Telecom's Fight Over Interoperability › Is the 5G open standard really open?
P. 5

Why Regulating AI Will Be Hard › Lessons learned from self-driving cars
P. 30

Microgrids for the Real World › A town in Puerto Rico shows what's possible
P. 36

FOR THE TECHNOLOGY INSIDER

OCTOBER 2023

IEEE Spectrum

How Radar Drones Reveal the Secret Life of Ice

IEEE

Boost Your Lab's Performance in Quantum Research

We support your application with

Performant instrumentation

→ High-fidelity qubit control, readout, and feedback up to 8.5 GHz

Intuitive software

→ Swift experiment design in a dedicated language

In-depth application support

→ Direct contact with specialists for your research field

Zurich Instruments

VOLUME 60 / ISSUE 10 OCTOBER 2023

Contents

ON THE COVER:
Photo by Eliza Dawson

HEATHER DUNCAN/ORNL

Ben Ollis, Maximiliano Ferrari, and Richard Rains of Oak Ridge National Laboratory [from left] stand next to the Solar Forest installation in Adjuntas, Puerto Rico.

BY HARRY GOLDSTEIN

Our Hands-On Citizen Scientist

David Schneider's DIY projects unveil nature's secrets

Senior Editor David Schneider, sporting one of his signature Hawaiian shirts, futzes with his homemade coffee-can radar.

GRACE SCHNEIDER

Whether chasing weather balloons with software-defined radio or searching for buried treasure with a DIY magnetometer, detecting solar flares and gamma-ray bursts for less than US $100 or tracking the movement of the spiral arms of the Milky Way with a radio telescope he built himself, Senior Editor David Schneider has a knack for satisfying his scientific curiosity with projects engineered to delight *Spectrum* readers.

With a doctorate in geophysics, Schneider has been helping to make the invisible visible for scientists and engineers for almost three decades at *Scientific American, American Scientist,* and for the last 15 years, *IEEE Spectrum*.

As he prepares to leave *Spectrum* to pursue more writing and assorted other projects, I asked Schneider to survey his DIY oeuvre and choose some highlights. He notes that his exoplanet detector, which he made back in 2014, still impresses him almost a decade on. That project reinforces a theme running through many of Dave's projects: The only way most nonspecialists can explore these pursuits themselves is by using materials at hand or scored on the cheap.

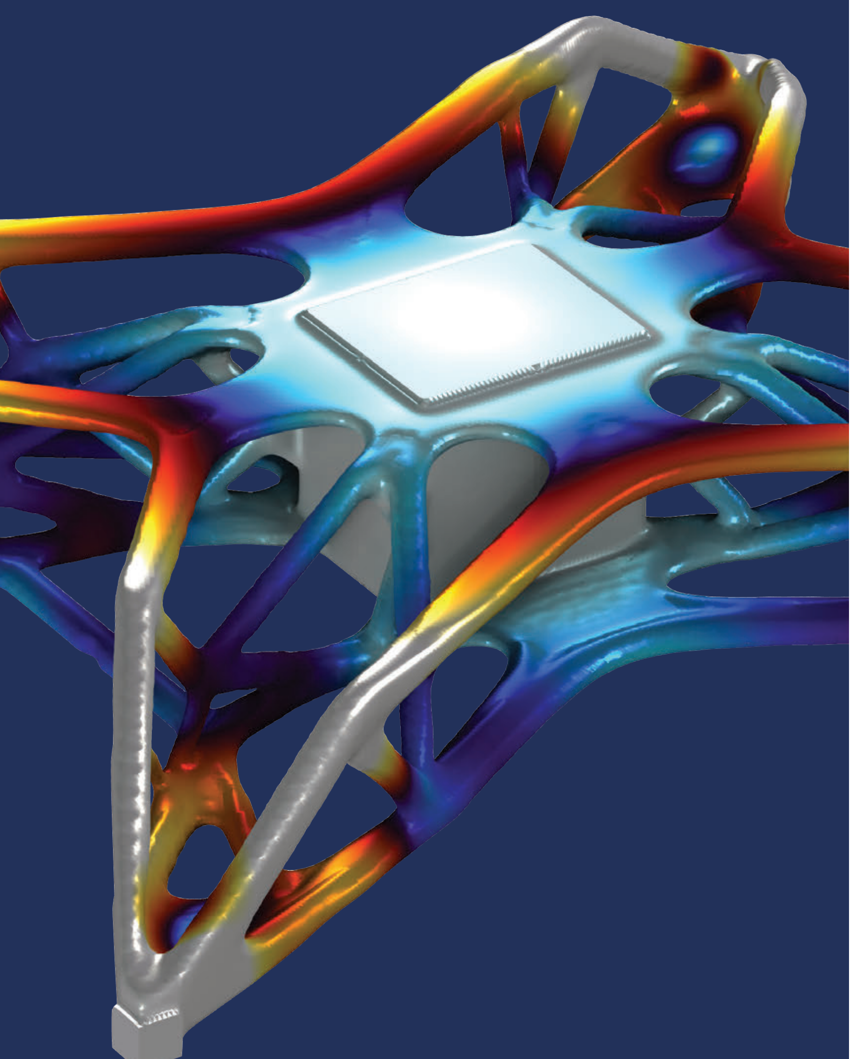
The exoplanet detector is an absolute delight for the frugal DIYer. Schneider took a camera he already had and added a vintage telephoto lens (which he bought on eBay for just $92, because by that time it was worthless to most photographers), then mounted it on a "barn door" tracker made of two pieces of plywood and gears yanked out of an old printer, along with some other odds and ends. With a contraption that would have earned him kudos from MacGyver, Schneider managed to track the transit of an exoplanet in the binary star system HD 189733, a type of observation that had escaped even professional astronomers until 2002.

"It's just that those vibrations, having traveled long distances through the Earth, have (thankfully) been too small to feel. If I had a suitably sensitive seismometer, though, I'd be able to measure them."

Schneider's training and experience as a geophysicist has informed not only his Hands On projects but also the hundreds of articles he has reported or edited over the years. A great example is his story about how BP dealt with a devastating well blowout in the Gulf of Mexico in 2010. "I was able to include a lot of detail in the midst of the Deepwater Horizon disaster at a time when nobody could interview the folks who were actually doing the work because they were too busy—and on a drill rig in the middle of the Gulf," Schneider recalls.

And the bits and pieces he picked up about seismology during his early career came in handy when he was building the Raspberry Pi seismometer for this month's Hands On, "Detect Quakes With 'Raspberry Shakes,'" on page 16. It's a new project driven by a familiar citizen-scientist investigative motive. "Many earthquakes have...vibrated the ground beneath my feet," he notes. "It's just that those vibrations, having traveled long distances through the Earth, have (thankfully) been too small to feel. If I had a suitably sensitive seismometer, though, I'd be able to measure them."

While this is Schneider's last Hands On as an editor for *Spectrum*, it won't be his last as a contributor. We're looking forward to his next project and the wonders Schneider will help us uncover for ourselves. ■



MARY L. "MISSY" CUMMINGS

Cummings, a roboticist at George Mason University who once flew fighter jets off aircraft carriers, knows well how people use and misuse automation. "Pilots figured out how to use air-to-ground bombing radar to manage the landing on the carrier by themselves," she says. "But the system didn't adjust for the pitching deck, so it set people up for much more lethal approaches." Cummings argues that such weird man-machine interactions plague self-driving cars and other forms of AI [p. 30].

MAXIMILIANO FERRARI

Ferrari and coauthors Ben Ollis and Michael Starke are with Oak Ridge National Laboratory. Ferrari and Ollis are members of the R&D staff, and Starke is an electrical-engineering system integrator. Coauthor Arturo Massol-Deyá is executive director of Casa Pueblo de Adjuntas, a community-based nonprofit in Puerto Rico. In "The Well-Connected Microgrid" [p. 36], they describe new technologies for creating a resilient network of solar-powered microgrids. After validating the approach at ORNL, they plan to conduct a field demonstration next year in Adjuntas.

TRACY H. SCHLOEMER & DANIEL N. CONGREVE

Schloemer is an Arnold O. Beckman Postdoctoral Fellow and Congreve is an assistant professor of electrical engineering, both at Stanford University. They describe their research in triplet-triplet annihilation upconversion, a process used to change the color of light, in "Smashing Photons" [p. 44]. Previously, Schloemer taught high school chemistry and was a lead contributor to the ChemEd X teaching blog. Congreve is a cofounder of Quadratic 3D, a startup that's commercializing 3D-printing technology.

THOMAS TEISBERG

Teisberg is a Ph.D. candidate in electrical engineering working in the Stanford Radio Glaciology lab. He and a friend built their first quadcopter, from scratch, in high school. It flew for about 10 seconds. He has worked on drone-based perception systems for a medical logistics company and is now developing the drone-based, ice-penetrating radar system described in "Seeing Secrets in the Ice" [p. 22].

IEEE Spectrum

EDITOR IN CHIEF Harry Goldstein, h.goldstein@ieee.org

EXECUTIVE EDITOR Jean Kumagai, j.kumagai@ieee.org

MANAGING EDITOR Elizabeth A. Bretz, e.bretz@ieee.org

CREATIVE DIRECTOR
Mark Montgomery, m.montgomery@ieee.org

PRODUCT MANAGER, DIGITAL Erico Guizzo, e.guizzo@ieee.org

EDITORIAL DIRECTOR, CONTENT DEVELOPMENT
Glenn Zorpette, g.zorpette@ieee.org

SENIOR EDITORS
Evan Ackerman (Digital), ackerman.e@ieee.org
Stephen Cass (Special Projects), cass.s@ieee.org
Samuel K. Moore, s.k.moore@ieee.org
Tekla S. Perry, t.perry@ieee.org
Philip E. Ross, p.ross@ieee.org
David Schneider, d.a.schneider@ieee.org
Eliza Strickland, e.strickland@ieee.org

ART & PRODUCTION
DEPUTY ART DIRECTOR Brandon Palacio, b.palacio@ieee.org
PHOTOGRAPHY DIRECTOR Randi Klett, randi.klett@ieee.org
ONLINE ART DIRECTOR Erik Vrielink, e.vrielink@ieee.org
PRINT PRODUCTION SPECIALIST
Sylvana Meneses, s.meneses@ieee.org
MULTIMEDIA PRODUCTION SPECIALIST
Michael Spector, m.spector@ieee.org

NEWS MANAGER Margo Anderson, m.k.anderson@ieee.org

ASSOCIATE EDITORS
Willie D. Jones (Digital), w.jones@ieee.org
Michael Koziol, m.koziol@ieee.org

SENIOR COPY EDITOR Joseph N. Levine, j.levine@ieee.org
COPY EDITOR Michele Kogon, m.kogon@ieee.org
EDITORIAL RESEARCHER Alan Gardner, a.gardner@ieee.org
CONTRACT SPECIALIST Ramona L. Foster, r.foster@ieee.org

AUDIENCE DEVELOPMENT MANAGER
Laura Bridgeman, l.bridgeman@ieee.org

CONTRIBUTING EDITORS Robert N. Charette, Steven Cherry, Charles Q. Choi, Peter Fairley, Edd Gent, W. Wayt Gibbs, Mark Harris, Allison Marsh, Prachi Patel, Julianne Pepitone, Lawrence Ulrich, Emily Waltz

THE INSTITUTE
EDITOR IN CHIEF Kathy Pretz, k.pretz@ieee.org
ASSOCIATE EDITOR Joanna Goodrich, j.goodrich@ieee.org

DIRECTOR, PERIODICALS PRODUCTION SERVICES Peter Tuohy

ADVERTISING PRODUCTION MANAGER
Felicia Spagnoli, f.spagnoli@ieee.org

ADVERTISING PRODUCTION +1 732 562 6334

EDITORIAL ADVISORY BOARD, *IEEE SPECTRUM*
Harry Goldstein, *Chair*; Robert Schober, Ella M. Atkins, Sangyeun Cho, Francis J. "Frank" Doyle III, Hugh Durrant-Whyte, Matthew Eisler, Shahin Farshchi, Alissa Fitzgerald, Benjamin Gross, Lawrence O. Hall, Daniel Hissel, Jason K. Hui, Michel M. Maharbiz, Somdeb Majumdar, Lisa May, Carmen S. Menoni, Ramune Nagisetty, Paul Nielsen, Sofia Olhede, Christopher Stiller, Mini S. Thomas, Wen Tong, Haifeng Wang, Boon-Lock Yeo

EDITORIAL ADVISORY BOARD, *THE INSTITUTE*
Kathy Pretz, *Chair*; Qusi Alqarqaz, Stamatis Dragoumanos, Jonathan Garibaldi, Madeleine Glick, Lawrence O. Hall, Harry Goldstein, Francesca Iacopi, Cecilia Metra, Shashi Raj Pandey, John Purvis, Chenyang Xu

MANAGING DIRECTOR, PUBLICATIONS Steven Heffner

DIRECTOR, BUSINESS DEVELOPMENT, MEDIA & ADVERTISING Mark David, m.david@ieee.org

EDITORIAL CORRESPONDENCE
IEEE Spectrum, 3 Park Ave., 17th Floor,
New York, NY 10016-5997 **TEL: +1 212 419 7555**
BUREAU Palo Alto, Calif.; Tekla S. Perry **+1 650 752 6661**

ADVERTISING INQUIRIES Naylor Association Solutions,
Erik Albin +1 352 333 3371, ealbin@naylor.com

REPRINT SALES +1 212 221 9595, ext. 319

REPRINT PERMISSION / LIBRARIES Articles may be photocopied for private use of patrons. A per-copy fee must be paid to the Copyright Clearance Center, 29 Congress St., Salem, MA 01970. For other copying or republication, contact Managing Editor, *IEEE Spectrum.*

IEEE SPECTRUM (ISSN 0018-9235) is published monthly by The Institute of Electrical and Electronics Engineers, Inc. All rights reserved. © 2023 by The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016-5997, U.S.A. Volume No. 60, Issue No. 10. The editorial content of IEEE Spectrum magazine does not represent official positions of the IEEE or its organizational units. Canadian Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40013087. Return undeliverable Canadian addresses to: Circulation Department, IEEE Spectrum, Box 1051, Fort Erie, ON L2A 6C7. Cable address: ITRIPLEE. Fax: +1 212 419 7570. INTERNET: spectrum@ieee.org. ANNUAL SUBSCRIPTIONS: IEEE Members: $21.40 included in dues. Libraries/institutions: $399. POSTMASTER: Please send address changes to IEEE Spectrum, ℅ Coding Department, IEEE Service Center, 445 Hoes Lane, Box 1331, Piscataway, NJ 08855. Periodicals postage paid at New York, NY, and additional mailing offices. Canadian GST #125634188. Printed at 3401 N. Heartland Dr, Liberty, MO 64068, U.S.A. Printed at Spenta Multimedia Pvt. Ltd., Plot 15, 16 & 21/1, Village Chikhloli, Morivali, MIDC, Ambernath (West), Dist. Thane. IEEE Spectrum circulation is audited by BPA Worldwide. IEEE Spectrum is a member of the Association of Business Information & Media Companies, the Association of Magazine Media, and Association Media & Publishing. IEEE prohibits discrimination, harassment, and bullying. For more information, visit https://www.ieee.org/about/corporate/governance/p9-26.html.

News

TELECOM

Is the Wireless Industry Opening Up Locked Systems? › Open RAN fight affects interoperability of wireless tech

BY MICHAEL KOZIOL

At a June meeting in Osaka, Japan, cellular-industry stakeholders gathered to propose solutions to a technical oddity with surprisingly far-reaching consequences. At stake was who calls the shots when it comes to defining interoperability: big-name vendors, smaller manufacturers of specialized components, cell-service providers, or a mixture across the entire industry.

The interoperability struggle has led to the Open RAN movement, whose supporters hope to disrupt the wireless-industry hierarchy and allow more companies to take more significant roles in network infrastructure.

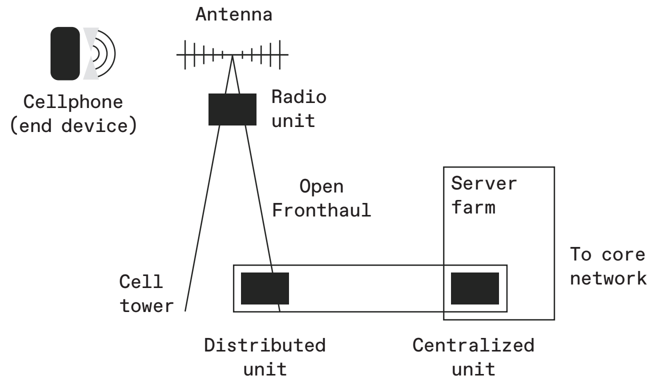
A radio access network (RAN) is the portion of a cellular network (think cell towers) that connects individual devices, like phones, to a central, wired core network. Open RAN wants to make the interfaces between individual RAN components "open"—capable of interacting with one another regardless of who made each component. The idea runs contrary to traditional RAN development, in which a vendor like Ericsson, Huawei, or Nokia would build an end-to-end network that would not interface with another vendor's components.

The Open RAN movement gained steam in 2018 with the formation of the O-RAN Alliance, based in Alfter,

CHRISTOPH DERNBACH/PICTURE ALLIANCE/GETTY IMAGES

This 5G cellphone tower undergoing installation in Germany in 2020, uses a proprietary system called a radio access network (RAN) that transmits, receives, and processes wireless data. Efforts are now underway to open up RANs to provide greater versatility and access.

Split Characteristics



A wireless network's RAN (radio access network) functions as a middleman, connecting end devices like cellphones to the larger world. Open RAN proponents want the interfaces between RAN components, notably the radio unit (RU), distributed unit (DU), and centralized unit (CU), to be standardized so that components from different companies can be mixed and matched. The most popular division, or "split," is called 7.2x and prioritizes creating a flexible (hence the "x") interface called Open Fronthaul between the RU and the DU.

Germany. Which is not to say the entire industry was on board immediately. Indeed, the industry was initially divided into two camps by the issue.

On one side were the vendors that build the network components and seek to bake in competitive advantage by making their systems incompatible with another vendor's equipment. On the other side were the network operators—think AT&T, Deutsche Telekom, Orange, or any other cell-service provider—that wanted the opportunity to mix-and-match components and avoid getting locked into one vendor's ecosystem, even across cellular generations.

There was also a hope that opening up the interfaces would allow smaller vendors to enter the market. These vendors would theoretically be able to focus on building one component really well and not have to worry about customers passing them over because they couldn't easily integrate their equipment into an end-to-end system.

Open RAN's progression over the past several years has seemed, at times, both breakneck and stuck in the mud. The O-RAN Alliance, for example, has gone from just five founding members to well over 300 participants in just half a decade, and the group already has 101 publicly available Open RAN specifications, with more being developed by the organization's technical groups.

Open RAN's progression over the past several years has seemed, at times, both breakneck and stuck in the mud.

Whereas half a dozen "splits"—ways to divide up RAN components to implement open interfaces—have already been explored across the industry, subsequent developments have zeroed in on a specific split called 7.2x that creates the Open Fronthaul Interface. Open Fronthaul moves data between two RAN components called the radio unit—such as the antennas at the top of a cell tower—and the distributed unit, which checks for errors and duplicated data, among other tasks.

Despite 7.2x's ascendancy, progress in other directions has slowed as vendors and operators disagree on what counts as a sufficiently "open" interface. Overall investment in Open RAN deployments has fallen: Analysts at Dell'Oro Group recently estimated that revenue from Open RAN will account for only 15 percent of the global RAN market by 2027, which is 5 percent less than they had previously projected. And while Vodafone in the United Kingdom announced earlier this year—following a 2020 order from the UK government to rip and replace Huawei components by 2027—that it would install Open RAN components in 2,500 cell sites, the company is opting to replace far more (3,500 sites) with Ericsson equipment.

Open RAN requires new cellular deployments, and outside of rip-and-replace scenarios, the wireless industry isn't eager for more. After all, the entire industry has just finished its monumental, multiyear effort of initial 5G rollouts. "Most operators that I'm familiar with in Western Europe and in the U.S. will probably not for the next five to seven years really start massively deploying something else," says Kim Larsen, a wireless-industry advisor who was previously the chief technology and information officer for T-Mobile in the Netherlands. That kind of timeline aligns with when many network operators will begin thinking about 6G deployments, which is why Open RAN may find a larger role in that generation.

Which brings us back to Osaka. There are still plenty of technical questions that require answers as Open RAN continues to take shape. On the agenda in Japan was a specific question about how to incorporate massive MIMO (short for multiple-input, multiple-output) antenna arrays, which incorporate large numbers of antennas to collectively beam precise signals to devices.

At issue was the fact that massive MIMO arrays weren't playing nicely with open fronthaul interfaces. The short version is that because of the particulars of split 7.2x, Open Fronthaul, when paired with massive MIMO, would have to handle too much data traffic. Vendors and network operators were seeing performance degradation up to 40 percent compared to single-vendor RAN installations.

Massive MIMO has seen widespread use in 5G networks and could play an even bigger role in 6G networks, so it's important to make sure it will work with Open Fronthaul. At the Osaka meeting, O-RAN Alliance members agreed to adopt two solutions to the problem as "operation modes" that could be selected, depending on the needs of a specific network operator.

The expectation is that the large vendors will just implement both operation modes into their RAN interfaces. The benefit is clear: Rather than developing, manufacturing, and selling two varieties of components, they can provide one solution to any network operator's needs. The trade-off is that the components on either side of Open Fronthaul have become more complex, with duplicated features and functions.

More notable than any specific technical agreement, however, is how the compromise in Osaka is indicative of the larger trend happening in Open RAN's development: After initially resisting the movement, large vendors are now actively engaged in the process. Regarding the Osaka agreement, analyst Caroline Gabriel at Analysys Mason wrote, "With the exception of Mavenir, the list of contributors could be related to any traditional RAN standards work." (Gabriel did not respond to requests for comment.)

Despite the influx of participation by big players, the O-RAN Alliance says that all players will continue to have an equal opportunity to contribute.

Larsen says it's not accurate to view the industry as entirely recoalescing around the usual vendors. "I don't think it necessarily means that if you have been a startup or a smaller player that everything is lost," he says. "I think you probably will see a segmentation. Some, and that might be the bigger, usual people on the block like Nokia, Ericsson, and Samsung, will focus on the big incumbent players. And the smaller startups will focus on private networks, which is a really growing business." ■

AEROSPACE

Whose Sky Is It Anyway? ›

Regulators ponder how to accommodate eVTOLs and delivery drones

IN JULY, THE U.S. Federal Aviation Administration released its *Advanced Air Mobility (AAM) Implementation Plan*. The 40-page document lays out the FAA's preliminary vision for how U.S. aviation will transition to an era in which a new breed of aircraft, called electric vertical takeoff and landing (eVTOL), will be able to safely share the skies with everything else that now flies: passenger airliners, light aircraft, business jets, helicopters, as well as cargo, delivery, and other kinds of drone aircraft.

