in gaining STEM skills and experience that can prepare them for attending graduate school in the US and other countries.

We would like to highlight the Pan-African School for Emerging Astronomers (PASEA), which we codirect. Formerly known as the West African International Summer School for Young Astronomers, our program consists of a week-long course intended to help African university students develop their STEM skills and interests and learn about careers in astronomy. Founded in 2013 and held every two years, it is designed and taught by an international collaboration of astronomers. PASEA's vision is to build a critical mass of astronomers in Africa and to exchange teaching ideas across continents. PASEA has been held in Nigeria, Ghana, and Zambia, and PASEA 2024 will be held in Tunisia this September.

PASEA has several notable aspects:

- ▶ Our curriculum, based in inquiry and education research principles, in which students ask and investigate their own mini research questions in small teams.[1]
- ▶ "Paired teaching," in which international partners coteach to learn new strategies.
- ▶ Our alumni community.
- ▶ Our evaluations of student learning.

Half of our instructors are women (unusual for physics and astronomy), and we hold special events to support women students and discuss gender-based barriers to participating in science. (More details on PASEA can be found at www.paseaafrica.org and in reference 2.)

PASEA now has almost 300 alumni from 20 African countries. Almost all continue in STEM. About one-third are pursuing astronomy; most of them did not have an astronomy program at their university, and PASEA helped bring them into the field. Nearly 90% of alumni say that their involvement with PASEA was a big influence on them when choosing what career to pursue or subject to study. Five alumni have returned to PASEA as instructors because they want to give back to the program and their fellow students from around the African continent. Two alumni instructors are now graduate students in Canada, and one is a postdoctoral researcher in France. Other alumni are graduate students in African countries, including Zambia, South Africa, and Mauritius. Some of them are supported by the Development in Africa with Radio Astronomy (DARA) program and the PanAfrican Planetary and Space Science Network (PAPSSN).

While many of our alumni have strong scientific backgrounds and would like to apply to graduate school, they need support in navigating the process—for example, figuring out where to apply, understanding the differences between master's and PhD programs, learning how funding for graduate students works in different countries, and writing competitive research statements. These are significant barriers for talented students around the African continent who are otherwise well qualified for graduate school. We would like to create a mentorship program to support PASEA alumni and other African students with guidance on careers and graduate-school applications, and we are currently seeking funding for such a program.

STUDENTS Niza Gladys Kamanga, Joseph Mukuka, and Martha Nambela (left to right) designing a way to measure the distance to a star cluster, facilitated by Pan-African School for Emerging Astronomers instructor and alumna Margaret Ikape (far right), in the inquiry activity at PASEA 2022 in Livingstone, Zambia. (Courtesy of Linda Strubbe/CC BY-NC 4.0.)

References

1. L. Strubbe et al., "Distances in the Universe: An Inquiry Lab Sequence Taught in West Africa and North America," CourseSource (2023).
2. L. Strubbe et al., *Nat. Astron.* **5**, 217 (2021).

Linda Strubbe
Strubbe Educational Consulting
Vancouver, British Columbia, Canada
James Chibueze
University of South Africa
Pretoria
Bonaventure Okere
Centre for Basic Space Science
Nsukka, Nigeria
Jielai Zhang
Swinburne University of Technology
Melbourne, Australia
(program email: paseacollaboration@gmail.com)

Roman Jackiw and the chiral anomaly

The morning after Roman Jackiw died on 14 June 2023, his widow, So-Young Pi, wrote to inform me, and I replied to her, "Roman left a great legacy of discoveries to the physics world, which are widely admired and appreciated, and which you shared in in his later years." I have since been surprised, and dismayed, to see old controversies resurfacing about the history of the chiral anomaly, a quantization-induced breaking of classical axial symmetry.

To appreciate Roman's legacy, it is not necessary to rewrite the history of the chiral anomaly, as Andrew Strominger does in his obituary on page 53 of the December 2023 issue of Physics Today. Strominger writes, "Famously, as a young postdoc working with John Bell at Harvard, Jackiw discovered the so-called Adler-Bell-Jackiw axial anomaly and used it to explain the decay of pions into photons. He turned lemons into lemonade by both interpreting and making physical predictions from what seemed to be an inconsistency in the Feynman diagrams." As I wrote to Strominger in response, the original Bell–Jackiw article[1] does not give a correct calculation of neutral pion decay into photons; it introduces a spurious regulator. The correct calculation was first given in my article finding the chiral anomaly, in the appendix I added after I learned of the Bell–Jackiw work.[2]

When I sent my draft appendix to Roman, he replied in a letter dated 25 August 1968, expressing skepticism that an unambiguous calculation of pion decay to photons was possible using the chiral anomaly, which he characterized as beset with intrinsic mathematical ambiguity. More to the point, and very candidly, John Bell wrote in a letter dated 2 September 1968, "The general idea of adding some quadratic electromagnetic terms to PCAC [the partially conserved axial current] has been in our minds since Sutherland's η problem. We did not see what to do with it." The originals of those letters reside in the Institute for Advanced Study Archives and can be viewed there by researchers in the history and philosophy of science.

Further detail on the history of the chiral anomaly is contained in my 2006 volume of commentaries and selected papers.[3] I consider the correct calculation of pion decay to photons to be the result of a fruitful scientific interaction between the Bell–Jackiw work and my own. I hope that people who are interested in the subject of anomalies will read (or reread) the original papers.

References

1. J. S. Bell, R. Jackiw, *Nuovo Cimento* **60**, 47 (1969).
2. S. L. Adler, *Phys. Rev.* **177**, 2426 (1969).
3. S. L. Adler, *Adventures in Theoretical Physics: Selected Papers with Commentaries*, World Scientific (2006).

Stephen L. Adler
(adler@ias.edu)
Institute for Advanced Study
Princeton, New Jersey

► **Strominger replies:** Stephen Adler has an incredible scientific legacy, including his central role in work on the chiral anomaly and beyond. To quote his letter, I "consider the correct calculation of pion decay to photons to be the result of a fruitful scientific interaction between the Bell–Jackiw work and [his] own." I also follow Adler in urging everyone to read the original papers.

In the very limited space I was allotted to cover the rich life and works of Roman Jackiw, I did not attempt to focus on scientific accomplishments beyond his own or to write a comprehensive history of others' contributions to developments he was involved in. I presumed that that would be understood and that my words would not be taken to diminish the significance of those contributions. I apologize for any misunderstandings in that regard.

Andrew Strominger
(andrew.strominger@gmail.com)
Harvard University
Cambridge, Massachusetts PT

CONTACT PHYSICS TODAY

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

Slow-motion spectroscopy paves the way for a nuclear clock

For the first time, the energy level of an atomic nucleus has been directly manipulated with a laser.

Jules Janssen was a 19th-century umbraphile: He traveled the world to witness solar eclipses. During one such event in 1868 in India, he saw something that would help to shape the understanding of both the Sun's inner workings and the periodic table. When he sent the highly attenuated sunlight through his spectrometer, he found a yellow line that no one had ever seen before. It turned out to be an atomic emission from an as-yet undiscovered element: helium.

The discrete colors of light absorbed and emitted by atoms have long been integral to our understanding of the physical world. A few decades after Janssen's discovery, the pioneers of quantum mechanics looked to atomic energy levels, deduced from their spectra, to work out the laws that govern the dynamics of electrons in atoms. And modern atomic clocks, the best of which waver by less than a second over the age of the universe, are the basis for GPS navigation, tests of general relativity, and the very definitions of the fundamental units of measure. (See the article by David Newell, Physics Today, July 2014, page 35, and the Quick Study by Emily Caldwell and Laura Sinclair on page 54 of this issue.)

Atomic nucleons, like electrons, organize into energy shells, and nuclei, like atoms, can be promoted between discrete energy states. But there's no nuclear optical spectroscopy, and there aren't yet any nuclear clocks. The problem is that almost all nuclear transitions lie at impractically high energies: in the realm of gamma rays, not visible light, and far beyond the reach of tabletop lasers.

In fact, out of all the known nuclides, only one—the rare, radioactive isotope thorium-229—hosts an excited state, denoted ^{229m}Th, that lies within a handful of electron volts of the ground state. Laser spectroscopy of nuclei will never span the periodic table the way atomic spectroscopy does. But for the purpose of building a nuclear analogue of an atomic clock, one transition of the right energy is all that's needed. And ^{229}Th will do the job.

Now a team of researchers led by Thorsten Schumm (Technical University of Vienna) and Ekkehard Peik (PTB, the National Metrology Institute of Germany) has succeeded in exciting ^{229}Th with a laser.[1] Meanwhile, at UCLA, Eric Hudson and colleagues have achieved similar results.[2] Laser excitation is just one step on the path to a nuclear clock, but it's an important one. And the hurdles encountered along the way—including finding the transition energy in the first place—highlight the differences between nuclear and atomic physics.

The nuclear option

The heart of an atomic clock is a spectroscopic frequency measurement: The clock's ticks are the electromagnetic cycles associated with an atomic energy transition. If precision timekeeping is the goal, it's necessary to tamp down all sources of uncertainty in the measurement, such as Doppler broadening, which can make the transition look bluer or redder depending on whether the atom is moving toward or away from the observer.

For those uncertainty-mitigation efforts to do any good, it's necessary to start with a transition with an inherently narrow linewidth. By the Heisenberg uncertainty principle, that means the excited state must have a long radiative lifetime. Most atomic transitions won't work: Their excited states relax back to the ground state in a fraction of a second.

In contrast, the ^{229m}Th state's radiative lifetime is comparatively long—minutes to hours, depending on the environment—but not extraordinarily so: Plenty of atomic states have even longer lifetimes. An excited state in ionic ytterbium, used in some state-of-the-art clocks, has a lifetime of more than a year. And the excited state of the microwave-frequency cesium transition that's the basis for the definition of the second has a lifetime so long that it defies measurement.

But the metrological appeal of the ^{229}Th transition—remarked on in 1996 by Eugene Tkalya and colleagues[3] and developed into a full-fledged nuclear clock proposal in 2003 by Peik and Christian Tamm[4]—goes beyond its narrow inherent linewidth. For example, nuclear energy levels are less susceptible than electronic states to distortion by stray electromagnetic fields.

Furthermore, the nuclei used in a nuclear clock can be embedded in a solid. That's not possible for atomic clocks: The chemical bonds of a crystal lattice would hopelessly alter the energy of the electronic transition. All atomic clocks, therefore, use atoms held station-

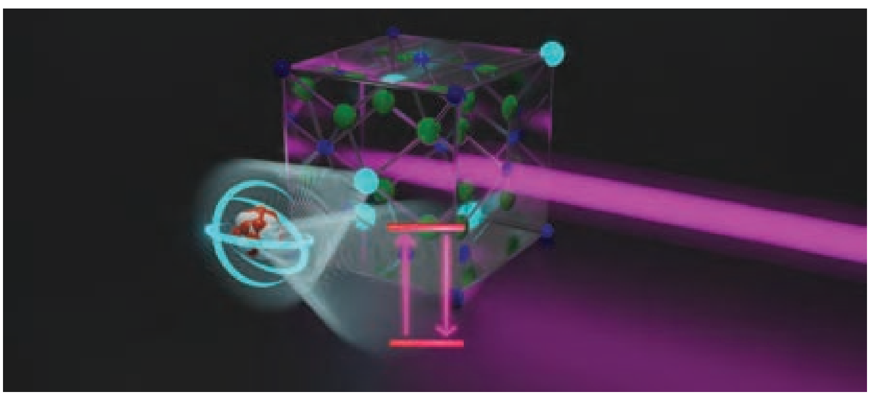


FIGURE 1. THORIUM-229 ATOMS, shown in light blue in this artist's impression, are held in the lattice of a calcium fluoride crystal. The isotope is prized for its unusually low-lying nuclear excited state, the transition to which can now be driven with a laser. (Image by Oliver Diekmann, Technical University of Vienna.)

ary in the gas phase, such as in electromagnetic or optical traps. The option to use a crystal instead makes it easier to hold large numbers of nuclei in a small space, and it could make nuclear clocks more portable.

Lost in transition

The prospect of a thorium nuclear clock faced a barrier that would be unheard of in the world of atomic clocks: Nobody knew the energy of the ^{229}Th nuclear transition, not even approximately.

It's been known for decades that the ^{229m}Th state exists. When ^{229}Th is produced from the radioactive decay of other elements—such as the alpha decay of uranium-233—the pattern of the gamma rays that are released makes it clear that the isotope is being produced in two states of nearly equal energy. But finding the energy difference between the two states would have meant measuring the difference between two large gamma-ray energies, and nuclear experiments aren't capable of measuring things so finely.

Theory, likewise, wasn't equipped to grapple with such small energies. So in the 1976 paper[5] that first recognized the low-energy state, the best the authors could say was that the energy difference was something less than 100 eV—that is, something less than the energy of a 12 nm soft x ray. That's not much help.

Experimenters would need the energy information to have any hope of exciting the nuclear transition directly. The usual spectroscopic approach of exciting atoms with broadband light and looking for photons absorbed or emitted wasn't going to work on ^{229}Th. The long radiative lifetime of the ^{229m}Th state means that only a trickle of photons is being emitted at any given time.

Meanwhile, radiative decay is competing with radioactive decay. The ^{229}Th isotope's 8000-year radioactive half-life dwarfs the ^{229m}Th state's radiative lifetime—but each radioactive decay produces thousands of photons, not just one. Only with a tightly concentrated laser frequency would researchers have a hope of exciting enough nuclei for the radiative signal to outcompete the radioactive background. Scanning the whole 100 eV range would take forever.

Happily, theory and experiment improved, and researchers started to home in on the ^{229}Th transition energy. Unhappily, for a while at least, they were homing in on the wrong answer. In the late 1990s and early 2000s, when Tkalya, Peik, and others were formulating ideas for a nuclear clock, the best guess was that the transition was somewhere around 3.5 eV, or 354 nm. If true, that would have been fortunate: Light of that wavelength propagates not only in air but also in plenty of crystals, including some pure thorium compounds, such as thorium oxide. But all searches for the transition in that vicinity turned up empty—because it wasn't there.

The actual energy, it would turn out, was more than twice as high: not in the near UV but in the vacuum UV. Thorium oxide is opaque in that range, as are most crystals. But a few wide-bandgap materials, such as calcium fluoride, are transparent, and they can be doped with thorium, as illustrated in figure 1.

Moreover, vacuum-UV light, as the name suggests, doesn't propagate in air. The whole experiment, from laser to sample to detector, would need to be placed under vacuum—a change that would prompt a rethinking of every aspect of the experimental design. "It becomes harder and harder to build a good laser when the wavelength becomes shorter," says Peik. "And you can't just reach in and tweak things with your hands."

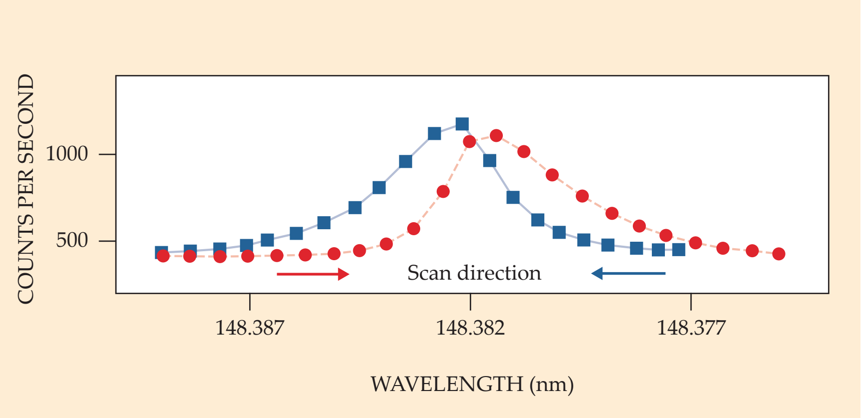


FIGURE 2. LASER EXCITATION of the thorium-229 nucleus looks different from different directions. To efficiently scan the range of possible wavelengths (only a small portion of which is shown here), Thorsten Schumm, Ekkehard Peik, and colleagues recorded these spectra at a rate of five minutes per data point. But the ^{229m}Th excited state's lifetime is close to 10 minutes, so the spectral line is skewed. (Adapted from ref. 1.)

Great catch

A breakthrough came last year, when the ISOLDE group at CERN published the first observations of photons emitted from the thorium radiative transition.[6] The ISOLDE researchers did it by peppering a calcium fluoride crystal with radioactive precursors to ^{229}Th. Some of the precursors would decay into the ^{229m}Th state, which then relaxed to the ground state. That approach had been tried before, but always using ^{233}U, which decays to ^{229}Th by emitting alpha particles that carry significant momentum and could have been damaging the crystal. The ISOLDE researchers instead created ^{229}Th through the beta decay of actinium-299, a gentler process that allowed them to finally observe ^{229m}Th's radiative decay.

The ISOLDE observation narrowed the transition energy down to 8.338 ± 0.024 eV, or a wavelength somewhere between 148.3 and 149.1 nm. At last, researchers could optimize their lasers and begin their search in earnest. But scanning even that narrow range was a painstaking process. Both Peik and Schumm's group and Hudson's used the same basic approach: Illuminate a ^{229}Th-doped crystal with a laser for a while, then switch off the laser and look for any photons emitted by nuclei that had been excited by the laser. Repeat the process with different laser wavelengths until the entire energy range is covered.

Both research groups used lasers with bandwidths of around 10 GHz, or 40 μeV, so they needed some 1000 data points to span ISOLDE's energy range. How long does that take? It depends on how patient you are. In a crystal, ^{229m}Th has a radiative lifetime of around 10 minutes, but that's the half-life—only half of the excited nuclei decay in that time. To reset all the nuclei back to the ground state, you'd need to wait for several half-lives.

Hudson and colleagues took that approach, doing their laser scan at a rate of an hour per data point, or 1000 hours of data-taking to cover the whole ISOLDE range. Schumm, Peik, and colleagues took a more expedient approach—although still sluggish by spectroscopic standards—of five minutes per data point. As a result, they were the first group to find the excitation. But because the signal they observed at each step included nuclei that had been excited in previous steps, the spectral line they observed was skewed, as shown in figure 2. To compensate, they scanned the range in both directions.

Schumm and Peik's team measured the resonance wavelength to be 148.3821 ± 0.0005 nm; Hudson and colleagues found it to be 148.3822 ± 0.0002 nm. The agreement lends credence to both measurements, especially because the two experiments used different crystal materials: calcium fluoride in Schumm and Peik's experiment, and lithium strontium aluminum fluoride in Hudson's. There was an outside chance that the spectroscopic feature could have been an artifact, perhaps the result of a crystal defect created by thorium's radioactivity. But the odds of a spurious signal at exactly the same wavelength in two different materials are slim to none.

Fine structure

There's still a long way to go on the path to a nuclear clock. To achieve precision of a part in 10^{18}, as the best atomic clocks do, researchers would need a wavelength measurement that's accurate to the 15th decimal place, not the 4th. The next step, therefore, is to repeat the excitation measurement using narrower-band lasers. There's plenty of room for improvement. The inherent linewidth of the ^{229}Th transition is thought to be smaller than 1 mHz, more than 13 orders of magnitude narrower than the 10 GHz laser bandwidth used in the initial search.

As researchers zero in on the transition wavelength, a problem that's plagued the search so far—background due to radioactive decay—should naturally resolve itself. With a 10 GHz bandwidth, the vast majority of the laser power is always wasted: Even when the laser is centered on the transition wavelength, most of the light is too far off resonance to excite any nuclei. With a narrower bandwidth, more light can contribute to exciting the nuclei. "For the same laser power, we get a much stronger signal," says Schumm. "Then we can reduce the doping concentration to reduce the radioactive background."

More than two decades after they were initially proposed, nuclear clocks are looking like an achievable goal. "But atomic clocks are a moving target," notes Peik. "Researchers have made so much progress in recent years, and they've solved a lot of the problems that the nuclear clock was designed to circumvent." From a purely timekeeping perspective, it remains to be seen whether nuclear clocks will ever get the upper hand.

But there may be more subtle benefits to a clock based on nuclear rather than atomic physics. For example, one way that researchers test whether the fundamental constants have changed over time is to compare two clocks based on different atomic transitions. If the fine-structure constant, say, isn't actually constant in time, the clocks' relative tick rates would slowly drift.

So far, no such drift has been observed—but that could be either because the fine-structure constant isn't changing or because its change is too small to be observed. Because of the overall larger energy scale of nuclear physics, a nuclear clock should be orders of magnitude more responsive than an atomic clock to any changes in the fundamental constants. So researchers could have a chance at spotting a drift in the fine-structure constant that they otherwise would never observe in a million years.

Johanna Miller

References

1. J. Tiedau et al., *Phys. Rev. Lett.* **132**, 182501 (2024).
2. R. Elwell et al., https://arxiv.org/abs/2404.12311.
3. E. V. Tkalya et al., *Phys. Scripta* **53**, 296 (1996).
4. E. Peik, C. Tamm, *Europhys. Lett.* **61**, 181 (2003).
5. L. A. Kroger, C. W. Reich, *Nucl. Phys. A* **259**, 29 (1976).
6. S. Kraemer et al., *Nature* **617**, 706 (2023).

Gravitational patterns reveal a tumultuous lunar past

Geophysical models and data analysis show that subsurface titanium causes anomalies in the Moon's gravitational field.

Twin spacecraft orbited the Moon on a three-month mission in 2012. Together, the two satellites of NASA's GRAIL (Gravity Recovery and Interior Laboratory) mission mapped the Moon's gravitational field to better learn about its internal structure. The extraordinarily detailed map that the mission produced contained mostly what the GRAIL scientists expected to find: Gravity is stronger above the mountains and weaker above the craters. Yet when they removed the topographical contributions, what remained were strangely linear features[1] stretching up to roughly 500 km, as seen by the blue lines in figure 1. At the time, it was unlike anything the team had seen.

Now a research team led by Jeff Andrews-Hanna (University of Arizona) has explained what was causing the patterns. Building on work by Nan Zhang (Peking University) and colleagues, Andrews-Hanna and his group have demonstrated that dense, titanium-rich materials below the surface could shape the gravitational field.[2] Their simulations and data, together with Zhang's geophysical analysis, provide evidence of the Moon's mantle overturning—that is, denser materials at the top migrating to lower depths—and when the event occurred during the Moon's early formation.

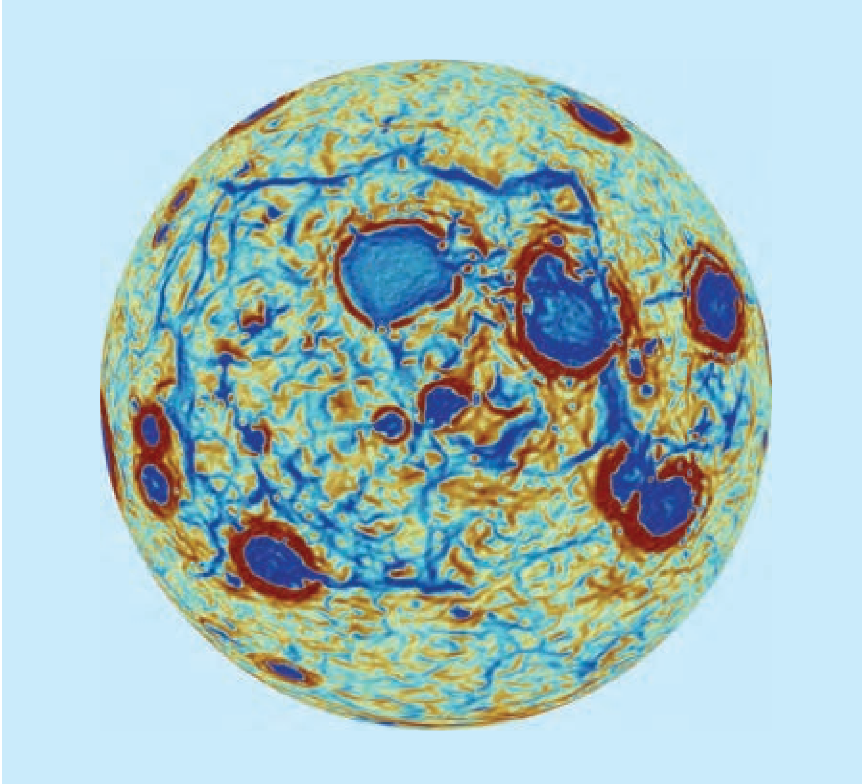


FIGURE 1. THE GRAVITATIONAL FIELD of the near side of the Moon. The map shows the gravitational strength (stronger in blue) after the contributions from the surface features are removed. Rough circles (blue) of gravitational gradients correspond to the craters of the region, but the strange, linear anomalies don't have a visual counterpart. Rather, they are the signatures of dense material beneath the surface. The presence of titanium-rich materials sinking below the surface can be explained by the overturning of the mantle early in lunar history. (Courtesy of Adrien Broquet/University of Arizona.)

A molten past

The Moon was born after another celestial body collided with a young Earth approximately 4.5 billion years ago. The debris coalesced to form a Moon that was largely molten, with a magma ocean in which the elements were well mixed. The prevailing theory is that the ocean cooled and crystallized to form the lunar mantle and crust. The last minerals to solidify likely contained metals such as titanium. They would've crystallized into the densest material higher in the mantle, counterintuitively than less dense materials that had already sunk. Because of that unstable arrangement, scientists suspect that the Moon's mantle overturned as the titanium-rich materials sank lower. Yet physical evidence of the event and its timing has been hard to come by.

But there's more than one lunar mystery: If the mantle did overturn, the event cannot explain the Moon's geophysical asymmetries. A combination of Apollo samples and reflectance data from previous lunar studies has shown that the near side of the Moon has higher quantities of titanium and other elements that are rare on the Moon. In 2022 Zhang and colleagues used numerical simulations to try to understand why. They hypothesized that a migration of materials throughout the interior was triggered by the giant impact that formed the largest lunar crater.

That crater, the South Pole–Aitken impact basin, is on the far side of the Moon. Zhang and colleagues modeled impact parameters, such as the velocity and impact angle of the projectile, to obtain the most accurate simulation of the impact event and compared simulation results with the current topography of the impact basin.[3] The group then investigated the effect that the impact would have had on the mantle's convection and the distribution of ilmenite, a titanium–iron oxide found in high concentrations on the near side of the Moon.

Ilmenite is less viscous than other material in the lunar mantle, and it would've been pushed by the force of the impact toward the near side of the Moon—opposite the impact. Zhang and his group use numerical simulations to conclude that the impact not only formed the largest lunar crater but instigated thermal upwellings, shown in figure 2.

The simulations suggest that those upwellings triggered the migration of titanium-rich materials to the nearside, where higher concentrations of ilmenite in the solid mantle and upper crust are found today.

The simulations show that after the ilmenite migrated to the near side, it formed sheet-like slabs and then sank deeper into the mantle during the overturning. The modeled abundance of titanium oxide after the vertical integration shows patterns of straight lines along the surface.

Connecting the lines

The pattern caught the eye of Andrews-Hanna, who had been part of the GRAIL team. He recognized that the simulation's straight lines may be connected to the linear anomalies seen in the gravity map of the same region. Some of the lines in the GRAIL data appear to form a border around the Ocean of Storms—a large, dark region on the near side of the Moon that appears on the left when viewed from the Northern Hemisphere.

When GRAIL scientists mapped the gravity of the region, they removed the effect of topographical surface features to try and learn about the Moon's interior. Topography accounts for 98% of lunar gravity, and regions of dense material below the surface are correlated to surface features. But the lines around the Ocean of Storms didn't have an obvious connection to the topography.

Andrews-Hanna and his research group collaborated with Zhang to investigate whether the titanium oxides could be influencing the gravity and leading to the strange patterns. At the University of Arizona, Weigang Liang and Adrien Broquet simulated different geometries, depths, and densities of the titanium-rich material to determine the geophysical scenario most consistent with the GRAIL data. They also worked from the opposite direction, using the GRAIL gravity data to calculate the most likely mass and density of the subsurface material. Reassuringly, the two independent approaches provided consistent results. Sheet-like slabs of dense, titanium-rich material could create the linear signatures seen in the GRAIL data.

The cohesion between the results lends credence to the theory of the lunar mantle overturning. The simulations demonstrate how a giant impact could have created thermal instabilities that caused an asymmetry in the distribution of materials such as titanium oxides. The slabs that sank into the mantle appear as the pattern of linear anomalies in the gravitational map.

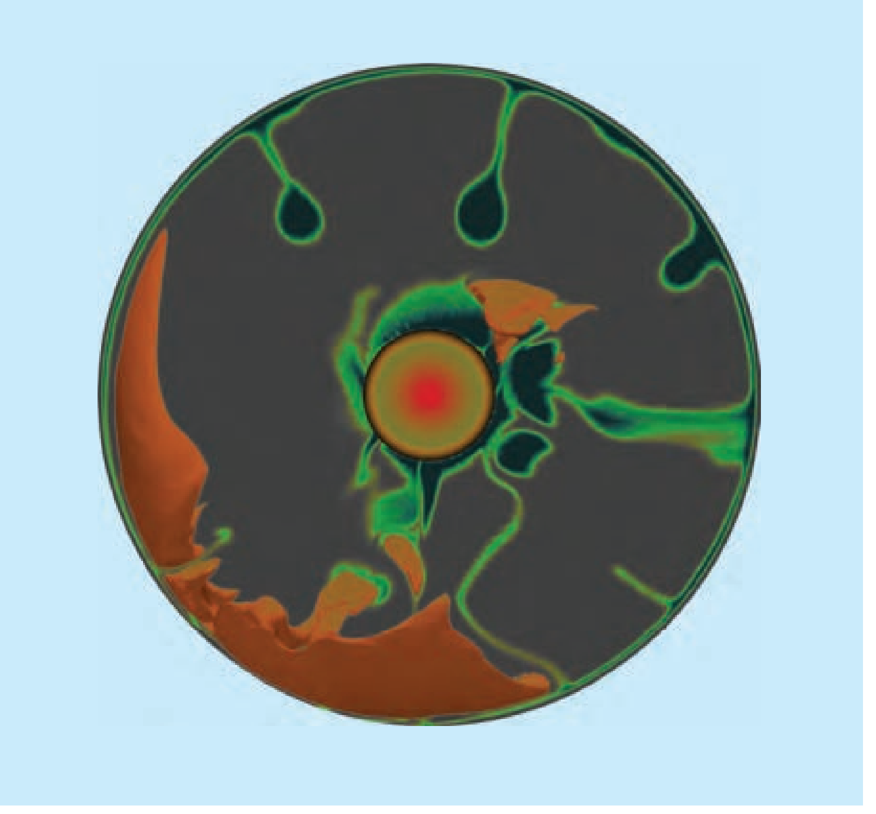


FIGURE 2. DENSE, TITANIUM-RICH MATERIALS (green) sank from the surface to the deep mantle shortly after the Moon's formation. A cross section showing Nan Zhang's simulation of the overturning event illustrates the effect of a large impact at the bottom left, which created temperature anomalies (brown) and downwellings of titanium-rich materials. Some remnants of the dense material are expected to be preserved throughout the Moon's geologic evolution; they appear as straight lines in the measured gravitational gradients seen in figure 1. (Courtesy of Adrien Broquet, University of Arizona/Nan Zhang, Peking University.)

Moreover, Andrews-Hanna and his team used the relationships between the gravity anomalies in the Ocean of Storms and impact basins of known ages in the region to show that the overturning event would've happened at least 4.2 billion years ago, merely a few hundred million years after the Moon was formed. In a few locations on the surface, it is especially clear that a physical sign of a sinking ilmenite slab was overwritten by an impact crater, which showed that the crater formed after the overturning of the lunar mantle.

The new work led by Andrews-Hanna provides the first connection between an observable piece of data and the overturning of the lunar mantle, a defining event in the Moon's history. The limits on the time frame have allowed the team to begin reexamining older lunar theories in light of the new parameters to better understand the timeline of the Moon's formation. As part of NASA's Artemis program, astronauts are scheduled to land at the rim of the South Pole–Aitken impact basin in late 2026. The plan is for them to collect samples from the impact that triggered the overturning, which may provide more data to fill in details of the lunar formation timeline.

Jennifer Sieben

References

1. J. C. Andrews-Hanna et al., *Nature* **514**, 68 (2014).
2. W. Liang et al., *Nat. Geosci.* **17**, 361 (2024).
3. N. Zhang et al., *Nat. Geosci.* **15**, 37 (2022).

Cosmic tau neutrinos uncovered

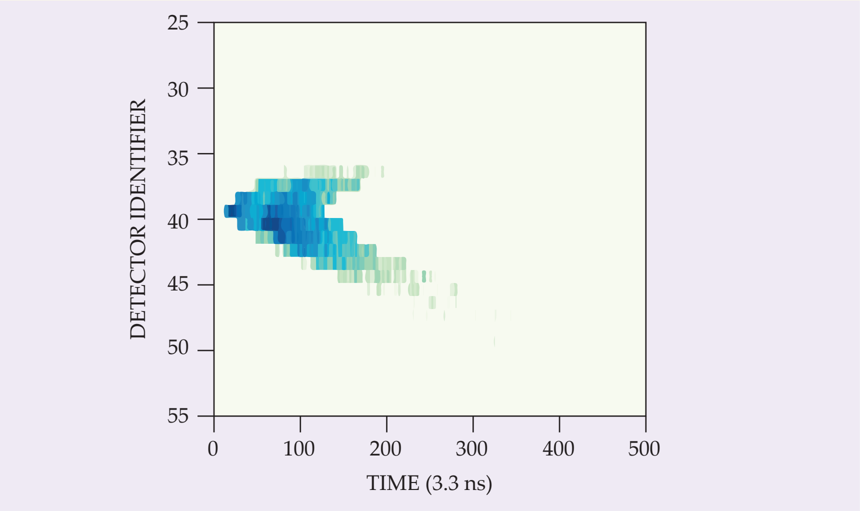
An image-scouring AI model assists researchers in finding signatures of the most elusive variety of neutrino.

LIGHT SENSORS FOR THE IceCube Neutrino Observatory are lowered into the Antarctic ice. (Image by Mark Krasberg, IceCube/NSF.)

It took two years for researchers using the vast IceCube Neutrino Observatory to uncover a population of the charge-less, featherweight particles that originated beyond the solar system. They would need a decade more, it turns out, to confirm that those energetic particles come in all three neutrino flavors. To pair with its previous measurements of cosmic muon and electron neutrinos, the IceCube collaboration now reports identifying seven candidate tau neutrinos, the most elusive variety of neutrino because its partner lepton decays so rapidly.

The candidates were plucked from nearly 10 years of data acquired from the IceCube detector, whose more than 5000 light sensors are attached to cables that are frozen within an Antarctic ice sheet. Whether produced in a laboratory or in a galaxy billions of light-years away, neutrinos are constantly oscillating between flavors until they interact with other matter. When such an interaction occurs, including in the dark icy depths monitored by IceCube, a charged muon, electron, or tau lepton is released. That particle, and the cascade of particles it yields from further collisions, briefly exceeds the speed of light in ice and emits a measurable blue glow as it slows down.

A tau neutrino announces its arrival at IceCube via a so-called double bang: a flash of light triggered by the neutrino's impact, followed by another from the decay of the tau lepton. But the lepton's lifetime is so short that the dual flashes are hard to spot. To do the job, IceCube researchers turned to convolutional neural networks, AI models with applications that include identifying objects in images. One model was trained to differentiate between single and double flashes; the others distinguish tau signals from those of muon neutrinos and of cosmic rays.



THIS CANDIDATE TAU NEUTRINO was detected at IceCube in 2019. The darker the blue, the brighter the light sensed by the detectors; the numbers on the vertical axis correspond with depth. The two regions of dark blue are presumably caused by the neutrino interaction and then the decay of the tau lepton. (Image adapted from the IceCube collaboration.)

Seven of IceCube's high-energy particle events aced the tau-ness test of all the neural networks, a number that reflects the rareness of cosmic neutrino detections and the fact that the analysis focused on only a small section of the vast detector. Still, the results establish that there is at least some flux of cosmic tau neutrinos. And they are consistent with theoretical predictions that muon, electron, and tau neutrinos should appear in roughly equal numbers from distant sources because of the particles' continuous oscillation en route.

The tau neutrino detections also provide verification that the speedy neutrinos that IceCube has been reporting since 2013 as sourced from beyond the solar system are indeed produced far from home. (The cosmic sources may include blazars, active galactic nuclei, and other violent objects.) Although particle interactions in Earth's atmosphere can produce energetic electron and muon neutrinos, they don't have time to oscillate into tau neutrinos before reaching the IceCube site. (IceCube collaboration, *Phys. Rev. Lett.*, **132,** 151001, 2024.)

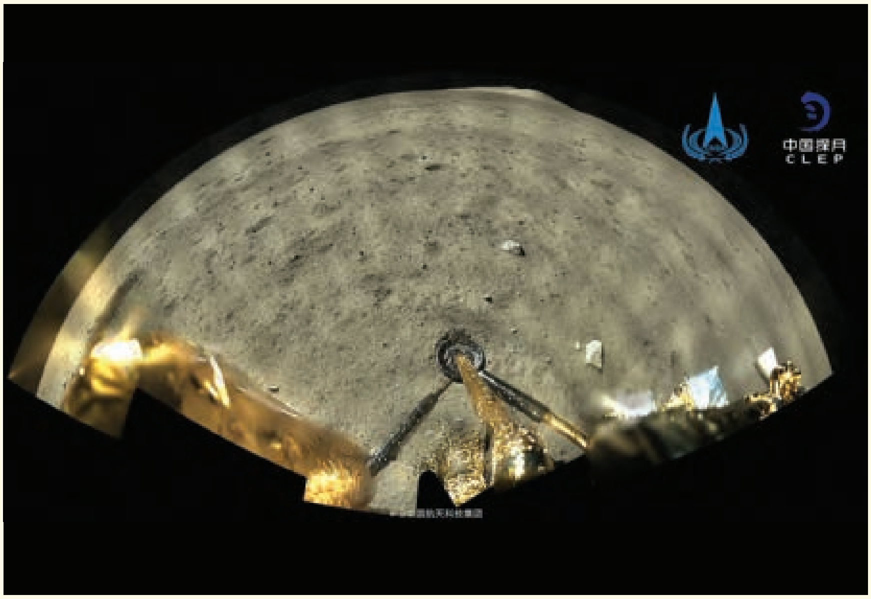
Andrew Grant

Two new minerals found on the Moon

A glass bead containing an impact crater just 9 μm across is the source of otherworldly titanium oxide compounds.

China's *Chang'e 5* lander returned to Earth on 16 December 2020 with the first sample brought back from the Moon since 1976. Within the roughly 1.7 kg sample, researchers found a glass bead with a pit about 9 μm across, formed by the impact of a piece of fast-traveling space dust known as a micrometeorite. On the rim of the tiny crater they found two titanium-based minerals—trigonal and triclinic Ti_2O—that had not been found on the Moon before and do not occur naturally on Earth. Those are now the seventh and eighth new minerals discovered on the Moon to date, as described in a recent *Nature Astronomy* paper by Xiaojia Zeng, Yanxue Wu, and colleagues.

Above Earth, the friction generated by meteors moving through the atmosphere slows them down and can burn them up, depending on their incoming size and speed. Above the Moon and other airless bodies, though, there is no atmospheric buffer between the debris of space and the rocky surface. The Moon is thus bombarded not only with the large meteors and asteroids that have given rise to its iconic cratered surface but also with tiny dust-sized meteors that travel at high speeds—as fast as 20 km/s, about 30 times as fast as an F-16 jet. Those micrometeorite impacts are crucial to the weathering of the lunar surface.



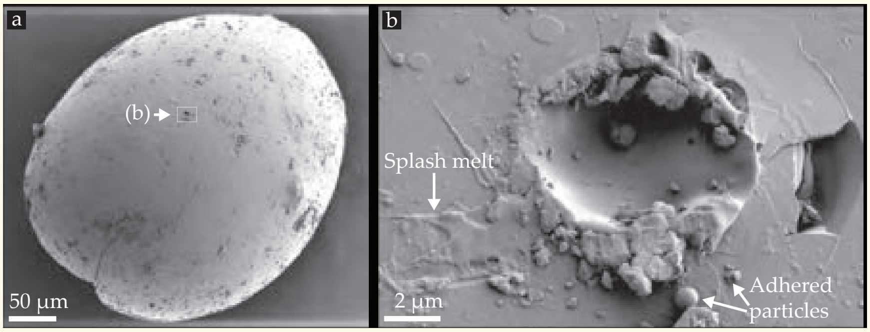
A PANORAMIC PHOTO of the lunar surface captured by China's *Chang'e 5* as it landed on the Moon in 2020. (Image from the China National Space Administration.)

The cratered glass bead, according to Zeng, Wu, and colleagues, is rich in iron and was likely formed from a larger meteorite collision in the lunar maria—vast, dark basaltic plains that fill ancient impact basins. Using scanning electron microscopy, the researchers identified the minerals ilmenite, troilite, and apatite on the surface of the bead. Ilmenite, which is common in the lunar mare regions, is a titanium iron oxide mineral and the likely source of the titanium that formed the two new minerals. To shed light on the chemical and physical reactions at play during the micrometeorite impact, the researchers homed in on the impact crater on the bead's surface.

Zeng, Wu, and colleagues used transmission electron microscopy to look at the tiny (100–300 nm) titanium oxide deposits on the crater rim and observe the lattice structures that define their mineral phases. The researchers believe that the minerals were formed by the vaporizing of ilmenite during impact, which ripped apart its constituent elements. Titanium and oxygen plasma then rapidly recombined and deposited onto the glass bead's surface. Such minerals have been artificially created in laboratories using pulsed-laser deposition, a similar process.

Zeng says there are two reasons for why we likely don't see the minerals occurring naturally on Earth: the lack of high-speed micrometeorites and the presence of water and oxygen in the atmosphere, which might alter the reactions that lead to deposition of the minerals postimpact. Given the abundance of both ilmenite and micrometeorite impacts on the lunar surface, the researchers expect that those products of space weathering should be present across the mare regions. But because the deposits are so small—only a few hundred nanometers—it's possible that they have been overlooked in previous samples taken from the Moon. (X. Zeng et al., *Nat. Astron.*, 2024, doi:10.1038/s41550-024-02229-4.)

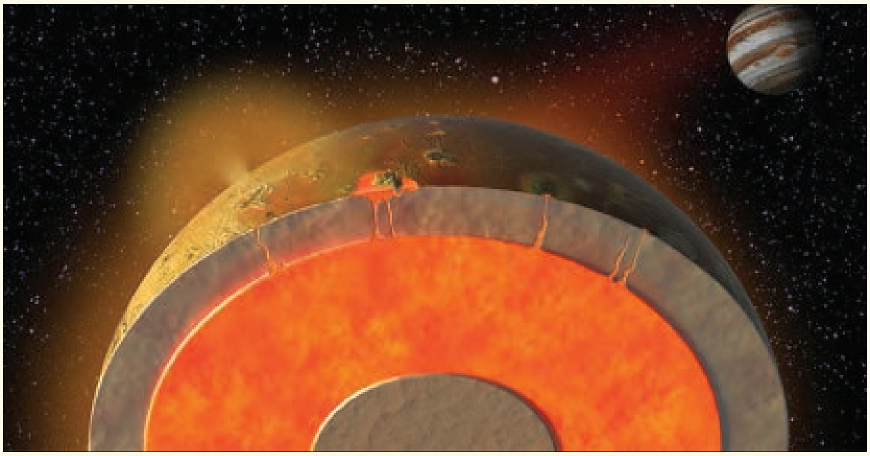
Laura Fattaruso



A GLASS BEAD (a) from the surface of the Moon contains a micrometeorite crater **(b)** produced by a high-speed impact of micron-scale space debris. The impact produced titanium minerals that do not occur naturally on Earth. (Adapted from X. Zeng et al., *Nat. Astron.*, 2024, doi:10.1038/s41550-024-02229-4.)

Io was always extremely volcanic, evidence indicates

The Jovian moon's abundance of a heavy sulfur isotope is higher than that of any other object in the solar system.



JUPITER'S MOON IO may have experienced large eruptions and lava flows throughout its 4.5-billion-year history. (Image by Chuck Carter and James Tuttle Keane/Keck Institute for Space Studies.)

The most volcanically active object ever observed in the solar system is Io. Its massive and sustained eruptions, extensive lava flows, and huge lava lakes are more substantial than any on Earth and have led researchers to wonder: Has Io always been that way? The question is hard to answer because its surface is constantly changing. In just a million years—which is short by geological standards—Io's entire surface has been repaved by volcanic deposits, destroying craters and other evidence of geologic history over its 4.5-billion-year existence.

Now Katherine de Kleer of Caltech and colleagues have developed a way to study Io's volcanic history by tracking its sulfur. With new isotope evidence and geochemical modeling, they found that Io's volcanism was likely even more pronounced in the past and may have been present over the moon's entire history.

The researchers observed sulfur gases in Io's atmosphere with the Atacama Large Millimeter/Submillimeter Array in Chile. When a gas molecule loses angular momentum, it transitions from a higher rotational energy state to a lower one and, in the process, emits microwave radiation. Each molecule has many unique spectral emission lines that can be observed. Even molecules made with different isotopes—say, with sulfur-32 and sulfur-34—are distinguishable. With the molecular emission data, the researchers determined the sulfur isotope composition of Io's atmosphere and found that it contains much more ^{34}S than ^{32}S.

In fact, no other measured body in the solar system has such a high ratio of $^{34}S/^{32}S$. To interpret Io's unusual sulfur enrichment, one of the paper's coauthors, Ery Hughes, led the development of a geochemical model of Io's interior and its atmosphere. The sulfur ratio likely reflects a geochemical process that begins when sulfur dissolved in magma rises through Io's mantle and erupts from volcanoes. The lighter ^{32}S diffuses into the upper atmosphere more readily than ^{34}S, and then the atmospheric sulfur collides with stray plasma from Jupiter's magnetosphere before being lost to space.

The high $^{34}S/^{32}S$ ratio may indicate an especially volcanic past. The geochemical modeling shows that Io could have been losing sulfur in the past at a rate as high as five times what it is today. The extreme volcanism necessary to cause that much sulfur loss comes from tidal heating. Jupiter's intense gravity flexes Io, and the resulting tidal energy dissipates in Io's interior as frictional heat. Io's tidal heating is possible only because of its elliptical orbit around Jupiter, caused by interactions with other Jovian moons. So by learning more about Io's volcanic history, researchers may unravel more about the formation of Jupiter and its moons—and potentially other systems subject to strong tidal forces. (K. de Kleer et al., *Science*, 2024, doi:10.1126/science.adj0625.)

Alex Lopatka

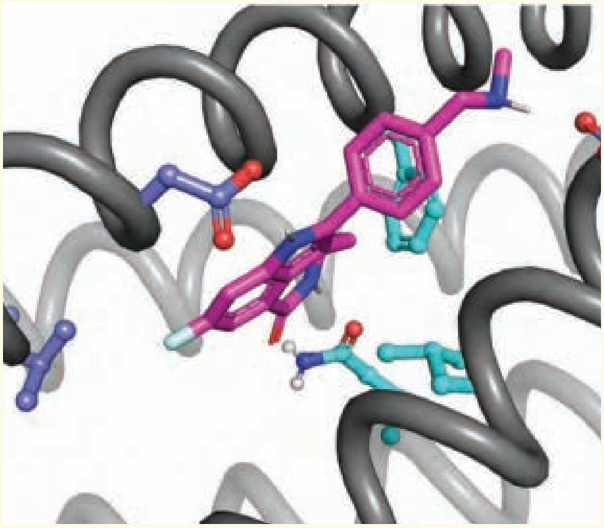
Designer proteins fit like a glove

The bespoke biomolecules interact with molecular targets in predictable, controllable ways.

The relationship between protein sequence and structure isn't as much of a mystery as it once was. In late 2020, researchers from DeepMind in London turned heads with their AlphaFold2 model, which uses artificial intelligence to predict the structures of natural proteins with stunning accuracy. (See Physics Today, October 2021, page 14.) But there's more to know about a protein than its structure. Proteins aren't static isolated objects—they're constantly interacting with other molecules around them. The energetics and dynamics of those interactions are fundamental to proteins' role as the building blocks of life.

Now researchers led by William DeGrado (University of California, San Francisco), his former postdoc Nicholas Polizzi (now on the faculty at Harvard University), and his current postdoc Lei Lu have unveiled a way to design proteins from scratch so that the proteins not only bind to a specified target molecule but do so with predictable binding energy. So-called *de novo*–designed proteins, made from amino-acid sequences that nature never exploited, are nothing new. But in most cases, the computed structures need to be experimentally refined with several rounds of mutation and screening before they're fit for purpose. Polizzi and DeGrado's proteins are notable exceptions: At least much of the time, they work on the first try.

The work builds on an underlying approach Polizzi and DeGrado developed several years ago, breaking down the protein–molecule binding problem



STARTING FROM A FEW standard backbone structures, such as the alpha helices shown in gray, protein designers have extraordinary leeway to tune a protein's chemical properties. Here, the amino acid side chains shown in purple and light blue are chosen to bind to a target molecule, shown mostly in pink. (Image adapted from L. Lu et al., *Science* **384**, 106, 2024.)

into pieces called van der Mers. That's not an obscure Dutch surname but rather a portmanteau of "van der Waals" and "rotamer." Rotamers are floppy parts of amino acids; although they can, in principle, adopt enormously many conformations, only a handful of possibilities ever show up in the real proteins in the Protein Data Bank. By limiting themselves to just the conformations that nature tends to favor, protein designers can radically simplify their computational searches.

As Polizzi and DeGrado discovered, the same strategy works for the van der Waals interaction between amino acids and other molecules: Given an amino acid and the nearest bit of the target molecule, the atoms only ever arrange themselves in a few discrete ways. By sifting through the ways that the protein–molecule pieces can be packed into a known protein backbone—such as the four-helix structure shown in the figure—the researchers create a *de novo* protein fitted to the target molecule. With judicious choices of amino acids, they can make the binding as strong or as gentle as they like. And they can keep it specific: The protein binds to the target molecule, but not to any others.

The problem the researchers tackled is the inverse of conventional drug design. Given a naturally occurring protein, drug designers want to identify a molecule that binds to it, perhaps to stop it from performing some harmful function in the body. So why design a protein to bind to a given molecule? One possible application is to create antidotes for drugs—to neutralize them and stop their effects. Another is as a first step toward designing artificial enzymes: Before an enzyme can catalyze a reaction, it first needs to bind to the reactant molecule. (L. Lu et al., *Science* **384**, 106, 2024.)