One of the issues regulators are now grappling with is how to manage the airspace in the vicinity of busy airports, which will be shared by most of those aircraft types. Several proposals have suggested dividing the airspace into vertical layers, with the highest, above 6,500 feet, reserved for commercial aircraft and the lowest, below about 400 feet, for drones. A middle layer would be reserved for Urban Air Mobility (UAM), including eVTOLs and personal air vehicles, or PAVs. **—Glenn Zorpette**

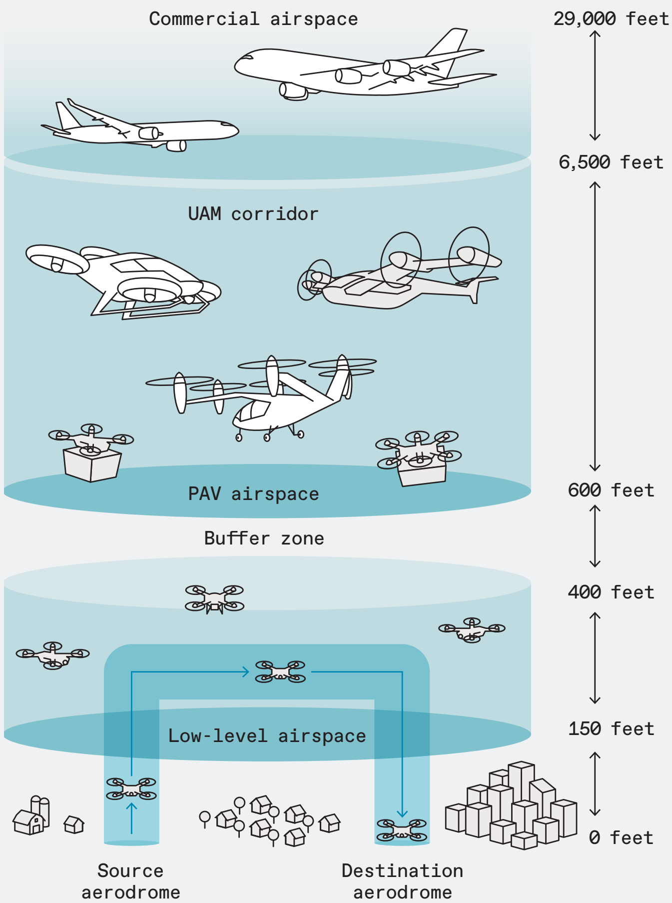


Illustration by James Provost

JOURNAL WATCH

Digital Twins Model Hydrogen's Growth

Hydrogen has great promise as a low-emission fuel source. It burns cleanly, generating only water as a by-product, and when it's produced through electrolysis by splitting water into hydrogen and oxygen, the entire life cycle can be very environmentally friendly.

Hydrogen, however, has never reached its potential as a renewable fuel, mostly because its production process drives up the cost. Most hydrogen today is still produced in conjunction with fossil-fuel refinement, but it releases methane with carbon monoxide as a result. Meanwhile, the hydrogen produced by electrolysis represents less than 1 percent of all the world's hydrogen production.

Sharaf AlSharif, a researcher at the Oldenburger OFFIS Institute for Information Technology in Oldenburg, Germany, believes digital twins could help bring down the cost of clean hydrogen production. Digital twins are computer-simulation programs that can track and optimize the operations of a physical device. Such close monitoring of hydrogen electrolyzers could help streamline the devices' operation and bring down the cost of electrolysis.

Digital twins can monitor components like electrodes, membranes, or pumps to detect imminent failure and proactively make recommendations for scheduling maintenance. AlSharif says that such predictive capabilities could save operators many hours of production time by avoiding downtime for unscheduled troubleshooting.

To encourage the adoption of digital twins for electrolysis, AlSharif and his colleagues at OFFIS presented a software architecture—a kind of architectural blueprint—for developing electrolysis-monitoring digital twins at ETG Congress 2023 held in Kassel, Germany, in June.

Future research, AlSharif says, might study the use of digital twins to enhance electrolyzers' manufacturing process. To truly scale up clean hydrogen production, large electrolyzer manufacturing, which is currently a semi-manual process, needs to scale up as well, he says. **—Tammy Xu**

Finely ground iron powder is more energy dense than gasoline, but it emits no greenhouse gases when combusted.

KROLS MEDIA

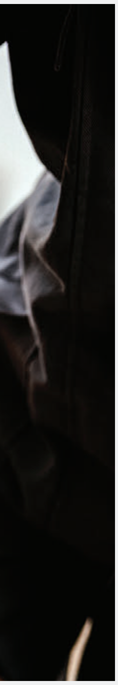
ENERGY

Iron Fuel Shows Its Mettle › It's abundant, recyclable, and has a higher energy density than gasoline

BY PRACHI PATEL

At the end of June, a large 1-megawatt plant that burns iron fuel was fired up, producing the heat needed to brew beer at the Swinkels brewery near Eindhoven, Netherlands. The plant, named Iron+, is a joint venture between three companies and uses technology first demonstrated as a 100-kilowatt system in 2020 by the Metal Power Consortium, which includes the Eindhoven University of Technology and startup Metalot, which was spun out of the university.

Because even metals can burn if you grind them into fine powders, and they do so without emitting toxic or planet-warming emissions, they're a potentially attractive fuel for producing clean power—one that can be easily stored and transported.



RIFT (Renewable Iron Fuel Technology), another spin-off out of Eindhoven, recently demonstrated that it could heat 500 homes using its own iron-fuel technology. Meanwhile, in Canada, startup Altiro Energy, launched by McGill University researchers, has run a prototype 10-kilowatt iron-fuel plant that it now plans to scale up.

Iron powder is an ideal alternative to carbon fuels, says Jeff Bergthorson, a mechanical engineering professor at McGill and the chief scientific advisor for Altiro. Bergthorson and collaborators at the European Space Agency developed the metal-fuel concept and published their report in the journal *Applied Energy* in 2015.

In addition to being the most abundant metal on Earth, iron has an energy density of about 11.3 kilowatt-hours per liter—better than gasoline. Burning iron powder produces heat that can be used directly or converted into electricity by a steam turbine, leaving behind iron oxide, or rust. This can later be reduced—by stripping the oxygen away—back into iron powder. “You can think of iron fuel as a clean, recyclable coal,” says Bergthorson.

Iron oxide can also be reduced to iron using hydrogen, which is already a carbon-free green fuel if produced by splitting water using renewable electricity. But hydrogen is also an ultralight, voluminous gas, so it must be converted into liquid using high pressures and extreme cold. The liquid then must be stored and transported in special containers. Iron, by contrast, is already moved in dry containers for a lower cost.

So, while both hydrogen and metals are essentially ways to store energy, using metals makes more sense, Bergthorson says. Technical assessments by Metalot and the Technical University of Darmstadt suggest that “it’s more efficient to produce iron from hydrogen gas than to produce liquid hydrogen.”

That’s not to say iron fuel doesn’t come with its own challenges. It does not ignite as easily as hydrocarbon fuels, and iron’s flame is unstable, slow, and prone to going out. Altiro gets around this problem by adding a little natural gas to ignite the iron powder when the boiler first starts up. It has also come up with a technology to better stabilize the flame so that it burns longer, Bergthorson says.

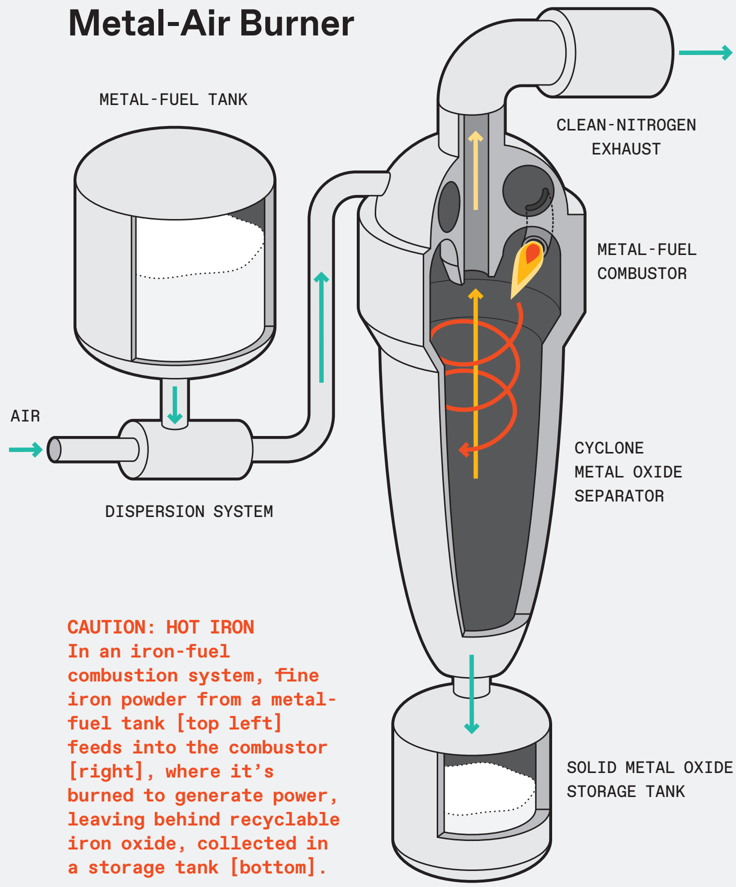
Collecting the resulting iron oxide can also be tricky. But Altiro’s technology ensures the formation of iron oxide particles that are large enough to easily capture “using cyclones and other methods without needing high-tech or costly equipment,” Bergthorson says.

Some of the iron powder inevitably evaporates to form iron oxide nanoparticles that cannot be readily collected and turned back to iron. Both Altira and Iron+ have worked out ways to minimize this nanoparticle formation in order to reduce metal loss. “We improved the boiler by increasing the efficiency of heat transfer,” says Philip de Goey, a mechanical engineering professor at Eindhoven, cofounder of Metalot, and chief technical advisor at RIFT. “The nanoparticle emissions have been decreased by a factor of 10, so [the loss] is less than 0.3 percent. Furthermore, the nanoparticles are not emitted in the atmosphere but captured in a HEPA filter.”

The next step for iron fuel is bolstering the technique’s conversion efficiency, De Goey says. Then there are market hurdles to jump, such as increasing iron-powder production and lowering the cost of producing green hydrogen.

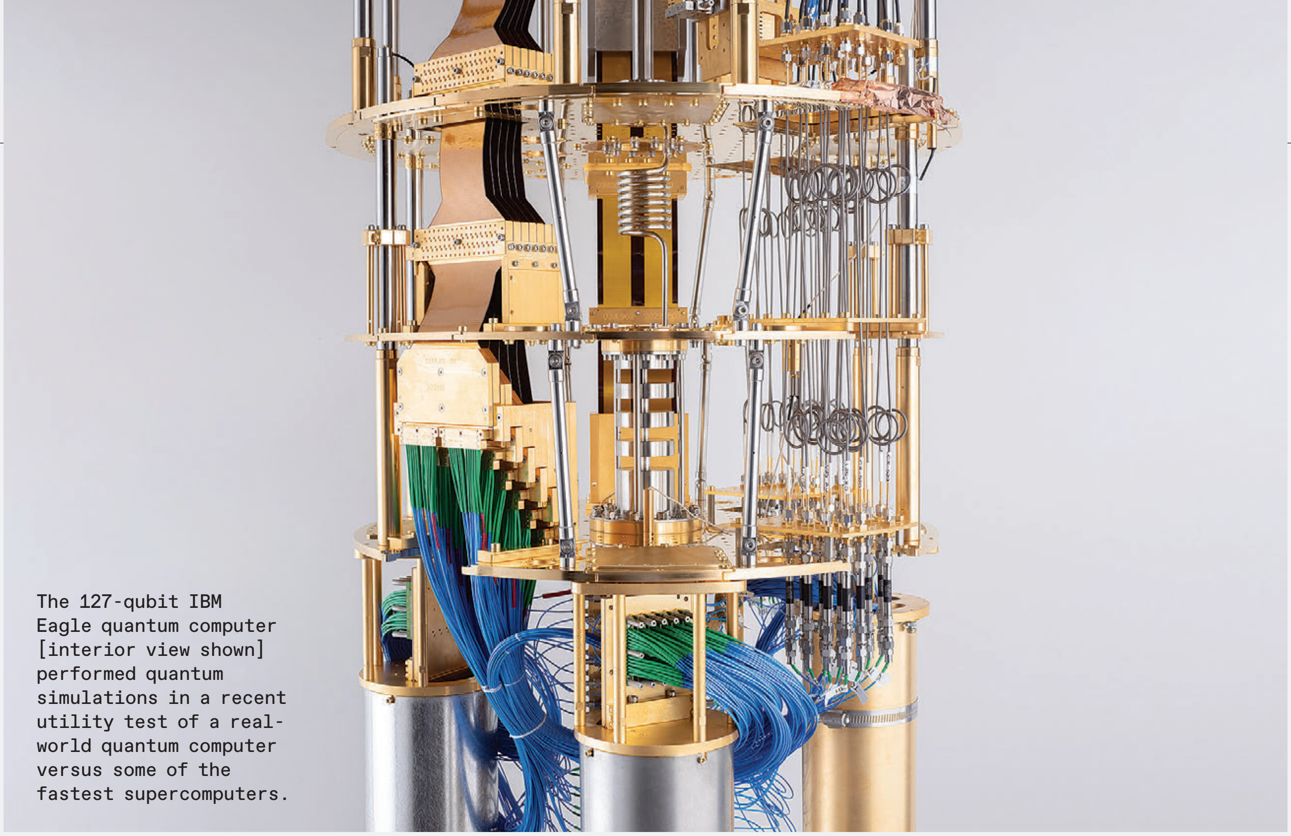
If these problems can be overcome, you could use renewable electricity to produce iron powder, store it as long as necessary, conveniently transport it to power plants, and then burn it when needed, says Bergthorson. “Places that have excess energy could make iron, and others can buy it. This way, you could commodify renewable energy so it can be globally distributed without the need for transmission lines. Metals can solve a big problem in the renewable energy transition: long-duration energy storage.” ■

Metal-Air Burner



CAUTION: HOT IRON
In an iron-fuel combustion system, fine iron powder from a metal-fuel tank [top left] feeds into the combustor [right], where it’s burned to generate power, leaving behind recyclable iron oxide, collected in a storage tank [bottom].

Illustration by Chris Philpot



The 127-qubit IBM Eagle quantum computer [interior view shown] performed quantum simulations in a recent utility test of a real-world quantum computer versus some of the fastest supercomputers.

COMPUTING

IBM's Quantum Computer Can Beat a Supercomputer—Sometimes › Its 127-qubit Eagle handles problems that defy classical supercomputers

BY CHARLES Q. CHOI

Quantum computers are currently so prone to error that their ultimate utility is often questioned. But IBM argues that quantum computing may be entering a new era of utility sooner than expected, with its 127-qubit Eagle quantum computer potentially delivering accurate results on useful problems beyond what even today's supercomputers can tackle.

Though quantum computers can, in theory, find answers to problems that classical computers would take thousands of years to solve, they're dogged by one key problem: Today's quantum computers are notoriously vulnerable to disruption from the slightest disturbance. These so-called noisy intermediate-scale quantum (NISQ) platforms typically suffer roughly one error every 1,000 operations, and many practical applications demand error rates lower by a factor of a billion or more.

Now, IBM reveals that its Eagle quantum processor can accurately simulate physical phenomena that regular computers find difficult to model past a certain level of complexity. Not only are these simulations of actual use to researchers, the company says, but the methods they developed could be applied to other kinds of algorithms running on quantum machines today.

Although quantum computing is still considered to be in its early stages, previous experiments by Google and others claimed that quantum computers may have entered the era of "quantum advantage," "quantum primacy," or "quantum supremacy" over typical computers. Critics in turn have argued that such tests showed only that quantum computers were able to outperform classical machines on contrived problems. Consequently, it remains hotly debated whether quantum computers are good enough to prove useful right now.

IBM's quantum computer modeled the dynamics of the spins of electrons in a material with the aim of predicting its properties, such as magnetization. This model is one that scientists under-

IBM

stand well, which made it easier for the researchers to validate the accuracy of the quantum computer's results.

At the same time, scientists at the University of California, Berkeley, performed versions of these simulations on classical supercomputers to compare how well the quantum computer performed. They used two sets of techniques. Brute-force simulations provided the most accurate results, but also demanded too much processing power to simulate large, complex systems. On the other hand, approximation methods could estimate answers for big systems, but they generally proved less and less accurate the larger a system got.

At the largest scale examined, the quantum computer was roughly three times as fast as the classical approximation methods, finding answers in 9 hours compared with the classical machines' 30. More importantly, the researchers found that as the scale of the models increased, the quantum computer matched the classical brute-force simulations, while the classical approximation methods became less accurate.

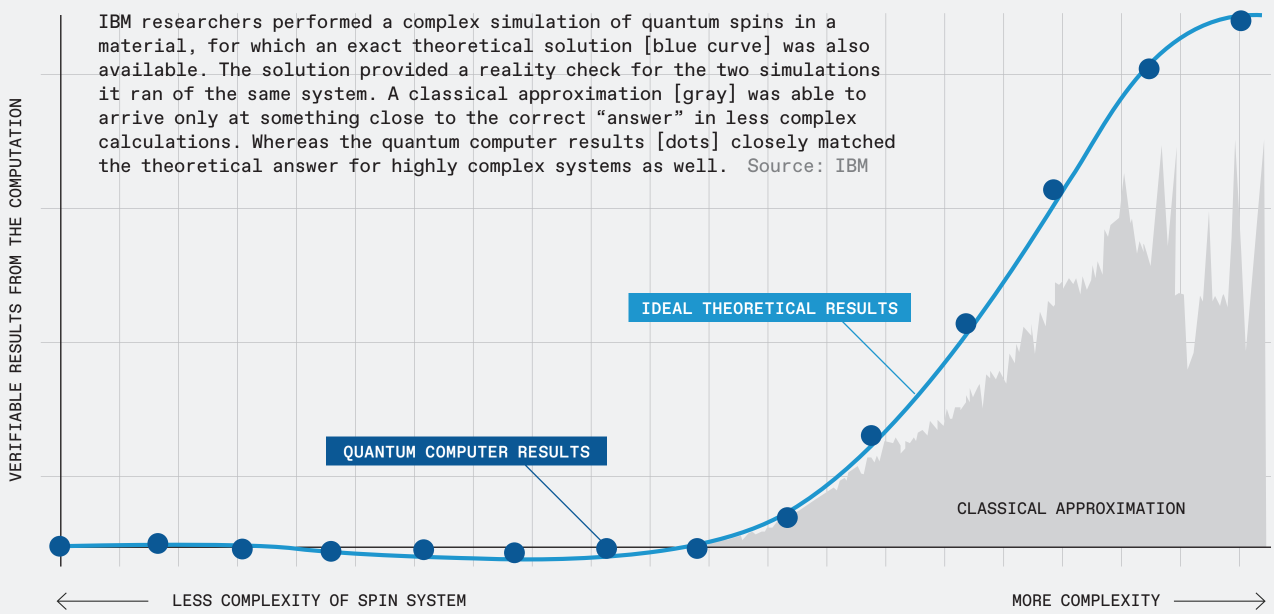
The IBM scientists caution that they are not claiming their quantum computer is better than classical computing. Future research may soon show that regular computers may find correct answers for the calculations used in these experiments, they say.

IBM notes that its quantum hardware displayed more stable qubits and lower error rates than it has previously. However, the new findings depended on what IBM calls "quantum error mitigation" techniques, which examine a quantum computer's output to account for and eliminate noise that its circuits experienced.

"Both our hardware and our error-mitigation methods are now at the level they can be used to start implementing the overwhelming majority of the near-term algorithms that have been proposed in the last 5 to 10 years, to see which algorithm actually provides a quantum advantage in practice," says Kristan Temme, a quantum physicist at IBM's Thomas J. Watson Research Center, in Yorktown Heights, N.Y.

One drawback of the quantum error mitigation strategy that IBM employed is that it requires a certain amount of redundancy. "For the zero-noise extrapolation method we have used here, we need to run the same experiment at three different noise levels," Temme says. "This is a cost that has to be paid for every data point in the calculation—that is, each time we use the processor."

Nevertheless, IBM says its quantum computers running both on the cloud and on-site at partner locations in Japan, Germany, and the United States will be powered by a minimum of 127 qubits over the course of the next year. ■



IBM researchers performed a complex simulation of quantum spins in a material, for which an exact theoretical solution [blue curve] was also available. The solution provided a reality check for the two simulations it ran of the same system. A classical approximation [gray] was able to arrive only at something close to the correct "answer" in less complex calculations. Whereas the quantum computer results [dots] closely matched the theoretical answer for highly complex systems as well. Source: IBM

ROBOTICS

Pneumatic Muscles Power Robotic Cheetah › Bursts of air could push legged robots to high speeds

BY EVAN ACKERMAN

Inspired by the high-speed maneuvering of cheetahs, roboticists at the University of Cape Town have started experimenting with the old-school sibling of hydraulic actuators: pneumatics. By using gas as a working fluid instead of a liquid, you can get a high force-to-weight ratio in a relatively simple and inexpensive form factor with built-in compliance—meaning the working fluid can be compressed and return to its original volume when the pressure is released—that hydraulics lack. Are pneumatics easy to control? Nope! But to make a robot run like a cheetah, it turns out that complicated control may not even be necessary.

Hydraulics are complicated and expensive. And while the noncompliant nature of hydraulics makes them easier to model and control, it also makes them less forgiving in real-world use. That's why, in the 1980s, when Marc Raibert was developing dynamic legged robots at MIT, he and his team relied on pneumatics rather than hydraulics, because pneumatics were much easier to implement.

Nowadays, pneumatics have fallen out of favor, mainly because air is compressible, which messes up most traditional control methods. "Fine force control is difficult with this actuator, and most have avoided it," explains Amir Patel, the associate professor who heads the Cape Town research group. "When looking at animals that require explosive motion from their limbs, we thought that pneumatics would be a good, and often overlooked, actuator."

Patel has done an enormous amount of research on cheetah biomechanics. He has learned enough to suggest a new approach to cheetah-inspired locomotion. "From our years studying cheetahs here in South Africa, it appears as if they're not really trying to do fine force control when accelerating from rest,"

GETTY IMAGES

Patel says. "They're just pushing off as hard as they can—which makes us think that a pneumatic on/off actuator [also known as a bang-bang controller] could do that job."

Patel and his coauthors have built a legged robot—in this case a two-legged model—called Kemba. The intent is to explore the kind of rapid acceleration and maneuverability that pneumatics can offer. Kemba's hips incorporate high-torque quasi-direct-drive electric motors for higher-fidelity positioning, with high-force pneumatic pistons attached to the knees. While the electric motors give the kind of precise control that we've come to expect from them, the pistons are controlled by simple (and cheap) binary valves, which can be either on or off.

With a boom for support, the 7-kilogram Kemba can repeatedly jump to 0.5 meter with a controlled landing. For a single burst, it reaches a maximum jump height of 1 meter. But maximizing jump height and top speed is not what the research is necessarily about, explains Patel. "With Kemba (and all the robots and animals we study in my lab) we focus on the transient phase of the locomotion—like rapid acceleration from a standstill or coming to rest once you're at a high-speed gait. Most papers don't really concentrate on that phase of the motion. I would love for more labs to be publishing their results in this area so that we can have some metrics (and data) to compare to."

"Getting Air: Modelling and Control of a Hybrid Pneumatic-Electric Legged Robot," by Amir Patel and coauthors was presented in June at the 2023 IEEE International Conference on Robotics and Automation, in London. ■

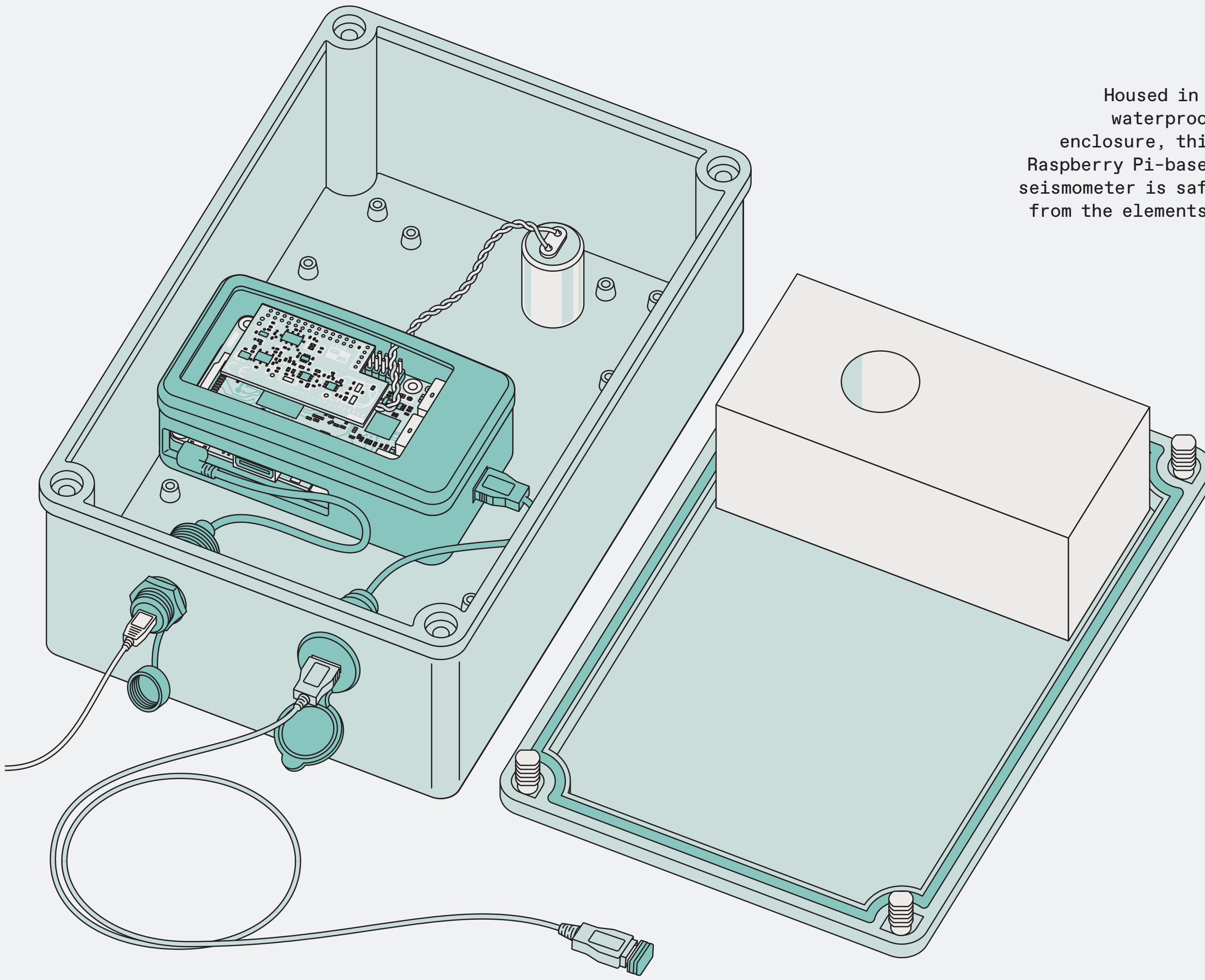
India's Solar Center

By Willie D. Jones

India is a beacon of light when it comes to pursuing the switch from fossil fuels to solar power. The country overachieved, having met its goal of reaching 20 gigawatts of installed solar capacity before its self-imposed 2020 deadline. The Indian government immediately moved the goal post, setting 100 GW as the new target. To symbolize the nation's commitment to solar power, the West Bengal Housing Infrastructure Development Corp., or HIDCO, has erected this dome in Eco Park, New Town, Kolkata. The 29-meter-high, 45-meter-diameter dome has 2,000 solar panels that together generate 180 kilowatts—enough to light and cool the facility and power nearby streetlamps. Inside, HIDCO maintains exhibits designed to inform visitors about the benefits of renewable energy sources such as tidal energy, geothermal, wind, and of course, solar.

PHOTOGRAPH BY DIBYANGSHU SARKAR/AFP/GETTY IMAGES

Hands On



Housed in a waterproof enclosure, this Raspberry Pi-based seismometer is safe from the elements.

Detect Quakes With "Raspberry Shakes" ›

A Raspberry Pi–based seismometer can detect even distant earthquakes

BY DAVID SCHNEIDER

I have only once felt an earthquake—in 1985, when a magnitude-4 temblor occurred just north of New York City. It wasn't until I heard the news reports later that I realized the vibration that had awakened me at 6 a.m. was, in fact, a small earthquake.

Many earthquakes have since vibrated the ground beneath my feet. It's just that those vibrations, having traveled long distances through Earth, have (thankfully) been too small to feel. If I had a suitably sensitive seismometer, though, I'd be able to measure them.

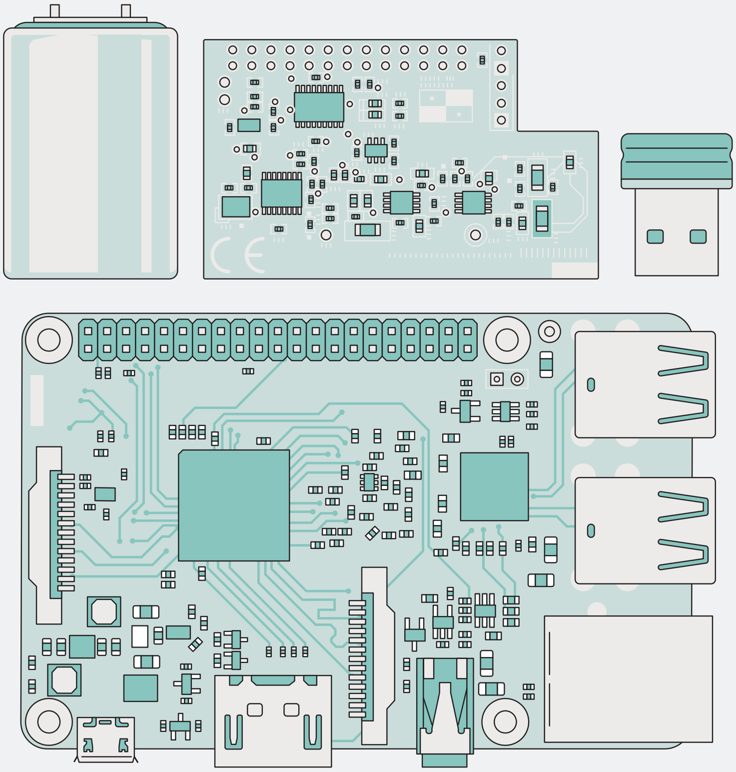
Illustrations by James Provost

I recently decided that I needed to give this a try. Searching the interwebs, I found no shortage of leads about how to build a DIY seismometer. Fundamentally, these consist of a magnet attached to a mass, with a nearby pickup coil. The mass is suspended so that it remains largely motionless when the ground shakes. The shaking does vibrate the coil, however, inducing a voltage in it due to its relative motion through the magnet's magnetic field. The problem is that the DIY seismometer designs I was seeing were large and ungainly contraptions. I wondered whether I could build a more compact one using a geophone.

Geophones are commonly used , in the oil and gas industry for seismic surveying, where the seismic waves are artificially generated to probe the ground below. On land, special trucks—called "thumpers"—do the job. The seismic waves they produce reflect back up from layers of rock and are sensed using geophones.

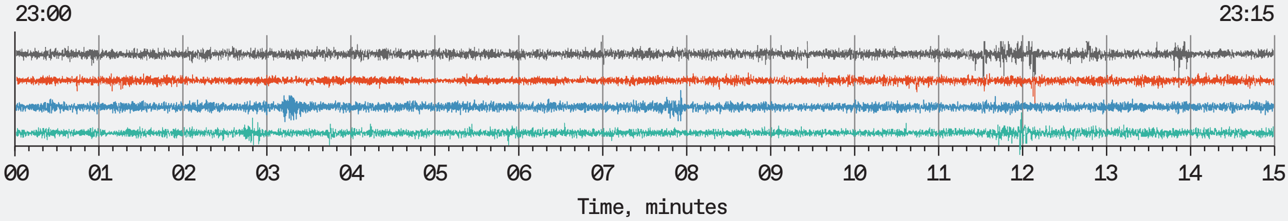
A search of eBay showed that geophones could be had inexpensively. The rub, I soon realized, is that geophones aren't designed to pick up the low frequencies found in teleseismic waves from distant earthquakes. These range from about one cycle per second (1 hertz) down to a fraction of a cycle per second. Most geophones are designed for measuring frequencies above 10 Hz. The lowest-frequency models generally available are for 4.5 Hz.

Further investigation, though, revealed that some clever electronic signal conditioning could extend the range of a geophone to lower frequencies. I was all set to pursue this strategy when I discovered that somebody had beat me to it. Actually, a whole community of (mostly) amateur seismologists had, using a Raspberry Pi–based device called a Raspberry Shake, developed in 2016 by a group in tectonically active Panama. The Raspberry Shake effort has grown to include users worldwide who share seismic data. Even some professional seismologists use Raspberry Shakes because they are inexpensive as seismometers go.



The required electronics consist of a geophone [top left], a signal-conditioning and A/D board [top middle], a Wi-Fi dongle [top right], and a Raspberry Pi Model 3B+ [bottom].

The Raspberry Shake folks offer a variety of configurations. I purchased the most bare-bones package for about US $175. This consists of a geophone and a sensor board that plugs into a Raspberry Pi. I used a Raspberry Pi Model 3B+.



The Raspberry Shake's Web interface makes it easy to view recorded data, presented in what seismologists call helicorder format. This portion of the data for 6 September 2023 includes the time at which teleseismic waves from a moderate-size earthquake in Chile would have reached the recording site, at about 23:59 UTC, which is shown at the 14-minute mark in the green trace [bottom]. No obvious earthquake signal is seen at that point, though.

I housed the unit in a waterproof enclosure, in which I had installed one bulkhead connector for 5-volt power and a second one for USB, so that I could plug in a Wi-Fi dongle that was physically separated from the Raspberry Pi. (The Raspberry Shake people advised not using the Model 3B+'s built-in Wi-Fi, which apparently causes data glitches.)

Setting up my Raspberry Shake, like most Raspberry Pi projects, involved a few magic incantations to the Linux gods. In this case, there were really just two challenges. The first was to get an SD card prepared with the operating system and the Raspberry Shake software. For me the primary approach described in the installation documentation flopped, but the alternative system offered worked just fine.

The second challenge was getting a Wi-Fi dongle set up. The first one I purchased, said to be suitable for Linux, proved a bust. But an older dongle I had on hand worked. Wary of Wi-Fi issues, I first tested my Raspberry Shake in my living room, wired directly to my router. The Raspberry Shake is designed to be used in a so-called headless configuration, which eliminates the need for a display: You can connect to it remotely using SSH or via a nifty Web interface. So in no time I was able to see data the Raspberry Shake was recording.

Letting it record the shaking caused by people walking around my house revealed mysterious data gaps. Investigating the cause, I discovered that the problem was the power adapter I was using, which couldn't deliver enough current. Once I replaced it, the data outages disappeared.

At this point, I installed the unit on the cement-slab floor of my home's detached garage, figuring that this location would be free of signals caused by anyone walking around the house. Then I left it to gather data until an earthquake was reported somewhere in the world sufficiently large to possibly be detectable.

On 29 August, there was a magnitude-5.5 earthquake in, fittingly, Panama, birthplace of the Raspberry Shake. I consulted a Web page that shows data recorded by a scientific-grade seismic station near my home in North Carolina. This revealed that faint signals from this earthquake had reached my area.

When I looked at the data recorded by my Raspberry Shake, though, it showed no matching signal. I was disappointed but not particularly surprised: Magnitude 5.5 is a pretty wimpy earthquake, after all, and it took place almost 3,000 kilometers away.

I investigated what some other Web-connected Raspberry Shakes had recorded during that earthquake. The farthest one from Panama that registered a good signal was in Puerto Rico. The seismic waves from the Panama earthquake were apparently too small to register on Raspberry Shakes in the continental United States.

Since that time, a larger (magnitude 6.2) quake took place in Chile. The earthquake-magnitude scale is logarithmic, so this was five times the size of the magnitude-5.5 Panama quake. But it was much farther (about 7,400 km) away. And my Raspberry Shake didn't register waves from it either.

So I'm still waiting for a big one. And I'm thankful that, from my East Coast location, I'll only be seeing it as signals on my garage seismometer, rather than as a bunch of rubble in the street. ■

BY EDD GENT

Careers: Jim Vanns

The Industrial Light & Magic engineer wins a Technical Oscar

Jim Vanns was one of 19 people who received a Technical Achievement Award this year for their contributions to the film industry over the last decade.

MAGZ CHIDOTHE

Winning an Academy Award is probably not something most engineers ever expect to do. But it was a fitting honor for Jim Vanns, who won a 2023 Technical Achievement Award—commonly referred to as a Technical Oscar—from the Academy of Motion Picture Arts and Sciences for his work on the high-performance computing systems behind the blockbuster films *Gravity* (2013) and *Guardians of the Galaxy* (2014). He shared the award with Mark Hills, a former colleague at the London-based visual-effects studio Framestore. The pair designed the software to manage the cutting-edge computing cluster that Framestore used to create its computer-generated imagery, or CGI. Their software is still being used today.

"The great thing about the Academy Awards is that they raise awareness of the technology being used, which would be impossible to make films without," Vanns says. "These awards show that the dependence on technology is understood and appreciated, rather than just assuming making movies is all about the creative teams."

Vanns, who now works at Disney-owned Industrial Light & Magic, in London, was an avid cinephile growing up. But he didn't realize until after college that his coding skills would be his entry into the film industry.

"I have always loved movies," he says, "but I never really thought there would be a chance to apply my computer science knowledge to making them." And with the growing importance of visual effects (VFX) in moviemaking, he says, there are increasing opportunities for the more technically minded to make their mark on the silver screen.

"The great thing about the Academy Awards is that they raise awareness of the technology being used, which would be impossible to make films without."

Vanns grew up in Tonbridge, England, and played guitar for several bands as a teenager. He began tinkering with the software used to produce music. The programs often crashed unexpectedly, so he dug into the underlying code to find out why. That was his first taste of computer science. "I got into reverse engineering and understanding how computers worked under the hood," he says.

Meanwhile, his father was gently nudging him to study something with a more "realistic" career path than music. Vanns decided to combine his interests, majoring in computer science and music at Canterbury Christ Church University, in Canterbury, England.

After graduating with a joint honors degree in both fields in 2001, Vanns worked for a time writing software, and seemed to be setting off on a fairly conventional IT career. But his love of cinema set him on a different path. While watching the bonus content on the DVD version of *The Lord of the Rings: The Two Towers* (2002), he was captivated by a discussion about the CGI in an iconic battle scene. The scene's large numbers of computer-generated characters were created on a dedicated cluster of high-performance computers—a render farm—that converted the data from a 3D computer model into a photo-realistic video.

Vanns had assumed that the only computer science jobs in the film industry were related to creating the graphics themselves, something he hadn't studied. But that DVD discussion opened his eyes to the extensive software development and systems management that went into producing CGI. Eager to break into the industry, he began applying for jobs, and in 2006 he was hired as a systems programmer at Framestore.

"It was that 'foot in the door' opportunity that I was after," he says. "I think a lot of people end up falling into my line of work, but for me it was very much a conscious decision to try to get into the industry."

After a few years of learning the ropes and working on a variety of projects, Vanns was assigned to revamp the management software for Framestore's render farm. The company had just been hired for the VFX-heavy *Gravity* and needed to switch to a far more computationally taxing form of rendering called path tracing. The approach simulates the physics of light more faithfully than other approaches do, leading to scenes with more realistic and dynamic lighting.

The project was going to stretch Framestore's render farm nearly to its breaking point, Vanns recalls, so it needed software that could squeeze every drop of efficiency out of the hardware. But the company also wanted to future-proof its systems. Vanns and Hills had to create software that would last a decade and could handle 10 times the workload required for *Gravity* as the company's render farm grew.

At its heart, the challenge they faced was one of resource management, Vanns says. "We had this render farm made up of 100,000 processor cores, but the company was often working on three different shows at the same time," he says. "It was all about how we divvied up the cores."

The goal was to ensure that the processors were used as efficiently as possible, and that different tasks running on the same machines didn't end up competing for resources, like memory. That required some clever scheduling and networking. The team was also given the job of creating a responsive user interface that could provide real-time updates on the progress of rendering jobs.

The resulting system, dubbed the FQ render-farm engine, is what earned Vanns and Hills their Academy Award. The software went live in 2010 and is still in use at Framestore today. "They have had to do very few code changes," Vanns says. "The system still runs just as it was designed."

After completing the FQ project, Vanns wanted a new challenge. He joined the London office of Industrial Light & Magic in 2014 as a senior production engineer. The first project he worked on involved researching whether the company could shift the bulk of its computing workload onto the cloud. That meant imagining how to build the company's entire global VFX infrastructure from scratch using the latest cloud technology.

"It was real 'the world is your oyster,' blue-sky-thinking kind of stuff," he says.

The project was fascinating, Vanns says, but the company concluded that a wholesale shift to the cloud didn't make sense. It would be difficult to migrate the legacy software, and the investments the company had already made in hardware meant it was unlikely to save much money. However, insights gained from the project led to improvements to the company's IT infrastructure. Vanns is currently developing a new data-storage system that resulted from the project.

When it comes to a software-development career in the VFX industry, graphics-related work gets most of the attention. But Vanns says there are plenty of other interesting roles available.

"I don't think it's really well understood how much work is involved in terms of building databases, operating systems, networks, and all that slightly less grand software," he says.

The requirements for the movie industry are similar to those for software engineering jobs in other sectors, Vanns says. An understanding of algorithms and data structures is important. The industry relies heavily on the Linux operating system, so relevant experience is also required.

Those who want to write software to support the creation of CGI will need to learn how the process works, Vanns says. It involves coordination among the various departments specializing in different aspects of VFX, including textures, animation, and lighting, which all have different requirements and workflows.

Most important, though, is the ability to find creative solutions to problems. Complications are inevitable when managing computer systems made up of thousands of devices and used by thousands of people, who are often scattered around the globe.

"I think an underestimated aspect of problem solving is creativity," Vanns says. "Being able to think not only analytically but also creatively about how you might solve a problem is a must." ■

Employer:
Industrial Light & Magic, London

Title:
Principal production engineer

Education:
Joint honors degree in computer science and music, Canterbury Christ Church University, Canterbury, England

5 Questions for Andrew Leland

The history of assistive technology for the blind

Seeing his words on the printed page is a big deal to Andrew Leland—as it is to all writers. But it's also more precious to him than for most people, because Leland is gradually losing his vision due to a congenital condition called retinitis pigmentosa, which slowly kills off the eyes' light receptors.

Leland is the author of the newly released book *The Country of the Blind: A Memoir at the End of Sight*, which presents a history of blindness and some of the related social, political, and technological advances alongside Leland's own experiences from three years of deteriorating sight. He answered five questions from *IEEE Spectrum* about the role technology has played in helping the visually impaired navigate the world around them.

What is the most important technology for visually impaired people in their day-to-day?
Andrew Leland: The fundamental technology of blindness is the white cane. I've heard from blind technologists who will often be pitched new technology that's like, "Oh, we came up with this laser cane, and it's got lidar sensors on it." There are tools like that that are really useful for blind people. But I've heard super-techy blind people say, "You know what? We don't need a laser cane. We're just as good with the ancient technology of a really long stick."

Andrew Leland is a writer, audio producer, editor, and teacher living in Western Massachusetts. His first book, The Country of the Blind: A Memoir at the End of Sight, *was published in July 2023.*

What about Braille—has that changed over the years?
Leland: Braille displays can hook up to a computer so that any text that appears on the screen you can read in Braille. And that brings up the screen reader. It's a piece of software that sits on your phone or computer and takes all the text on the screen and turns it into synthetic speech or into Braille. These days, the speech is a good synthetic voice. Imagine the Siri voice or the Alexa voice; it's like that, but rather than being an AI that you're having a conversation with, it moves all the functionality of the computer from the mouse to the keyboard.

How would writing your book have been different before screen readers and Braille displays existed?
Leland: If you had asked me that question in the 1960s or '70s, then I might have had to write the book in longhand with a really big Magic Marker and fill up hundreds of notebooks with giant print. Or I might have had to use a Braille typewriter. I'm so slow at Braille that I don't know if I actually would have been able to write the book that way. Maybe I could have dictated it. I would then have to have that transcribed and hire someone to read the manuscript back to me as I made revisions. That's not too different from what John Milton had to do [for *Paradise Lost*].

What technological advances made screen readers possible?
Leland: One really important one is optical character recognition (OCR). There's been versions of it stretching back shockingly far—even to the early 20th century, when they created a device in the 1920s called the optophone. The technique was known as musical print. Some blind people read entire books that way.