Johanna Miller PT

Advanced conductors could double power flows on the grid

Widespread reconductoring of the US transmission system could ease bottlenecks that are preventing renewable energy sources from hooking up to power grids.

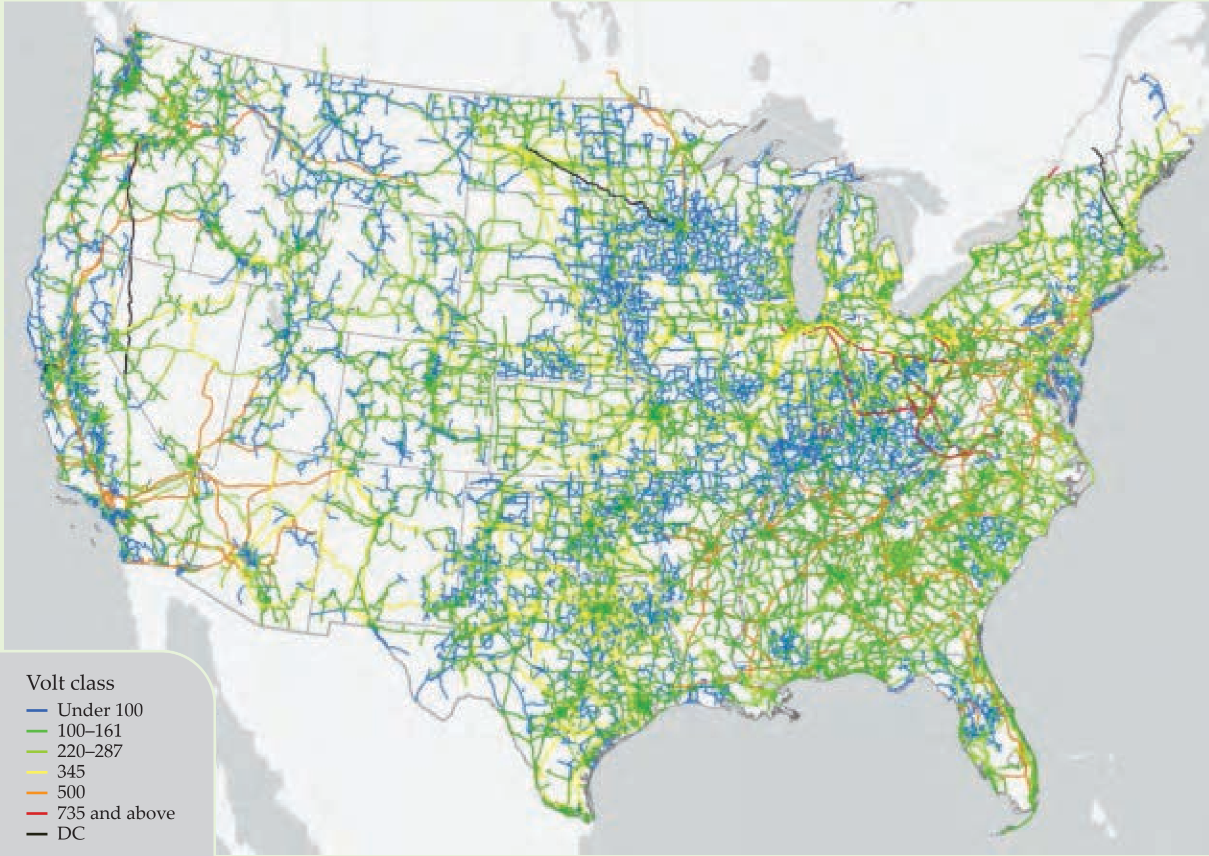
Following 15 years of stagnating electricity consumption in the US, demand is now on the rise, driven by a proliferation of power-hungry data centers and by the increasing adoption of electric vehicles and heat pumps. As those technologies become more widely adopted and industrial processes are electrified over the coming decade, the US will need to increase its electricity-generation capacity by more than the amount currently available in Texas, according to Jeffrey Brown, managing director of the nonprofit EFI Foundation.

An assessment by Lawrence Berkeley National Laboratory released in April

TS CONDUCTOR

THE ARIZONA UTILITY SRP is shown carrying out a reconductoring of a 230 kV double-circuit transmission line with composite-core conductors made by TS Conductor. The circuit on the right side of the tower is live. As lineworkers string the new conductor, it is anchored to heavy concrete blocks on the ground using the gray clamps shown. The blocks are then moved to the next section for reuse.

US TRANSMISSION LINES



THE VAST MAJORITY of US transmission lines are 235 kV or below. Their capacity could double if they were refitted with advanced conductors, researchers say.

identified 2600 GW of proposed wind, solar, and energy storage projects that are awaiting connection to the grid. Their capacity is more than double that of the entire nation's 1250 GW installed power generation. Renewable energy and batteries make up 95% of the queue, and 1200 GW were proposed since passage in August 2022 of the Inflation Reduction Act, which included a range of tax incentives and other provisions to encourage clean-energy production. History indicates that nearly three-quarters of those interconnection requests will eventually be withdrawn, the assessment says.

The consulting firm Grid Strategies published a report last year saying that transmission congestion costs exceeded $20 billion nationwide in 2022, a 56% increase from 2021. The primary cause, it says, is the failure of transmission expansion to keep up with the growth of low-cost wind and solar energy.

The growth of data centers is already straining the grid in some locations, notably northern Virginia, which is home to 150 hyperscale centers, more than one-third of the world's total, according to the Virginia Economic Development Partnership. Hyperscale centers are generally considered to comprise 5000 servers or more, and they are typically operated by tech giants such as Google, Amazon, Meta, and Microsoft. Dominion Energy, the utility that services most of them, experienced load growth of 500 MW annually from 2019 to 2022, says Grid Strategies.

The Department of Energy's 2023 National Transmission Needs Study finds that to rapidly electrify the US economy and decarbonize the grid, intraregional transmission must double and interregional transmission capacity must increase by a factor of six by 2040.

Building new transmission lines, however, can take 10 or more years. The process involves acquiring new rights of way, determining how to allocate the costs to the utilities that will benefit, acquiring permits from regulators, and building the towers and associated infrastructure.

On 25 April DOE announced a final transmission-permitting reform rule that commits the agency to coordinate and complete within two years—about half the current timeline—all the interagency reviews required for approval of new transmission corridors. At the same time, the agency announced a $331 million investment, from the 2021 Infrastructure Investment and Jobs Act, in the construction of a new 459 km transmission line that will carry 2000 MW of wind energy from Idaho to southern Nevada and southern California.

A quick doubling

Even before new transmission corridors are built, the capacity of today's grid could quickly be doubled if cables were replaced with advanced conductors on a wide scale, says an April report by researchers at the

University of California, Berkeley, and GridLab. Although advanced conductors may be two or three times as costly as conventional ones, upgrading existing lines with them could be accomplished at one-quarter to one-half the cost of building new lines, and the jobs could be completed in 18–36 months, without a need to obtain permits, the researchers found.

Of the approximately 1.3 million kilometers of existing transmission lines in the US, only 1% are reconductored or rebuilt each year, the report says. And even though advanced-conductor alternatives have been available for more than two decades, many rewiring projects in the US are still completed with conventional conductors, which don't appreciably increase capacity.

Most transmission lines in the US today deploy a more than century-old type of cable known as aluminum-conductor steel-reinforced (ACSR). Those conductors are designed to operate continuously at temperatures of up to 93 °C and at higher temperatures for brief periods during emergency conditions. As line temperature increases in concert with the current carried, the ACSR steel core expands, increasing the sag between towers.

The 2003 blackout in parts of the midwestern and northeastern US and Ontario, Canada, resulted in 11 deaths and cost the economy an estimated $6 billion. A federal commission identified the initial trigger of the event as an Ohio transmission line sagging and contacting a tree. In the aftermath, the North American Electric Reliability Corp issued a requirement that lines be surveyed and clearances be maintained.

Advanced conductors have composite cores, made of material such as carbon fiber and ceramics, that are stronger than steel and thermally expand at far lower rates—carbon fiber has one-tenth the thermal expansion of steel, says Gary Sibilant, a senior program manager at the Electric Power Research Institute. That allows much higher operating temperatures and current flow.

The smaller diameter of advanced conductor cores allows more aluminum to be packed into a cable of the same diameter as an ACSR cable. Advanced conductors also lose less power during transmission, so generating stations can produce less power to satisfy the load. Those advantages add up to cables with twice the current-carrying capacity of ACSR ones.

Advanced conductors have been underutilized for a variety of reasons, according to the Berkeley–GridLab report. They include workforce shortages, cost, lack of space needed to perform the reconductoring, uncertainty over permitting requirements, age and condition of existing structures, and concerns that the increased capacity will require transformer upgrades at substations.

Many other nations in Europe, South America, and Asia are reconductoring, some to a greater extent than in the US. Globally, more than 144 000 km of advanced conductors have been deployed since 2003, the Berkeley–GridLab report says.

Elia Transmission Belgium, which operates the country's national grid, has been reconductoring since 2009. Kristof Sleurs, the company's head of grid development, says the utility expects to reconductor its entire transmission backbone of 380 kV lines over the next decade. In combination with other grid-enhancing technologies, such as dynamic line rating—calculating actual conductor conditions in order to match current capacity to real-time conditions—and flow-directing tools, reconductoring should accommodate most of Belgium's transmission needs for the next decade, he says. Still, new circuits, including high-voltage DC lines, must be built for electricity generated by new offshore wind farms to be carried, he adds.

Thermal constraints

Opinions differ on the extent to which advanced conductors could be deployed in the US. According to a 2023 report by Idaho National Laboratory, around 20% of the US transmission system is ripe for advanced conductors. But the Berkeley–GridLab report says that all but around 2% of the nation's transmission lines could benefit. An accompanying report by GridLab and Energy Innovation says wide-scale advanced conductors could help quadruple the capacity of the new transmission lines that are projected to be built by 2035. That should enable a 90% clean electricity system and could save $85 billion in energy system costs, it says.

An April report from DOE says advanced reconductoring is most economically viable on lines that are approaching end of life or in places where the increase in capacity is so significant that it warrants early replacement. Capacity benefits are most applicable to lines with lengths less than 48 km and voltages less than 500 kV, the report says. For reducing line losses, however, advanced reconductoring is more broadly applicable, it adds. Supporting structures and poles must have sufficient remaining life and structural integrity to be candidates, the report notes.

Sibilant says lines carrying 138–230 kV are thermally constrained and would benefit the most from advanced conductors. They consist of a single line for each of the three AC phases carried, and they make up the bulk of the nation's grid (see the figure on page 22). Transmission at 345 kV or greater is not thermally constrained, because it uses two or more conductors—and much more aluminum—for each phase, Sibilant says.

Jason Huang, CEO of TS Conductor, which manufactures carbon-fiber-core cable, says that savings from reduced line losses could be monetized to finance the cost of reconductoring without any cost to utilities. He notes that advanced conductors also allow for transmission lines with longer spans and fewer towers. At $250 000 or more per tower, those savings can add up, he says.

Still, many utilities remain hesitant to install composite-core conductors over perceptions that they are fragile and easy to damage. Some advanced conductors can't be bent as easily as those with steel cores, notes Sibilant, and those require different handling procedures. Composite-core conductors are also considerably more expensive; they cost as much as four times the price of ACSR cables, according to the Idaho National Laboratory report. But given that the conductor material is less than 5% of the total cost of a new transmission line, the premium for a more expensive conductor is easily justified, the report adds.

Upgrading the voltage of transmission lines could also increase their capacity, says Sibilant. High-voltage DC provides the greatest capacity, but that requires expensive conversion equipment at both ends of the line. DC is used for transmission over very long distances, such as those involved in bringing hydroelectric power from northern Quebec to the US and bringing offshore wind power to land, for which line-loss savings justify the additional expense.

David Kramer

Q&A: Ernest Moniz on the nuclear weapons threat

Nuclear and nonnuclear nations need to act to prevent the growing risk of accidental or deliberate use, says the former US secretary of energy.


NTI

As Hollywood was preparing for the Academy Awards in early March, a publicity campaign that had nothing to do with advocating for a best actor or best cinematography nominee became increasingly visible in Los Angeles. Posters plastered across the city warned that although "[J. Robert] Oppenheimer is history, nuclear weapons are not."

The campaign advocating for the end of nuclear weapons was the creation of the nonprofit Nuclear Threat Initiative (NTI). The organization was founded in 2001 to lobby for policies to prevent the proliferation of nuclear weapons. It then broadened its portfolio to cover biological and chemical weapons. It has increasingly become a voice for nuclear disarmament. Ernest Moniz, a former MIT physics professor and secretary of energy for President Barack Obama from 2013 to 2017, is the organization's CEO.

Moniz spoke with Physics Today a few weeks after the Oscars to discuss the NTI's nuclear security efforts and the challenges of elucidating the nuclear threat to the public.

PT: How significant is the threat of nuclear weapons use today?

MONIZ: When the Cold War ended, almost everyone exhaled and thought the age of nuclear weapons was over. The reality is that in the 30 years since then, we have now come back to a place where the risk of nuclear use, either accidentally or deliberately, is probably at least on the scale of the Cuban missile crisis. Vladimir Putin's implicit and explicit threats on nuclear weapons use around Ukraine have violated all kinds of norms, including Russian norms: The P5 [the five permanent members of the United Nations Security Council, all declared nuclear weapons states, which include Russia] all agree that we would not use a nuclear weapon, or threaten the use of a nuclear weapon, against a non-nuclear-weapons state.

Now we have China building its arsenal: US intelligence predicts on the order of 1500 weapons in China by 2035 [roughly the same number of deployed warheads that the 2011 New START Treaty imposes as a limit for the US and Russia]. Then there is North Korea and its nuclear saber-rattling. And you have the whole world of emerging technology cybersecurity and AI.

The history of US–Soviet and US–Russia relations has always been built around a bilateral control architecture. Let's say that China comes in at a similar scale to the US and Russia 10 years from now. In simple-minded physics terms, two-body negotiations don't fit. You can't do the fundamental algorithm of US and Russia having equal numbers if there's three. The hawks will say we have to have as many deployed weapons as the sum of Russia and China. That's great, except Russia and China would never accept that. That is a formula for an arms race. We don't need 1550 deployed weapons, let alone another couple thousand, to deter Russia and China at the same time.

As you know in physics, the two-body problem is inherently stable, and the three-body problem is inherently not. We need a whole new algorithm for strategic stability in a multipolar world. The non-nuclear-weapons states have to be in this conversation. They also signed the Nuclear Non-Proliferation Treaty, and there is a lot of unhappiness with the weapons states for not making adequate progress toward the agreed-to goal [of world disarmament].

PT: Can you discuss the NTI's involvement with the fail-safe review underway at the Department of Defense?

MONIZ: Fail-safe means that if there is a failure of command-and-control systems, it does not lead to catastrophe. We've had several incidents of incorrect information reaching the presidents of the US and Russia that there were incoming strikes. [See "Nuclear weapons dangers and policy options," by Steve Fetter, Richard Garwin, and Frank von Hippel, Physics Today, April 2018, page 32.] The last and only systematic fail-safe review done in the US—and we don't know of any that have been done elsewhere—was during the George H. W. Bush administration. Today's geopolitical and technological world bears no resemblance to the world of 1990. In this world of cyberattacks and new technologies such as AI, everything should be on the table for evaluation.

[NTI cofounder and US former senator] Sam Nunn and I went directly to the House and Senate Armed Services Committees, and the chairs of the two committees mandated a fail-safe review in the 2022 National Defense Authorization Act. That review is now being carried out, and the expectation is that it will be finished this fall.

If all the countries with nuclear arsenals would do their own fail-safe reviews, it would lower the risk of accidentally blundering into the use of nuclear weapons. It's in their self-interest, as well as in the interests of everybody else.

PT: Is accidental use still your biggest concern?

MONIZ: Twenty years ago we would have said the primary risk was accidental use. A second risk, especially since 9/11, NTI

A BILLBOARD parked in front of the Hollywood Walk of Fame in the days before *Oppenheimer* received seven Academy Awards. The sign was part of a blitz campaign by the Nuclear Threat Initiative in support of global nuclear disarmament.

is terrorists getting hold of a weapon or weapons materials. But the deliberate use of weapons by nuclear weapons states was until recently thought to be a sort of thing of the past. NTI is partway through a major project looking *de novo* at the risk today of deliberate use and how to construct norms, and ideally agreements, to lower or eliminate those risks.

PT: What was the inspiration for the Academy Awards publicity campaign calling for an end to nuclear weapons?

MONIZ: Recognizing that we have limited capacity, we've always had an eye on how and when we can help influence public opinion, particularly in a way that would ultimately create more political imperatives in the policy world to work toward that vision. We had been told directly by members of Congress who were quite involved in nuclear security policy that there's only so much they could do because their constituents are not exactly writing letters saying they are worried about nuclear threats.

Oppenheimer, its popularity, and its 13 Academy Award nominations provided an unusual opportunity to engage in a strategic communications initiative. About five weeks before the awards, we got a significant anonymous gift to do just that.

We had a five-week blitz that included an op-ed by me and [former California governor] Jerry Brown and an open letter with signatories from the entertainment industry. We put up a thousand posters in L.A., particularly on the route to the awards ceremony, and we commissioned a pop artist to do some outdoor art where there is a huge amount of foot traffic. Cillian Murphy made a brief statement in his acceptance speech for best actor. And [director] Christopher Nolan made some strong statements in interviews about listening to Oppenheimer's warnings in the postwar period.

Oppenheimer's postwar warnings were about not getting into an arms race and not developing evermore powerful nuclear weapons. Today it's pretty routine to have weapons with one or two orders of magnitude more yield than those in World War II.

PT: You talk a lot about educating the public. But nuclear weapons have been around nearly 80 years, and everyone surely is aware of the threat they pose.

MONIZ: I don't agree with that premise. Much of the population was born since the end of the Cold War. They kind of know, but they don't view it as a major threat until maybe very recently, driven mainly by Putin's statements. North Korea is far away. "It doesn't threaten us directly," is the public's attitude. "India and Pakistan, well you know they're always fighting." That kind of thing.

PT: But couldn't increasing public awareness of the nuclear danger reinforce support for arms buildup?

MONIZ: That's where we have to guide education in our way, that [more weapons] will just increase the risk. At NTI, we are also looking at updating the potential effects of a nuclear war. There's the blast, radiation, and nuclear winter. I'm not saying it isn't bad enough already, but we saw painfully how COVID-19 completely discombobulated global supply chains. Imagine what would happen in a nuclear war. The scale of the effects in the modern world are probably much worse than was anticipated 30 years ago.

We've been doing some work, but it's very much on our punch list for trying to attract the resources to make a big effort here. It's all consistent with our thinking that if the public understands in the modern context the extreme implications of nukes-use exchange, that can be part of building up the pressure on the will to address these risks in a serious way.

David Kramer

Firearms forensics is becoming more quantitative

Science over subjectivity could increase juries' confidence in gun identification.

Someone drops a dead body at a hospital. Bullets from the decedent and from a wall where a shooting took place are found to possess matching toolmarks—striations and other markings in the softer bullet material made by imperfections in the gun's barrel. When a suspect's gun is determined to create such toolmarks, prosecutors argue that the suspect fired those bullets. But if, say, a bullet retrieved from a child has toolmarks that differ from those created by a suspect's gun, "You know it's not the same gun," says Michael Haag, who worked for 25 years in a police crime department and continues his long-time gig as an independent forensics consultant.

For more than a century, examiners have matched guns to crimes by comparing the markings made on fired bullets and spent casings with those test-fired from a known gun. Matching bullets or the cartridge cases they are held in—which often remain near the shooting site—to guns is a primary source of evidence in hundreds of cases every day across the US.

Determining that a bullet was shot by a particular gun is not easy, says Alicia Carriquiry, a statistics professor and director of the Center for Statistics and Applications in Forensics Evidence at Iowa State University. "To the naked eye, the toolmarks look like cat scratches. There's a lot of subjectivity. The question is, Can you do something more scientific?"

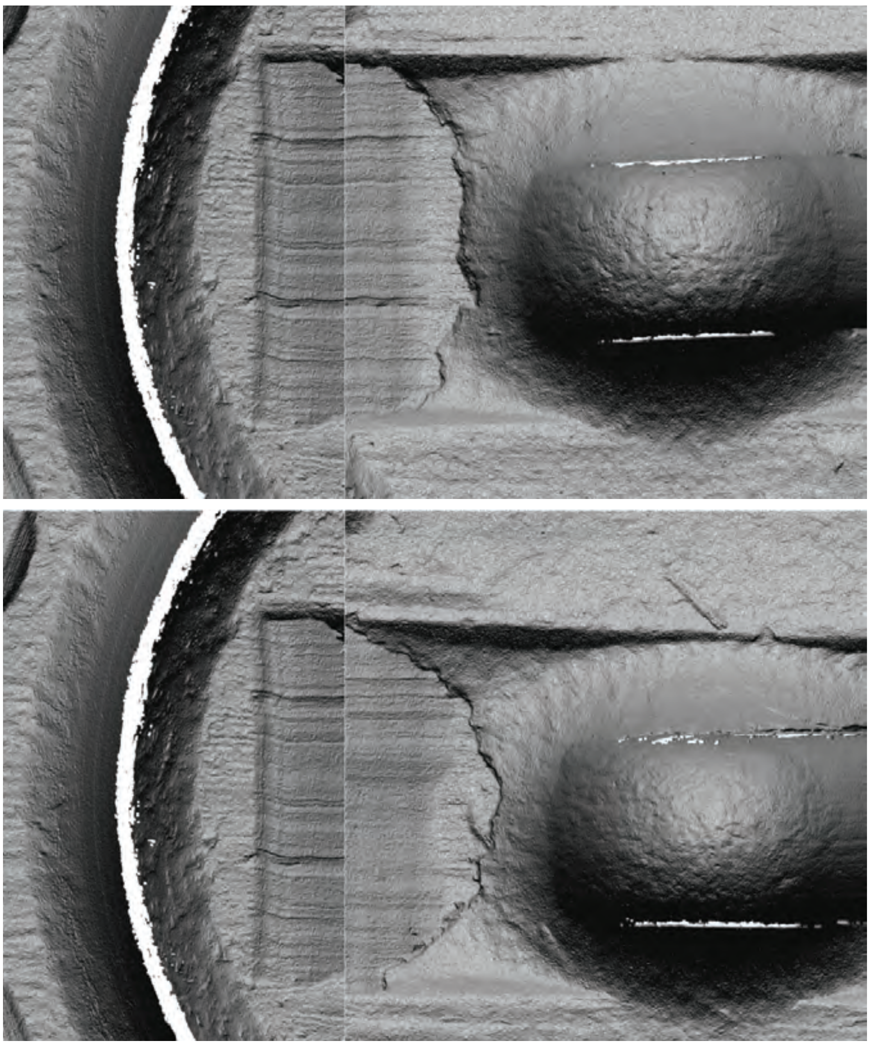
Carriquiry and her colleagues are working to improve toolmark analysis. She calls the results "promising" and expects to see "some changes in the way evidence is evaluated in the next few years."

Meanwhile, scientists at NIST, the Federal Bureau of Investigation, and the Netherlands Forensic Institute are creating a database to help gauge the likelihood that a casing or bullet was fired by a given gun. Xiaoyu Alan Zheng is a mechanical engineer and the NIST lead on the database. He says that the project was "springboarded" by the 2009 National Research Council report *Strengthening Forensic Science in the United States: A Path Forward*. "They wanted more objective comparisons, with scores and numbers, for what is a match or not." The database is expected to go live in three to five years, Zheng says, and will be an ongoing project.

3D advantages

Firearms examiners use traditional microscopy to compare bullets and casings from crime scenes with ones shot from known guns. "There are no real standards on how much similarity there has to be," says Zak Carr, a firearms and toolmark examiner who is now at Cadre Forensics, where he is working on improving and quantifying toolmark identification. When examiners evaluate samples from known sources, the accuracy is high, he adds, noting that research studies show error rates to be around 1%.

Despite the generally high accuracy of examiners' results, says Carr, "it would be nice if you could come up with



NIST

A KNOWN MATCH (top), for which two cartridge cases (left and right) were from the same firearm. The bottom image shows two cartridge cases from different firearms. Three-dimensional imaging and improved algorithms are in development to help examiners quantify their assessments on whether a bullet or casing was fired by a specific gun. Such evidence is commonly used in criminal trials.

RACHAEL GOMINSKY/FBI LABORATORY

ANALYSIS FROM THIS CARTRIDGE CASE (left), shot with a 9 mm Luger pistol, is part of a database that the FBI, NIST, and the Netherlands Forensic Institute are working on to help assign a likelihood to whether a bullet or casing was fired by a given gun. The features that would catch an examiner's attention are the lines, the firing pin impression at the center, the drag at 3 o'clock, and the crescent-shaped toolmark running around the firing pin from 7 to 11 o'clock. The center image shows a scan of the cartridge case, which is used to obtain the coordinates of toolmarks. The right-hand image is a map of the depths of the toolmarks made with focus variation canning; black represents deep spots and yellow represents high spots.

a quantified score for whether toolmarks were from the same or different firearms. That's what we are working toward."

A further impetus for quantifying certainty about evidence, says Carriquiry, is second opinions: Different examiners may reach different conclusions about the same data, or the same examiner looking at the same data again six months later may have a different opinion. And, she says, in recent years "examiners have been taking a beating. Lawyers are waking up to the fact that you can argue against subjective examination. Firearms evidence has been excluded in some cases, which has been a shock."

To set probabilities for source determinations, Carriquiry, scientists at Cadre, and researchers elsewhere are working on analyses of 3D images of bullets, from which "you can look at the depth and position of striations," says Carriquiry. She notes that the algorithms do not always work. And some guns don't mark. "But the crappy guns that most criminals use mark nicely, and the accuracy in identification is high."

Using 3D virtual comparison microscopy has many advantages over traditional examinations, says Carr. "The algorithm can analyze similarities between items in high-volume cases faster. It's huge." The resolution is better, he adds, and the lighting is more consistent. Sharing evidence is also easier. Defense attorneys would like to see criminal evidence, but approval to access and move it can be difficult to obtain. With virtual comparison microscopy, Carr says, "the raw data can be shared without risk of compromise or loss of evidence."

During his nearly two decades as an examiner, Carr says, he "spent hours hunched over the microscope [and] developed callouses and neck pain. It takes a lot of time to see the details."

Bullet analysis with 3D images "may not make a big difference in results," says Robert Thompson, the senior forensic science research manager at NIST, "but it will augment examiners' subjective opinions with objective data. It will increase confidence for juries."

Not surprisingly, it's often easier to be certain about nonmatches than about matches. And then there are the cases that can't be called. In Carr's experience, some 5–10% of cases have been inconclusive.

Erich Smith is the technical leader for firearms and toolmarks at the FBI Laboratory. In looking through six years of in-house data, he found that just under 14% of cases were reported as inconclusive, with the rest split roughly evenly between identifications and eliminations. Inconclusive is a valid finding, Smith notes. Bullets get damaged, for example. Improved image analysis can make a dent in the rate of inconclusive findings, he and others say.

The algorithms developed in Carriquiry's center are starting to be tested in forensics labs, she says. Some labs already have 3D microscopes, but for others, ponying up a few hundred thousand dollars for one can be a challenge. Introducing 3D microscopes and algorithmic analysis on a large scale "will take buy-in," she says.

Reference guns and algorithms

NIST, the FBI, and the Netherlands Forensic Institute are building their database from reference guns. The FBI has so far collected scan data for four types of firearms. It has tested about 1000 individual firearms, doing 11 test-fires per gun and using bullets and casings made of different metals. It creates 3D scans of casings and bullets—at about 60 minutes per scan—and uses custom software to score known matches and known nonmatches. "We are building a population with information on the type of gun, manufacturer, surface material, caliber, and more," says Smith. Most important, he adds, the database builds on empirical knowledge about same-source and different-source toolmark comparisons. "In teaching an algorithm to do what humans do, the accuracy of the discipline is increased."

The database project goes back to 2012. When it opens for use, says Smith, examiners will be able to send the 3D images of crime-scene and reference bullets and casings to the FBI team. The team will take those data and assign a score-based likelihood ratio from reference data. Just how the scoring will be done is yet to be determined, says Smith. "That's next on our list."

However the scoring is defined, for a known nonmatch the score is low, and for a known match the dispersion is wider. Importantly, Zheng says, the false-positive rate is low. "You don't want to put the wrong person in jail."

Toni Feder

Young physicists excited to network through the International Association of Physics Students

The student-run organization seeks to increase outreach in Asia, Latin America, and the US.

"The International Association of Physics Students will change your life," is something Ruhi Chitre kept hearing at a conference for women in physics in 2019. Eager to see what the fuss was about, later that year Chitre traveled from her base in the UK to Germany for the annual International Conference of Physics Students (ICPS), the flagship event of the association, known as IAPS. For her, the prediction proved true.

The ICPS showcased possible career avenues for physics students. Through discussions there, Chitre became hooked on the idea of working in science policy. Being around other physics students who were also trying to find their career paths was inspiring, she says. She realized that IAPS offered opportunities for making friends and sharing advice about career options. She served as its secretary in 2020–21 and president in 2021–22. In April, she began working as a consultant in artificial intelligence at UNESCO.

During her time as IAPS president, Chitre and other executive committee members created a server on the instant messaging platform Discord for members to make friends. "We made a structured place where people could ask about different things," she says. The topics include sharing information about events, posting funny memes, and even asking if someone has a couch to crash on.

The inaugural conference was held in Hungary in 1986 by physics students there who wanted to collaborate with their peers from around the world. It led to the formation of IAPS a year later. The organization is run by students. It gets advice and financial support from the International Union of Pure and Applied Physics and from the European Physical Society, which also provides space at its headquarters in Mulhouse, France.

IAPS comprises more than 90 000 students from 45 countries, and the number of members—and countries—continues to grow. The membership got a bump during the height of the COVID-19 pandemic, spurred by a combination of widespread isolation and easy access with the click of a button. Students can join IAPS during their undergraduate or graduate studies, and they can stay for a year after they earn their highest degree. IAPS is organized into territorial and local committees, but if there is no committee in a student's region, they can join IAPS as an individual member.

Eleven executive committee members work together to connect students around the world. Committee member Dimitris Gkavakos, a physics bachelor's student in Greece, is the public relations manager. He describes the organization as a "knowledge-transfer pipeline" that helps physics students succeed in the field.

About 400 students attend the ICPS conference each year. In 2022 Mexico was the first non-European country to host the event, but it was held remotely that and the previous year in Denmark because of the COVID-19 pandemic. The Philippines hosted the conference in person in 2023.

The IAPS executive committee wants

IAPS

CONTESTANTS GREET STEPHEN HAWKING at the opening ceremony of the 2014 Physics League Across Numerous Countries for Kick-ass Students in the Netherlands.

STEFANO UGLIANO

STUDENTS FROM AROUND THE WORLD meet in Cologne, Germany, for the 2019 annual conference of the International Association of Physics Students.

to know how they can help individual members, says current president Cyrus Walther, who this year earned his master's degree in experimental particle physics in Germany. The organization contributes financially toward events, fosters professional networking, connects committees to other associations, and offers other support when it can, he says. Through IAPS, he adds, members can develop their leadership skills and form international networks.

Creating community

The executive committee aims to increase IAPS membership in Asia, Latin America, and the US. It's been difficult to make inroads in the US, in particular, Walther says, because students there are well served for career building and networking by the Society of Physics Students (SPS, which is affiliated with the American Institute of Physics, the publisher of Physics Today). SPS is the US committee of IAPS. A couple of SPS student council members attend the ICPS every year. Still, despite having automatic membership, many SPS members are unfamiliar with IAPS, says Brad Conrad, who was SPS's director until August 2023. For the foreseeable future, he adds, SPS cannot afford the several hundred thousand dollars that it would cost to host large IAPS events in the US.

Thara Caba became involved with IAPS in 2021. At the time, she lived in her native Dominican Republic, and astronomy, her field of choice, is not taught in colleges there. Through IAPS, she formed a local committee to bring astronomy outreach to her town. With a grant from the organization, she hosted a month-long astronomy event for women and girls. Participants attended astronomy lectures and learned about potential career options. Caba says that the event inspired one high school student to create an astronomy club at her school. That sense of community is what makes IAPS special, says Caba. In 2021–22 she served as IAPS public relations manager, and in 2022–23 she was secretary. She is pursuing her master's degree in astrophysics and space science in Serbia.

In Singapore, Soe Gon Yee Thant stumbled onto IAPS in 2020 when she took part in the Physics League Across Numerous Countries for Kick-ass Students (PLANCKS), another popular IAPS event. The three-day exam-based physics competition takes place in a different country every year. In 2021 she organized a preliminary competition to select whom to send to PLANCKS from Singapore. Next, she and others formed a territorial IAPS committee. Working on the competition gave her a sense of purpose and helped her build a community, she says. She served on the IAPS executive committee in 2022–23 as the membership and advocacy manager. She graduated with a bachelor's degree in physics and is currently a researcher working on complex systems, machine learning, and neural networks. She remains involved with IAPS through an external-relations working group.

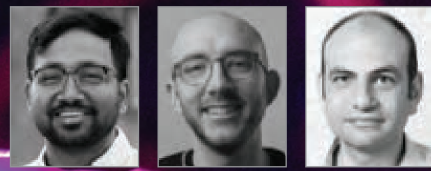
This August IAPS will be hosting the 38th ICPS in Tbilisi, Georgia.

Hannah H. Means PT

GROUND PICTURE/SHUTTERSTOCK.COM

Putting microLED technology on display

Vikrant Kumar is a PhD candidate in the department of electrical engineering at Columbia University in New York City. **Keith Behrman** is an optoelectronics engineer at Fathom Radiant in Colorado. **Ioannis Kymissis** is the Kenneth Brayer Professor of Electrical Engineering at Columbia.



Vikrant Kumar, Keith Behrman, and Ioannis Kymissis

After some two decades of advances in manufacturing processes, microLEDs have the quality and capabilities necessary for many display applications.

When the Nintendo Virtual Boy was released in 1995, it was perhaps the earliest consumer product to use LEDs in a display. It used only a 1D row of 224 red-colored pixels for its monochrome, stereoscopic 3D display. The display's oscillating mirror scanned the row of pixels through 384 lines, resulting in a resolution of 384 × 224 pixels.[1] The Virtual Boy, however, was a commercial failure—it is Nintendo's only game console to sell fewer than a million units—and the development of LED display technology stagnated.

LEDs were not traditionally used for displays, lighting, or any of their other modern applications. Rather, they were limited to simple indicator lighting in electronics. The history of LEDs dates back to the 1960s, with red and green LEDs made from the semiconductor materials of gallium arsenide and gallium phosphide, respectively. Higher costs, inefficient energy consumption, and low brightness limited the usefulness and adoption of early LEDs. Color displays need red, green, and blue (RGB) subpixels at varying brightness to combine into single pixels that can cover the color spectrum set by the International Commission on Illumination, an authority on light, illumination, and color.

Blue LEDs were not possible to manufacture with appreciable brightness levels until Shuji Nakamura's invention of them in 1993 and the subsequent vast improvements that he and others made to green LEDs. The advances made it possible to combine all three primary colors to emit light across the entire color spectrum.[2] The pioneering work won Nakamura, Isamu Akasaki, and Hiroshi Amano the 2014 Nobel Prize in Physics (see Physics Today, December 2014, page 14).

After the blue LED puzzle was solved, researchers in 1998 at Kansas State University proposed the idea of LED miniaturization, termed microLED. In a patent, the researchers outlined the potential usage of microLEDs as bright-light elements for making minidisplays and as detectors or sensors.[3] Unlike LEDs, microLEDs range from single-digit microns to 100 microns in size.[4] MicroLEDs emit light when current is injected by applying a positive voltage on the anode and a negative voltage on the cathode. During that process, electrical energy is converted to optical energy by electronic carriers (electrons and holes) that move through the active semiconductor material, where they recombine radiatively and emit photons.

Improvements to microLED technology came in 2001, when the Kansas State group demonstrated a blue monochrome microLED microdisplay.[5] And 10 years later, the same research group, now at Texas Tech University, used indium gallium nitride and gallium nitride in the first blue and green microLED display with a 640 x 480 resolution and video-graphic capabilities.[6] Since then, microLED brightness, efficiency, lifetime, and manufacturing have advanced considerably, primarily driven by improvements in the material qualities of InGaN

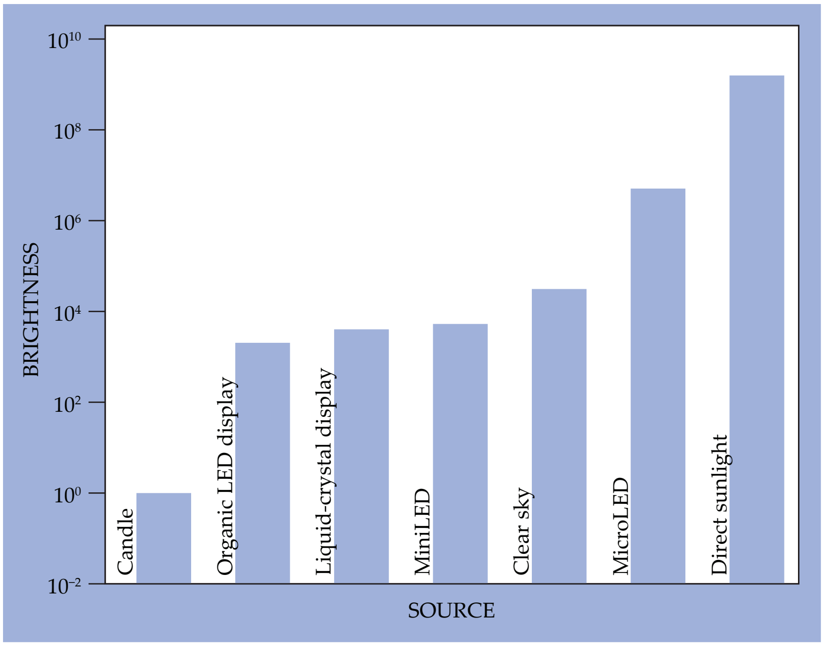


FIGURE 1. MICROLED TECHNOLOGY is brighter than other light-emitter technologies. MicroLEDs are more than two orders of magnitude as bright as clear daylight, which makes them suitable for displays that can be used outside.

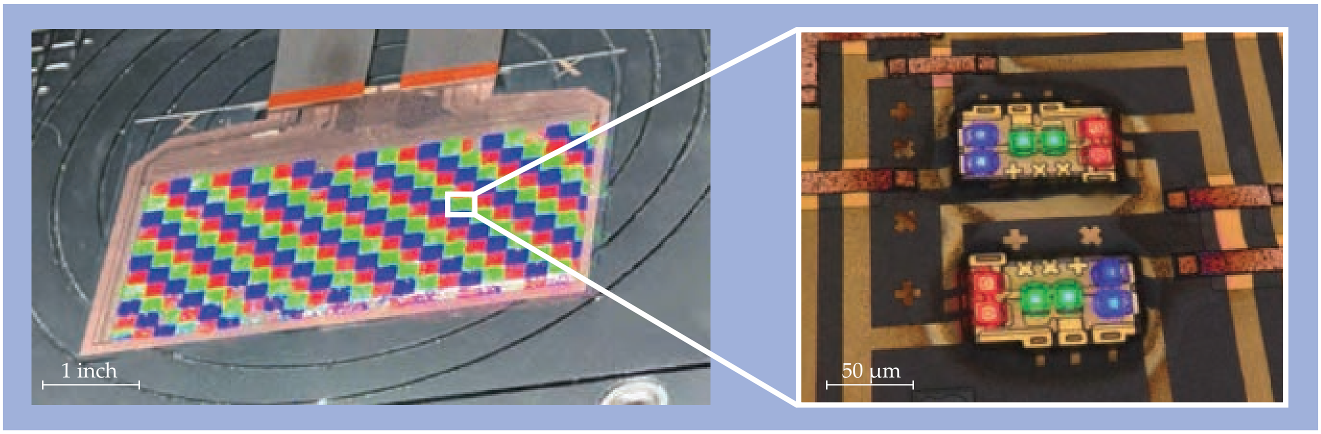


FIGURE 2. THIS RGB DISPLAY, made by X Display Company, is 5.1 inches and has a resolution of 320 × 160 pixels. The inset shows the microLED display's RGB (red, green, and blue) subpixels, which are assembled using an elastomer stamp mass-transfer process. (Adapted from ref. 15.)

and GaN. The display industry is interested in microLEDs because of its growing focus on augmented reality and virtual reality.

Why microLEDs?

The key advantages of microLEDs are their ultrahigh brightness, high efficiency, and long operational lifetimes of more than 100 000 hours.[7] Ultrahigh brightness is particularly relevant for applications in augmented-reality displays that compete with the Sun's brightness in outdoor environments. Figure 1 compares other light-emitting sources with microLEDs, which show brightness capabilities that are three orders of magnitude higher than liquid-crystal displays (LCDs) and organic LEDs (OLEDs). Some of the biggest technology companies, including Meta (with the formation of Reality Labs) and Google (with its acquisition of Raxium in 2022), have put microLEDs at the forefront of next-generation display technology. Other applications include small displays, such as for smartwatches and smartphones; heads-up and infotainment displays in the automotive industry; and pico projectors, which are small, portable projectors that require high brightness.

MicroLED displays are often directly compared with LCDs and OLED displays, but each technology offers its own set of advantages and disadvantages, depending on the specific application. In conventional displays, microLED technology shouldn't be confused with miniLED technology. MiniLEDs provide better contrast and localized dimming zones for traditional LCDs by using many smaller LEDs as backlight sources. MicroLEDs represent a more significant technological leap forward because they offer true self-emission properties. Self-emission microLED displays, like OLED displays, are defined by each pixel intrinsically generating light of its respective color. Self-emission results in true black levels and high contrast ratios because each pixel fully turns off when not in use.

In contrast, LCD technology constantly emits white light and applies color filters to achieve RGB subpixels. Although LCDs are cost-effective, their reliance on a backlight prevents them from achieving high contrast ratios and thin form factors. LCDs and microLED displays are durable and have longer lifespans than OLED displays because they are less susceptible to pixel burn-in. OLED displays, although not as cheap to produce as LCDs, have seen considerable price drops, especially in mobile form factors, and they're still much more cost-effective than microLEDs. Additionally, OLED displays can be built on flexible and conformable substrates for folding and curved displays. Compared with OLED displays and LCDs, microLED displays stand out for their combination of high performance, durability, and energy efficiency.

To produce colors across the entire visible spectrum, RGB subpixels are spaced closely together and programmed with different intensities. When viewing a display from a sufficient distance, the human eye detects the subpixels as one light source, and the individual colors appear to mix. It is desirable to have all three colored subpixels made from the same semiconductor material to simplify manufacturing. The most commonly used materials for making blue LEDs are InGaN sandwiched between layers of GaN. The LED is grown epitaxially on a sapphire substrate: A crystal layer of each material is deposited one atomic layer at a time on a seed layer with a well-defined orientation in specialized deposition chambers.

One measure of manufacturing success is the improvement in external quantum efficiency (EQE), which refers to the ratio of the number of photons emitted to the number of electrons injected into the semiconductor material. Whereas blue microLEDs' EQE can be high at over 40%, achieving the same efficiency with red and green microLEDs has been challenging when using InGaN for dimensions of less than 20 μm. The difficulty in achieving efficient green emission from InGaN microLEDs, termed the "green gap" phenomenon, is primarily because of the reduced crystal quality of the grown InGaN.

Blue to green to red

Light emission from LEDs depends on the active material's bandgap—the energy difference between the material's electronic states, separated into conduction and valence bands. A material's bandgap is determined by crystal-structure

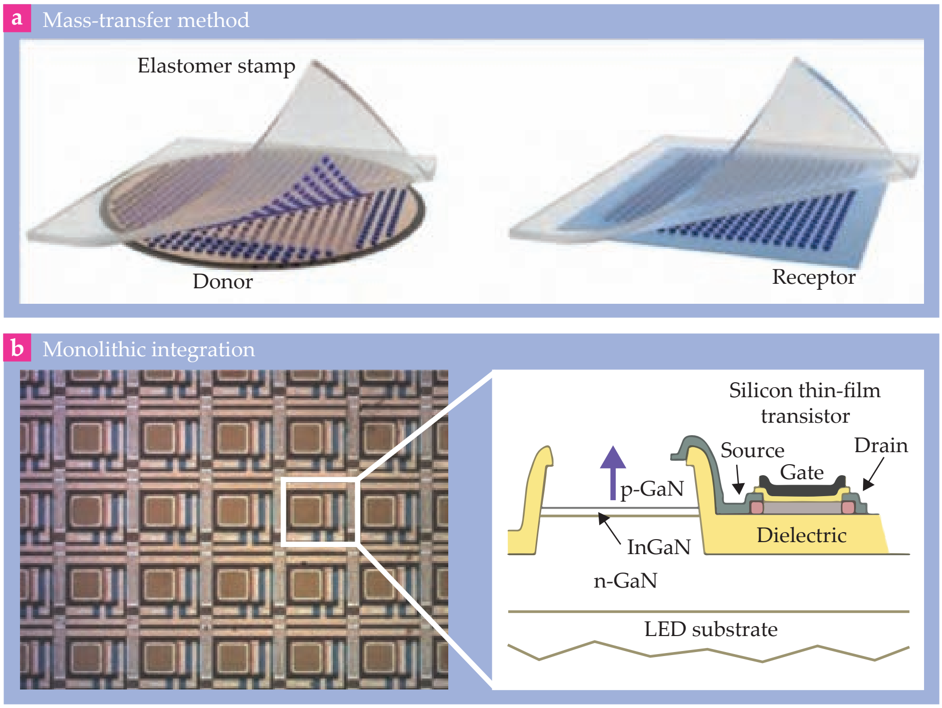


FIGURE 3. FULL-COLOR DISPLAYS made with microLEDs are typically developed with two approaches. **(a)** In the mass-transfer method, an elastomer stamp removes individual microLEDs from a donor substrate and then prints them onto a receptor substrate. (Adapted from ref. 16.) **(b)** For the monolithic-integration approach, a silicon thin-film transistor is fabricated side by side with a microLED onto the same substrate, so no mass transfer is required. The microLED consists of a p-type gallium nitride layer, a layer of indium gallium nitride, and an n-type GaN layer. The purple arrow shows the direction of light emission. The transistor's three terminals—the source, the drain, and the gate—connect to the electronic circuit, and the dielectric layer provides isolation from the GaN layers. (Adapted from ref. 17.)

parameters, such as the lattice constant. The presence of indium in InGaN leads the material to have a higher lattice constant and a smaller bandgap than GaN.

The smaller bandgap results in the emission of lower-energy light, or longer wavelengths. As more indium is added and the bandgap narrows, the color of emitted light shifts from blue to green and eventually to red. The larger lattice-constant mismatch with the underlying GaN layer creates a compressive stress in the active InGaN layer, which causes crystal defects and decreases the efficiency of the light emission from the material. Recent methodologies involving InGaN and GaN nanowires have demonstrated a remarkable closing of the green gap by achieving an EQE that exceeds 25% for green microLEDs.[8]

Similar challenges exist in achieving red emission from InGaN. Compared with green microLEDs, red microLEDs require a higher indium content, which causes an even more significant lattice mismatch. To incorporate more indium into the InGaN layer, the fabrication temperature must be lower, but that leads to higher defect densities and decreases the overall efficiency of the microLEDs. Red microLEDs made using the AlGaInP—rather than the InGaN—material system are more efficient but still undergo nonradiative recombination, in which electrons and holes recombine but do not exhibit light emission. Nonradiative recombination emits energy thermally, and the increased device temperature can further reduce the efficiency of microLEDs if they're not properly heat-sinked.

The efficiency of red microLEDs made from InGaN has improved in recent years. Researchers have achieved an EQE value of about 8% for red microLEDs by employing nanowires of InGaN and GaN, and that improvement has started to close the performance gap with blue microLEDs.[9] Although red microLED efficiency has not yet reached the level of blue microLEDs, research is ongoing, and several technologies—including nanowires, strained quantum-well growth, and double quantum-well technologies—are showing promising improvements.

Given the reduced efficiency of green and red microLEDs and the integration challenges of combining three materials in one display, researchers have developed color-conversion technologies as an alternative solution to achieving full-color displays. One such technique uses quantum dots to obtain green and red colors from blue microLEDs.[7]

Quantum dots are tiny semiconductor particles, typically 2–10 nm in diameter. They are so small that they have unique optical and electronic properties that differ from those of larger particles because of quantum mechanical effects. One of the most critical properties of quantum dots is their ability to emit light of different colors depending on their size, thus making it possible to achieve green and red emissions. (For more on quantum dots, see Physics Today, December 2023, page 16, and the article by Dan Gammon and Duncan Steel, Physics Today, October 2002, page 36.)

Another color-conversion technique uses phosphors. The luminescent substances generally consist of two materials: a host material of wide-bandgap oxides or sulfides and an activator material of transition metals. Phosphors emit light of longer wavelengths when exposed to a radiant energy of shorter wavelengths. Exposing the phosphors to a UV or a blue light source excites the electrons in the material to a higher energy state. The excited electrons emit light of a specific color when they return to the lower energy state. The choice of the activator in the phosphor material determines the wavelength of the light emitted.

For displays, both color-conversion techniques use highly efficient blue microLEDs and down-convert the wavelength to

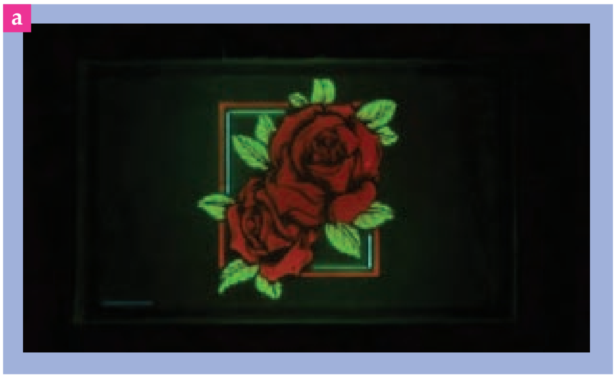


FIGURE 4. MICROLED DISPLAYS can be used in different formats. **(a)** At Display Week 2023, Porotech demonstrated its 0.26-inch 1280 × 720 monolithic microdisplay with dynamic pixel tuning, which enables a full-color display without the need for three distinct subpixels.[14] (Courtesy of Porotech.) **(b)** Jade Bird Display showcased 0.13-inch 640 × 480 monochrome microdisplays. The exceptional brightness is advantageous in augmented-reality and virtual-reality applications. It won a Society for Information Display 2023 Display of the Year award. (Courtesy of Jade Bird Display.)

emit green and red light to achieve full color. Integrating quantum dots or phosphors with displays, however, adds more manufacturing steps and complexity. The overall efficiency of color conversion with quantum dots is diminished—by as much as 50%—because of optical loss caused by inefficient photon travel.[10] Color-conversion techniques can also be limited by cross talk to adjacent subpixels, which leads to color inaccuracies and blurred images.