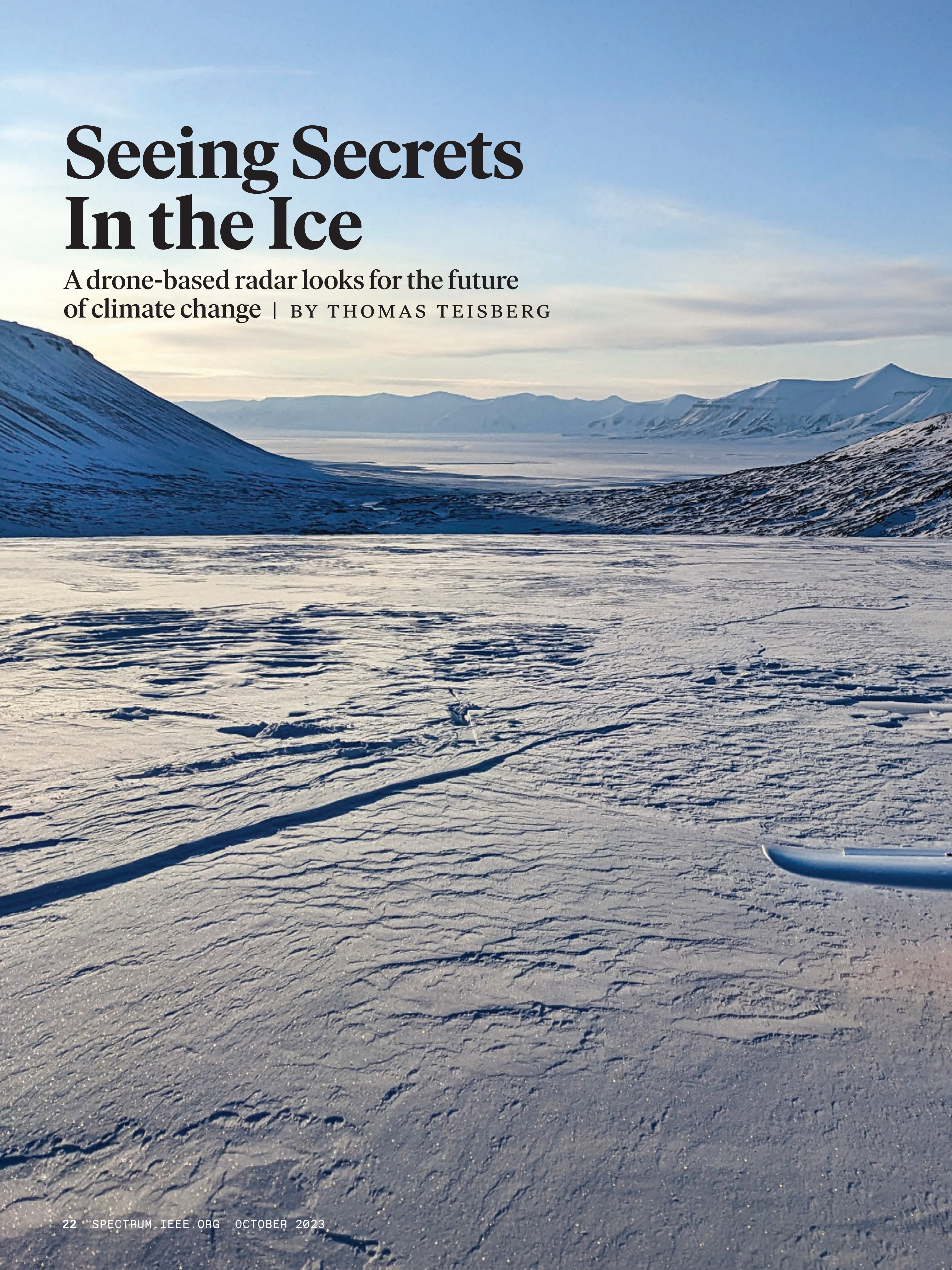
In the 1970s, Ray Kurzweil invented the flatbed scanner and perfected the OCR technology that was nascent at the time. The Kurzweil Reading Machine was not instantaneous but instead converted text to synthetic speech. Suddenly the entire world of print really started to open up to blind people. Now I can take my iPhone and snap a picture of a restaurant menu, and it'll OCR that restaurant menu for me automatically.

What's next in this technological progression?
Leland: Now you have ChatGPT machine vision, where I can hold up my phone's camera and have it tell me what it's seeing. There's a visual interpreter app called Be My Eyes that has partnered with OpenAI, so now a blind person can hold their phone up to their refrigerator and say, "What's in this fridge?" It'll say, "You have three-quarters of a 250-milliliter jug of orange juice that expires in two days; you have six bananas and two of them look rotten." ■

Photo-illustration by Stuart Bradford

Seeing Secrets In the Ice

A drone-based radar looks for the future of climate change | BY THOMAS TEISBERG



Peregrine, a drone-based ice-penetrating radar system, was tested over Norway's Slakbreen glacier in March.

I'M STANDING ON TOP OF 100 METERS OF ICE, watching a drone crisscross the Slakbreen glacier on Norway's Svalbard archipelago, more than 600 kilometers north of the mainland. I'm part of a team testing Peregrine—a fixed-wing unmanned aerial vehicle (UAV) equipped with miniaturized ice-penetrating radar, which can image the glacial ice all the way down to the bedrock below. • It's –27 °C, dipping below –40 °C with wind chill—well below the operating temperature of most of the commercial equipment we brought for this expedition. Our phones, laptops, and cameras are rapidly failing. The last of our computers that is still working is sitting on top of a small heating pad inside its own little tent. • Harsh as the weather is here, we intend for Peregrine to operate in even tougher conditions, regularly surveying the Antarctic and Greenland ice sheets. These great masses store enough water to raise global sea level by 65 meters should they melt entirely. Although neither ice sheet is expected to melt completely anytime soon, their incredible scale makes even small changes consequential for the future of our planet. And the data that Peregrine will gather will help scientists to understand how these critical areas will respond to climate change.

Thomas Teisberg, an electrical engineering Ph.D. candidate at Stanford University, launches Peregrine [opposite page] with fellow Stanford student Danny May at the laptop controller, while a group of students from the University Centre in Svalbard (UNIS) looks on. As Peregrine climbs into the air over the Slakbreen glacier, the system's red antennas are clearly visible under the wings.

SCIENTISTS HAVE LONG LOOKED at changes in the surface height of ice sheets, using data collected from satellite-borne laser altimeters. This data has come in large part from ICESat, launched in 2003, and its successor, ICESat-2, launched in 2018. With information from these NASA satellites, scientists measure the change in elevation, which they use to infer the net impact of surface processes such as snowfall and melting and the rates at which the ice sheets release icebergs into the ocean.

These measurements are important, to be sure, but laser altimetry provides no direct information about what's happening beneath the surface, including how the ice deforms and how it slides over the underlying rock.

And as we try to understand how ice sheets are responding to new climate extremes, these processes are key. How will changes in temperature impact the rate at which ice deforms under its own weight? To what extent will liquid water reaching the bottom of a glacier lubricate its bed and cause the ice to slide faster into the ocean?

Getting answers to these questions requires seeing beneath the surface. Enter ice-penetrating radar (IPR), a technology that uses radio waves to image the internal layers of glaciers and the bed beneath them. Unlike other more labor-intensive methods, such as drilling bore holes or setting up arrays of geophones to collect seismic data, IPR systems from their earliest days have been flown on aircraft.

In the 1960s, as part of an international collaboration, a U.S. Navy Lockheed C-130 Hercules transport was converted into an IPR-data-collection aircraft. The project (which I'll discuss in a little more detail in a moment) showed that it was possible to rapidly collect this type of data from even the most remote parts of Antarctica. Since then, IPR instruments have gotten better and better, as has the means of analyzing the data and using it to predict future sea-level rise.

Meanwhile, though, the aircraft used to collect the data have changed comparatively little. Modern instruments are often flown on de Havilland Canada DHC-6 Twin Otters, which are two-engine turboprops, or Basler BT-67s, which are modified Douglas DC-3s. (Some Baslers flying missions in Antarctica today flew World War II missions in their past life.) And while support for these operations varies by country, the demand for new data is outpacing the ability of crewed aircraft to collect it—at least with a price tag that doesn't put it out of reach for all but the most well-funded operations.

Collecting such data today just shouldn't be that hard.

That's why I and other students in Dustin Schroeder's Stanford Radio Glaciology lab are developing several novel ice-penetrating radar systems, including Peregrine.

Peregrine is a modified UAV carrying a miniaturized ice-penetrating radar that we designed around a software-defined radio. The radar system weighs under a kilogram—featherweight compared with conventional IPR systems,

PREVIOUS PAGES AND THESE PAGES: ELIZA DAWSON

Thomas Teisberg huddles over a laptop computer [left] that's partly shielded from the cold by a small tent. The tripod supports the radio used to communicate with the drone. Later, Teisberg carries Peregrine back to the team after a test flight.

ELIZA DAWSON (2)

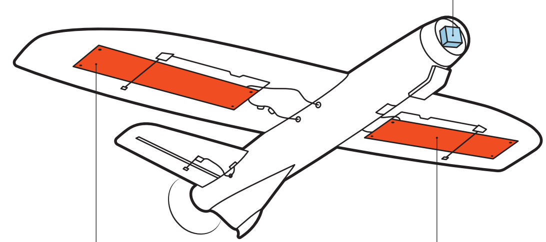
which take up entire equipment racks in crewed aircraft. The whole package—drone plus radar system—costs only a few thousand dollars and packs into a single ruggedized case, about the size of a large checked bag.

But to truly understand why we felt we need to get Peregrine out into the world now, you need to know a bit about the history of data gathering with ice-penetrating radar.

THE FIRST LARGE-SCALE IPR surveys of Antarctica began in the late 1960s when a group of American, British, and Danish geoscientists mounted a set of radar antennas under the wings of a C-130. Predating GPS, the project recorded flight paths using internal navigation systems and known ground waypoints. The system recorded radar returns using a cathode-ray tube modified to scan over a passing reel of optical film, which the researchers supplemented with handwritten notes. This effort produced hundreds of rolls of film and stacks of notebooks.

After the project ended in 1979, various national programs began carrying out regional surveys of both Antarctica and Greenland. Although they were initially limited in scope, these programs grew and, crucially, began to collect digitized data tagged with GPS coordinates.

In the late 2000s, IPR surveying got an unexpected boost. ICESat lost one laser altimeter after just 36 days of data collection in 2003, and by late 2009 all the satellite's lasers had stopped working. Laser altimetry's problems would seem to have no connection to aircraft-based IPR surveys. But with ICESat-2 still years away from launching and a favorable political environment for public earth-science funding in the United States, NASA organized Operation IceBridge, a large-scale aircraft-based campaign to cover the laser-altimetry data gap in Greenland and Antarctica.



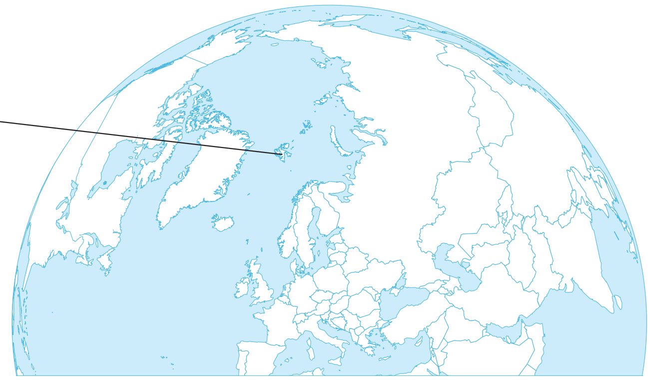
The software-defined radio and other electronics that make up the ice-penetrating radar, shielded to avoid interference with GPS signals, sits in the nose.

Although the primary purpose was collecting laser altimetry, the use of aircraft instead of satellites meant that other instruments could be easily added. At the time, two U.S. institutions—the University of Texas Institute for Geophysics and the Center for Remote Sensing and Integrated Systems (CReSIS) at the University of Kansas—had been developing improved IPR instruments, so IPR was ready to get on board.

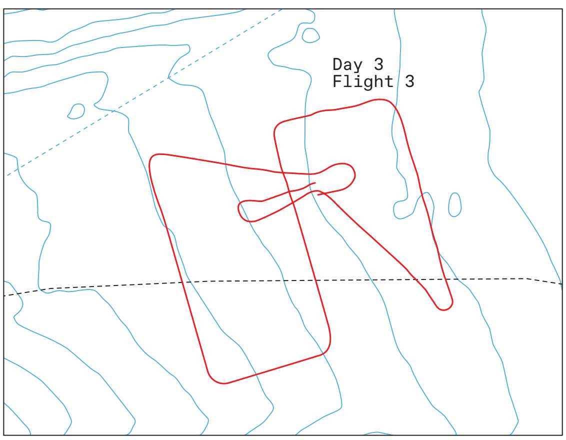
Between 2009 to 2019, the aircraft of Operation IceBridge flew more than 350,000 kilometers over the Antarctic while collecting IPR data. During this same period, the National Science Foundation's Investigating the Cryospheric Evolution of the Central Antarctic Plate (ICECAP) program funded more than 250,000 km of additional Antarctic IPR data.

Operation IceBridge enabled a huge jump in the amount of IPR data collected worldwide. While other organizations around the world also gathered and continue to gather IPR data, particularly the British Antarctic Survey and the Alfred

Illustrations by Chris Philpot



The Slakbreen glacier, located on Norway's Svalbard archipelago [see enlarged view of map] in the coldest part of the country, was selected for testing Peregrine because it was unlikely to contain liquid water, which would interfere with imaging of the bedrock below.



Peregrine flew a pattern [left, red line] spanning an area roughly 0.6 square kilometers over the Tellbreen glacier, also on the Svalbard archipelago. The drone's ice-penetrating radar mapped the ground below the glacier and also the layers within it. The 3-D visualization [right] created from the data shows those layers as faint lines and the bedrock as a brighter line.

BOTTOM RIGHT: STEPHAN MANTLER AND THOMAS TEISBERG

Wegener Institute, IceBridge took U.S.-led data collection from being almost negligible in most years to being the main source of data while the project was in operation.

In 2018, IceSat-2 launched, heralding the end of Operation IceBridge. Some IPR surveying continued, but the rate of data collection since 2018 has significantly lagged the scientific demand for such observations.

Adding to the need for better ice-monitoring tools is a recent shift in the type of IPR data that scientists see as important. Historically, these radar measurements have been used to identify the thickness of the ice above its bed of rock or sediment.

Bed topography, with some exceptions, does not change on time scales relevant to people. So collecting this kind of IPR data could generally be a one-time—or at least infrequent—exercise, ending once enough data was gathered to build a sufficiently detailed map of the bed of a glacier or ice sheet.

But the depth of the ice to the bed isn't the only important information hidden below the surface. For one, IPR data reveals internal layering in the ice caused by changes in the composition of the snow that fell. The shape of these internal layers provides hints about the current and past flows of the ice.

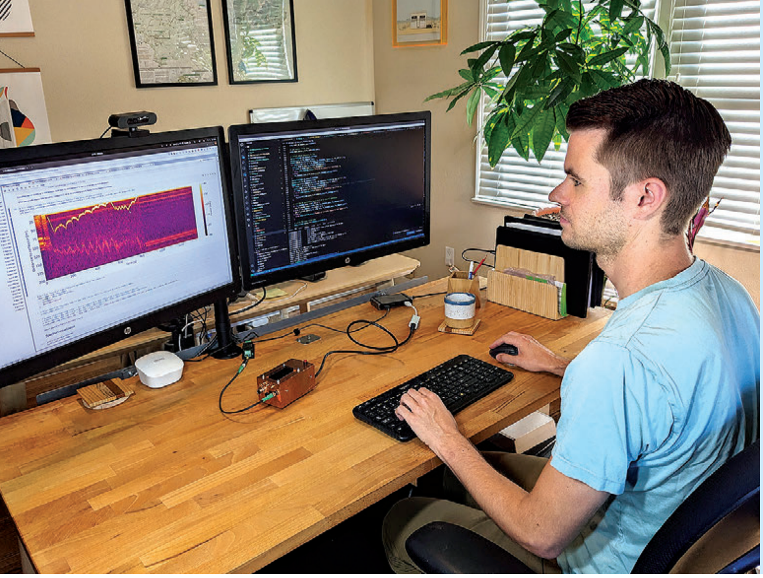
Scientists can also look at the reflectivity of the bed, which can reveal the likelihood of liquid water being there. And the presence of water can give indications about the temperature of the surrounding ice. The presence of water plays a crucial role in how fast a glacier flows, because water can lubricate the base of the glacier, causing more rapid sliding and, consequently, faster mass loss.

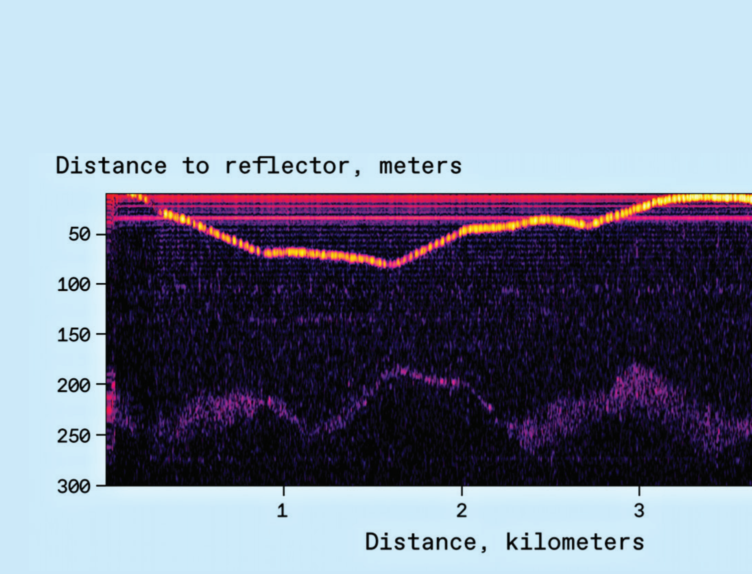
All of these are dynamic observations that may change on an annual or even seasonal basis. So having just one radar survey every few years isn't going to cut it.

Gathering more frequent data using just crewed flights is difficult—they are expensive and logistically challenging, and, in harsh environments, they put people at risk. The main question about how to replace crewed aircraft is which direction to go—up (a constellation of satellites) or down (a fleet of UAVs)?

A handful of satellites could provide global coverage and frequent repeat measurements over many years, but it isn't the ideal platform for ice-penetrating radar. To get the same power per unit area on the surface of the ice as a 1-watt transmitter on a UAV flying at an altitude of 100 meters, a satellite in orbit at 400 km would need a roughly 15-megawatt transmitter—that's more than three times the maximum power for which SpaceX's Starlink satellites have been licensed by the Federal Communications Commission.

Another challenge is clutter. Imagine you have an antenna that emits power primarily within a 10-degree cone. You're





trying to observe the bottom of the ice sheet 1.5 km below the ice surface, but there's a mountain range 35 km away. From 400 km up, that mountain range is also being illuminated by your antenna and reflecting energy back much more strongly than the echo from the bottom of the ice sheet, which is attenuated by the 1.5 km of ice it passed through each way.

At the other end of the spectrum of options are UAVs, flying even closer to the ice than crewed aircraft can. Researchers have been interested in the potential of UAV-borne radar systems for imaging ice for at least a decade. In 2014, CReSIS fielded a 5-meter-wingspan radio-controlled aircraft with a miniaturized version of its IPR system. The design made clever use of the existing wing geometry to provide low-frequency antennas, albeit with a small bandwidth that limited data quality.

Since this pathfinding demonstration, much of the research focus has shifted to higher-frequency systems, sometimes called snow radars, designed to image the near surface to better understand mountain snowpacks, snow cover on sea ice, and the layering structure in the top few meters of ice sheets. CReSIS has tested its snow radar on a small autonomous helicopter; more recently, it partnered with NASA and Vanilla Unmanned to fly its snow radar on a massive 11-meter-wingspan UAV that can stay aloft for days at a time.

There's still a need, though, for IPR imaging through ice sheets, with a high enough bandwidth to distinguish internal layers and a price tag that allows for widespread use.

HERE'S WHERE PEREGRINE comes in. The project was started in 2020 to build a smaller and more affordable system than those attempted previously, now made possible by advances in fixed-wing UAVs and miniaturized electronics.

We knew we couldn't do the IPR with off-the-shelf systems. We had to start with a blank slate to develop a system that was small and light enough to fit on an inexpensive UAV.

We decided to use software-defined radio (SDR) technology for our radars because these RF transmitters and receivers are highly customizable and shift much of the complexity of the system from hardware to software. Using SDR, an entire radar system can fit on a few small circuit boards.

From the start, we looked beyond our first project, developing software built on top of Ettus's USRP Hardware Driver application programming interface, which can be used with a variety of software-defined radios, ranging in cost from US $1,000 to $30,000 and in mass from tens of grams to multiple kilograms.

We added a Raspberry Pi single-board computer to control our software-defined radio. The Raspberry Pi also connects to a network of temperature sensors, so that we could be sure nothing in our system gets too hot or too cold.

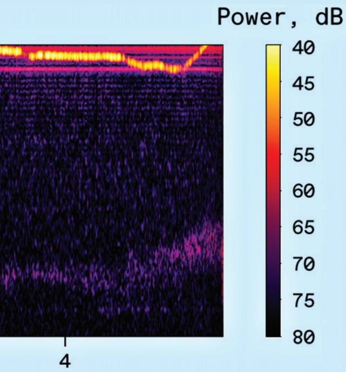
The SDR itself has two sides to it, one for transmitting the radar signal and one for receiving the echoes, each connecting to our custom-made antennas through amplifiers and filters. This entire system weighs a little under 1 kilogram.

Those antennas were tricky to design. IPR antennas require relatively low frequencies (because higher frequencies are more significantly attenuated by ice) and have relatively wide bandwidths (to achieve sufficient range resolution). Normally, these criteria would mean a large antenna, but our small UAV couldn't handle a big, heavy antenna.

I started by considering a standard bow-tie antenna, a type commonly used in ground-based radar systems. The initial design was far too large to fit even one antenna, much less two, on our little UAV. So using a digital model of the antenna, I adjusted the geometry to find an acceptable compromise between size and performance, at least according to the simulation software I was using.

I also built several prototypes along the way to understand how real antenna performance might differ from my simulations. The first of those I made from copper tape cut and pasted onto sheets of plastic. The later and final versions I fabricated as printed circuit boards. After a few iterations, I had a working antenna that could be mounted flat under each wing of our diminutive aircraft.

TOP LEFT: MAI BUI; TOP RIGHT: THOMAS TEISBERG



Thomas Teisberg [opposite page] reviews some of the data recorded by Peregrine during a test flight [bottom] in Norway. The small box on his desk with wires attached is part of Peregrine's payload, a package that includes a software-defined radio, a Raspberry Pi, and other electronics wrapped in copper shielding. The surface of the ice and shape of the bedrock [left] are clearly visible in this two-dimensional tracing of the data.

For the drone, we started with a kit for an X-UAV Talon radio-controlled plane, which included a foam fuselage, tail assembly, and wings. We knew that every piece of conductive material in the aircraft would affect the antenna's performance, perhaps in undesirable ways. Tests showed that the carbon-fiber spar between the wings and the wires to the servo motors in each wing was creating problematic conductive paths between the antennas, so we replaced the carbon-fiber spar with a fiberglass one and added ferrite beads on the servo wiring to act as low-pass filters.