Full-color integration

Manufacturing a display out of micron-scale light sources requires assembling millions of pixels on a backplane—the electronic circuitry for the logic and the driving current. A full, high-definition, 1920 × 1080 display needs about 6 million microLEDs. But the display technology is unforgiving: Even a single dead pixel is visible to the end user, so an exceptionally high yield is required to make a fully functioning display. The method for assembling microLED displays should be both fast and accurate on an industrial scale. With those factors in mind, the industry has developed two approaches: mass-transfer technology and monolithic integration, both of which have their advantages and challenges.

Mass-transfer technology is more suitable for larger mobile displays, computer monitors, and digital signs. Individual RGB subpixels are picked up and transferred from the native donor substrate to a target substrate. The method provides the freedom to choose a substrate more suitable for the driving backplane. Using different donor substrates enables the use of different material systems for RGB subpixels such that each subpixel is matched to its most efficient microLED material. That approach eliminates the need for color-conversion layers, such as quantum dots or phosphors.

Perhaps the most successful approach to mass transfer uses a stamp to move RGB subpixels to a driving backplane.[11] Figure 2 shows a microLED display fabricated using the mass-transfer method. MicroLEDs are first fabricated on top of a sacrificial layer that is subsequently dissolved, which leaves the microLEDs suspended above an air gap and held to the substrate via thin, breakable tethers. A transfer mechanism—such as an elastomer stamp or printhead—uses van der Waals forces, suction, or adhesives to remove the microLEDs batch by batch from their donor substrate.

The stamp or printhead then moves the microLEDs to the target substrate, where they are aligned and attached at the desired location. The target substrates establish electrical contact with the transferred microLEDs through the use of metal layers that are deposited with conventional lithography processes. Researchers have implemented robotic pick-and-place and roll-to-roll techniques to achieve high-resolution displays with yields that are good enough for industrial scales, although the costs may still be high.[1]

Another method for mass transfer uses fluidic self-assembly.[12] It's a process in which microLED subpixels are assembled onto the target substrate under the influence of fluidic forces. The concept is based on the principle that complementary components will spontaneously assemble into stable structures when they are brought into contact with each other in a fluid environment. Researchers have used the process to assemble GaAs LEDs onto a silicon backplane.[11] The method is simple to implement, low cost, and scalable, but obtaining high yields and assembling each of the RGB subpixels together into one pixel is challenging. Figure 3a shows a schematic of the elastomer stamp technique.

Homegrown pixels

Rather than move microLED pixels from one substrate to another, monolithic-integration techniques address microLED pixels directly on their native substrate. The driving circuitry—that is, the electronics used to manipulate the display pixels—are made available to the native LED substrate without any need for transferring individual pixels. MicroLED pixels with sizes as small as a few microns can make extremely pixel-dense microdisplays, with more than 5000 pixels per inch. Three major approaches are available to implement monolithic integration: microLED epitaxial growth on silicon, transistor fabrication on a microLED epitaxially grown on sapphire, and flip-chip bonding of a microLED substrate to CMOS chips. Figure 3b shows a monolithic-integration method that involves the fabrication of a thin-film transistor alongside the microLED pixel, all on a single substrate.

Although microLEDs that are grown on their native substrates produce the most efficient light emitters, they can also be grown on silicon substrates, and efficient blue LEDs can be color converted to obtain other colors. MicroLEDs grown on silicon are ideal for backplane fabrication because of the maturity of transistors built around silicon materials.[13]

The quality of microLEDs grown on silicon and their emission efficiency has traditionally been poor because of the large lattice mismatch. Using buffer layers such as AlN between GaN and silicon has improved their efficiency. But until microLEDs grown on silicon exhibit efficiency improvements that match that of their native substrate counterparts, they will not be tenable for product deployment. Another strike against microLEDs grown on silicon is that other techniques, such as selected-area epitaxy and strained quantum wells, can produce RGB pixels natively on one substrate without the need for color conversion.

For microLED displays on sapphire substrates, the driving circuits—which are needed for selecting desired pixels in a sequence, also known as pixel addressing—are implemented by fabricating thin-film transistors on microLEDs. Materials such as amorphous silicon and indium gallium zinc oxide are used as the semiconductor layer for fabricating the necessary transistors. But it remains challenging for researchers to achieve uniform electrical characteristics that are as good as those made with CMOS technology. Additionally, such displays are inherently monochromatic or require color-conversion techniques to achieve RGB pixels as complex as the ones typically made with blue microLEDs.

Recent progress in porous GaN technology has paved the way for creating dynamically tunable pixels. With porous GaN, formed by electrochemical etching, higher amounts of indium can be incorporated into the InGaN crystals of the microLEDs because of a decrease in the strain of the lattice. That enables efficient red emission with material systems such as InGaN and GaN. A pixel that's tuned for color with porous GaN technology emits a spectrum of wavelengths ranging from blue to IR. Color tuning eliminates the requirement for multiple subpixels of distinct colors to be grown and subsequently transferred. At Display Week 2023, the company Porotech demonstrated, using porous GaN, the first monolithically integrated, single-panel, and full-color microdisplay.[14]

Another approach for integration is fabricating GaN on sapphire and then bonding it to a conventional CMOS backplane by using such techniques as flip-chip bonding and thermocompression. Flip-chip bonding uses metal bumps to connect two electronic devices. It's appropriate for small-size displays with higher pixel density, such as smartwatches and augmented- and virtual-reality displays.

For active-matrix addressing—in which the individual pixels are connected to a transistor and controlled by applying voltages to it—each pixel needs a bond site on the microLED terminal and another on the CMOS backplane. It's effective for high-resolution displays, but the increased complexity of the bonding methods makes it less suited for mass-scale production. The simpler passive-matrix addressing connects pixels in rows and columns and then applies the voltage directly to the entire structure. That setup is easier to fabricate but lacks the high refresh rate needed in high-resolution displays. Still, passive-matrix addressing is useful for lower-resolution displays, which makes it a popular method to implement in academia for demonstration purposes.

What's next for microLEDs?

In addition to the scaling, packaging, and driving challenges, another issue microLEDs face is the relatively high cost of the source materials and of their subsequent fabrication processing. One solution is to move to larger wafer sizes. GaN is often grown on sapphire, which yields reasonably high LED performance but does not scale to substrate areas larger than about 200 mm. Significant recent work has been applied to the growth of high-performance GaN microLEDs on silicon substrates, which allows for 300 mm wafers to be built and processed.

The higher production efficiency per unit area in LED growth and semiconductor processing offers a road map: A significantly improved cost per LED can help lead to mass-market applications for microLED technologies. Figure 4 shows some microLED displays for different formats and uses that have been made at an industrial scale.

Although microLED displays are in their infancy, many of the technology's technical advantages, such as luminance, lifetime, color quality, and device scaling, have been demonstrated. As one would expect in a maturing technology, the commercial challenges have now transitioned to issues of cost and of scaling the manufacturing process to industrial levels. The future for microLED technologies is bright, and soon we expect to see microLEDs in many display applications, including augmented- and virtual-reality headsets, smartwatches, and smartphones.

REFERENCES

1. A. Wilcott, "Virtual Boy review," *Micro-64* (1 July 2014).
2. Y. Nanishi, *Nat. Photon.* **8**, 884 (2014).
3. H. Jiang et al., "Micro-size LED and detector arrays for minidisplay, hyper-bright light emitting diodes, lighting, and UV detector and imaging sensor applications," US Patent 6,410,940 (25 June 2002).
4. K. Behrman, I. Kymissis, *Nat. Electron.* **5**, 564 (2022).
5. H. X. Jiang et al., *Appl. Phys. Lett.* **78**, 1303 (2001).
6. J. Day et al., *Appl. Phys. Lett.* **99**, 031116 (2011).
7. V. Kumar, I. Kymissis, *Appl. Phys. Rev.* **10**, 021306 (2023).
8. A. Pandey et al., *Nano Lett.* **23**, 1680 (2023).
9. A. Pandey et al., *Appl. Phys. Lett.* **122**, 151103 (2023).
10. B. Karadza et al., *Nano Lett.* **23**, 5490 (2023).
11. S.-I. Park et al., *Science* **325**, 977 (2009).
12. H.-J. J. Yeh, J. S. Smith, *IEEE Photonics Technol. Lett.* **6**, 706 (1994).
13. C. J. Hun et al., *Appl. Phys. Express* **13**, 026501 (2020).
14. T. Zhu, *Photonics Spectra*, May 2023, p. 38.
15. C. A. Bower et al., *IEEE J. Sel. Top. Quantum Electron.* **29**, 2000111 (2023).
16. J.-E. Ryu et al., *Adv. Mater.* **35**, 2204947 (2023).
17. B. R. Tull et al., *SID Symp. Dig. Tech. Pap.* **46**, 375 (2015).

PT

Science and Industry

Scientists for scientists: an approach to make a difference

Stories about spin-offs whose technologies change the course of science are inspiring to both researchers and entrepreneurs. But why does the next step – operating at scale while keeping an innovative edge – matter to scientists? Sadik Hafizovic and Andrea Orzati share some insights.

Sadik Hafizovic co-founded Zurich Instruments in 2008 as a spin-off of the Swiss Federal Institute of Technology (ETH) that launched novel lock-in amplifiers. Under his leadership, the company grew into a leader of test & measurement and quantum computing instrumentation. Andrea Orzati worked as an ASIC designer at u-blox, and in commercial and executive roles at Sensirion. He succeeded Sadik as CEO of Zurich Instruments in May 2024.

ANDREA ORZATI (right) takes over the management of Zurich Instruments from its co-founder and previous CEO, Sadik Hafizovic.

What motivates you in your career?

SADIK: I'm a technical person and I get inspired when complicated systems work to the benefit of mankind. But I was also always intrigued by entrepreneurial adventures; to take ideas to practical products.

ANDREA: By working in several start-ups, I saw the frontiers of different technologies. For me, the most exciting part is to help grow innovative people and products.

What is your advice to scientists moving to industry?

SADIK: Don't be afraid of competition, the world outside of science is vast, there is a place for everybody, and find the spot you can best make an impact in.

ANDREA: Figure out what you are excited about; developing new things, making them widely accessible, or both. Connect with people in the industry and explore industry-academia collaborations!

What did it take to start a company?

SADIK: My PhD research was on building microelectrode array interfaces for natural neuronal networks. I onboarded students in summer and diploma projects to explore new lock-in amplifiers needed for the research. This side project received my full attention during my postdoctoral period. Together with my co-founders, Flavio Heer and Beat Hofstetter, we commercialized the project and established Zurich Instruments.

ANDREA: Always focusing on what is possible and reasonable does not necessarily get you to the best results. In a high-tech industry, we need culture that balances being reasonable (market-driven) and being disruptive (technology-driven). In 2018, this enabled Zurich Instruments to launch the first commercially available quantum controlling system, while continuing the development and production of lock-in amplifiers.

Why would "scaling" matter to scientists?

SADIK: I see two challenges: the "scale-up", and the "scale-out". The scale-up is growing the company, inventing new instruments, and improving them. I feel most at home in the scale-up: finding new ways to increase the customer value of our lock-in amplifiers and developing quantum computing control systems that control ever more and better quantum bits. In the scale-out, aspects like quality and footprint come into play. Zurich Instruments is pushing the limits of technology and provides instrumentation in a consistent and reliable way to thousands of customers. This is where innovative companies differ from truly outstanding companies. With the support of Rohde & Schwarz, as well as our internal organization, we master both scaling challenges.

ANDREA: Scaling is also about increasing responsibility for people and society. People at Zurich Instruments continuously acquire new skills and roles while getting involved with scientific activities such as MuniQC-SC, SuperMOOSE, or SuperQuLAN. So, when transferring to industry, evaluate the potential for your growth, and how this brings value to scientific communities.

How much can a company grow and still maintain a relationship with science?

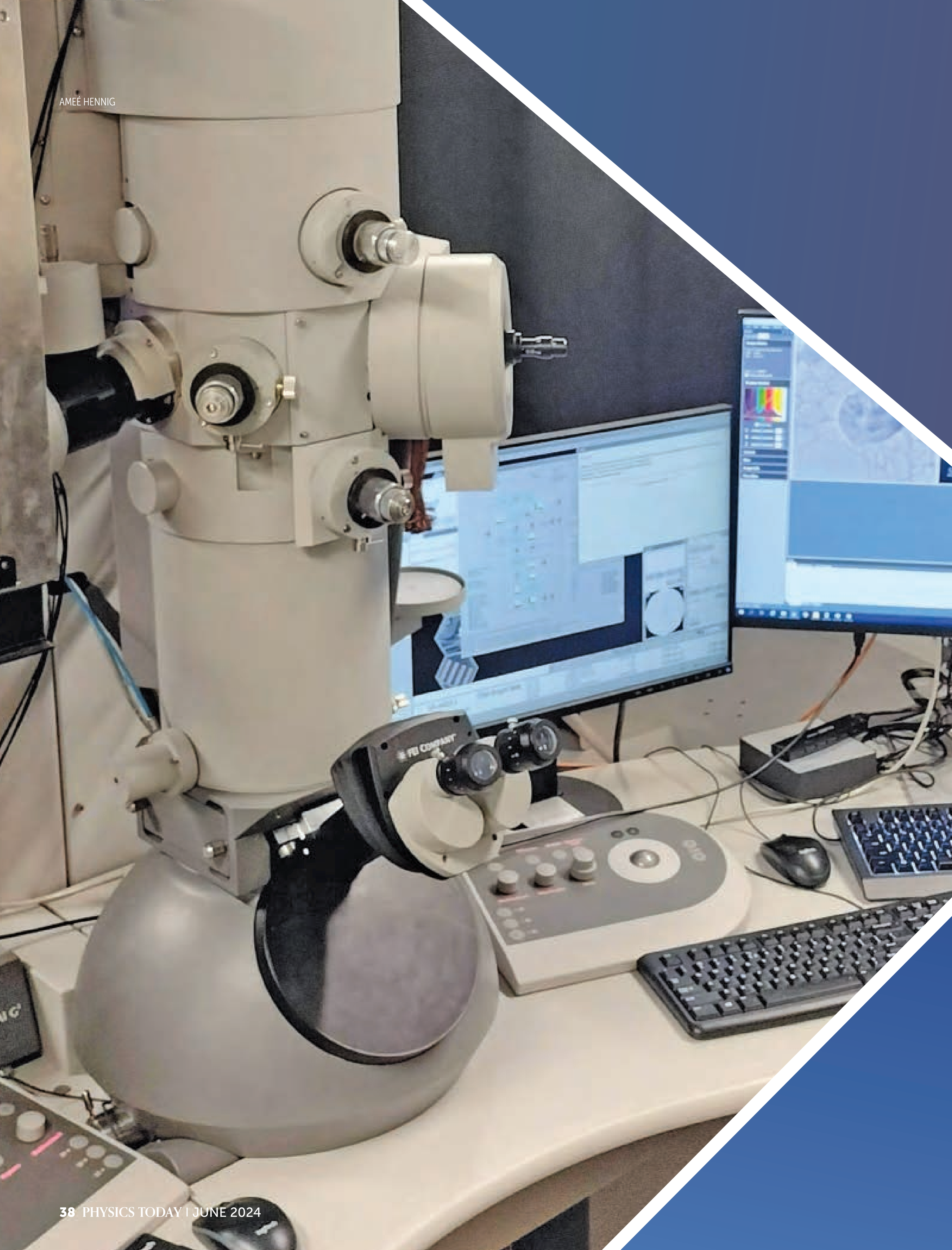
ANDREA: Fantastic success over the last sixteen years comes with a responsibility to serve a large customer base, enable our worldwide teams, and support more scientific collaborations. With Rohde & Schwarz, the company has all the resources to effectively accelerate the second quantum revolution.

SADIK: Achieving the right balance between innovation, efficiency, and effectiveness demands excellence in developing the organization, a task I don't know anyone more suitable for than Andrea.

ANDREA: We are a company of scientists for scientists. It is fantastic to see thousands of citations of Zurich Instruments in scientific papers. The world of science and technology is moving quickly, and we continue to be a reliable partner amidst many uncertainties. As in the past, we continue to deliver – at all scales.

Zurich Instruments

AMEÉ HENNIG



Mohammed Hassan is an associate professor of physics at the University of Arizona in Tucson. (Photo courtesy of Ameé Hennig.)

ELECTRON MICROSCOPY FOR ATTOSECOND SCIENCE

Mohammed Hassan

Some of the fastest processes in physical and biological systems can be studied by generating ultrashort electron pulses.

The transmission electron microscope (TEM) is one of the most powerful imaging instruments. It works similarly to an optical microscope, but instead of using visible light, a TEM generates an accelerated electron beam that passes through a thin sample. The transmitted beam then interacts with a set of lenses that magnifies the sample image onto a camera detector. Because of diffraction limits, an optical microscope's resolution can be no better than one-half the wavelength of light, or about 200 nm. But accelerated electrons in a TEM have much shorter wavelengths, making it possible to see 3D, atomic-scale features.

After the invention of the TEM in the early 1930s, for which some of its inventors were awarded the 1986 Nobel Prize in Physics, its imaging capabilities have been expanded significantly. In the TEM family of techniques, electron diffraction can be used to image crystallographic structure at subatomic resolution. Electron energy-loss spectroscopy (EELS) allows for imaging a sample's core electronic structure and mapping its chemical elements. And cryoelectron microscopy—the development for which the 2017 Nobel Prize in Chemistry was awarded—enables researchers to determine the high-resolution structure of biomolecules in a cryogenic state (see Physics Today, December 2017, page 22).

On the temporal-resolution front, the TEM is limited by the video camera's millisecond recording rate. The ultrafast electron microscope (UEM) overcomes that limit by using the TEM in a pump–probe approach. In that case, the camera's recording rate becomes irrelevant to the temporal resolution, and the UEM is limited only by the duration of the ultrafast electron pulse.

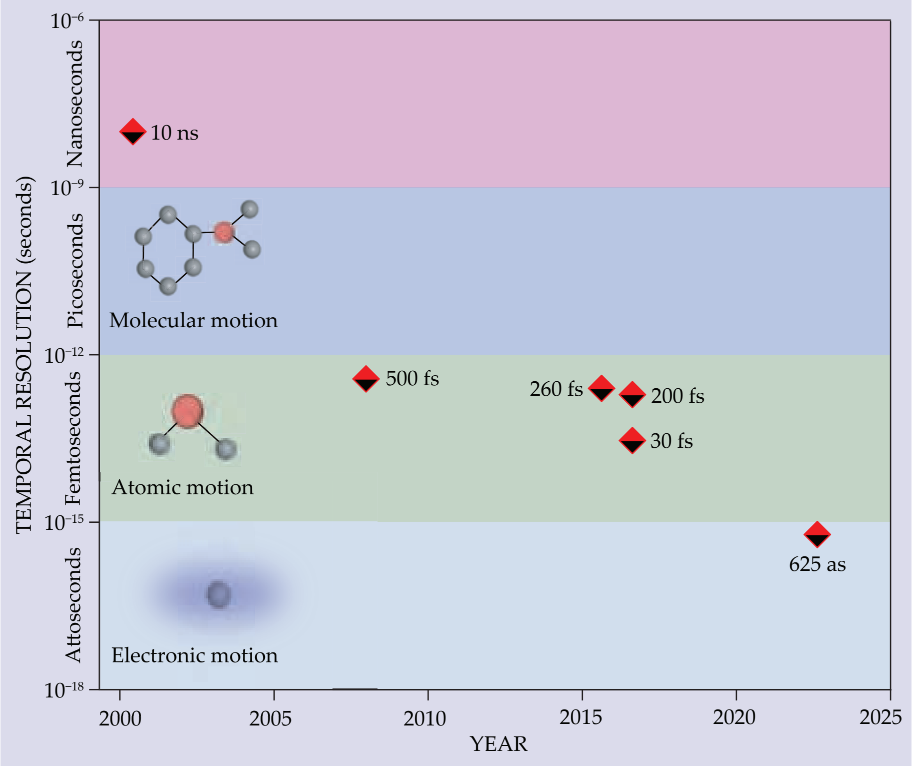


FIGURE 1. TEMPORAL RESOLUTION ENHANCEMENTS. Over the past several decades, the time scale at which the ultrafast electron microscope can operate has improved several orders of magnitude, from nanoseconds to hundreds of attoseconds. The developments open the door for imaging ultrafast molecular, atomic, and electron motions in action. (Sources, from longest: ref. 6; B. Barwick et al., *Science* **322**, 1227, 2008; T. T. A. Lummen et al., *Nat. Commun.* **7**, 13156, 2016; ref. 4; ref. 12; ref. 15.)

The UEM is a tool that crosses disciplinary boundaries. It provides access to studying electronic, atomic, and molecular systems and their ultrafast dynamical processes, some of which range from picoseconds to a few hundred attoseconds.

Ultrafast electron microscopy

A typical ultrafast stroboscopic pump–probe measurement can capture processes that take place at the scale of hundreds of femtoseconds. The dynamics of a sample under study are triggered by the ultrafast pump pulse. Snapshots of the dynamics at different instants can then be recorded using ultrafast probe pulses that arrive at a sample at slightly different times after the triggering pump pulse. For the UEM, the pump and probe pulses are laser and electron pulses, respectively. To develop the UEM to image ultrafast processes, the design of the TEM needed modifications.

The continuous electron beam inside a TEM is usually generated from thermionic emission: Similar to how light is emitted from an incandescent bulb, electrons are generated by heating a filament. The UEM, to produce the ultrafast electron pulses it needs, replaces thermionic emission with photoemission. In the UEM, a photocathode of lanthanum hexaboride, when illuminated by UV laser pulses, will emit a beam of electron bursts, and the electron pulses have a duration comparable to that of the UV laser pulses.

The resulting temporal resolution of the UEM is a few hundred femtoseconds to a picosecond, defined mainly by the duration of the electron probe pulse, the duration of the pump pulse, and the timing jitter between the electron and laser pulses. To maintain the UEM's temporal resolution, the jitter must be shorter in time than the duration of the electron pulse.

The duration of ultrafast electron pulses increases as the electrons travel farther from the photocathode to the sample. That's because of the space-charge effect: As the electrons in the wavepacket move through the microscope, Coulomb repulsion among them causes some to travel faster than others and results in a longer pulse. In the UEM, the distance from the photocathode to the sample is fixed, which limits the temporal resolution to no better than a few hundred femtoseconds.

To overcome the space-charge effect, recent efforts have focused on implementing techniques that could compress the electron pulses and thus enhance the temporal resolution of the UEM.[1] One proposed approach is RF compression, which is based on slowing down the faster electrons and speeding up the slower ones. The output pulse, therefore, would have its electrons traveling closer to the same speed. The compressed electrons that exits the RF compressor would travel to the sample in a pulse that's shorter than a few hundred femtoseconds.

A proposed alternative to RF compression would use a deflection cavity inside the TEM. An electron beam passing through the cavity could be chopped into pulses of a few hundred femtoseconds for the UEM.[1]

Figure 1 shows the improvement in temporal resolution since the first UEM was demonstrated in 2000. The catch is that although the spatial resolution of the TEM can reach subangstrom scales,[2] the space-charge effect ruins the spatial resolution of the UEM, which when uncorrected, is on the order of a few tens of nanometers. Many research groups are working to improve the spatial resolution by developing brighter photoemission electron sources.

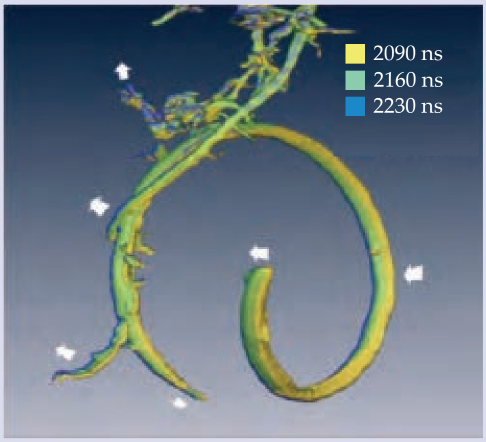


FIGURE 2. A CARBON NANOTUBE, roughly shaped like a ring, wiggles over a time scale of nanoseconds. Each color corresponds to a different snapshot in time, with white arrows indicating the local direction of motion. It's one example of the imaging applications that have become possible because of advances in ultrafast electron microscopy. (Adapted from ref. 7.)