I THOUGHT WE WERE READY. But when we took our UAV out to a field near our lab, we discovered that we could not get a GPS fix on the drone when the radar system was active. After some initial confusion, we discovered the source of the interference: our system's USB 3.0 interface. To solve this problem, I designed a plastic box to enclose the Raspberry Pi and the SDR, 3D-printed it, and wrapped it in a thin layer of copper tape. That shielded the troublesome USB circuitry enough to keep it from interfering with the rest of our system.

Finally, we were able to fly our tiny radar drone over a dry lake bed on the Stanford campus. Although our system cannot image through dirt, we were able to get a strong reflection off the surface, and at that point we knew we had a working prototype.

We carried out our first real-world tests six months later, on Iceland's Vatnajökull ice cap, thanks to the help and generosity of local collaborators at the University of Iceland and a grant from NASA. That was a good spot, because from time to time, a nearby volcanic eruption spews volcanic material known as tephra over the surface of the ice cap. That tephra eventually gets buried under new snow and forms a layer under the surface. We figured these strata would serve as a good stand-in for the internal layering found in ice in Greenland and Antarctica. Although an abundance of liquid water in the relatively warm Vatnajökull ice prevented our system from probing more than tens of meters below the surface, the tephra layers were apparent in our radar soundings.

But these first trials did not go uniformly well. After one of our test flights, I discovered that the data we had collected was almost entirely noise. We tested every component and cable, until I found that the shield on one of the coaxial cables had broken and was only intermittently making a connection. With a spare cable and a generous application of hot glue, we were able to complete the rest of our testing.

For our next round of tests, we were aiming to image bedrock under a glacier, not just internal layers. And that's why, in March of this year, we ended up on a glacier in the coldest part of Norway, where liquid water within the ice was less likely to interfere with our measurements. There we were able to image the bed of the glacier, as much as 150 meters below the surface where we were flying. Crucially, we also convinced ourselves that our system will work properly in the harsh environments we expect it to face in Antarctica and Greenland.

OUR PRESENT SYSTEM is relatively small. It was designed to be inexpensive and portable so that research teams can easily bring it along on expeditions to far-flung spots. But we also wanted it to serve as a test bed for a larger UAV-borne IPR system with an operational range of about 800 km, one that is inexpensive enough to be permanently deployed to Antarctic research stations. With the 11 existing research stations as bases, at least one member of such a drone fleet could access nearly every part of coastal Antarctica. Though larger and more expensive than our original Peregrine, this next-generation UAV will still be far cheaper and easier to operate than crewed airborne systems are.

Operating a larger UAV, much less a fleet of them, is beyond what a few Ph.D. students alone can reasonably do, so we are launching a collaborative effort between Stanford University, the Scripps Institution of Oceanography, and Lane Community College, in Eugene, Ore., to get this new platform off the ground. If all goes well, we're hoping we can have IPR UAVs surveying the Antarctic and Greenland ice sheets within three years. Doing so would no doubt help scientists studying the responses of Earth's ice sheets to climate change. With permanently deployed UAVs able to cover most areas of active study, requests for new data could be fulfilled within days. Surveys could be repeated at frequent intervals over dynamic areas. And when rapid and unpredictable events occur, such as the collapse of an ice shelf, a UAV could be deployed to gather real-time radar data.

Such observations are just not possible today. But Peregrine and its successors could make that possible. Having the ability to collect this kind of radar data would help glaciologists resolve fundamental uncertainties in the physics of ice sheets, improve projections of sea-level rise, and enable better decision making about mitigations and adaptations for Earth's future climate. ■

BELOW: ELIZA DAWSON

5 CONCLUSIONS FROM AN AUTOMATION EXPERT WITH FIRSTHAND KNOWLEDGE OF HIGHWAY REGULATION >

By MARY L. "MISSY" CUMMINGS

This self-driving Cruise robotaxi got stuck at a crossroads in San Francisco in 2019, inconveniencing pedestrians.

ANDREJ SOKOLOW/ PICTURE-ALLIANCE/AP

< WHAT SELF-DRIVING CARS TELL US ABOUT AI RISKS

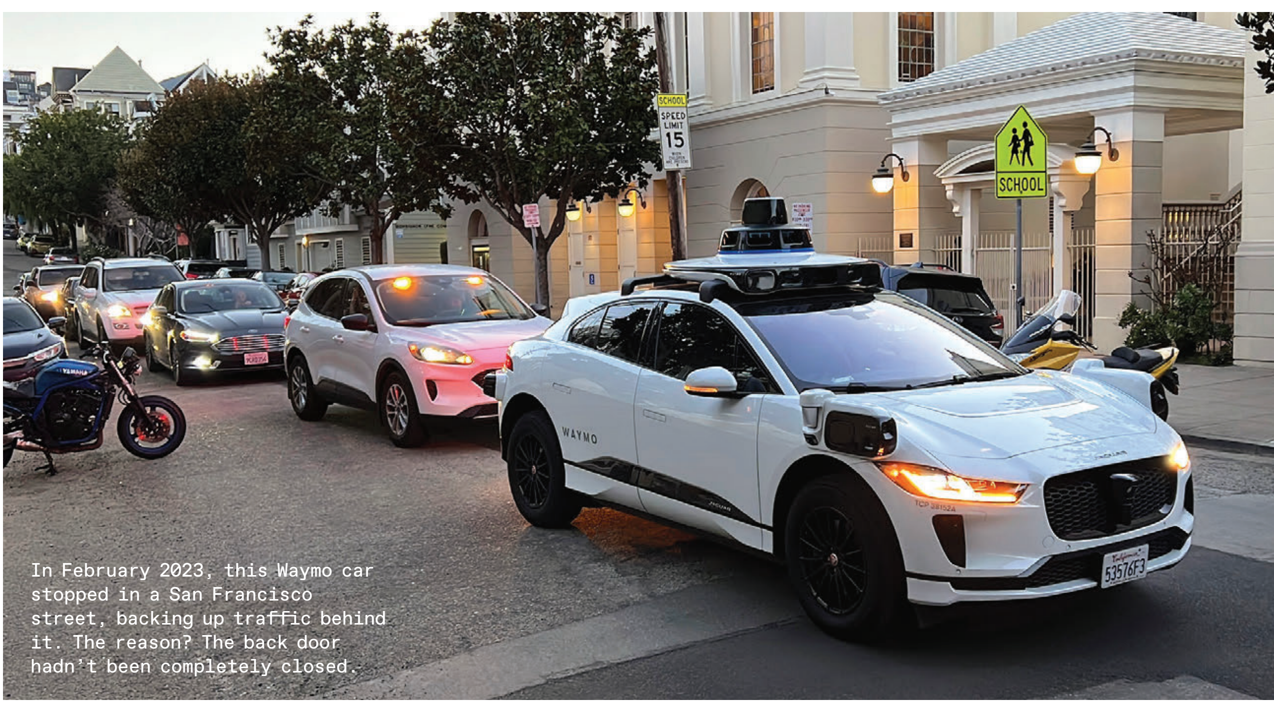
In 2016, just weeks before the Autopilot in his Tesla drove Joshua Brown to his death, I pleaded with the U.S. Senate Committee on Commerce, Science, and Transportation to regulate the use of artificial intelligence in vehicles. Neither my pleading nor Brown's death could stir the government to action. • Since then, automotive AI in the United States has been linked to at least 25 confirmed deaths and to hundreds of injuries and instances of property damage.

The lack of technical comprehension across industry and government is appalling. People do not understand that the AI that runs vehicles—both the cars that operate in actual self-driving modes and the much larger number of cars offering advanced driving-assistance systems (ADAS)—are based on the same principles as ChatGPT and other large language models (LLMs). These systems control a car's lateral and longitudinal position—to change lanes, brake, and accelerate—without waiting for orders to come from the person sitting behind the wheel.

Both kinds of AI use statistical reasoning to guess what the next word or phrase or steering input should be, heavily weighting the calculation with recently used words or actions. Go to your Google search window and type in "now is the time" and you will get the result "now is the time for all good men." And when your car detects an object on the road ahead, even if it's just a shadow, watch the car's self-driving module suddenly brake.

Neither the AI in LLMs nor the one in autonomous cars can "understand" the situation, the context, or any unobserved factors that a person would consider in a similar situation. The difference is that while a language model may give you nonsense, a self-driving car can kill you.

In late 2021, despite receiving threats to my physical safety for daring to speak truth about the dangers of AI in vehicles, I agreed to work with the U.S. National Highway Traffic Safety Administration (NHTSA) as the senior safety advisor. What qualified me for the job was a doctorate focused on the design of joint human-automated systems and 20



In February 2023, this Waymo car stopped in a San Francisco street, backing up traffic behind it. The reason? The back door hadn't been completely closed.

TERRY CHEA/AP

years of designing and testing unmanned systems, including some that are now used in the military, mining, and medicine.

My time at NHTSA gave me a ringside view of how real-world applications of transportation AI are or are not working. It also showed me the intrinsic problems of regulation, especially in our current divisive political landscape. My deep dive has helped me to formulate five practical insights. I believe they can serve as a guide to industry and to the agencies that regulate them.

1 Human errors in operation get replaced by human errors in coding

Proponents of autonomous vehicles routinely assert that the sooner we get rid of drivers, the safer we will all be on roads. They cite the NHTSA statistic that 94 percent of accidents are caused by human drivers. But this statistic is taken out of context and inaccurate. As the NHTSA itself noted in that report, the driver's error was "the last event in the crash causal chain.…It is not intended to be interpreted as the cause of the crash." In other words, there were many other possible causes as well, such as poor lighting and bad road design.

Moreover, the claim that autonomous cars will be safer than those driven by humans ignores what anyone who has ever worked in software development knows all too well: that software code is incredibly error-prone, and the problem only grows as the systems become more complex.

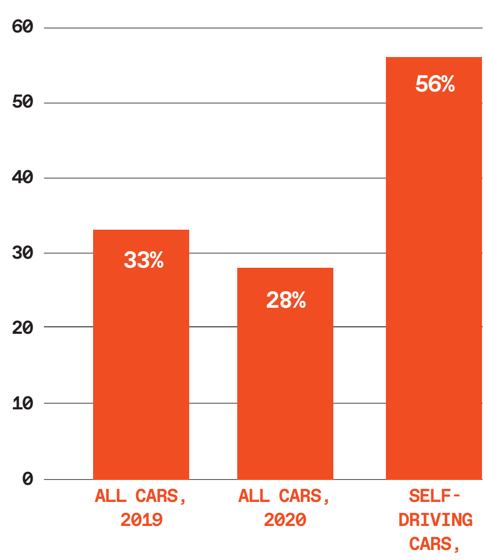
Consider these recent crashes in which faulty software was to blame. There was the October 2021 crash of a Pony.ai driverless car into a sign, the April 2022 crash of a TuSimple tractor trailer into a concrete barrier, the June 2022 crash of a Cruise robotaxi that suddenly stopped while making a left turn, and the March 2023 crash of another Cruise car that rear-ended a bus.

These and many other episodes make clear that AI has not ended the role of human error in road accidents. That role has merely shifted from the end of a chain of events to the beginning—to the coding of the AI itself. Because such errors are latent, they are far harder to mitigate. Testing, both in simulation but predominantly in the real world, is the key to reducing the chance of such errors, especially in safety-critical systems. However, without sufficient government regulation and clear industry standards, autonomous-vehicle companies will cut corners in order to get their products to market quickly.

Phantom braking: when a self-driving car stops cold for no apparent reason

REAR-END CRASH PERCENTAGES

Cars equipped with self-driving systems had worse rear-end collision rates than those not so equipped.



Source: National Highway Traffic Safety Administration

2 AI failure modes are hard to predict

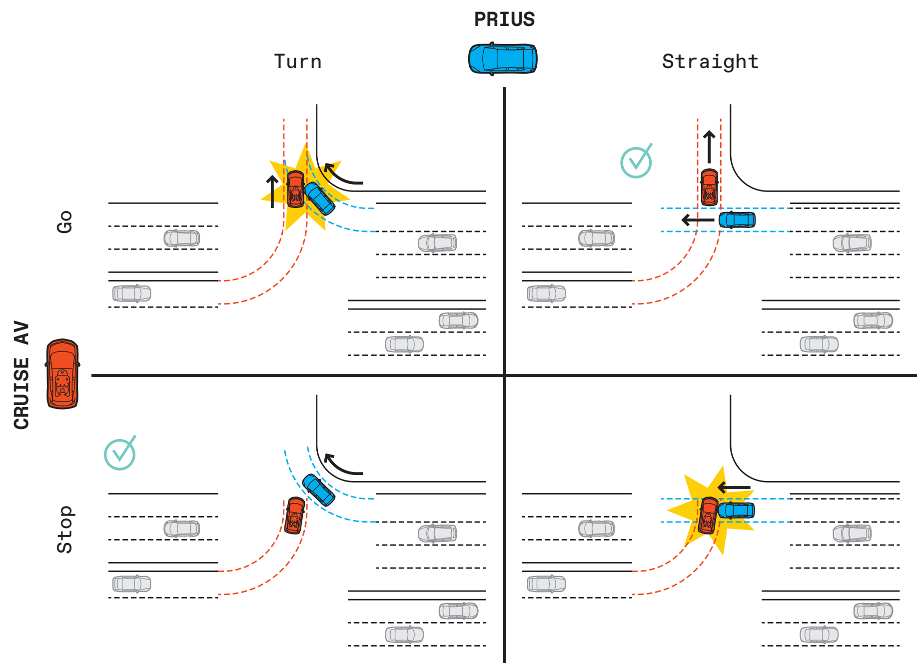
A large language model guesses which words and phrases are coming next by consulting an archive assembled during training from preexisting data. A self-driving module interprets the scene and decides how to get around obstacles by making similar guesses, based on a database of labeled images—this is a car, this is a pedestrian, this is a tree—also provided during training. But not every possibility can be modeled, and so the myriad failure modes are extremely hard to predict. All things being equal, a self-driving car can behave very differently on the same stretch of road at different times of the day, possibly due to varying sun angles. And anyone who has experimented with an LLM and changed just the order of words in a prompt will immediately see a difference in the system's replies.

One failure mode not previously anticipated is phantom braking. For no obvious reason, a self-driving car will suddenly brake hard, perhaps causing a rear-end collision with the vehicle just behind it and other vehicles farther back. Phantom braking has been seen in the self-driving cars of many different manufacturers and in ADAS-equipped cars as well.

The cause of such events is still a mystery. Experts initially attributed it to human drivers following the self-driving car too closely (often accompanying their assessments by citing the misleading 94 percent statistic about driver error). However, an increasing number of these crashes have been reported to NHTSA. In May 2022, for instance, the NHTSA sent a letter to Tesla noting that the agency had received 758 complaints about phantom braking in Model 3 and Model Y cars. This past May, the German publication *Handelsblatt* reported on 1,500 complaints of braking issues with Tesla vehicles, as well as 2,400 complaints of sudden acceleration. It now appears that self-driving cars experience roughly twice the rate of rear-end collisions as do cars driven by people.

Clearly, AI is not performing as it should. Moreover, this is not just one company's problem—all car companies that are leveraging computer vision and AI are susceptible to this problem.

As other kinds of AI begin to infiltrate society, it is imperative for standards bodies and regulators to understand that AI failure modes will not follow a predictable path. They should also be wary of the car companies' propensity to excuse away bad tech behavior and to blame humans for abuse or misuse of the AI.



On 3 June 2022, a Cruise AV in San Francisco sensed an oncoming Prius, considered two scenarios [checked in green] out of a possible four, and chose a path. When things did not go according to plan, the Cruise stopped cold, right in the middle of the lane, causing a collision. SOURCE: RETROSPECT

3 Probabilistic estimates do not approximate judgment under uncertainty

Ten years ago, there was significant hand-wringing over the rise of IBM's AI-based Watson, a precursor to today's LLMs. People feared AI would very soon cause massive job losses, especially in the medical field. Meanwhile, some AI experts said we should stop training radiologists.

These fears didn't materialize. While Watson could be good at making guesses, it had no real knowledge, especially when it came to making judgments under uncertainty and deciding on an action based on imperfect information. Today's LLMs are no different: The underlying models simply cannot cope with a lack of information and do not have the ability to assess whether their estimates are even good enough in this context.

These problems are routinely seen in the self-driving world. The June 2022 accident involving a Cruise robotaxi happened when the car decided to make an aggressive left turn between two cars. As the car safety expert Michael Woon detailed in a report on the accident, the car correctly chose a feasible path, but then halfway through the turn, it slammed on its brakes and stopped in the middle of the intersection. It had guessed that an oncoming car in the right lane was going to turn, even though a turn was not physically possible at the speed the car was traveling. The uncertainty confused the Cruise, and it made the worst possible decision. The oncoming car, a Prius, was not turning, and it plowed into the Cruise, injuring passengers in both cars.

Cruise vehicles have also had many problematic interactions with first responders, who by default operate in areas of significant uncertainty. These encounters have included Cruise cars traveling through active firefighting and rescue scenes and driving over downed power lines. In one incident, a firefighter had to knock the window out of the Cruise car to remove it from the scene. Waymo, Cruise's main rival in the robotaxi business, has experienced similar problems.

While a language model may give you nonsense, a self-driving car can kill you.

These incidents show that even though neural networks may classify a lot of images and propose a set of actions that work in common settings, they nonetheless struggle to perform even basic operations when the world does not match their training data. The same will be true for LLMs and other forms of generative AI. What these systems lack is judgment in the face of uncertainty, a key precursor to real knowledge.

4 Maintaining AI is just as important as creating AI

Because neural networks can only be effective if they are trained on significant amounts of relevant data, the quality of the data is paramount. But such training is not a one-and-done scenario: Models cannot be trained and then sent off to perform well forever after. In dynamic settings like driving, models must be constantly updated to reflect new types of cars, bikes, and scooters, construction zones, traffic patterns, and so on.

In the March 2023 accident, in which a Cruise car hit the back of an articulated bus, experts were surprised, as many believed such accidents were nearly impossible for a system that carries lidar, radar, and computer vision. Cruise attributed the accident to a

Illustration by Chris Philpot

faulty model that had guessed where the back of the bus would be based on the dimensions of a normal bus; additionally, the model rejected the lidar data that correctly detected the bus.

This example highlights the importance of maintaining the currency of AI models. "Model drift" is a known problem in AI, and it occurs when relationships between input and output data change over time. For example, if a self-driving car fleet operates in one city with one kind of bus, and then the fleet moves to another city with different bus types, the underlying model of bus detection will likely drift, which could lead to serious consequences.

Such drift affects AI working not only in transportation but in any field where new results continually change our understanding of the world. This means that large language models can't learn a new phenomenon until it has lost the edge of its novelty and is appearing often enough to be incorporated into the dataset. Maintaining model currency is just one of many ways that AI requires periodic maintenance, and any discussion of AI regulation in the future must address this critical aspect.

5 AI has system-level implications that can't be ignored

Self-driving cars have been designed to stop cold the moment they can no longer reason and no longer resolve uncertainty. This is an important safety feature. But as Cruise, Tesla, and Waymo have demonstrated, managing such stops poses an unexpected challenge.

A stopped car can block roads and intersections, sometimes for hours, throttling traffic and keeping out first-response vehicles. Companies have instituted remote-monitoring centers and rapid-action teams to mitigate such congestion and confusion, but at least in San Francisco, where hundreds of self-driving cars are on the road, city officials have questioned the quality of their responses.

Self-driving cars rely on wireless connectivity to maintain their road awareness, but what happens when that connectivity drops? One driver found out the hard way when his car became entrapped in a knot of 20 Cruise vehicles that had lost connection to the remote-operations center and caused a massive traffic jam.

Of course, any new technology may be expected to suffer from growing pains, but if these pains become serious enough, they will erode public trust and support. Sentiment toward self-driving cars used to be optimistic in tech-friendly San Francisco, but now it has taken a negative turn due to the sheer volume of problems the city is experiencing. Such sentiments may eventually lead to public rejection of the technology if a stopped autonomous vehicle causes the death of a person who was prevented from getting to the hospital in time.

Software code is incredibly error-prone, and the problem only grows as the systems become more complex.

So what does the experience of self-driving cars say about regulating AI more generally? Companies not only need to ensure they understand the broader systems-level implications of AI, they also need oversight—they should not be left to police themselves. Regulatory agencies must work to define reasonable operating boundaries for systems that use AI, and issue permits and regulations accordingly. When the use of AI presents clear safety risks, agencies should not defer to industry for solutions and should be proactive in setting limits.

AI still has a long way to go in cars and trucks. I'm not calling for a ban on autonomous vehicles. There are clear advantages to using AI, and it is irresponsible for people to call for a ban, or even a pause, on AI. But we need more government oversight to prevent the taking of unnecessary risks.

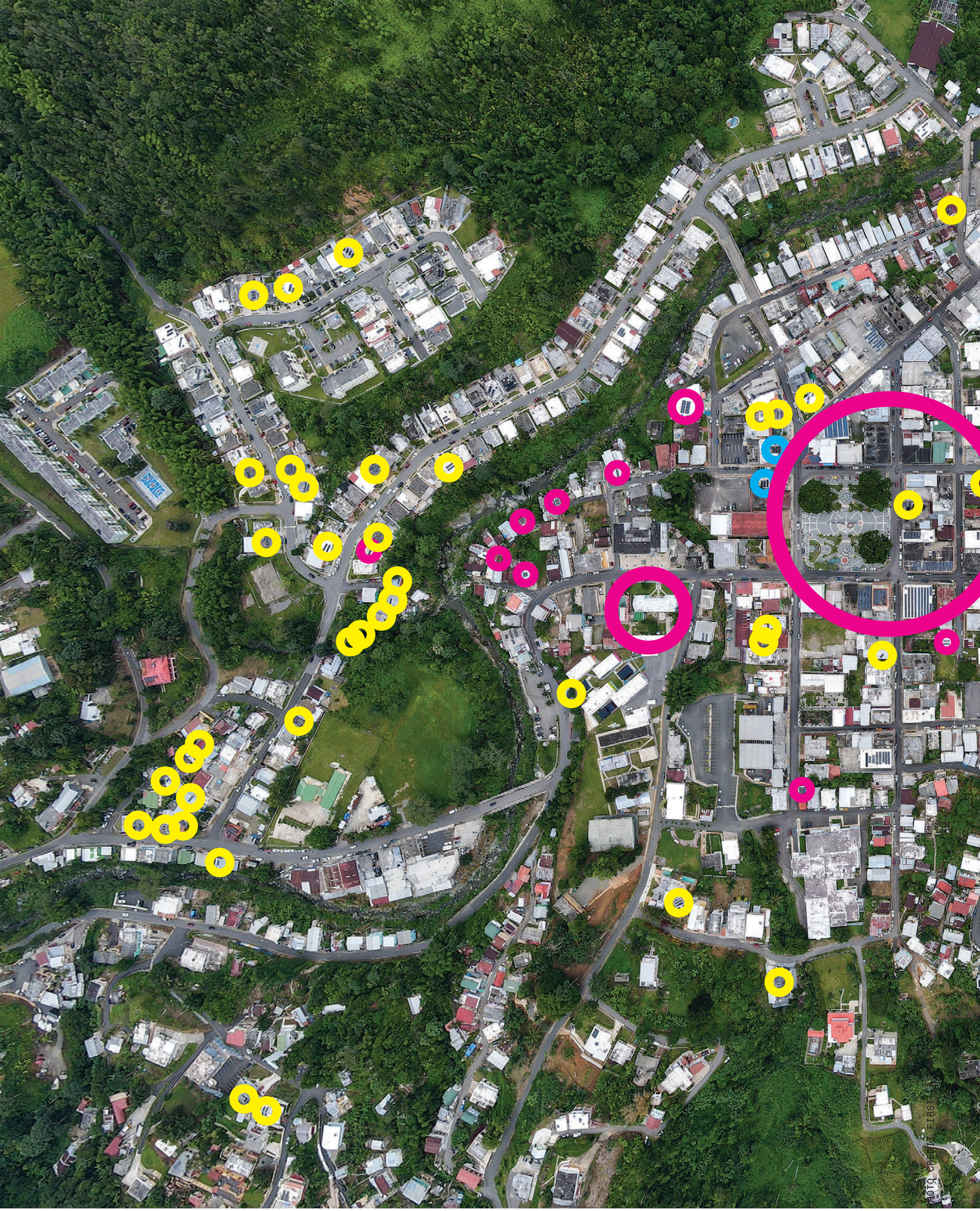
And yet the regulation of AI in vehicles isn't happening yet. That can be blamed in part on industry overclaims and pressure, but also on a lack of capability on the part of regulators. The European Union has been more proactive about regulating artificial intelligence in general and in self-driving cars particularly. In the United States, we simply do not have enough people in federal and state departments of transportation that understand the technology deeply enough to advocate effectively for balanced public policies and regulations. The same is true for other types of AI.

This is not any one administration's problem. Not only does AI cut across party lines, it cuts across all agencies and at all levels of government. The Department of Defense, Department of Homeland Security, and other government bodies all suffer from a workforce that does not have the technical competence needed to effectively oversee advanced technologies, especially rapidly evolving AI.

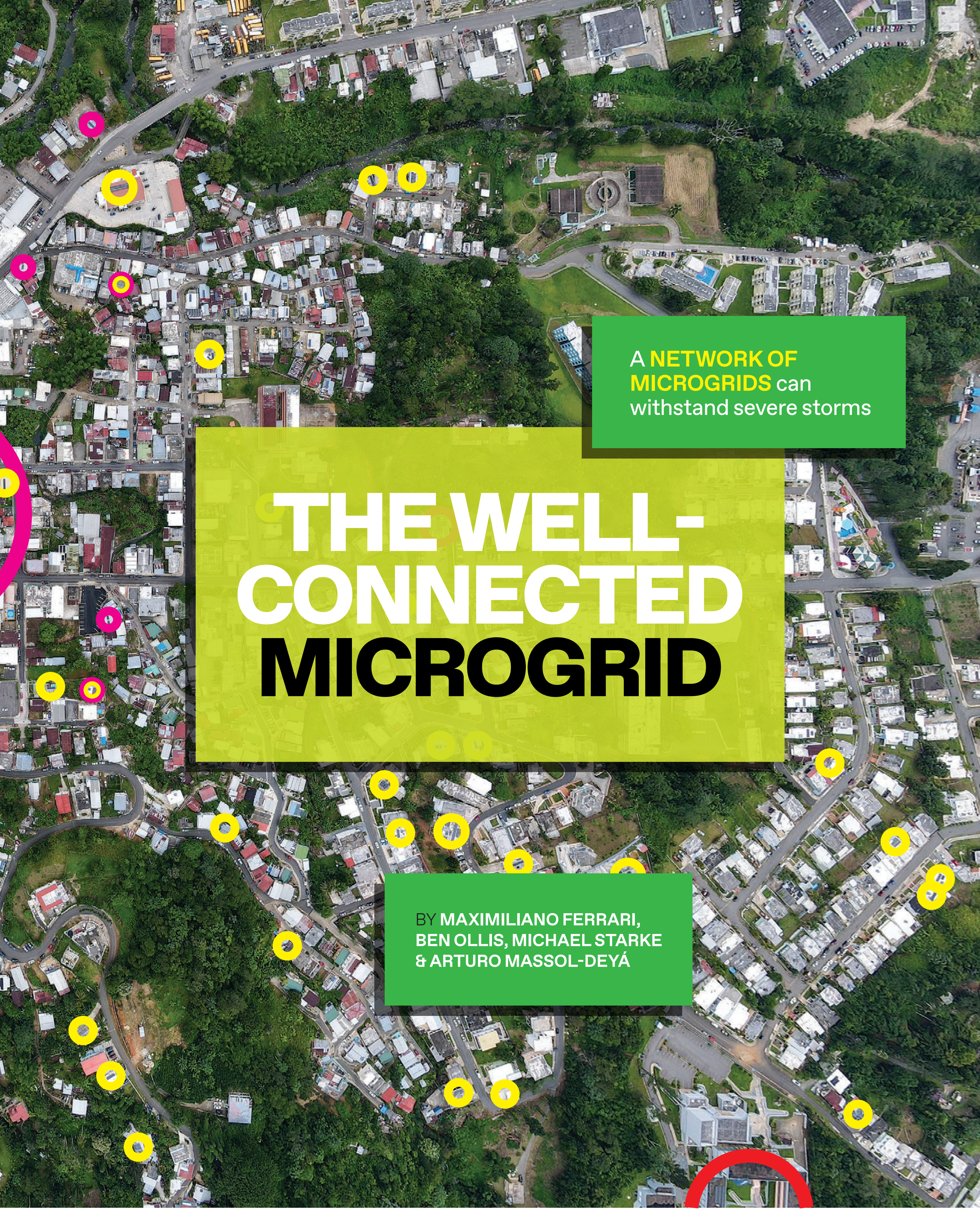
To engage in effective discussion about the regulation of AI, everyone at the table needs to have technical competence in AI. Right now, these discussions are greatly influenced by industry (which has a clear conflict of interest) or Chicken Littles who claim machines have achieved the ability to outsmart humans. Until government agencies have people with the skills to understand the critical strengths and weaknesses of AI, conversations about regulation will see very little meaningful progress.

Recruiting such people can be easily done. Improve pay and bonus structures, embed government personnel in university labs, reward professors for serving in the government, provide advanced certificate and degree programs in AI for all levels of government personnel, and offer scholarships for undergraduates who agree to serve in the government for a few years after graduation. Moreover, to better educate the public, college classes that teach AI topics should be free.

We need less hysteria and more education so that people can understand the promises but also the realities of AI. ■



The people of Adjuntas, in central Puerto Rico, are aiming to make it an all-solar town. The circles show houses and



A NETWORK OF MICROGRIDS can withstand severe storms

THE WELL-CONNECTED MICROGRID

BY MAXIMILIANO FERRARI, BEN OLLIS, MICHAEL STARKE & ARTURO MASSOL-DEYÁ

businesses where rooftop photovoltaic panels have been installed. JOSÉ “TITO” ALMODÓVAR/CASA PUEBLO

↑ Alexis Massol González, founder of the environmental nonprofit Casa Pueblo de Adjuntas, stands next to a rooftop PV installation, one of more than 400 deployed in the region.

IN SEPTEMBER 2017, Hurricane Maria devastated Puerto Rico. It was one of the deadliest and most damaging natural disasters to ever make landfall in the United States. Maria destroyed infrastructure and crops and initially claimed 64 lives. The aftermath was even more calamitous: More than 4,000 people are estimated to have died as a result of the shortage of medical care and basic resources during the extended power outages that followed the storm, according to research reported in *The New England Journal of Medicine*. Low-income and rural communities suffered the most, with some remaining without power for almost a year.

The winds had barely died down when electric-power specialists and others began calling for a radical overhaul of Puerto Rico's electricity networks, one that would emphasize renewable energy, distributed generation, and, critically, microgrids.

A microgrid is like a miniaturized, tightly controlled version of a power grid. Each microgrid includes generation, loads, transformers, distribution lines, protective devices, and, typically, energy storage. Several factors combine to make Puerto Rico an ideal place for solar power and microgrids. The island gets on average nearly 3,000 hours of sunshine per year, putting it on a par with Honolulu and Brisbane, Australia. And it is vulnerable to precisely the kind of destructive storms that microgrids are designed to ride out with minimal interruption in electric service.

Although most microgrids today operate independently, one way to get the maximum benefit from them is to connect them, letting them share resources wherever and whenever possible. To that end, engineers at the U.S. Department of Energy's Oak Ridge National Laboratory (including three of us—Ferrari, Ollis, and Starke) are investigating distributed controllers that will allow microgrids to form a larger network that will increase their resiliency and facilitate the integration of more renewable energy. Such networked microgrids can be isolated for longer periods from the main grid, which is crucial during long-term outages caused by severe weather and other natural disasters.

We have designed such a controller, which we call a microgrid orchestrator,

CLOCKWISE FROM LEFT: THE HONNOLD FOUNDATION; RICARDO ARDUENGO/THE HONNOLD FOUNDATION; FABIO ANDRADE/UPRM

←↑→↙↙↓ **Rooftop solar power totaling about 200 kilowatts, complemented by 1 megawatt-hour of battery storage, supplies electricity to 14 businesses in Adjuntas.**

→ The Solar Forest, a novel installation of PV panels, doubles as a public charging station in the town square of Adjuntas.

CLOCKWISE FROM TOP: FABIO ANDRADE/UPRM; THE HONNOLD FOUNDATION; HEATHER DUNCAN/ORNL; RICARDO ARDUENGO/THE HONNOLD FOUNDATION

and are now developing facilities to thoroughly test it in the lab. Next year, we plan to install and further test our controller on-site in the central Puerto Rican community of Adjuntas. Ultimately, such technology could benefit any part of the world where the electricity network is vulnerable to outages and where increasing amounts of solar and wind power are being integrated into the power grid.

IN THE DARKNESS caused by Hurricane Maria, a grassroots environmental nonprofit called Casa Pueblo de Adjuntas provided hope for the people of Adjuntas. (Coauthor Massol-Deyá is executive director of the organization.) The group is a pioneer in bringing solar energy to Puerto Rico, having installed the first photovoltaic (PV) panels in Adjuntas in 1999. During and after Maria, Casa Pueblo continued to operate a solar-based microgrid it had installed in 2017. The microgrid provided vital services, including running dialysis machines, refrigerating medicines, and charging cellphones. Adjuntas, a region that shares the same name of its main town, is known for its many coffee plantations and has a population of about 18,000 spread among 16 towns and villages. Casa Pueblo's solar-powered radio transmitter was the main means of disseminating news across central Puerto Rico after the 2017 storm.

Maria was not an isolated event—since it struck, eight major storms have either hit the island directly or passed close enough to have caused flooding or power losses. Vulnerable communities in Puerto Rico and elsewhere must decide how they will cope with future disasters, intensified by the effects of a changing climate. The residents of Adjuntas have chosen a path toward self-sufficiency and a power system that is relatively clean, decentralized, and resilient.

The Adjuntas model is built around solar-powered microgrids. A key feature of a microgrid is the option of operating it connected to the main grid—a mode called grid-connected—or isolated from the grid, in islanded mode. Islanding lets microgrids continue to supply electricity to users even when the main grid is down. In 2019, in partnership with the Honnold Foundation and other organizations, Casa Pueblo started connecting the Adjuntas town square to two solar-powered microgrids. Construction was completed this past March, and the microgrids will be fully operational by the end of the year.

Puerto Rico's government has committed to transitioning to 100 percent renewable energy by 2050, with interim goals of 40 percent by 2025, the phaseout of coal-fired plants by 2028, and a 60 percent renewable system by 2040. Meanwhile, though, much of the island continues to rely on a centralized grid that is fed mainly by power plants that burn natural gas, petroleum, or coal. From 2014 to 2022, the Puerto Rican power company, PREPA, increased its share of renewable energy from 1.7 percent to 2.8 percent. During this same period, distributed residential solar connected to the main grid saw an incredible 50-fold increase, from 4.4 megawatts to 224.6 MW. The growth in residential solar capacity has been 2.5 times as great as the combined growth in the commercial and industrial sectors.

Solar-based microgrids have also seen an uptick in adoption since Maria. During prolonged outages, they can provide a cleaner and less expensive alternative to emergency diesel generators. Diesel generators emit harmful pollutants and are expensive to operate, especially in the aftermath of a devastating storm, when fuel is scarce. Following Maria, businesses in Adjuntas struggled to obtain enough diesel and spent thousands of dollars on fuel to keep their generators operating—an expense that sorely tested them.

↑ The pizzeria in Adjuntas is connected to a microgrid. It should still be able to operate even if the main power grid goes down.

←↑↗↗ In a hurricane-prone region like Puerto Rico, it's good to be prepared. The pharmacy, hardware store, bakery, and

Sunlight, though, is a free and abundant energy source in Puerto Rico, and solar-based microgrids have proven to be a reliable alternative to the central grid. The two microgrids in Adjuntas encompass more than 700 PV panels installed on the rooftops of 14 small businesses in the center of the town, including the bakery, the pizzeria, and critical facilities such as the pharmacy. The microgrids, which have a total capacity of about 200 kilowatts, as well as two large battery-storage systems with more than 1 megawatt-hour of capacity, are designed to power these businesses.

Most microgrids run in grid-connected mode whenever the main grid is available. The microgrids in Adjuntas, however, are designed to run in islanded mode, disconnected from the main grid. They're owned, operated, and maintained by a nonprofit cooperative, which functions as an independent system operator. The businesses thus purchase electricity that is generated locally. Each business is equipped with its own energy meter, and at the end of the month it's charged for the electricity it consumed.

Business owners pay a rate that's lower than the main grid rate on the island, which is one of the highest in the United States. The average cost of electricity for commercial users in Puerto Rico is about 29 U.S. cents per kilowatt-hour; in May, the rate on the United States mainland was about 13 cents per kilowatt-hour. Revenue generated by the microgrids will be reinvested into the community in several ways, including operation, maintenance, and future expansion of the microgrids, and as an emergency fund to bring solar energy to other low-income communities in the area.

The experience of regularly operating the system in islanded mode will be crucial when the next natural disaster shuts down the main grid. During such extreme events, the connected businesses will serve the entire community by providing refrigeration for medicines and perishables, power to recharge smartphones and laptops, a community kitchen, and a laundry, just to name a few essential services. In Adjuntas, even a pizzeria can become a critical emergency facility. Having access to electricity will help to

CLOCKWISE FROM TOP: HEATHER DUNCAN/ORNL; MAXIMILIANO FERRARI/ORNL; ERIKA P. RODRIGUEZ/THE NEW YORK TIMES/REDUX

↓ Last March, Adjuntas celebrated its commitment to self-sufficient solar power. The microgrid technology being deployed there could be a model for the rest of Puerto Rico and other parts of the world.

barbershop are connected to the Adjuntas microgrids, and in extreme events, they will continue to serve the community.

HEATHER DUNCAN/ORNL (3)

maintain a sense of normality during the crisis. It will also enable businesses to continue operating, mitigating the economic impact that a long-term outage would otherwise inflict. It's a powerful example of how electricity and community resiliency are intertwined.

MICROGRIDS OFFER great promise for energy resilience, but they have some limitations. Traditionally, islanded microgrids have rigid boundaries, creating energy silos that can't communicate with one another or share resources. As a result, power generation and demand may be unevenly distributed across a region: One microgrid where demand is low or generation is high may have excess generation that goes to waste, while another microgrid nearby may have to disconnect electrical loads due to insufficient generation or high demand.

To tackle this challenge, our team at Oak Ridge National Laboratory (ORNL), in Tennessee, with funding from the Department of Energy's Solar Energy Technologies Office, is designing a microgrid orchestrator that can interconnect islanded microgrids. At present, there's no single standard for microgrid architecture, and so the orchestrator is designed to be compatible with any type of microgrid controller. The orchestrator works with local microgrid controllers to optimize and coordinate the distribution of the available electricity supply. It does so by creating an artificial electricity market among the microgrids. In this market, the microgrids can sell or buy power from neighboring microgrids based on their local needs.

The orchestrator looks at the power imbalance—the difference between generation and loads—for each microgrid. Its pricing algorithm then adjusts the pricing signal, which is sent out to each microgrid. The pricing signal isn't the actual amount that a customer will pay for electricity. Rather, it's a transactive signal used within the network of microgrids to determine if a specific microgrid will buy power from, or sell power to, its neighbors. The pricing signal is periodically updated based on the power imbalance and other factors.

Usually, the optimization of a microgrid is centralized, with the central controller gathering information from and controlling each device, which could be, among other things, a PV inverter, battery inverter, or controllable load. Based on forecasts of loads and generation, it determines the optimal way to distribute power and sends instructions to each device. For example, it will prioritize a critical load like life-support equipment over a noncritical load like office lighting to ensure that the critical load remains powered. While a centralized approach is simpler to formulate than a distributed one, it requires full access to data and struggles to handle a large number of devices.

The orchestrator's distributed approach, by contrast, is highly scalable, as the optimization problem is divided into microgrid-size chunks. The orchestrator doesn't control the individual devices in the microgrid network. Instead, the local microgrid controllers maintain full control of their assets, such as PV panels, and they react to the pricing signal and make their own decisions about buying or selling power to one or more neighboring microgrids. This approach preserves the autonomy and privacy of each microgrid.

Adjuntas will be an ideal location to demonstrate our microgrid orchestrator. Currently, the region has more than 400 solar installations, capable of generating over 600 kW in bright sunlight. The fire station, barbershop, and several grocery stores are all powered by solar energy. The energy landscape is thus a glimpse of the transformation that needs to happen in Puerto Rico if the island is to reach its goal of 100 percent renewable energy by 2050. Microgrids with energy storage will help stabilize the larger grid as it integrates more distributed, intermittent generation coming from solar and wind power.

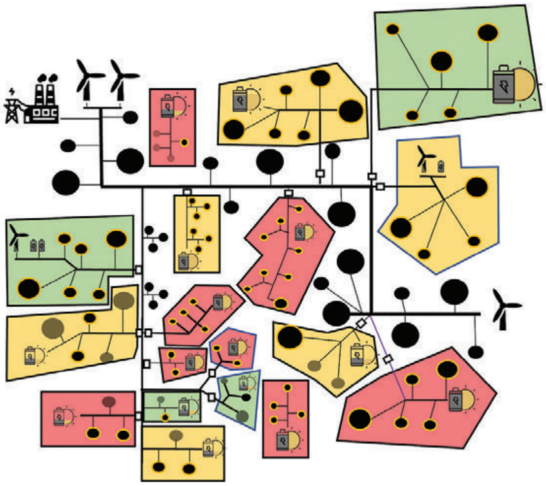
Adjuntas's two microgrids are designed to operate in islanded mode. In a demonstration next year, they will be networked together with a microgrid orchestrator.

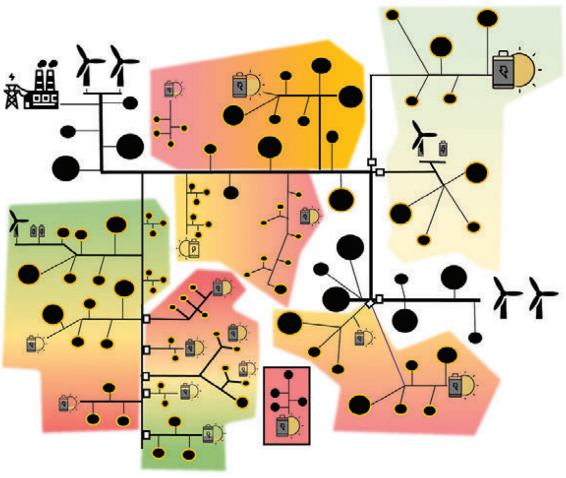
BEFORE DEPLOYING the microgrid orchestrator in the field, we are conducting a comprehensive validation of the technology in the lab. ORNL has two state-of-the-art facilities dedicated to microgrid research. One uses a hardware-in-the-loop (HIL) platform, which mimics a real system by integrating physical equipment such as external controllers into a simulation. The other is a network microgrid facility based on actual hardware.

At ORNL, the HIL simulator runs a detailed model of the Adjuntas microgrids, including PV inverters, energy storage, and protective systems, and it uses actual measurements of electricity consumption from each business, as well as PV inverter generation. This detailed simulation is thus validating the operation of the orchestrator in a safe and controllable virtual environment.

The hardware testbed consists of three microgrids, each of which integrates commercial PV inverters, energy-storage inverters, and synchronous machines to emulate conventional backup generation. The microgrids can be configured to operate in grid-connected or islanded mode, or as a network. The hardware testbed uses emulators for PV generation and loads, and we'll use data collected in Adjuntas through weather stations and data loggers to replicate those conditions in the lab.

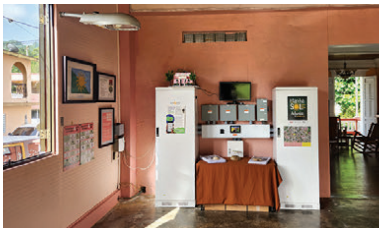
In the scenarios we have been evaluating, we're seeing that during normal conditions, when the main grid is healthy, having a network of microgrids facilitates the exchange of power between the microgrids, reducing the need to shed loads on cloudy days and





Islanded microgrids [left] can fail if generation and loads become unbalanced. Networking the microgrids together [right] using technology such as ORNL's microgrid orchestrator allows them to share resources and makes them more resilient.

MAXIMILIANO FERRARI/ORNL (2)



At Oak Ridge National Laboratory, researchers are testing microgrid hardware and running simulations of microgrid operation based on real-world data collected in Adjuntas, Puerto Rico.

maximizing the economic benefits. In Adjuntas, that means that surplus power generated in the northeast section of town will help support the power-hungry critical loads in the west, such as at the pharmacy.

During a prolonged power outage following a storm, the testbed shows that a microgrid that's been damaged can use both PV generation and battery resources from its neighbors. In that extreme scenario, the orchestrator will move electricity between the microgrids and allow every watt produced by the remaining solar panels to be used, minimizing the need to cut electricity to critical loads.

THE ORNL MICROGRID orchestrator is the centerpiece of a group of related technologies. Another component we're working on with the University of Tennessee, Knoxville, are controls that will allow a microgrid that is operating in islanded mode to extend its reach and pick up additional loads and generation that are outside the microgrid's boundaries.

Setting up a microgrid with dynamic rather than fixed boundaries is particularly important during long-term power outages. In those instances, grid-connected PV installations in the affected part of the network will go off line because their commercial PV inverters require AC voltage to operate. But if a microgrid has dynamic boundaries, its voltage could reach PV installations in the affected area—which could be part of another microgrid or not connected to any microgrid—and bring them back on line.

Although we plan to test this concept in the lab, implementing it in the field will require permission from the main utility. Similar technology has been successfully tested in an airport microgrid in Chattanooga, Tenn. In a community like Adjuntas, which already has hundreds of distributed PV installations, dynamic boundaries would let each microgrid serve more customers while benefiting from their additional generation. And in general, this approach will help facilitate the integration of more renewable energy sources into the main grid.