One of the most promising techniques uses a laser-driven emission gun: It produces a bright electron beam by illuminating with laser pulses a nanoscale tip. In another technique, adopted from conventional TEMs, cold field-emission guns that use the apex of a nanoemitter have been utilized in UEMs to generate bright electron pulses.[1,3] In both approaches, more-coherent electrons are generated by the photoemission process because it is localized at a nanotip emitter. In addition, researchers can control the photoemission process to be in the linear regime—in which each electron is emitted as a result of single-photon absorption from the UV pulse. That reduces the space-charge effect because all the electrons are emitted simultaneously and have the same energy distribution. The electron pulses that are generated using those techniques, therefore, have better spatiotemporal resolution than the typical UEM, which uses a lanthanum hexaboride source.[4] Because of the UEM's advancement, many physical systems can be imaged at high temporal resolution.

Ultrafast imaging applications

UEM measurements connect ultrafast dynamics in a sample with the structural morphology of the sample.[5] The UEM, therefore, is a valuable tool in ultrafast science and has found different applications in chemistry, physics, materials science, and biology. In 2000, for example, a nanosecond electron microscope captured the laser-induced melting of a nickel–phosphorus metal thin film and the changing morphology of bulk cobalt in real space over time.[6] The images showed ultrafast behavior in the metals that wasn't easily seen before with more typical optical pump–probe techniques. Moreover, the UEM has been used for imaging laser-triggered motion, such as the wiggling of ring structures of carbon nanotubes,[7] as shown in figure 2.

At the femtosecond time scale, the UEM has been used to look at ultrafast demagnetization.[8] In another study, the magnetic dynamics of bilayer films composed of iron and silicon nitride were probed, and the researchers demonstrated that stable magnetic nanoscale networks of vortices and antivortices could be produced.[9]

Another imaging technique in the UEM family is EELS, which provides access to the electronic structure of matter at the valence and core levels with nanometer–femtosecond spatiotemporal resolution. The electronic-structure dynamics can be retrieved by tracing the EELS spectra as they change over time. One of the applications is the study of the photoinduced chemical-bonding dynamics in graphite. Researchers have found that even when the graphite's crystal lattice contracts, some of the carbon–carbon bonds lengthen, which could explain various properties of the 2D system.[10]

The applications of ultrafast electron microscopy extend to the study of dynamics in aqueous solution through the development of liquid-cell technology. Demonstrations include imaging the photoinduced rotational motion of gold nanoparticles in the liquid phase. With that development, the UEM could be helpful in studying ultrafast dynamics of biological systems in their native environments. In another example, cryo-electron microscopy was adopted into a UEM system to image picometer movements of photoresponsive biological structures embedded in ice. All the examples mentioned thus far have focused on processes that are no faster than hundreds of femtoseconds. To resolve still-faster processes, researchers had to exploit the subfemtosecond laser pulses of a recently developed microscopy imaging technique to generate equally short electron pulses.

Photon–electron coupling

In a pioneering experiment published in 2009, the interaction between laser photons and free electrons was imaged by a UEM.[11] The photons that scatter off a nanostructure surface can exchange momentum with the microscope's ultrafast pulse of free electrons, which lose and gain energy in integer multiples of photon quanta. The photon–electron coupling is resolved in the electron energy spectrum, which consists of discrete peaks that are spectrally separated by multiples of the photon energy. The approach, known as photon-induced near-field electron microscopy (PINEM), opens the door for various microscopy applications.[10]

The PINEM technique, however, has another use. Recall that the temporal resolution of the UEM depends on the electron pulse duration inside the microscope, which has been limited to no better than a few hundred femtoseconds because of the space-charge effect. PINEM can break that limit and generate even-shorter electron pulses by what's known

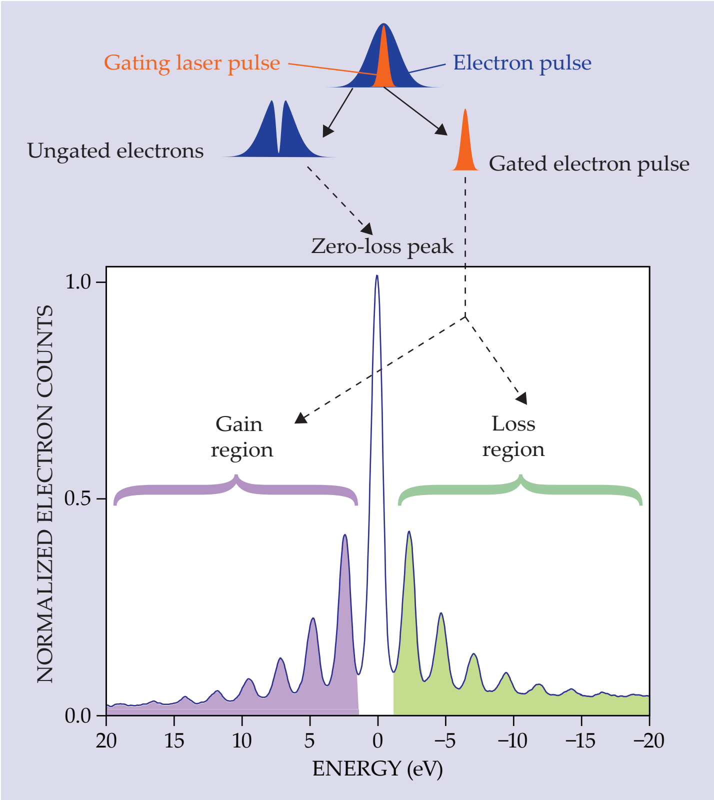


FIGURE 3. MAKING ULTRAFAST ELECTRON PULSES. In photon-induced near-field electron microscopy, an ultrafast laser pulse interacts with an electron pulse such that when the laser photons scatter off a nanostructure surface, they can exchange momentum with free electrons. The energy of electrons that interact elastically is at the zero-loss spectral peak. But the electrons coupled to photons have discrete electron energy peaks spectrally separated by multiples of the photon energy. Electrons that have gained and lost energy are correlated to the laser pulse and thus share its ultrafast temporal profile. That optical gating approach has been used to generate ultrashort electron pulses of a few hundred femtoseconds. (Adapted from ref. 13.)

as optical gating,[5,12,13] as shown in figure 3. In PINEM, the electrons in a pulse that gain or lose photons because of photon–electron coupling are separated in the energy domain and are created only during the laser pulse. Those gated electrons, therefore, have a temporal profile intrinsically similar to the laser pulse despite the long duration of the main electron pulse.

The gated electron pulses can be used in a UEM to image dynamical processes that occur faster than hundreds of femtoseconds. The approach has been demonstrated, for example, to generate 30 fs electron pulses inside a microscope[12] and to image the phase transitions of vanadium oxide nanoparticles.[13]

Attosecond electron microscopy

The optical-gating approach, in combination with laser pulses of a few hundred femtoseconds, has been exploited to generate a train of attosecond electron pulses inside a microscope. The laser pulses each consist of hundreds of half-cycle laser pulses, and each half-field cycle can optically gate electrons. That means that the gated electrons can form a train of electron pulses, and each pulse is confined to the subfemtosecond time scale for as long as the carrier frequency of the laser pulse is suitably high, in the visible or near-IR spectral region. In addition, a continuous-wave laser beam has been used to generate a train with an essentially infinite number of attosecond electron pulses. That approach has limited utility, though, and has been used for imaging only the periodic oscillations of scattered light.[14]

As mentioned earlier, the temporal resolution of the UEM depends on the electron pulse's duration, which defines the time window in which a dynamical process is probed. For a train of n electron pulses, the dynamical process will be probed n times during one pumping event. The recorded snapshots can be stacked to form a video of the average dynamics, as illustrated in figure 4a.

That sort of movie, however, won't show the evolving motion produced by the pumping event in real time. Instead, the train of attosecond electron pulses is limited to imaging repeatable dynamics, such as the periodic oscillations of scattered light. The temporal resolution of the UEM remains the same as the temporal profile of the entire train of pulses. To resolve nonperiodic attosecond dynamics, such as electronic and atomic motions, a single attosecond electron pulse must be generated inside the microscope (see figure 4b).

In 2023 my research group generated a single attosecond electron pulse for imaging, which we call attomicroscopy by the polarization-gating approach.[5,15] We used it to probe the electron motion of carbon atoms in multilayer single-crystal graphene.[15] In polarization gating, a series of waveplates varies the laser's polarization from linear to circular and then back to linear. At the instant of linear polarization, a strong, isolated attosecond pulse in the extreme UV can be generated via high-harmonic generation.[16] (For more on high-harmonic generation, see Physics Today, December 2023, page 13. To learn about laser polarization and recollision theory, see the article by Paul Corkum, Physics Today, March 2011, page 36.)

We used the polarization-gated laser pulse as an optical-gating pulse inside the microscope to generate the attosecond electron pulse. But polarization gating isn't the only way to generate attosecond electron pulses. An optical attosecond pulse, which we demonstrated with a light-field synthesis method,[17] can also be used for attosecond gating of electrons inside the microscope.[5,17,18]

Regardless of the method, the significance of generating attosecond temporal resolution in the electron microscope is the ability to image electron motion dynamics in real time. Although the spatial resolution of the UEM is still limited because of the space-charge effect, some indirect imaging approaches—such as attosecond time-resolved electron diffraction and EELS—can sidestep the space-charge effect and

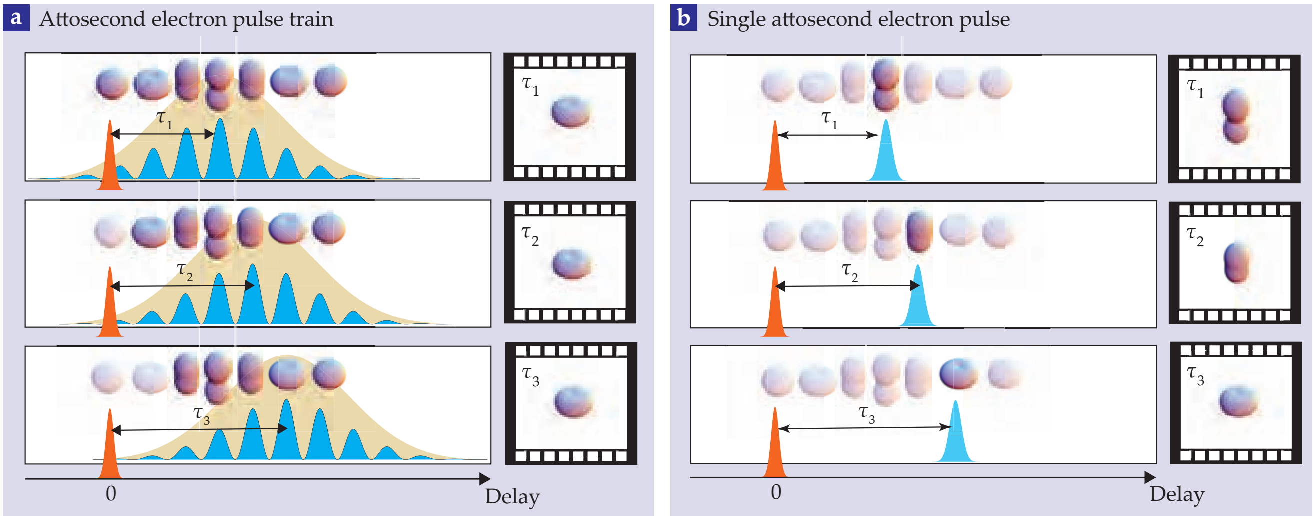


FIGURE 4. ATTOSECOND PULSES. (a) A femtosecond electron pulse (yellow) is optically gated by a laser pulse to generate a train of attosecond pulses (blue), as described in figure 3. After a laser pump (red) excites electron motion in an atom (illustrated with the geometric shapes), the electron pulses probe the ultrafast dynamical motion, seven times in this case. But because the duration of the entire attosecond pulse train is longer than the time scale of a particular electron dynamical event, the averaged image looks the same at each time delay τ. **(b)** A single attosecond pulse is so short that it can image specific electron dynamical motions and resolve the changes at different time delays as the motions evolve in real time.

trace the electron motion in reciprocal space and the movement of valence electrons, respectively.

One of the most important advantages of attomicroscopy is the possibility of studying the electron dynamics of matter in neutral systems, which is difficult to do with other attosecond tools, such as extreme-UV laser pulses, because of the high photon energy and strong electric-field interaction that they require. Another advantage is that multiple measurements can be made on the same sample in one experiment: Attomicroscopy can observe ultrafast dynamics, and the normal TEM mode can directly image and characterize the sample's morphology in real space. Thus I anticipate that attomicroscopy could be a powerful imaging tool for research in the development of nanoscale optoelectronics.

Viewing the microscopic world

Over the past two decades, the development of the UEM and the related technical advancements in spatiotemporal imaging resolution have led to many new findings in physics, chemistry, materials sciences, and biology. The full effect of the UEM on science and technology, however, remains unknown.

With the development of the optical-gating approach, I am optimistic that many groups will use the technology to study the ultrafast dynamics of a few tens to hundreds of femtoseconds. The new capability of attomicroscopy to probe field-induced electron motion dynamics would extend the capability of ultrafast electron microscopy to see the microscopic world of molecules, atoms, and electrons and their motion in real time and space.

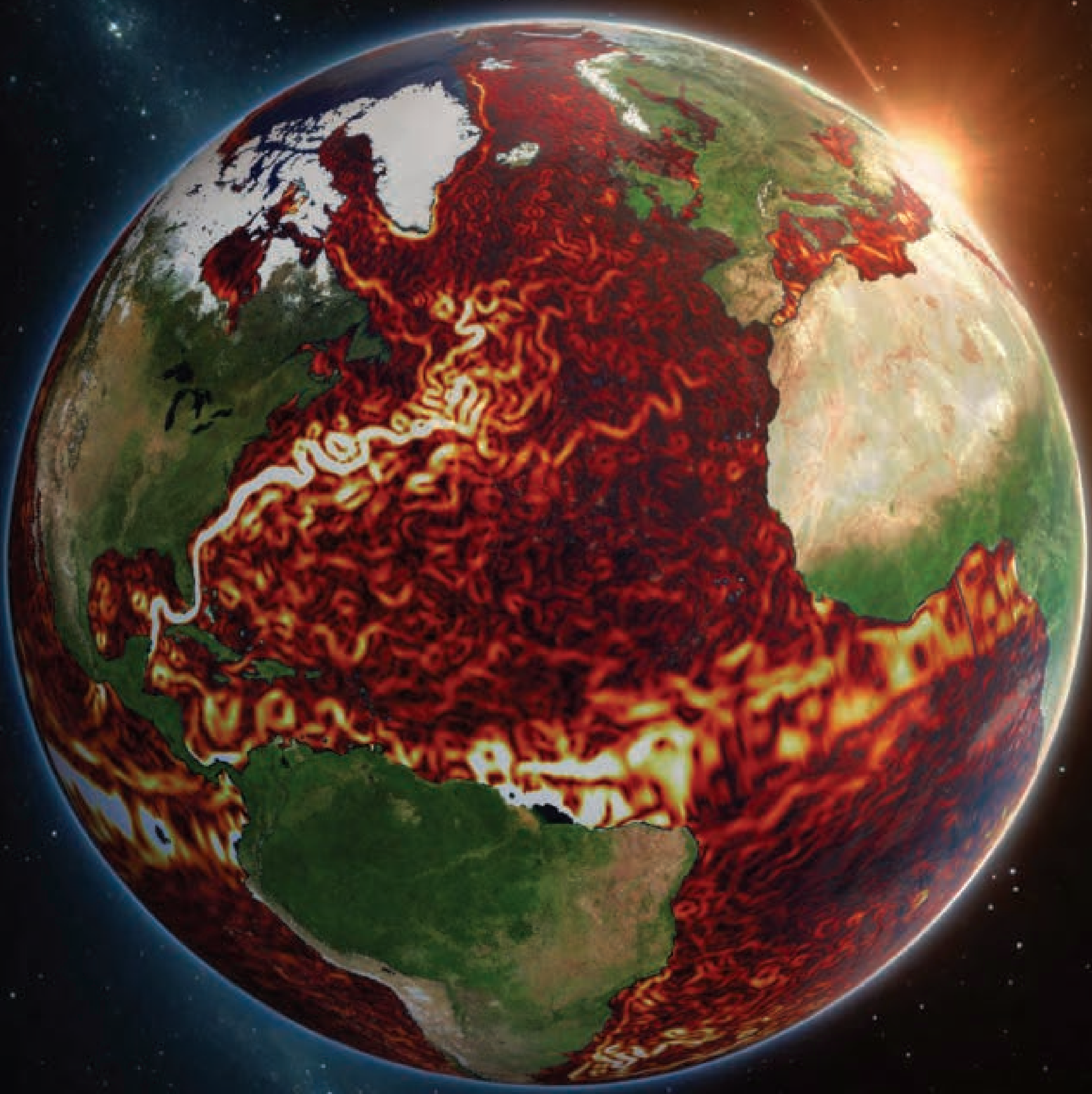
Furthermore, attomicroscopy could be used in combination with cryoelectron microscopy to study electronic responses in biological systems, such as electron tunneling in DNA. Such imaging would help clarify the mechanism of electron tunneling between DNA strands and how it could change the structure of DNA. Attomicroscopy imaging, therefore, could be important to understanding carcinogenesis and mutagenesis mechanisms and the processes that repair damaged DNA.

Attosecond electron microscopy is already helping researchers investigate other biological phenomena, such as the electron movement that takes place between chains of cells through nanosized cable bacteria. Imaging the cables at attosecond temporal resolution could provide valuable insight for the design and construction of medical-compatible nanodevices. If implemented in the human body, the devices may be able to help detect, diagnose, and treat various diseases.

REFERENCES

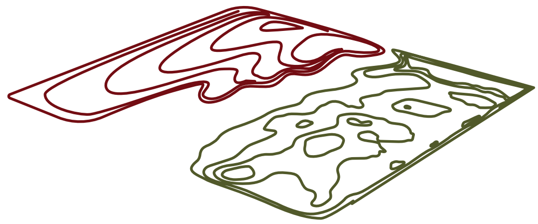
1. S. Borrelli et al., in *Structural Dynamics with X-ray and Electron Scattering*, K. Amini, A. Rouzée, M. J. J. Vrakking, eds., Royal Society of Chemistry (2024), p. 557.
2. P. E. Batson, N. Dellby, O. L. Krivanek, *Nature* **418**, 617 (2002).
3. G. M. Caruso et al., *Appl. Phys. Lett.* **111**, 023101 (2017).
4. A. Feist et al., *Ultramicroscopy* **176**, 63 (2017).
5. M. T. Hassan, *J. Phys. B* **51**, 032005 (2018).
6. O. Bostanjoglo et al., *Ultramicroscopy* **81**, 141 (2000).
7. O.-H. Kwon, A. H. Zewail, *Science* **328**, 1668 (2010).
8. K. B. Schliep et al., *Appl. Phys. Lett.* **110**, 222404 (2017).
9. T. Eggebrecht et al., *Phys. Rev. Lett.* **118**, 097203 (2017).
10. E. Pomarico et al., *MRS Bull.* **43**, 497 (2018).
11. B. Barwick, D. J. Flannigan, A. H. Zewail, *Nature* **462**, 902 (2009).
12. M. T. Hassan et al., *Nat. Photonics* **11**, 425 (2017).
13. M. T. Hassan et al., *Proc. Natl. Acad. Sci. USA* **112**, 12944 (2015).
14. D. Nabben et al., *Nature* **619**, 63 (2023).
15. D. Hui et al., https://arxiv.org/abs/2305.03014.
16. G. Sansone et al., *Science* **314**, 443 (2006).
17. M. T. Hassan et al., *Nature* **530**, 66 (2016).
18. D. Hui et al., *Nat. Photonics* **16**, 33 (2022); H. Alqattan et al., *APL Photonics* **7**, 041301 (2022). PT

DEEP CONVECTION DRIVES OCEANIC OVERTURNING



(IMAGE CREATED BY KAUSHIK MISHRA AND BISHAKHDATTA GAYEN WITH DATA FROM THE PHYSICAL OCEANOGRAPHY DISTRIBUTED ACTIVE ARCHIVE CENTER AND NASA'S VISIBLE EARTH, BLUE MARBLE COLLECTION.)

Bishakhdatta Gayen is an Australian Research Council Future Fellow and an associate professor in mechanical engineering at the University of Melbourne in Australia and an assistant professor at the Centre for Atmospheric and Oceanic Sciences at the Indian Institute of Science. His research explores the impact of small-scale turbulence and convection on global ocean circulation and the melting of Antarctic ice shelves. **Andreas Klocker** is a senior researcher in the Earth systems group at NORCE Norwegian Research Centre and the Bjerknes Centre for Climate Research, both in Bergen, Norway. His research focuses on understanding the dynamics of polar oceans and the global overturning circulation.



Bishakhdatta Gayen and Andreas Klocker

Research that combines fluid dynamics and climate science is uncovering the inner workings of the North Atlantic's overturning circulation. Future changes to that circulation system could trigger major disruptions to global weather patterns.

In 1751 Henry Ellis, captain of the British slave-trading ship *Earl of Halifax*, was sailing across the North Atlantic Ocean when he made a surprising discovery. At 25° N, the latitude of modern-day Miami, he measured the temperature of the ocean waters. Taking measurements as deep as a mile, he found very cold water sitting below the warm tropical waters at the surface. Those water masses could have been cooled to such low temperatures only in an extremely cold environment, which indicated that they originated in polar regions.[1]

It wasn't until decades later, in 1797, that Count Rumford (Benjamin Thompson) recorded, in his essay "Of the propagation of heat in fluids," the important implication of those measurements: that they could be explained only through a North–South circulation system in the ocean. In the high latitudes of the North Atlantic, cold air during the winter months cools and densifies surface waters, causing them to descend to great depths in the ocean interior. At depth, that cold water spreads equatorward, inducing an opposing surface-level current that transports warm water from the equator to high latitudes. The result is an overturning circulation that moves like a conveyor belt.

Today we know a lot more about the global ocean's overturning circulation, in particular the overturning circulation in the Atlantic Ocean. The Atlantic Meridional Overturning Circulation (AMOC) redistributes a significant amount of heat, fresh water, and carbon; as a result, it plays a crucial role in the climate system.[2] Poleward heat transport by the AMOC, for example, is the reason why temperatures in northern Europe are much higher than those at similar latitudes elsewhere. It has also stabilized Europe's climate for thousands of years.

Earth's climate has not always been stable. For example, in the last glacial period, which ended 11 500 years ago, Earth experienced pronounced temperature swings on time scales of decades or less, much shorter than the time scales of climate change's known orbital drivers, which operate over periods of 26 000 to 100 000 years. Those abrupt changes in the last glacial period were linked to variability of the AMOC.[3] Theory and climate models suggest that the AMOC is a nonlinear system

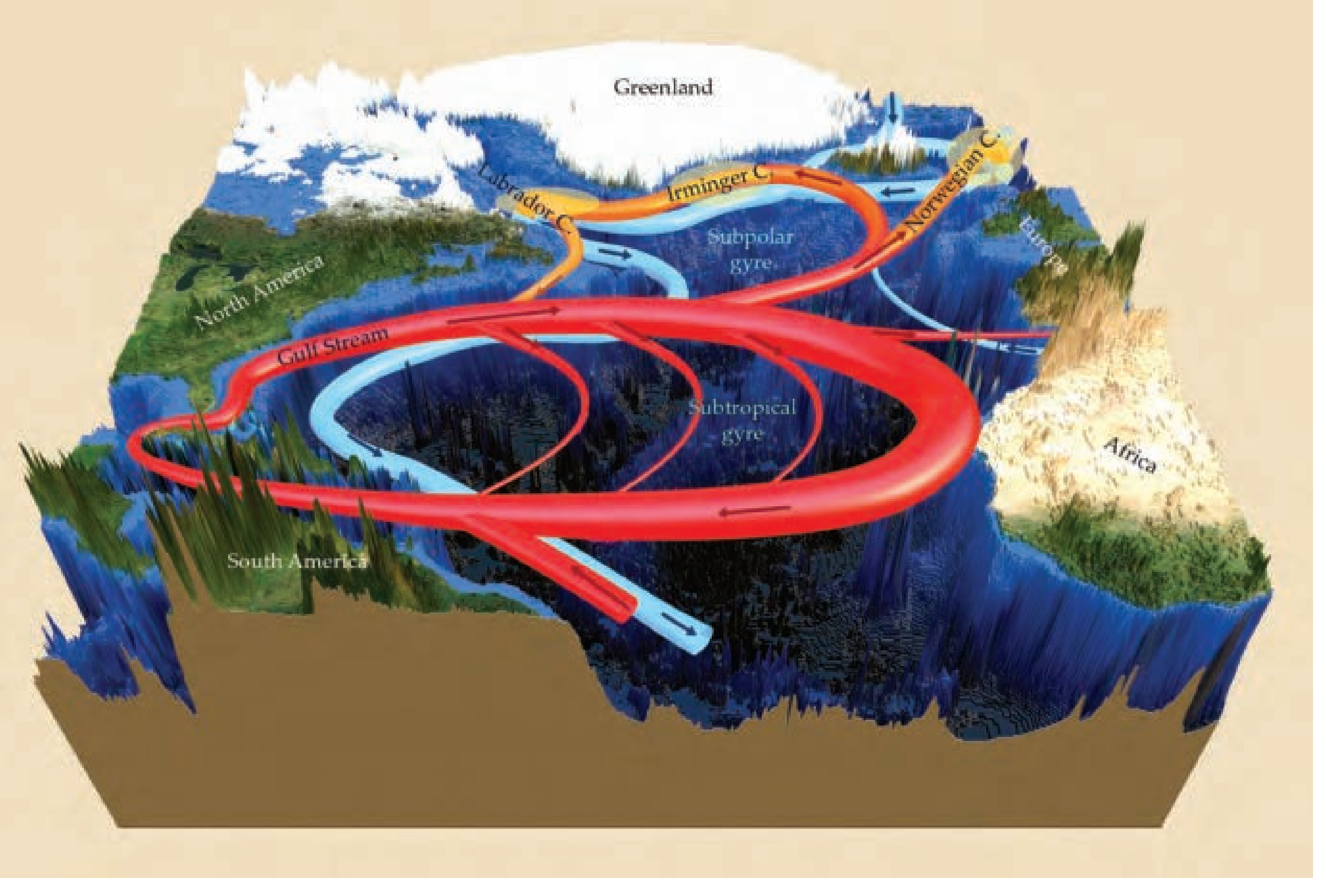


FIGURE 1. THE ATLANTIC MERIDIONAL OVERTURNING CIRCULATION is composed of currents that move warm, less-dense water (red arrows) at the ocean surface and cold, denser water (blue arrows) at depth, vertically linked by regions of deep convection in the Labrador, Irminger, and Norwegian Seas (yellow patches). Warm ocean currents flow toward the North Pole and are continuously cooled by the atmosphere along the way (illustrated by transition from red to orange in the Labrador, Irminger, and Norwegian currents), and then at depth, they return toward the equator. (Adapted from ref. 9.)

with multiple equilibrium states.[4] Climate scientists consider disruptions to the AMOC one of the most powerful and likely tipping points in Earth's climate system,[5] and they've extensively debated whether we are nearing a major upset of the circulation system due to changing climate.

One of the key uncertainties obscuring scientific understanding of the dynamics of that circulation system, including what drives its variability on different time scales, is the elusive link between deep convection and the mass transport of water associated with the AMOC.[6] As illustrated in figure 1, the AMOC consists of a shallower branch that brings warm waters northward (red arrows) and a deeper branch that brings cold waters southward (blue arrows). The two branches are thought to be connected through regions of deep convection (yellow patches).

The apparent conundrum is that the shallower and deeper branches are associated with a transport of water of close to 20 sverdrup (1 sverdrup is equal to 10^6 m^3/s; for comparison, the total global input of fresh water from rivers into oceans is ~1.2 sverdrup). But in the classic oceanographic view, the vertical exchange processes acting in regions of deep convection are not associated with any net mass transport.[7] Instead, deep convection is conceived of as the mixing of water parcels vertically, resulting in only up- and downward scalar transport of properties like temperature, salinity, and nutrient load. Given that view of ocean convection, how can an overturning circulation of 20 sverdrup move through regions of deep convection without a net transport of any water taking place?

Unnoticed by much of the oceanographic community, the missing link can be found in a process known as rotating horizontal convection. It is a complex form of convection that accounts for differential heating and cooling along the ocean surface, as well as Earth's rotation. The character of rotating horizontal convection has been uncovered by research into geophysical fluid dynamics using both laboratory experiments and turbulence-resolving direct numerical simulations.[8,9] Realistic scaled physical models, based on dynamic similarity arguments (see box on page 50), mimic the complexities of large-scale ocean circulation and help to generate a clearer picture of the AMOC's turbulent dynamics. Such research at the interface of climate science and fluid dynamics advances our understanding of the evolution of our climate system, especially related to critical tipping points.

The simplest form of convection

Convection is a captivating phenomenon that occurs when a fluid moves because of differences in buoyancy. At its most basic level, that movement is driven by the interplay between buoyancy and gravitational forces. Imagine a shallow pan filled with water, with its base heated and its surface exposed to a cooler environment, as modeled in figure 2. As the water warms at the base, it becomes less dense and rises, only to be replaced by the cooler water descending from the surface. That sets up a continuous cycle of rising and falling, a dance of fluids known as Rayleigh–Bénard convection (see the article by Leo Kadanoff in Physics Today, August 2001, page 34), and it's the foundation for the more complex convection patterns observed in vast natural systems such as the ocean, atmosphere, and mantle.

Convection in fluids is defined by the dimensionless Rayleigh number *Ra*, which accounts for buoyancy forces and viscous resistance. Beyond a critical *Ra* (~10^3), convection cells known as rolls emerge and arrange into symmetric patterns. With increasing *Ra*, flow becomes turbulent and shows chaotic movements. At higher levels ($Ra > 10^{10}$), smaller plumes self-organize into megaplumes that resemble large-scale atmospheric currents. Megaplumes add complexity by introducing shear forces that affect the behavior of smaller plumes and create a cycle of interactions.

Self-organization is a key process in which local interactions spontaneously transform disordered systems into structured patterns, such as mantle plumes and atmospheric convective storms, which are observable across various physical domains. (See the article by Caroline Muller and Sophie Abramian, Physics Today, May 2023, page 28.) In that process, complex behaviors and structures can emerge from simple rules inherent to the system. The phenomenon challenges the traditional top-down approach to understanding complexity.

The classic view of deep oceanic convection[7] is consistent with idealized Rayleigh–Bénard convection, in which fluid's vertical movements exchange water parcels between the heated and cooled parts of the domain without any net mass exchange. But that view cannot explain the mass transport associated with large-scale vertical and horizontal components of the overturning circulation observed in the ocean.

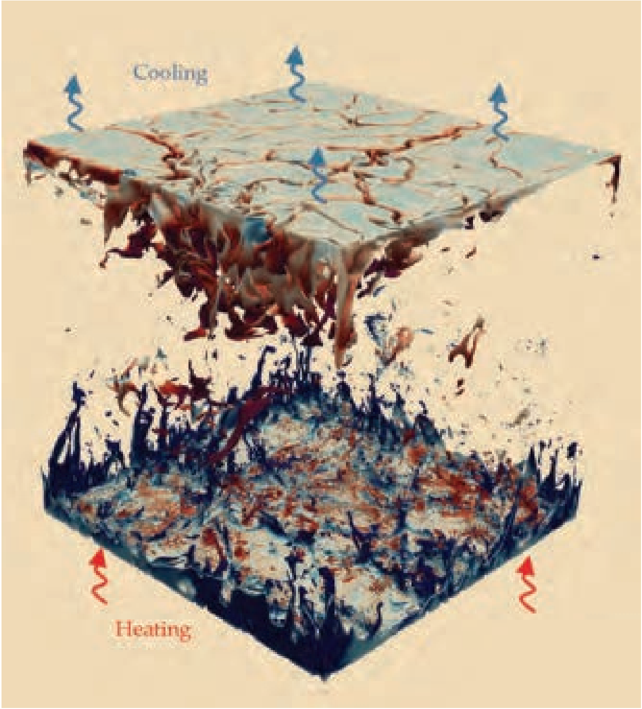


FIGURE 2. THE FLOW FIELD from a numerical experiment of Rayleigh–Bénard convection in a 3D rectangular box illustrates convective motions between a heated surface and a cooled surface, with colors representing vertical velocities (blue for upward and red for downward motions).

Going horizontal

In Rayleigh–Bénard convection, the flow is forced by heating from below and cooling from above, with a vertical buoyancy gradient parallel to the force of gravity. In the ocean, temperature changes are driven by air temperatures and solar radiation along the horizontal ocean surface. At equatorial latitudes, solar radiation heats the ocean, while near the poles, the ocean is cooled, setting up large-scale buoyancy gradients perpendicular to the force of gravity. Additional buoyancy gradients arise from differences in salinity, which are controlled by regions of evaporation and precipitation, the freezing and melting of sea ice, and freshwater discharge by rivers. Cooling and heating along a horizontal surface leads to a class of convective flows known as horizontal convection,[8] as shown in both a laboratory experiment and a numerical simulation in figure 3.

In Rayleigh–Bénard convection, both heating and cooling impose destabilizing buoyancy forcing, resulting in up and down motions concurrently. In contrast, in horizontal convection, it is only cooling that promotes destabilizing buoyancy forcing, because the heating from the top of the surface means the warmer, less dense water at the surface is gravitationally stable. That arrangement sets up overturning spanning the whole domain. As the cool water falls, a horizontal flow that forms along the surface moves water from the heated domain to the cooled domain, as shown from right to left in figure 3a.

That horizontal convection model is analogous to horizontal flow in the ocean, in which water moves from the low latitudes where it is heated to polar latitudes where it is cooled. In the ocean, the horizontal flow is confined to a stable boundary layer, generally a few hundred meters thick, at the ocean surface. To replenish the water at the surface, water rises at tropical and subtropical latitudes. The horizontal surface current is stabilized by the large vertical buoyancy gradients at the base of the boundary layer, with the dense, cold water mass naturally remaining below the warm water mass above.

Large horizontal buoyancy gradients thus create the conditions for an overturning circulation. For horizontal convection with $Ra < 10^{10}$ (the length scale is determined by the box's horizontal length, which is different from *Ra* values determined by the domain height, commonly used in the classical Rayleigh–Bénard definition), a laminar overturning circulation forms. Laminar circulation features a narrow and weak horizontal flow across the upper boundary and is marked by a single weaker end-wall plume in the domain of the unstable boundary.[10]

For settings with larger *Ra*, such as the ocean (where the value ranges between 10^{20} and 10^{24}), flow is more energetic and generates convective eddies. In cold regions, where cooling of the surface water is destabilizing, convective plumes emerge in the stable, warm boundary-layer flow and lead to the formation of mixed-temperature layers in mid- to high latitudes.[10]

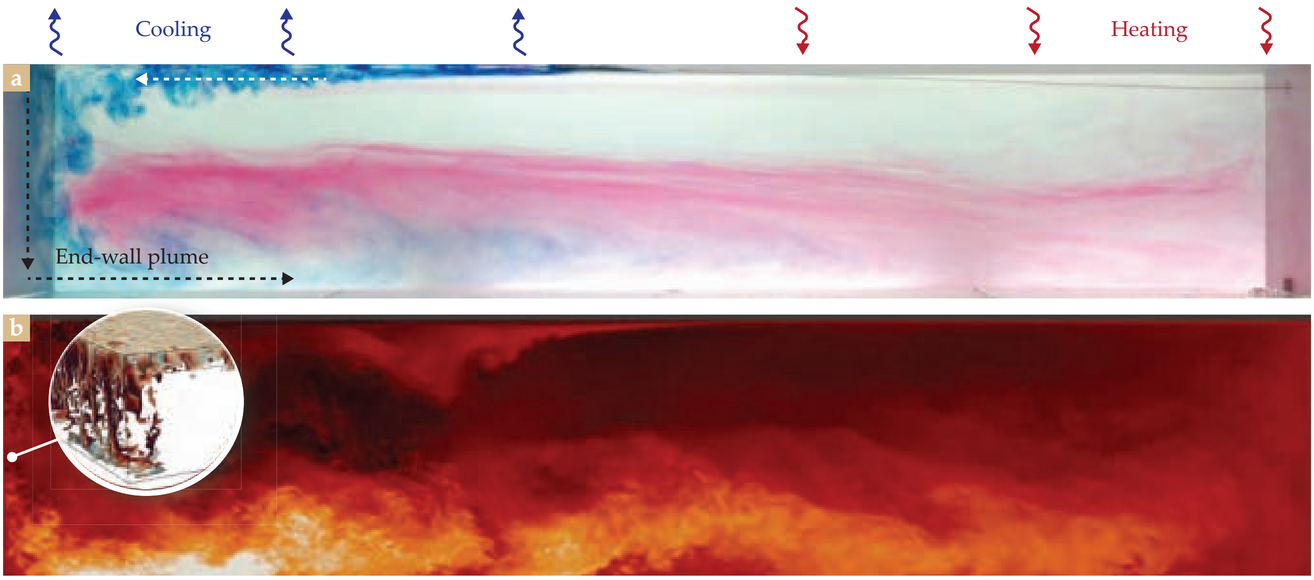


FIGURE 3. HORIZONTAL CONVECTION is illustrated by the flow field from both **(a)** a laboratory experiment (Rayleigh number 10^{12}) and **(b)** a numerical simulation (Rayleigh number 10^{15}) of horizontal convection in a 3D rectangular box domain. In panel a, dye visualization shows the flow path from the heated region to the cooled region (right to left) along the upper boundary and broader bottom flow in the opposite direction. A full-depth end-wall plume is observed against the wall at the cooled end of the domain. (Adapted from ref. 8.) As shown in panel b and in the inset, flow-field data from a large eddy simulation with a Rayleigh number three orders of magnitude higher than the lab experiment demonstrate that the flow is increasingly chaotic at larger buoyancy forcing. (Adapted from ref. 10.)

Farther toward polar latitudes, continuous cooling makes those convective plumes penetrate increasingly deeper into the water column.

In lab experiments, those dynamics change once the flow approaches the end wall. There, because of the combined effects of boundary-induced turbulence and the lack of lateral space for the plumes to spread, the plumes merge through self-organization and penetrate the entire depth of the water column. That phenomenon leads to the creation of large convective activity against the end walls, like the left wall in figure 3. In contrast to Rayleigh–Bénard convective plumes, which produce net-zero transport, horizontal convection produces convective plumes with a net-downward mass transport.

Away from the boundaries and the end walls, in the more quiescent interior of the domain, a broad upward flow is established to restore mass conservation in the system. After slowly moving upward, fluid moves horizontally from the warm end to the cool end and completes the box-scale circulation loop that is an overturning circulation.

Adding a twist

Just when it might seem that the complexity of oceanic convection has been fully unraveled, another dimension emerges: the rotation of the planet. As Earth rotates on its axis, it bestows upon every moving entity a subtle yet profound pseudoforce, known as the Coriolis force. Invisible yet omnipresent, the Coriolis force imparts a preferential spin to fluid bodies in motion, including the vast swaths of oceanic currents.

When the principles of horizontal convection are superimposed onto a rotating framework, the dynamics are known as rotating horizonal convection,[9] and the resulting circulation patterns resemble those observed in the ocean. One key change that occurs with the addition of the Coriolis force is that instead of the flow from heated to cooled regions spreading homogeneously across the full width of the domain, poleward flow is concentrated in narrow, fast-flowing boundary currents, as shown in a numerical simulation in figure 4.

In the heated region appears an anticyclonic gyre, akin to the subtropical gyre shown in the schematic in figure 1. Most of the modeled horizontal, poleward transport is concentrated in a current along the western boundary shown in figure 4. That style of boundary current is part of what makes up the Gulf Stream in the Atlantic Ocean, while wind forcing also contributes to the strength of the Gulf Stream.

Once the western boundary current reaches the point of negligible heat input, where the flow transitions from heating at equatorial latitudes to cooling at polar latitudes, the current separates from the western boundary and flows east across the basin. The east–west zonal current is supported by a north–south buoyancy difference caused by temperature forcing and is influenced by the Earth's rotation, a phenomenon commonly referred to as the thermal wind balance.[11] Often that balance breaks down and flow becomes unstable, which leads to the emergence of mesoscale eddies, the oceanic equivalents of atmospheric storms.

Once the flow approaches the eastern boundary, it continues northward and, when it reaches the northern end wall, turns westward (see figure 4). In moving north, it reaches the cooled region, where surface buoyancy forcing is destabilizing, and generates cyclonic gyre circulation akin to the subpolar gyre shown in the schematic in figure 1.

As with horizontal convection, surface cooling leads to downward convective plumes (commonly known as convective chimneys when discussed in the context of a rotating environment) that reach increasingly deeper into the water column until they eventually penetrate the entire column depth. In the presence of rotation, the plumes are located along the narrow boundary current, with an end-wall plume toward the northeast corner of the domain (see the figure 4 inset). Net-downward mass transport occurs against the eastern and northern boundaries due to surface cooling. These boundaries also create a lateral pressure gradient that redirects flow into the vertical direction. The overturning circulation is then balanced by broad upwelling in the heated part of the domain.

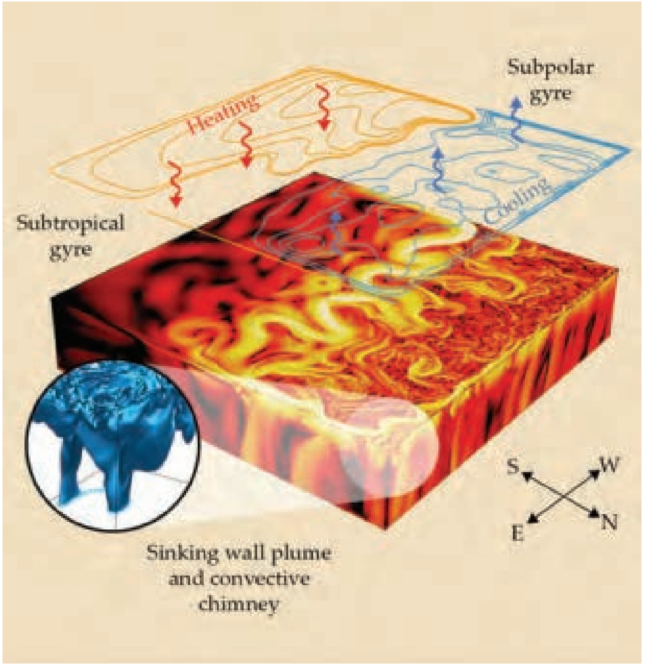


FIGURE 4. ROTATING HORIZONTAL CONVECTION is numerically simulated in a domain similar to the Northern Hemisphere section of the Atlantic basin. Circulation is shown as kinetic energy (brighter regions have higher energy). The inset illustrates a descending boundary-layer current, wall plumes, and convective chimneys as an isosurface of the kinetic-energy field. A horizontal stream function, shown as contours above the model domain, highlights the anticyclonic and cyclonic gyres associated with the heated and cooled regions, respectively.

Closing the circle

The idealized dynamics of rotating horizontal convection show the connection between oceanic deep convection and the AMOC. To visualize that connection, researchers use output from a high-resolution simulation based on the Community Earth System Model,[12] with a focus on the area of the subpolar gyre and deep convection of the northern Atlantic Ocean, as shown in figure 5.

One difference between idealized experiments and Earth is the roughness of the ocean floor, as shown in figure 5a. For the AMOC, the Greenland–Scotland Ridge and the Labrador Basin (southwest of Greenland) form a boundary much like the end wall in physical experiments, with strong ocean currents forming along its edges, as shown in a map of surface velocity in figure 5b. Because the ridge spanning from Greenland to Scotland does not consistently reach the ocean surface, a shallow (and much weaker) Arctic Meridional Overturning Circulation[13] exists farther north.

Oceanic convection can be visualized by the mean mixed-layer depth—the depth over which convection homogenizes the water column—over the winter months, as shown in figure 5c, and the maximum mixed-layer depth estimated across all seasons, as shown in figure 5d. The maps of those values highlight convective regions with two distinct properties. One style of convection is seen in the Labrador Basin and the western region of the Irminger Basin, off the southeast coast of Greenland. Those regions exhibit highly intermittent deep convection that's less sustained in the winter months but reaches depths of more than 2000 m. Those regions of highly turbulent deep convection resemble the end-wall plumes

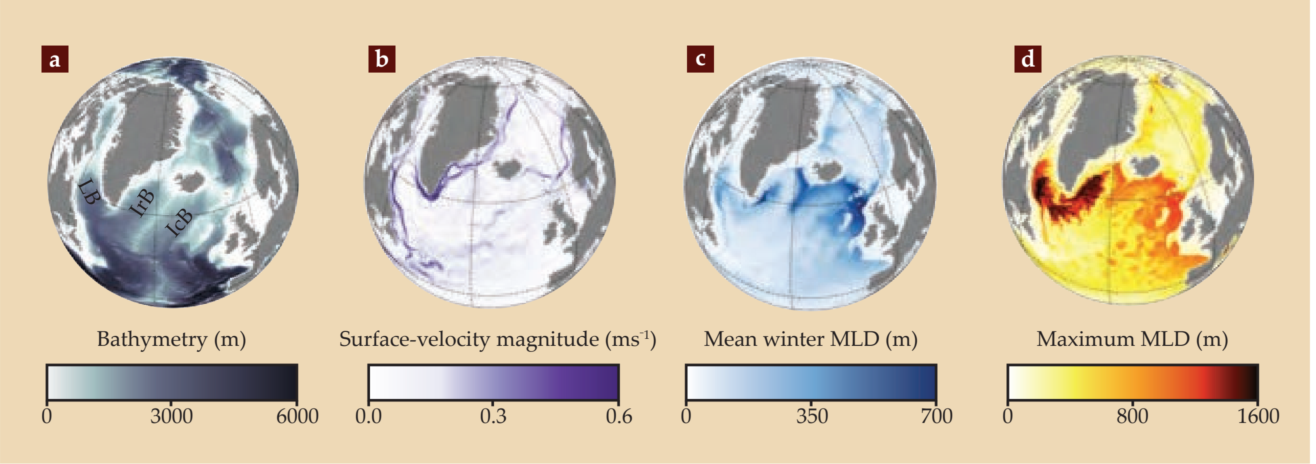


FIGURE 5. AN OCEAN SYSTEM MODEL derived from the Community Earth System Model. **(a)** Ocean basins relevant to regions of deep convection are apparent in a map of the bathymetry of the North Atlantic, including the Labrador Basin (LB), Irminger Basin (IrB), and Iceland Basin (IcB). Models that incorporate bathymetry and ocean eddies are used to calculate **(b)** the magnitude of velocities at the ocean surface, **(c)** the mean mixed-layer depth (MLD) over the winter months (December to March), and **(d)** the maximum MLD.

Using dynamic similarity to simulate ocean convection

The principle of dynamic similarity posits that if two fluid flows—regardless of their absolute length scale or each fluid's specific properties—have the same values for all relevant dimensionless numbers, they are dynamically similar. In essence, they behave in the same way. As a result, researchers can study complex Earth-scale systems by building dynamically similar lab-scale experiments.

The dimensionless numbers relevant to large-scale ocean circulation include the Rayleigh number, which signifies the driving force of buoyancy against dissipative forces; the Rossby number, which reflects the ratio of inertial forces to Coriolis forces; and the Ekman number, which indicates viscous forces in relation to the Coriolis effect. Not only do those numbers provide a means of producing systems with dynamic similarity, but they also encapsulate the fundamental forces at play in oceanic circulations. By carefully designing experiments or simulations in which those nondimensional numbers match the conditions of the real-world system of interest, researchers can gain invaluable insights into the behavior of vast ocean circulations. Those values can be tuned in many ways, such as by using materials of different viscosities, or altering the dimensions of the experimental setup. That approach not only makes the otherwise insurmountable task of studying global systems feasible, but it also bridges the gap between theory and observation, which opens the door to more accurate predictions and a deeper understanding of our world's oceans.

observed in lab experiments and numerical simulations and are associated with the lowest surface buoyancy (densest surface water) that induces circulation toward that area (see figure 5b).

The other style of convection can be seen in narrow bands against steep ocean floor topography in the Iceland Basin and northeastern region of the Irminger Basin (south and east of Iceland, respectively), with mixing extending up to 700 m in depth during the winter months. It is those northeast corners where the circulation, in the presence of planetary rotation, can build up a pressure gradient, which creates narrow regions with a net-downward mass transport,[14] consistent with observations,[15] idealized ocean models,[16] and experiments of rotating horizontal convection.[9]

The reason for those two distinct convective regions is planetary rotation, which separates end-wall-style plumes and regions of net-downward mass transport. Both of those convective regions are crucial components of the overturning circulation, and neither of them controls AMOC variability on its own.

The main features of the AMOC are the result of large-scale buoyancy gradients at the ocean's surface due to both surface heating in equatorial regions and cooling in polar regions, with salinity and winds also playing important roles that are not discussed here. Like the atmosphere (see the article by Martin Singh and Morgan O'Neill, Physics Today, July 2022, page 30), the AMOC acts as a heat engine that produces mechanical energy through the transport of heat from warmer regions to cooler regions. The resulting circulation, described by the dynamics of rotating horizontal convection, leads to the large oceanic heat transport that dictates the mild European climate. Thus, understanding the convection processes at play in ocean circulation is crucial to our understanding of the ocean's role in a future climate.

Laboratory experiments and numerical simulations are immensely valuable basic-research tools to understand geophysical fluid dynamics and our complex climate system. Already used by the atmospheric science community, that research approach could be applied in future work to build understanding about the role of self-organization of convective plumes in oceanic overturning circulation. Idealized physical experiments make complex physical processes into measurable phenomena, opening the door to a clearer understanding of those systems.

Bishakhdatta Gayen is supported by an Australian Research Council Future Fellowship. Numerical simulations were performed on the National Computational Infrastructure at the Australian National University, with support from the Commonwealth of Australia. Gratitude is extended for valuable discussions with colleagues, particularly Ross Griffiths, Graham Hughes, Catherine Vreugdenhil, Andy Hogg, and Fabien Roquet. Andreas Klocker appreciates the support of the Bjerknes Centre for Climate Research through its backing from the Norwegian Ministry of Education and Research, and he thanks his colleagues at the Bjerknes Centre's DYNASOR (DYnamics of the North Atlantic Surface and Overturning ciRculation) project for useful discussions.