Another enabling technology we're working on is a distributed architecture that lets the microgrid orchestrator communicate securely with the microgrid controllers, which in turn communicate with field devices, such as PV inverters and battery storage. We call this architecture ORNL CODAS, which stands for Control and Optimization using Distributed Agent-based System.

As the name suggests, this distributed platform is based on agents, which are software entities that can make decisions based on their programmed logic. The CODAS agents rely on peer-to-peer communications, the peers in this case being the PV inverters, energy-storage controllers, and other microgrid devices and assets. This arrangement is critical when the orchestrator loses communication with one or more of the microgrid controllers.

With a centralized microgrid architecture, the loss of communication between the microgrid controller and the microgrid devices can lead to the collapse of the entire microgrid. In an agent-based architecture like CODAS, the agents will temporarily take control of the microgrid assets to maintain power until the communication link is restored. For example, the energy-storage agent can communicate with the PV agents to prevent the state of charge of the battery from going beyond the maximum or minimum threshold. This agent-based approach brings another layer of resiliency to keep the lights on in extreme events.

Meanwhile, the residents of Adjuntas continue to support the push for energy self-sufficiency. On 18 March 2023, thousands of people came to Casa Pueblo for the inauguration of the two microgrids. The parade through town was not only a celebration—it was a community statement demanding clean and resilient energy for the island of Puerto Rico. The residents want to be among the first all-solar towns in the world, an ambitious goal but one that's already in progress and within reach. ORNL is developing the controls that will support this vision in Adjuntas and beyond, creating the foundation for a highly integrated power grid based on renewable energy. ■

FROM LEFT: HEATHER DUNCAN/ORNL; MAXIMILIANO FERRARI/ORNL; CARLOS JONES/ORNL/U.S. DEPARTMENT OF ENERGY

ACKNOWLEDGMENTS: The authors thank Cynthia Arellano of the Honnold Foundation; Guodong Liu, Aditya Sundararajan, Mohammed Olama, Ben Dean, Heather Duncan at ORNL; Yang Chen at North Carolina Agricultural and Technical State University; Fabio Andrade and Gerson Beauchamp at the University of Puerto Rico Mayagüez; Aleksandar Dimitrovski at the University of Central Florida; and Fred Wang at the University of Tennessee, Knoxville.

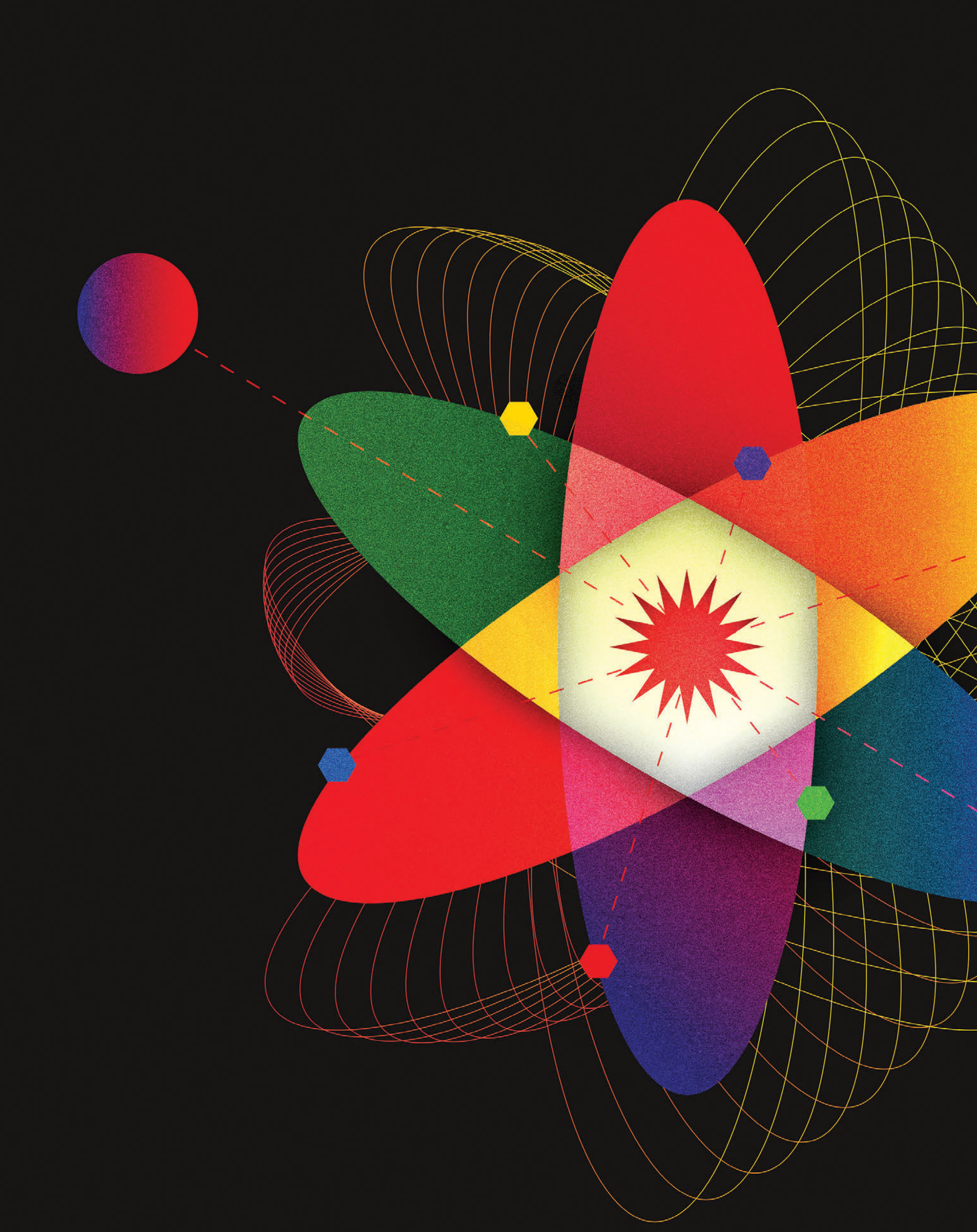


Illustration by Chad Hagen

Changing the color of light supercharges solar energy, 3D printing, and night vision

By Tracy H. Schloemer & Daniel N. Congreve

SMASHING PHOTONS

Color plays a huge role in most of our lives. It signals danger or caution, like the patterns on a poisonous frog or the color of a stoplight. It evokes joy and inspiration through nature, art, and fashion. It can even trigger the recollection of a favorite memory, through pictures of family and friends. ¶ In the technical world, the color of light is no less important. It affects the efficiency of solar cells, how far we can see inside our bodies, and the speed of 3D printing. ¶ But light can be useful only if you can actually get it where it needs to go; many materials will absorb or scatter the light long before it can reach its intended destination. ¶ We in the Congreve Lab at Stanford University became interested in changing the color of light for exactly this reason: It can help us get the right kind of light to the right place. ¶ Initially, we focused on developing color-changing technology for use in solar energy, where its usefulness is obvious. Photovoltaic cells harvest energy from only a limited range of energies—that is, colors. That range differs depending on the material used to produce the solar cell, but it's always limited. One approach to improving solar-cell efficiency has been to produce cells with multiple layers tuned to different energy levels. But there is another way to think about it: It could be simpler and more efficient to change the light to fit the cell.

Before we tell you more about how this process can boost solar power as well as revolutionize 3D printing and enable some other exciting applications, we'll explain how this technology works.

Traditionally, the color of a photon (defined by its energy or wavelength) is a given. Yet it turns out that we can turn two low-energy photons into a single higher-energy photon using a process called upconversion.

Upconversion has been observed in experiments for more than 50 years, but low efficiencies kept it a laboratory curiosity until materials that more efficiently emit upconverted light were discovered. Even then, issues with extracting a practical number of upconverted photons, how to incorporate the substances into the solid materials needed for real-life applications, and the availability of workable wavelengths have blocked upconversion's path to commercialization.

Lately, however, a flurry of effort from scientists around the world has led to substantial advances in each of those challenging areas. Most notably, researchers have discovered new materials, made up of inorganic nanoparticles and metal-organic compounds, to increase the range of input and output wavelengths.

In our lab, we use these materials to perform a type of upconversion process called triplet-triplet annihilation, which we'll explain in a moment. There are other approaches that use the intrinsic abilities of some rare heavy elements to conduct upconversion. But we chose the triplet-triplet annihilation path because the input energies required are low, so we don't need expensive pulsed lasers. Instead, we can use low-power lasers or even light-emitting diodes, whose intensity is similar to that of laser pointers. Just as important, the materials we are using are more abundant. Together, these characteristics put our technology on an easier path to commercialization.

To understand how triplet-triplet annihilation works, you first need to grasp the concept of electron spin states in organic molecules. Electrons in a molecule are organized in discrete locations. Think of the molecule as a multistory house. Each floor has a single room representing one of these locations, or molecular orbitals. Each room can hold two electrons, but they don't make for good roommates unless the electrons possess opposite qualities, called spin states. Electrons first fill the bottom floors—the lowest energy locations—until all electrons have a room. If a photon hits this molecule (the house), it can excite one of the electrons to a higher energy state, pushing it to an unoccupied room on a higher floor. The electron remains there for only a couple nanoseconds or so before it falls back to the ground state—that is, its original room.

When an electron moves back to the ground state, the molecule releases the same amount of energy—in the form of heat or light—that it absorbed to excite the electron. The short-lived excited state, in which the electron is up in its higher orbital, is called a singlet exciton.

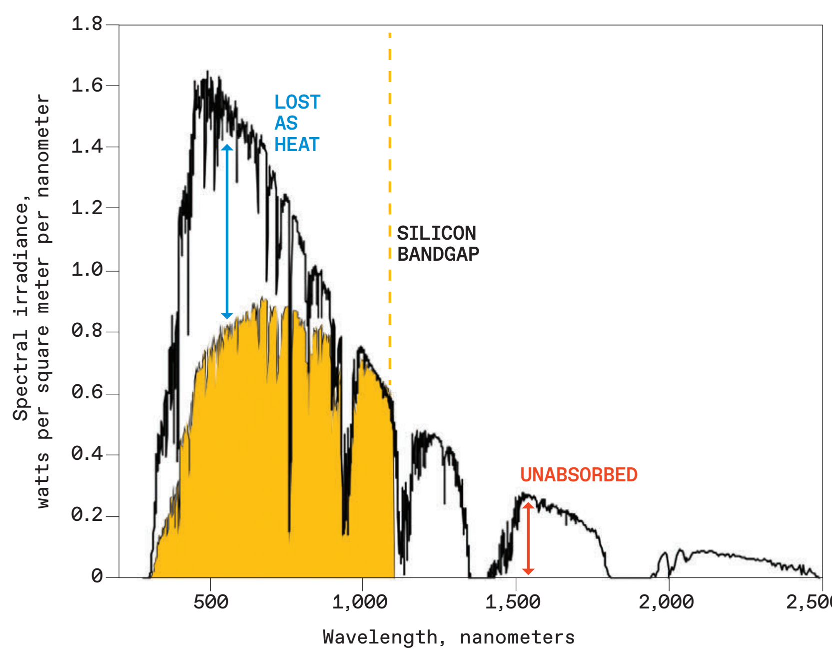
There are other types of excitons. For example, there is a state in which the spins are unpaired (both spin up). Here, once one of the electrons is kicked up to a higher room, it cannot easily relax back to its ground state because that room is already occupied by an electron with the same spin. Still, it does eventually get there. Paint, stickers, and toys that glow in the dark after a period of exposure to light exploit this time lag.

Apart from its use in novelty products, this species, called a triplet exciton, is usually viewed as a nuisance. For instance, in organic light-emitting diodes, it's the singlet excited states that emit light. But both singlet and triplet excited states form in an OLED, with the triplet states reducing the light we see and generating excess heat, both things you don't want in a display technology.

When you're trying to manipulate the color of light, however, these triplet states have a silver lining. If two molecules in triplet states collide, they can sometimes combine their energy. This process is called triplet-triplet annihilation.

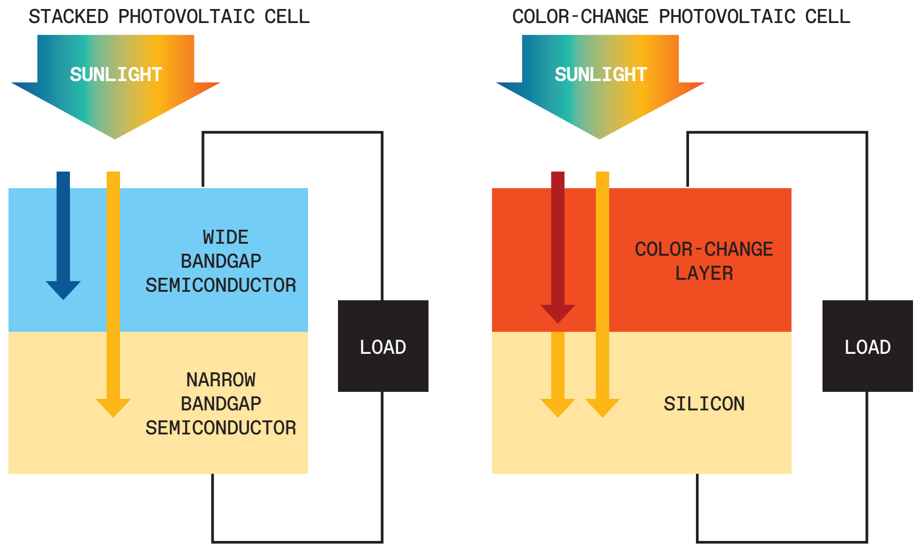
What interests us is that after combining, the resulting molecule can emit a photon at a shorter wavelength—with higher energy—just as if the molecule had been excited with higher-energy light directly. Here's how we make that happen.

We start by generating the triplet excited state, which is a challenge. While several classes of molecules, called organic semiconductor annihilators, can allow triplets to combine, they don't form triplets themselves when directly hit with



A silicon solar cell can efficiently convert photons to electricity only if the photons have energies close to silicon's bandgap. The cell loses much of the energy from shorter-wavelength (higher-energy) photons as heat, and it cannot absorb photons with lower energy. The authors are developing technology that can convert some of those unabsorbed wavelengths into colors close to silicon's sweet spot. SOURCE: ASTM INTERNATIONAL

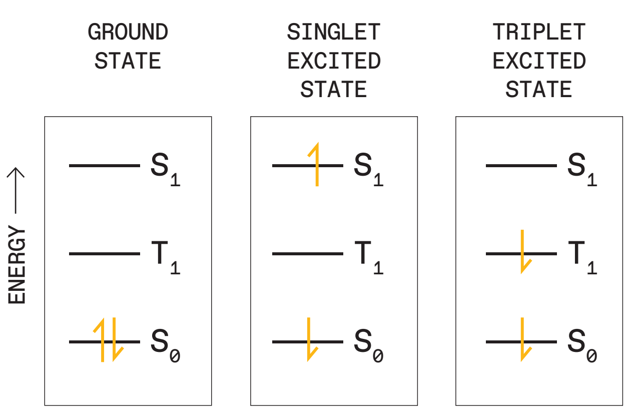
Sunlight contains many wavelengths that silicon solar cells cannot use efficiently. Short wavelengths [blue arrow] will be absorbed, but their excess energy is lost as heat. And long wavelengths [red arrow] aren't absorbed at all. Today, researchers try to capture more wavelengths by stacking multiple types of electricity-generating semiconductors atop one another, but this can be expensive and tricky to design. Instead, a layer of color-change material could convert the long wavelengths to colors that silicon can absorb, thereby simplifying the design and potentially reducing the cost.



light. Instead, we need to use a material called a triplet sensitizer. Triplet sensitizers typically contain a heavy metal like palladium, iridium, or platinum. This heavy metal serves as a mediator, creating a path for the molecule to move a singlet excited state to a lower-energy triplet excited state instead of falling directly to the ground state.

The sensitizer can then donate its triplet to an annihilator molecule, which possesses a triplet excited state slightly lower in energy than the sensitizer's. When its energy is transferred to the annihilator, the sensitizer returns to its ground state without releasing light. The annihilator molecule will eventually emit light—but not just yet.

To get the energy released as light, we need two annihilators in the triplet excited state. So we keep pumping low-energy light into the sensitizers, allowing them to repeat this process over and over, generating multiple excited annihilators and increasing the chances of two of these excited annihilators colliding.



Electrons occupy the ground state [S_0] in pairs with opposite "spins" [up and down arrows, left]. A photon can kick one of the electrons into the singlet excited state [S_1, center]. Usually, the electron will quickly fall back to the ground state and emit a photon. But sometimes the electron's spin can flip, and it gets stuck at a lower energy level, the triplet excited state [T_1, right].

When such a mash-up happens, the annihilators transfer energy in a process called triplet-triplet annihilation, converting one molecule into the singlet excited state and the other molecule into the ground state.

That singlet, however, has the combined energy of two triplets. So when it relaxes into the ground state, the photon it emits is higher in energy than the original photon absorbed by the sensitizer. We have upconverted two low-energy photons into one-high energy photon. In terms of colors, that means we can take two red photons and turn them into a blue one, for example, or take two infrared photons and turn them into a yellow one.

And that's how we change the color of light. Now to get back to the reason we started doing this: photovoltaics.

Sunlight offers abundant photons spanning a wide range of energies, from the ultraviolet through the visible spectrum and into the infrared. Yet we use only a fraction of the available photons. That's why a typical single-junction solar cell—a cell made of one layer of light-absorbing material—has a theoretical efficiency limit of just 34 percent; typical commercial solar cells today are only 15 to 20 percent efficient. The single largest source of this loss is a mismatch between the colors of incoming light and the colors of light that can be used by a solar cell.

To understand this situation, remember that photovoltaic cells are made of semiconductors, materials that possess an energy bandgap. When energy is applied, electrons will move from the valence band (the ground state) to the conduction band (the excited state) and can be harnessed as electrical energy.

If a photon whose energy matches the bandgap of the semiconductor hits a solar cell, this process proceeds smoothly: The incident photon generates an excited electron that is effectively captured to generate electricity. If a photon has an energy greater than the bandgap of the material—as is the case for all visible light for most photovoltaic materials in use—the incident photon generates an electron higher in energy. This excited electron then rapidly relaxes to an energy equal to the bandgap, and all the excess energy is lost as heat, a waste for the solar cell. Even worse, photons with less energy than the bandgap cannot do anything productive at all, and simply pass through the semiconductor unabsorbed.

This presents a difficult trade-off for the solar-cell designer: Wider bandgaps will lose less as heat but absorb fewer photons, while narrower bandgaps will absorb more of the available photons but lose more as heat. Silicon, the ubiquitous light-absorbing photovoltaic material that makes up more than 90 percent of today's photovoltaic market, sits in the sweet spot of this trade-off. Still, even the best experimental silicon solar cells leave almost three-quarters of the available sunlight power unharvested.

This frustrating state of affairs has long inspired scientists and engineers, including us, to search for a better approach.

One promising idea is to use multiple absorber materials to create a stack of solar cells in which each semiconductor is paired with a particular portion of the solar spectrum. Designing these cells can be tricky. For instance, in one configuration each subcell must output the same amount of current; otherwise, efficiency will be limited to that of the worst-performing subcell. Currently, the most efficient device made using three light absorbers under standard illumination—that is, without using lenses or other concentrators—has an efficiency of 39.5 percent.

But we think that changing the color of light can further boost efficiencies: Instead of trying to match the cell to the incoming light, we can match the light to the cell.

That means we convert the photons below the solar cells' bandgap to harvestable, higher-energy photons. In the last few years in our lab at Stanford and in collaboration with other scientists around the world, we have successfully upconverted low-energy infrared photons—which often can't be used by today's solar cells—into productive yellow photons that can. And we translated this chemistry, originally developed in a beaker, into a thin-film material.

We are studying how to improve these gains by controlling how energy moves inside our materials, how the singlet and triplet states interact, and how the light is emitted from the thin film to a solar array. Scientists around the world, including us, are working to develop materials that will enable more-efficient upconversion systems that harvest even further into the infrared. This technology isn't being used commercially yet, but we believe it will get there.

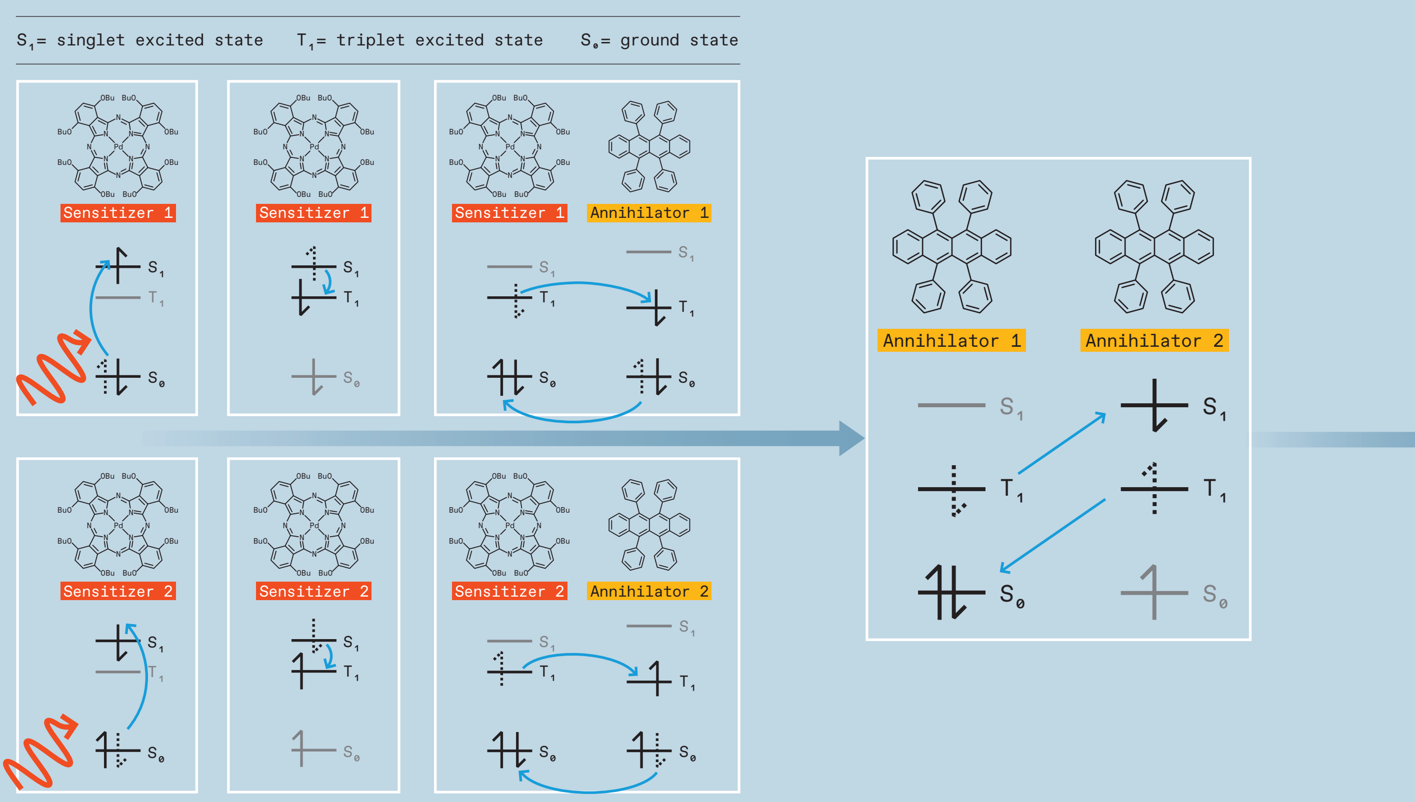
Improving the efficiency of solar cells is far from the only exciting use for changing the color of light through upconversion. This technology can also be used to target light to a precise location, solving a problem common to biology, chemistry, and additive manufacturing.