REFERENCES

1. B. A. Warren, in *Evolution of Physical Oceanography: Scientific Surveys in Honor of Henry Stommel*, B. A. Warren, C. Wunsch, eds., MIT Press (1981), p. 7.
2. M. W. Buckley, J. Marshall, *Rev. Geophys.* **54**, 5 (2016).
3. J. Lynch-Stieglitz, *Annu. Rev. Mar. Sci.* **9**, 83 (2017).
4. W. Weijer et al., *J. Geophys. Res. Oceans* **124**, 5336 (2019).
5. T. M. Lenton et al., *Proc. Natl. Acad. Sci. USA* **105**, 1786 (2008).
6. M. F. de Jong, L. de Steur, *Geophys. Res. Lett.* **43**, 7106 (2016).
7. J. Marshall, F. Schott, *Rev. Geophys.* **37**, 1 (1999).
8. G. O. Hughes, R. W. Griffiths, *Annu. Rev. Fluid Mech.* **40**, 185 (2008).
9. B. Gayen, R. W. Griffiths, *Annu. Rev. Fluid Mech.* **54**, 105 (2022).
10. B. Gayen, R. W. Griffiths, G. O. Hughes, *J. Fluid Mech.* **751**, 698 (2014).
11. H. U. Sverdrup, *Proc. Natl. Acad. Sci. USA* **33**, 318 (1947).
12. P. Chang et al., *J. Adv. Model. Earth Syst.* **12**, e2020MS002298 (2020).
13. A. Bretones et al., *J. Climate* **35**, 109 (2022).
14. J. Marotzke, J. R. Scott, *J. Phys. Oceanogr.* **29**, 2962 (1999).
15. M. S. Lozier et al., *Science* **363**, 516 (2019).
16. M. A. Spall, R. S. Pickart, *J. Phys. Oceanogr.* **31**, 810 (2001).

PT

NEW PRODUCTS

Focus on test, measurement, quantum metrology, and analytical equipment

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.

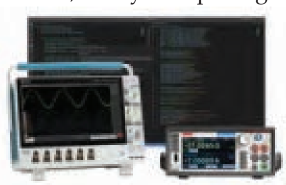
Andreas Mandelis

High-speed oscilloscope

Pico Technology designed its PicoScope 6428E-D oscilloscope for applications that require superior performance in signal analysis. It extends the capabilities of the existing PicoScope 6000E series and thus is suitable for users working in high-energy physics, lidar, visar, spectroscopy, accelerators, and other high-speed applications. The PicoScope 6428E-D offers up to 16 digital channels, 4 analog channels, and a maximum 3 GHz bandwidth with up to a 10 GS/s sampling rate in 8-bit mode. In dual-channel mode, the sample rate reduces to 5 GS/s and 2.5 GS/s on all four analog channels. It offers flexible resolution up to 12 bits (using the company's FlexRes technology), capture memory up to 4 GS, and a data transfer rate exceeding 300 MS/s. A streaming mode allows for continuous data capture directly to a PC's RAM or hard disk. Featuring four ranges between ±50 and ±500 mV, the compact PicoScope 6428E-D integrates into various systems that require 50 Ω measurements. ***Pico Technology***, *320 N Glenwood Blvd, Tyler, TX 75702, www.picotech.com*

Python driver for measurement automation

Tektronix has introduced an open-source Python instrument driver package, called tm_devices, for its test and measurement instrumentation. Available free of charge, tm_devices provides a native Python user capability for fast, seamless instrument automation. It works across a wide range of Tektronix and Keithley devices to facilitate ongoing development and updates. By integrating tm_devices into daily workflows and using it with a preferred IDE (integrated development environment), users can access auto-complete, precise type hinting, comprehensive built-in help, real-time syntax checking, and efficient debugging capabilities. According to Tektronix, its Python package simplifies instrument setup and use while other companies' packages often require users to install complex driver software and interface layers. The tm_devices is available via the Python Package Index at https://pypi.org/project/tm-devices. ***Tektronix Inc***, *14150 SW Karl Braun Dr, PO Box 500, Beaverton, OR 97077, www.tek.com*



RF signal-source analyzer

Keysight's SSA-X signal-source analyzer portfolio now extends to 26.5, 44, and 54 GHz. The integrated, one-box series provides comprehensive signal-source analysis for advanced wireless communications, radar, and high-speed digital applications. Those include phase and residual noise measurement, transient measurement, spectrum and network analysis, and voltage-controlled oscillator characterization. The platform features a clean signal that is enabled through a direct digital synthesis source and proprietary cross-correlation channels. According to the company, the SSA-X offers best-in-class phase noise sensitivity. The built-in signal source and two RF inputs enable residual-noise measurements without additional equipment and reconfiguration. The application software has been enhanced to address more measurement needs, including precision clock jitter analysis. With sensitivity of 2 fs at 10 GHz, the SSA-X is suitable for advanced high-speed digital communications applications. ***Keysight Technologies Inc***, *1400 Fountaingrove Pkwy, Santa Rosa, CA 95403-1738, www.keysight.com*

Electronics for scalable quantum computing

Quantum Machines has made two additions to its quantum electronics product family. Designed for quantum computing at scale, the low-frequency voltage generator and the breakout box support high channel density. The QDAC-II Compact is a versatile, stable, ultralow-noise, 24-channel voltage source for tuning superconducting and spin qubits for optimal performance. It offers all the features of the company's QDAC-II model but fits into one-fourth the space. The QSwitch is an easy-to-use, software-controllable breakout box with 240 relays. It saves time for researchers and R&D users by preprogramming experiments and quickly switching between setups and instruments. The QDAC-II Compact and the QSwitch can be used stand-alone or connected in series via a single 24-pin cable. Because they can be stacked and operated with multiple units of the company's OPX1000 ultrahigh-speed quantum controller, they allow for the expansion of control systems. ***Quantum Machines***, *HaMasger St 35, Tel Aviv-Yafo, 6721407, Israel, www.quantum-machines.co*

System for high-fidelity qubit control

Zurich Instruments has launched a new system for the control and readout of superconducting qubits. The SHF+ product line includes the SHFSG+ signal generator, the SHFQC+ qubit controller, and the SHFQA+ quantum analyzer. The SHF+ products deliver high-fidelity qubit control and quality performance. They allow for the reliable manipulation of fragile quantum states, let users run more powerful algorithms, and accelerate the development of longer-coherence-time qubits and scaling up of quantum processing units. The SHFSG+, SHFQC+, and SHFQA+ offer among the highest signal-to-noise ratios (SNRs) available on the market, according to the company. The high SNR reduces thermal qubit excitation and maximizes quantum computing algorithm fidelity. The improved phase noise allows for the suppression of phase errors in the control of long-lived qubits. In addition to a noise-optimized signal chain, the SHF+ products offer the ability to mute the output for measurements on even the most sensitive qubits. ***Zurich Instruments AG****, Technoparkstrasse 1, 8005 Zürich, Switzerland, www.zhinst.com*

Analyzers for measurement characterization

A new family of Rohde & Schwarz power analyzers, now available in three compact models, meets the requirements for measuring voltage, current, power, energy, and total harmonic distortion on DC and AC sources. The R&S NPA101 power meter provides all basic measurements; the R&S NPA501 power analyzer adds enhanced measurement functions and graphical analysis; and the R&S NPA701 compliance tester includes evaluation functions in line with IEC 62301 and EN 50564 standards for power consumption and EN 61000-3-2 for EMC harmonic emission testing. All R&S NPA models perform power measurements at levels from 50 μW to 12 kW, at potential differences from 1 mV to 600 V, and at currents from 1 mA to 20 A. To detect even the shortest transient ripples in output, they have a high sampling rate of 500 ksample/s. The 16-bit resolution A/D convertor ensures an accuracy of ±0.05% for current and voltage readings. ***Rohde & Schwarz GmbH & Co KG****, Mühldorfstrasse 15, 81671 Munich, Germany, www.rohde-schwarz.com* PT

QUICK STUDY

Emily Caldwell and **Laura Sinclair** are researchers at NIST in Boulder, Colorado.

Comparing clocks by using pulses of light

Emily Caldwell and Laura Sinclair

Microwave atomic clocks can easily be synchronized across long distances using RF methods. Their more precise optical cousins require a subtler approach.

Look carefully, and you will see clocks all around you: The plant wilting on your desk tells you that a week has passed since you watered it; the setting Sun tells you that another day has gone by; your smartphone tells time with its GPS-calibrated clock set by pinging satellites in space. Although we tend to experience time on the scale of seconds to years, scientists have developed ways to measure time at the level of 1 part in 10^{18} by using optical atomic clocks, which determine time by probing atomic resonances with lasers. That's equivalent to measuring the distance from Earth to the Moon with an uncertainty of less than half a nanometer. Such extremely precise clocks offer new ways for physicists to, for example, test general relativity, search for dark matter, and probe variations in fundamental constants.

To do many of those experiments, researchers look for unexplained variations between two physically separated clocks—an ideal future testbed would be one in space and one on Earth. The ability to compare time across such distances is the domain of what's known as time transfer. Time transfer is the basis for GPS, which operates at the submicrosecond level and uses RFs. Although it can synchronize communications networks and power grids, GPS is not precise enough for fundamental physics tests. Instead, optical clocks supported by optical time-transfer methods are needed for those experiments.

Free-space optical time transfer is a technique that uses frequency combs—trains of equally spaced ultrashort pulses—to compare time between optical clocks at the femtosecond level across hundreds of kilometers of air. Until recently, the technique was limited to tens of kilometers, due to diffractive and turbulence-induced losses. But with the invention of the time-programmable frequency comb (TPFC), that range was pushed to 300 km with margin to spare. Because a laser beam traveling through 300 km of turbulent air experiences the same loss expected for one traveling from the ground to a geosynchronous satellite, the door is now open to future ground-to-space time transfer.

Time-programmable frequency comb

Lasers abound in the metrology world, but perhaps the most useful is the optical frequency comb (see Physics Today, June 2000, page 19). It is a pulsed laser made of many discrete colors that in the frequency domain look like a series of equally spaced lines, much like teeth in a comb. The utility of the frequency comb comes from the ability to lock the location of those lines and generate an ultrastable pulse train. Two numbers control the location of every frequency line: the offset of the first line from zero frequency and the spacing between lines. If you can control the location of two frequency lines, you can stabilize the comb. The work of John Hall and Theodor Hänsch in fixing the comb lines led in part to their receiving half of the 2005 Nobel Prize in Physics (see Physics Today, December 2005, page 19).

Physicists usually think of comb stabilization in the frequency domain, but it also fixes the two degrees of freedom in the time domain: pulse-to-pulse spacing and the pulse phase—the relative phase between the carrier wave and the pulse envelope. In a conventional frequency comb, these two degrees of freedom are stable but rigid. The TPFC combines an optical fiber frequency comb with digital electronics running new firmware that maintains the underlying stability but lets the user select the pulse timing and phase, as shown in figure 1a. A user's command to delay the pulse-output time by 5 fs, for example, is converted to a phase shift and applied to the comb's two phase locks. That phase shift, in turn, alters the pulse timing.

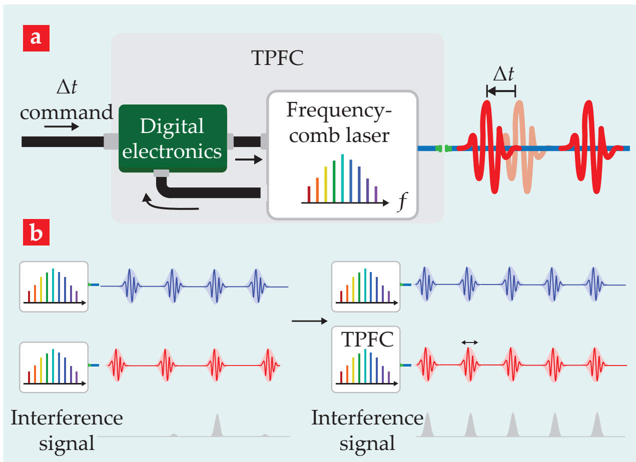


FIGURE 1. A TIME-PROGRAMMABLE FREQUENCY COMB (TPFC), which is based around a conventional frequency-comb laser, **(a)** converts a user-input timing command, Δt, to a timing shift in output pulses. **(b)** Two schemes measure the timing. In both, the timing of the comb under test (blue) is measured by mixing it with a second frequency comb (red) to generate interference signals on a photodiode (gray). In a conventional dual-comb timing measurement (left), the combs run at different rates, and timing shifts of the interference peaks directly map onto timing shifts in the comb under test. In tracking-style detection (right), interference in the signal amplitude is used as feedback to force the TPFC measurement to track the comb under test. Tracking-style detection requires 1/10 000 as much light as the conventional measurement technique.

FIGURE 2. THE MAUNA LOA OBSERVATORY (foreground) on the island of Hawaii sends a laser pulse as part of an optical time-transfer demonstration over the 300 km round-trip distance between it and a cat's-eye retroreflector (inset) on Maui. **(a)** A frequency comb referenced to a stable laser sends its pulses to the other clock's site over a bidirectional free-space link. Combining the pulses at each site yields the timing offset between the two clocks. **(b)** Time-programmable frequency combs enable quantum-limited timing detection, which reduces the power required for time transfer by 10 000. (Adapted from E. D. Caldwell et al., *Nature* **618**, 721, 2023.)

Tests of the TPFC have demonstrated that pulses can be moved over a 5 ns gap between pulses to within a 2 as accuracy.

Although the ability to move light around is striking, the real power of the TPFC comes from how it changes one's ability to measure other comb pulses. Generally, the best way to measure the timing of a comb pulse is with another frequency comb. In a conventional dual-comb timing measurement, one offsets the spacing of pulses between the comb being tested and a measurement comb. When the mixed light from the two lasers hits a photodiode, it generates a peak in the interference signal when the pulses overlap in time, as shown in figure 1b.

The problem with that technique is the long dead time between the interference signal peaks. With the TPFC, however, there no longer is any dead time. Instead, it tracks the pulses of the comb under test in time with the control signal because the tracking also acts as the timing measurement.

With no dead-time penalty, the tracking-style detection operates at the quantum limit—also called the photon-shot-noise limit. The noise in the measurement is then dominated by the quantized nature of light, and the signal-to-noise ratio of the measurement is set by the number of received photons. The quantum limit represents the best measurement possible in a classical system for a given optical power.

Quantum-limited time transfer

The advantage of quantum-limited detection becomes clear when you consider it in an application such as free-space optical time transfer. In the system shown in figure 2a, for instance, frequency-comb lasers are locked to a clock laser at each site. The clock laser, which is ultrastable, has been steered in frequency to resonate with the clock's atomic transition. After locking, the timing of the frequency-comb pulses follows the frequency of the clock laser light. Think of the comb pulses as ticks of a clock that speed up or slow down with changes to the reference-laser frequency.

To compare two clocks, then, one needs to compare the timing of the pulses in the two frequency combs. That's done by sending comb light bidirectionally through the atmosphere and detecting when the incoming pulses reach both sides. The bidirectional link is critical because combining the measured time signals on both sides cancels the time-of-flight noise that comes from atmospheric optical turbulence or slow platform motion.

Substituting TPFCs for conventional combs and using the tracking-style detection scheme yields striking results for time transfer. Making those changes reduces the required received power by 10 000 and increases the working distance from tens to hundreds of kilometers for the same amount of launched laser light.

Figure 2b illustrates a demonstration of TPFCs that can transfer time at the femtosecond level across 300 km of air. Although it represents the longest terrestrial experiment performed to date, it was perhaps more exciting because the tolerable power loss shown was close to that expected for sending light from the ground to a satellite in geosynchronous orbit 35 786 km away. Putting optical atomic clocks in space remains a work in progress, but the results discussed here prove the ability of future ground-to-space time-transfer missions to push the bounds of fundamental and applied physics.

Additional resources

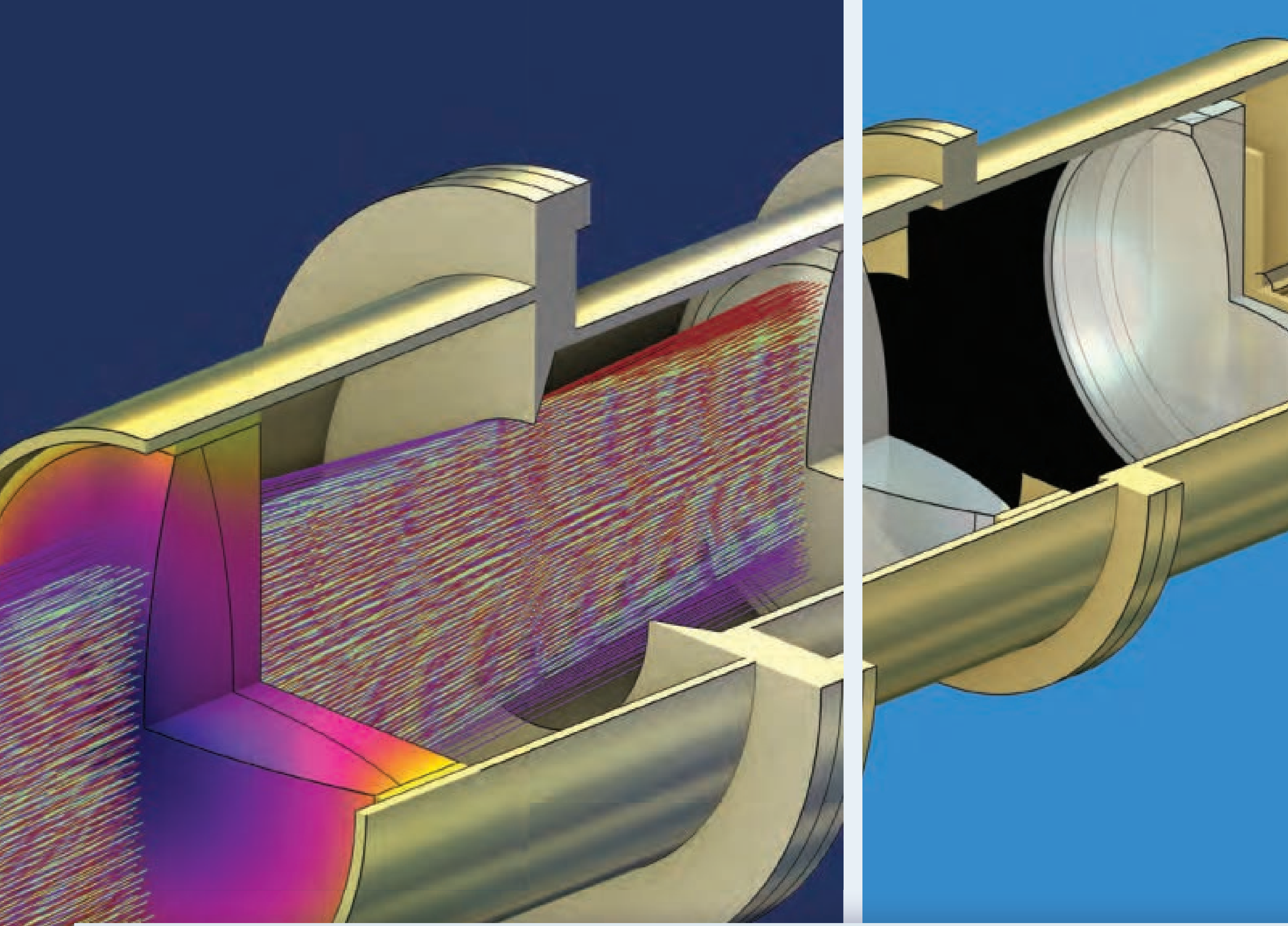
► T. Fortier, E. Baumann, "20 years of developments in optical frequency comb technology and applications," *Commun. Phys.* **2**, 153 (2019).
► A. Derevianko et al., "Fundamental physics with a state-of-the-art optical clock in space," *Quantum Sci. Technol.* **7**, 044002 (2022).
► E. D. Caldwell et al., "The time-programmable frequency comb and its use in quantum-limited ranging," *Nature* **610**, 667 (2022).
► E. D. Caldwell et al., "Quantum-limited optical time transfer for future geosynchronous links," *Nature* **618**, 721 (2023). PT

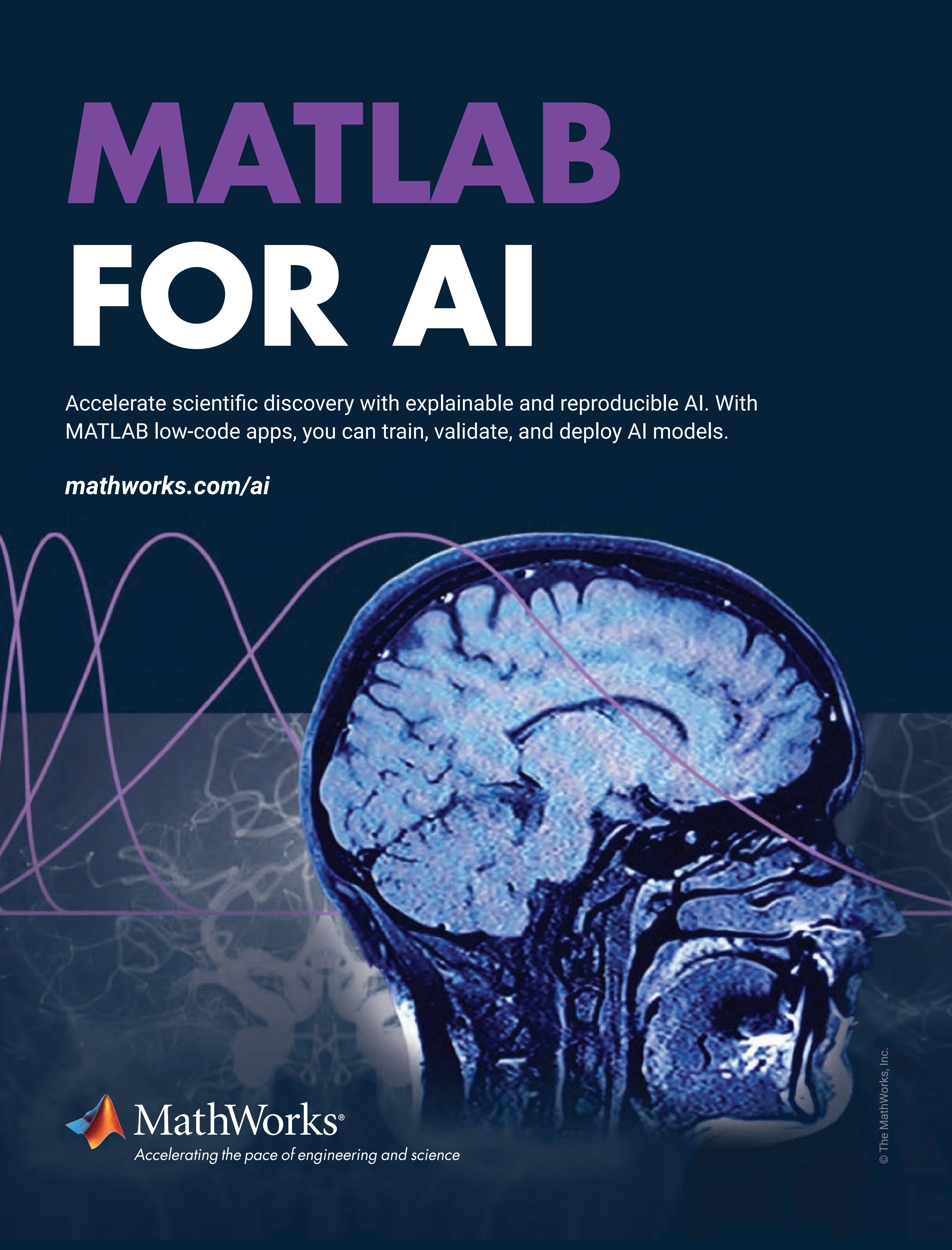
Making biodegradable plastic with bacteria

In a park or along a beach, you've almost certainly seen plastic trash. Despite being exposed to the weather, plastic lingers in the environment. One solution is to infuse it with plastic-eating bacteria. But a big problem is that heat and other conditions of the manufacturing process are often inhospitable for living organisms. Researchers from BASF Corp, the University of Georgia, and the University of California, San Diego, solved that issue by evolving successive populations of the safe soil bacteria *Bacillus subtilis* in the lab to endure conditions of up to 135 °C—the temperature at which many plastics are made. During manufacturing, the bacteria survive in metabolically dormant spores—natural defense structures that are reactivated by nutrients in the soil.

The bacteria-incorporated plastic, which the researchers made through standard industrial processes, degrades faster than typical plastic. In the left column are 48-mm-long strips of thermoplastic polyurethane, an elastic and durable plastic that's commonly used in medical devices, footwear, and other applications. The strips are increasingly degraded depending on how long they were buried in compost. The bottom sample had no exposure to compost, and the top one was buried for five months. The right column shows strips of bacteria-incorporated plastic with the same relative exposure, and about 90% of it degraded within five months, even in conditions without additional bacteria populations. (H. S. Kim et al., *Nat. Commun.* **15**, 3338, 2024; photos courtesy of David Baillot, UC San Diego Jacobs School of Engineering.) —AL

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




MATLAB
FOR AI
Accelerate scientific discovery with explainable and reproducible AI. With MATLAB low-code apps, you can train, validate, and deploy AI models.
mathworks.com/ai
MathWorks®
Accelerating the pace of engineering and science
© The MathWorks, Inc.