Preventing undesirable absorption or scattering of light is important in applications as diverse as activating a drug at a tumor site, lighting up a neuron to study brain function, and—perhaps surprisingly—precisely building structures through

UPCONVERTING PHOTONS

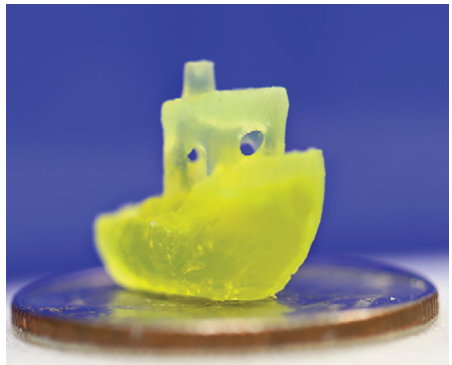
TWO LOW-ENERGY PHOTONS can be turned into a single higher-energy photon by means of triplet-triplet annihilation.

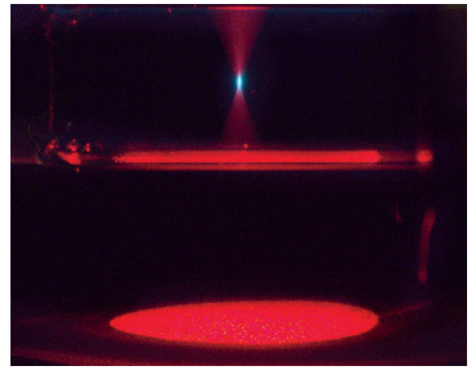
First, a low-energy photon [red wave] strikes a sensitizer molecule. This boosts [blue arrow] an electron from the ground state to the singlet excited state. The excited electron in the sensitizer subsequently falls to the lower-energy triplet state and flips its spin. The sensitizer molecule then exchanges an electron for an electron in the annihilator molecule, rubrene in this case. After this electron transfer, the sensitizer's electrons have returned to the ground



To print this 3D boat [right], materials that can change the color of light were dispersed in a pool of resin. Focusing a red laser [far right] at a point in the pool triggered upconversion, creating a dot of blue light that cured the resin at that spot. Moving the laser in three dimensions built the boat dot by dot.

STANFORD UNIVERSITY/HARVARD UNIVERSITY/NATURE





additive manufacturing. The way we solve this issue is similar in each case, but additive manufacturing (3D printing) is particularly promising and perhaps the easiest to explain.

If one were to imagine the best way to print a shape in three dimensions, without using today's technology, it's easy to picture curing individual points at their *x*, *y*, and *z* coordinates within a vat of resin. Yet curing a single target point without curing the space around it is difficult. Shining a laser beam into the resin, for instance, cures it along the entire laser path.

But we can get to this level of precision by changing the color of light. Here's how that works.

Inside our upconversion 3D printer we use a resin containing dispersed nanoparticles with sensitizers and annihilators. In 3D printing, blue or UV photons are typically used to drive the curing of resin, but we don't start with blue light. Instead, we shine a red laser beam toward our target.

Then we take advantage of the fact that upconversion happens only at certain intensities of light: We use a lens to focus our red beam on a particular point in the resin pool, increasing its intensity at that spot. Upconversion creates a small dot of blue light at the focal point of the red light, curing the resin at the dot. By moving the focal point around, we can create arbitrary shapes deep in our resin pool. What's exciting is that this entire process can be run with a laser no more powerful than a typical laser pointer. We have already made a number of sample objects, including a toy boat, a gear, and some Stanford and Harvard University logos.

Moving forward, we are particularly excited about using this technique to rapidly print many objects at the nanoscale in parallel—something that has been difficult to do, since focusing too many high-powered lasers into one pool of resin simply breaks the resin down before it can be transformed into a solid plastic. The low-power lasers used for upconversion don't do this.

Still further promising developments remain beyond these applications: Upconversion could allow for near-infrared beams, which penetrate deep into living tissue, to generate high-energy photons useful for deep-tissue imaging, optogenetics, and local chemical reactions.

Finally, we're also exploring applications like passive night-vision systems and robust anticounterfeiting schemes. Each of these applications would require a thin coating of upconversion materials on a surface, in the same way we're using our technology with solar cells. Imagine purchasing a pair of glasses with an added upconversion coating that allows you to see infrared photons, improving night vision without the bulky power source required in today's night-vision goggles. Or, if you embedded upconversion nanoparticles in currency or ID cards, distinguishing real from counterfeit would be as easy as shining a red laser pointer at the surface and seeing the light turn blue.

Although each application requires different types of materials, getting high-energy photons to the right place through upconversion can be used to kick-start each one. We're only beginning to scratch the surface of what this technology can do. ■

state and the annihilator is in the triplet excited state. The process repeats itself, generating a second annihilator in the triplet excited state.

Then, annihilator 1 exchanges electrons with annihilator 2. This boosts an electron to the singlet excited state in annihilator 2 and allows annihilator 1 to return to the ground state.

Finally, the energized annihilator emits a photon [yellow wave] of a shorter (higher-energy) wavelength than the ones that were initially absorbed by the sensitizer molecules.

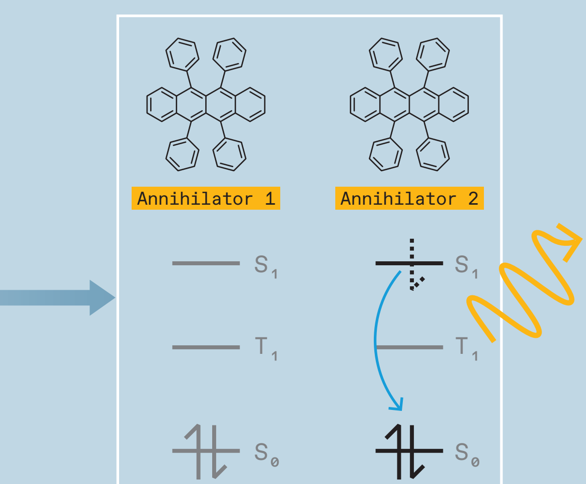


Illustration by Chris Philpot

MIT EECS

Electrical Engineering | Computer Science | Artificial Intelligence + Decision-making

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Cambridge, MA
FACULTY POSITIONS

The Massachusetts Institute of Technology Department of Electrical Engineering and Computer Science (EECS) in Cambridge, Massachusetts seeks candidates for faculty positions starting July 1, 2024, or on a mutually agreed date thereafter. We welcome outstanding applicants with research and teaching interests in any area of electrical engineering, computer science, and artificial intelligence and decision making. EECS believes that the intellectual, cultural and social diversity of our faculty, staff, and students is vitally important to the distinction and excellence of our academic and research programs. The Department seeks candidates who support our institutional commitment to ensuring that MIT is inclusive, equitable, and diverse.

Appointment will be at the assistant or untenured associate professor level. In special cases, a senior faculty appointment may be possible, commensurate with experience. Faculty duties include teaching at the undergraduate and graduate levels, research, and supervision of student research. Candidates should hold a Ph.D. in electrical engineering and computer science or a related field by the start of employment. Employment is contingent upon the completion of a satisfactory background check, including possible verification of any finding of misconduct (or pending investigation) from prior employers.

Candidates must register with the EECS search website at https://faculty-searches.mit.edu/eecs, and must submit application materials electronically to this website. Applications must include a cover letter, curriculum vitae, a research statement (2-4 pages) and a teaching statement (1-2 pages). In addition, candidates should provide a statement regarding their views on diversity, inclusion, and belonging, including past and current contributions as well as their vision and plans for the future in these areas. Each application should include the names and addresses of three or more individuals who will provide letters of recommendation. Letter writers should submit their letters directly to MIT, preferably on the website or by mailing to the address below. Complete applications should be received by December 1, 2023. Applications will be considered complete only when both the applicant materials and **at least three letters of recommendation are received**.

It is the responsibility of the candidate to arrange reference letters to be uploaded at https://faculty-searches.mit.edu/eecs by December 1, 2023.

Send all materials not submitted on the website to:
Professor Asu Ozdaglar
Department Head, Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Room 38-403
77 Massachusetts Avenue
Cambridge, MA 02139

MIT is an equal employment opportunity employer. All qualified applicants will receive consideration for employment and will not be discriminated against on the basis of race, color, sex, sexual orientation, gender identity, religion, disability, age, genetic information, veteran status, ancestry, or national or ethnic origin. MIT's full policy on Nondiscrimination can be found at the following: https://policies.mit.edu/policies-procedures/90-relations-and-responsibilities-within-mit-community/92-nondiscrimination.

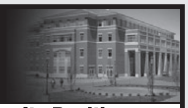
上海交通大学 | 溥渊未来技术学院

Global Institute of Future Technology (GIFT hereafter) at Shanghai Jiao Tong University invites applications for tenure-track or tenured professoriate and research scientist positions, mainly at Associate and Full levels, related to Sustainable Energy.

GIFT manages endowment funds from a variety of sources, including a recent donation of more than 200 million US dollars. GIFT's new building will be completed within two years, adding 50,000 square meters to the current 25,000 square meters. With strong public and university support, we will rapidly expand our faculty, including high-level hiring (e.g. NAE/NAS members and Fellows of multiple professional societies). Currently, five research centers have been established: Center for Advanced Energy System and Reliability, Future Battery Research Center, Future Photovoltaics Research Center, Center for Green Energy and Future Agriculture, and Intelligent Extreme Manufacturing Research Center. Applicants must hold a doctorate degree in a relevant field. Candidates are expected to establish a vigorous research program and contribute to undergraduate and graduate education.

For full consideration, please send a CV, teaching plan and research plan, copies of three representative publications and contact information of five referees as a single PDF file to the GIFT Search Committee (email: gift-facultysearch@sjtu.edu.cn). More information is available at http://gift.sjtu.edu.cn.

UNIVERSITY OF NORTH CAROLINA CHARLOTTE



Tenure Track Faculty Positions

The UNC Charlotte Department of Electrical and Computer Engineering (ECE) invites applications for multiple 9-month tenure-track or tenured faculty positions at Assistant, Associate or Full Professor ranks starting from Spring 2024. Candidates are sought to fill positions in the areas of: **(1) Electromagnetics** with expertise in electromagnetics, metamaterials, photonics, acoustics, massive MIMO antennas, beamforming, antenna design, RF/millimeter/THz devices and systems, electromagnetic modeling and simulation, and time-variant and non-reciprocal electromagnetic devices; **and (2) Semiconductor devices** with expertise in semiconductor device physics, electronics, optoelectronics, power electronics, nanoscale and novel devices, and semiconductor device design and fabrication techniques. Exceptional candidates in related areas are encouraged to apply.

Candidates should be committed to excellence in teaching at the undergraduate and graduate levels, development of sponsored research programs, supervision of student research, student mentoring, and academic services. Priority will be given to candidates having original and promising research as demonstrated by peer-reviewed publications, potential for success in extramural funding, and a plan for future research that integrates well with existing and aspirational department strengths. Hires at the rank of Associate Professor and Full Professor are also expected to have demonstrated success in receiving extramural research funding, supervising/mentoring of graduate students, and teaching. A Ph.D. degree in Electrical Engineering and related fields is required by the time of contract.

Applications must be made electronically at **https://jobs.charlotte.edu**. Review of applications will begin September 15, 2023, and the position will remain open until filled. *EOE*

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

The Department of Electrical and Computer Engineering in the Samuel Ginn College of Engineering at Auburn University invites applications for tenure-track Assistant Professor positions in the areas of image processing, computer engineering and microelectronics. Please visit **www.eng.auburn.edu/elec** for details about these positions and application instructions.

Auburn University is an EEO/Vet/Disability Employer. Auburn University is understanding of and sensitive to the family needs of faculty, including dual career couples.

ECE — ELECTRICAL & COMPUTER ENGINEERING — UNIVERSITY OF MICHIGAN

The Electrical and Computer Engineering (ECE) Division of the Electrical Engineering and Computer Science Department at the University of Michigan, Ann Arbor invites applications for junior and senior faculty positions.

Successful candidates will have a relevant doctorate or equivalent experience and an outstanding record of research in academia, industry and/or at national laboratories. They will have a strong commitment to teaching at undergraduate and graduate levels, to providing service to the university and profession, and to broadening the intellectual diversity of the ECE Division; and have expansive world-views on the potential impact of their research.

We invite diverse candidates across all research areas to apply.

The highly ranked ECE Division prides itself on the mentoring of junior faculty toward successful careers.

Ann Arbor is highly rated as a family friendly best-place-to-live.

Please see application instructions at:
https://ece.engin.umich.edu/people/faculty-positions

Applications will be reviewed as they are received. Submission of applications prior to Nov 30, 2023 is strongly recommended for full consideration. The application site will remain open until the positions are filled.

Michigan Engineers are world-class educators, researchers, students and staff who strive to build a people-first future. As part of the nation's number one public research institution, Michigan Engineering's mission is to provide scientific and technological leadership to the people of the world, develop intellectually curious and socially conscious minds, create collaborative solutions to societal problems, and promote an inclusive and innovative community of service for the common good.

Our vision, mission and values are supported by a people-first engineering framework that guides our work. As Michigan Engineers, we strive to apply excellent engineering fundamentals, integrated expertise and equity-centered values to reimagine what engineering can be, close critical gaps, and elevate all people. Information about our vision, mission and values can be found at: http://strategicvision.engin.umich.edu.

The University of Michigan has a storied legacy of commitment to Diversity, Equity and Inclusion (DEI). Michigan Engineering models that commitment in our research, culture and collaborations. We seek to recruit and retain a diverse workforce as a reflection of that commitment. Learn more about DEI at Michigan Engineering: **https://www.engin.umich.edu/culture/diversity-equity-inclusion**

The University of Michigan is an Affirmative Action, Equal Opportunity Employer with an Active Dual-Career Assistance Program. The College of Engineering is especially interested in candidates who contribute, through their research, teaching, and/or service, to the diversity and excellence of the academic community.

RIT | Rochester Institute of Technology

Tenure-Track/Tenure Assistant /Associate Professor

The Department of Electrical and Microelectronic Engineering at the Rochester Institute of Technology invites applications for tenure/tenure-track position from candidates in the area of Control/Robotics. Applicants must have a Ph.D. in Electrical Engineering or a closely related field. The candidate's research must specialize in Control/Robotics, with a record of refereed journal & conference publications. Candidates must have a strong aptitude and interest in undergraduate and graduate teaching, strong potential for establishing and conducting sponsored research, excellent written and oral communication skills, desire to provide service to the university and professional community and, ability to contribute in meaningful ways to the university's continuing commitment to cultural diversity, pluralism, and individual differences. Highly qualified candidates will be considered for an endowed chair.

Apply online at https://apptrkr.com/4489121.
Keyword Search: 8108BR.
Inquiries can be sent to Dr. Sohail Dianat (sadeee@rit.edu).

RIT does not discriminate. RIT is an equal opportunity employer that promotes and values diversity, pluralism, and inclusion. For more information or inquiries, please visit RIT/TitleIX or the U.S. Department of Education at ED.Gov.

BOSTON UNIVERSITY

The College of Engineering at Boston University has embarked on a bold new strategic plan aimed at excellence and impact along convergent and collaborative research themes. **The Department of Electrical & Computer Engineering** (ECE) at Boston University (BU) anticipates multiple openings for a **Tenure-Track Assistant Professor** in the area of Computer Systems. We seek candidates in operating systems, compilers, cybersecurity, or software engineering to build future cloud and edge computing systems for a smart, secure and connected society. Potential application areas include healthcare, communications, transportation, finance, and scientific computing.

Candidates with research programs that transcend the traditional boundaries of ECE may explore affiliated appointments in appropriate departments and divisions, such as Computer Science, Mathematics and Statistics, Systems Engineering, or the newly created Faculty of Computing and Data Sciences.

BU ECE attracts exceptional undergraduate and graduate student and faculty talent at all levels. Research activity by primary faculty is approximately $26M per year. The College of Engineering is currently ranked 35th in the nation by US News and World Report, and 15th among private universities. The College is 5th in the nation in total funding from NSF among engineering schools at private universities. BU ECE faculty lead and participate in several high-profile, multidisciplinary research centers, including the Center for Information and Systems Engineering, the Hariri Institute for Computing and Computational Science and Engineering, the Center for Systems Neuroscience, the Rajen Kilachand Center for Integrated Life Science and Engineering, and the Photonics Center.

We are looking for outstanding candidates who have earned or are expected to earn a Ph.D. in the relevant search area before Fall 2024, demonstrate potential for leading an independent and vibrant funded research program in their area of expertise, can teach effectively at the graduate and undergraduate levels, and can utilize their expertise to strengthen collaborative research within the department and beyond.

For more information about BU ECE, please visit:
http://www.bu.edu/ece/

We encourage candidates to apply early. Applications received by December 15, 2023 will be given full consideration.

For more information and to apply please visit:
https://academicjobsonline.org/ajo/jobs/25188

Boston University is an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability status, protected veteran status, or any other characteristic protected by law. We are a VEVRAA Federal Contractor.

BY ALLISON MARSH

Past Forward

This ingenious solar-system display was designed by musician-turned-engineer Clair Omar Musser.

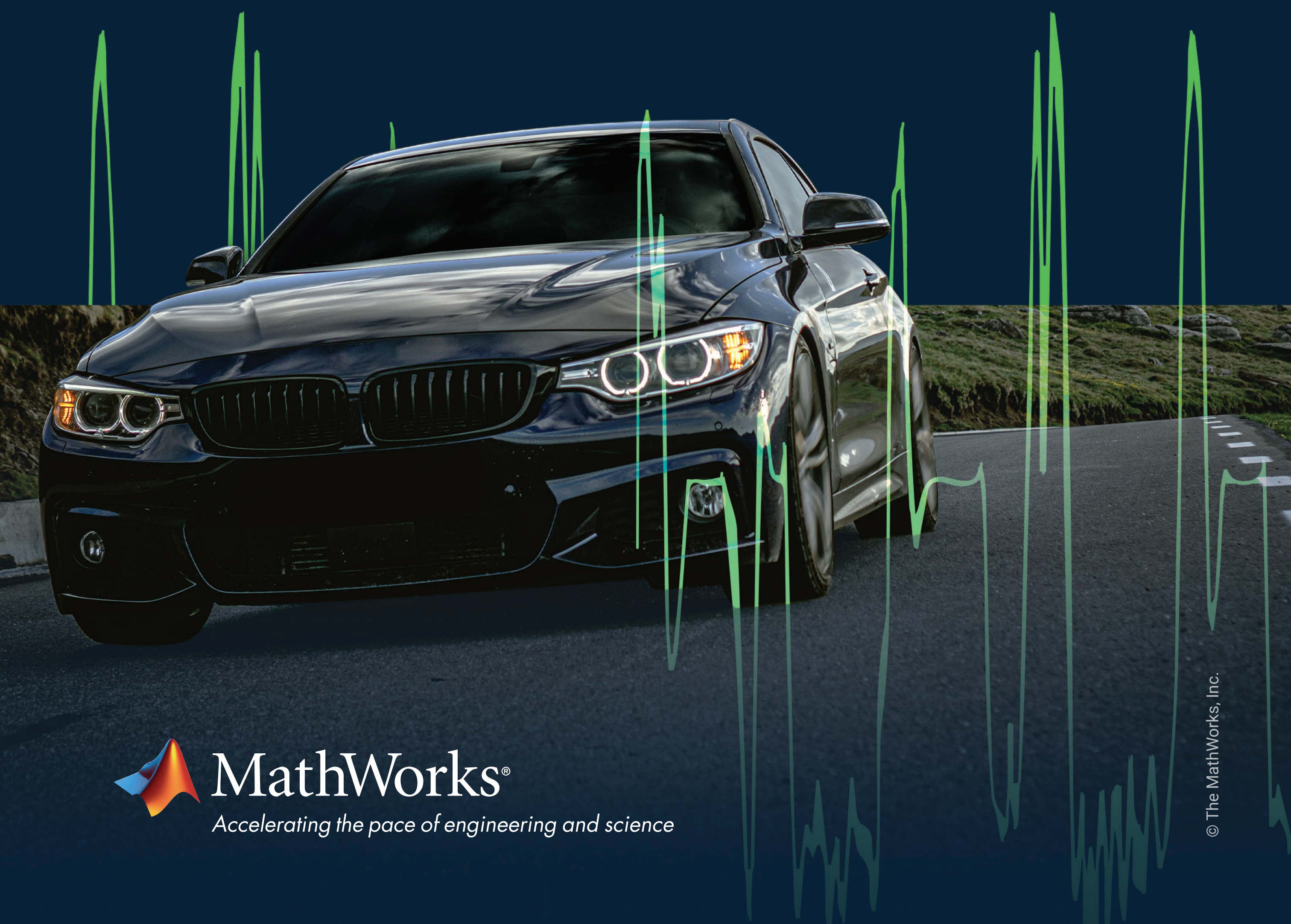
WHIPPLE MUSEUM OF THE HISTORY OF SCIENCE/UNIVERSITY OF CAMBRIDGE

No Ordinary Orrery

An orrery is a physical model of the solar system that shows the movement of the planets, moons, and other celestial bodies. For centuries, orreries were mechanical displays that ran on intricate clockwork mechanisms. In the early 1960s, though, the orrery got a makeover that aligned with the public's growing interest in all things space. In the electronic desktop version shown here, a red lightbulb represents the sun, surrounded by brightly colored planets that the user could manipulate using dials at the bottom of the screen. Different overlays map the moons, stars, and even Halley's Comet. The orrery's designer, Clair Omar Musser, was an engineer with Hughes Aircraft Co. who also happened to be a world-renowned marimba virtuoso. ■

FOR MORE ON MUSSER AND HIS INSTRUMENTS, see spectrum.ieee.org/pastforward-oct2023


BILLS DON'T STOP FOR DISABILITY.
WHY SHOULD YOUR INCOME?
Help protect your livelihood and lifestyle.



MathWorks®
Accelerating the pace of engineering and science
© The MathWorks, Inc.