IEEE
SPECTRUM

FOR THE TECHNOLOGY INSIDER | 09.18

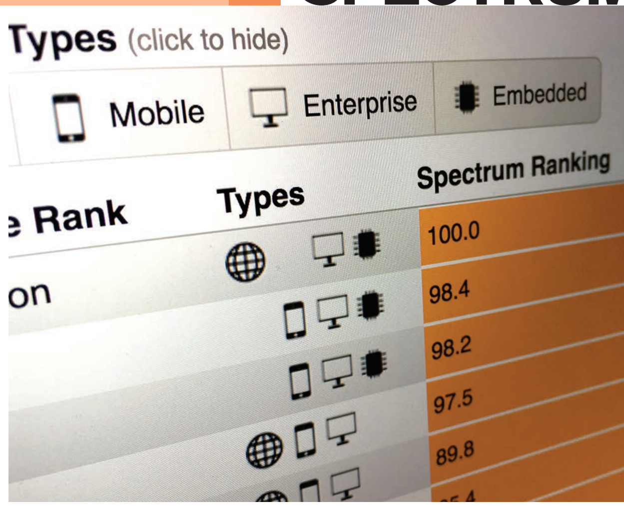
DROWNING *in* DATA

Our best hope for coping with the data deluge may be **magnetic tape**—it never went away, and it's getting better faster than any other medium P. 32

IEEE

On the cover and this page **Photographs for *IEEE Spectrum* by Victor Prado**

IEEE SPECTRUM



ADDITIONAL RESOURCES

Tech Insider / Webinars

Available at spectrum.ieee.org/webinars

- Magnetic Transmission Design Using Multiphysics Modeling

White Papers

Available at spectrum.ieee.org/whitepapers

- How to Optimize the Scalability and Performance of a Multicore Operating System
- Advantages of Using Silicone Adhesives for Medical Device Assemblies
- How to Create an IoT Edge System Prototype

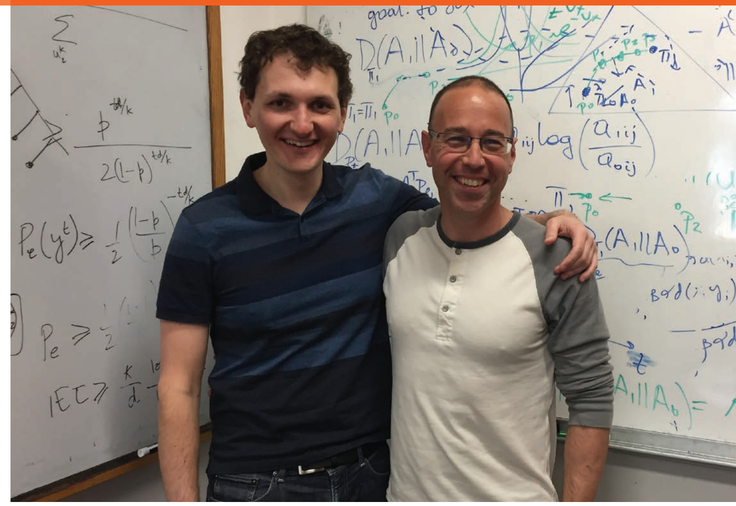
The Institute

Available at theinstitute.ieee.org

- **HOW TECH IS COMBATING CRIME** Learn about technologies being developed to help reduce illegal activities such as human trafficking.
- **THE HISTORY OF POLICE RADIO COMMUNICATION** One- and two-way radios developed in the late 1920s and early 1930s gave police a reliable way to dispatch patrol cars to the crime scene and allow communication between the cars and the police station.
- **NEW FORUM FOR ETHICS IN AUTONOMOUS AND INTELLIGENT SYSTEMS** This open community, established by the IEEE Standards Association, is for those organizations that want to develop and use standards to address ethical matters as well as explain how standards can help to solve problems in ways that go beyond technical solutions.

FROM LEFT: ESA; CARLTON SOOHOO; RANDI KLETT

IEEE SPECTRUM
(ISSN 0018-9235) is published monthly by The Institute of Electrical and Electronics Engineers, Inc. All rights reserved. © 2018 by The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016-5997, U.S.A. Volume No. 55, issue No. 9, International edition. 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, c/o 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 120 Donnelley Dr., Glasgow, KY 42141-1060, U.S.A. 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 http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html.



HBO'S DATA COMPRESSION STARS

FOR FANS OF THE HBO TELEVISION SHOW "Silicon Valley," the name "Weissman" and "data compression technology" go together like Heinz and ketchup. The show has prominently featured the Weissman Score, a metric that rates the power of a compression algorithm.

Stanford professor of electrical engineering Tsachy Weissman [right] helped create that metric for the show, which premiered in 2014. Although the algorithm was made for TV, academics in the real world picked up on it.

The creators of "Silicon Valley" tapped Weissman as they were developing the first season and casting about for a technology to feature. They had settled on a universal compression algorithm, but they needed an expert to come up with the specifics of a technology that would be plausible but not possible today. They reached out to Weissman, and he laid out what he calls the holy grail of the compression world: a form of powerful and efficient lossless compression that can work on any type of data and is searchable.

That worked out nicely for the show, but given that such a compression scheme won't exist in the real world anytime soon, Weissman continues to pursue his research. Most recently, his work has focused on genomic compression algorithms, which he and Dmitri Pavlichin [left], a Stanford postdoctoral fellow, describe in "The Quest to Save Genomics" [p. 26].

Incidentally, Pavlichin took over Weissman's original gig on "Silicon Valley" in 2017, coming up with plausible technology for season 4. He contributed sketches of whiteboards and technical documents and edited snippets of dialogue. A favorite moment of his came when the show's characters attempt to move their server to the Stanford campus and give a shout-out to Sherlock, a computing cluster that Pavlichin uses daily.

Recently, the show's fictional startup pivoted from compression technology to decentralizing the Internet. But the Weissman Score has already earned its place in tech history. ■

TEKLA S. PERRY

CONTRIBUTORS_

Carl Bugeja

Bugeja, an embedded-software developer based in Malta, works on open-source-hardware robotics projects in his spare time. In this issue, he describes how he built a small motor that uses coils made from the traces of a standard multilayer printed circuit board [p. 18]. Usually, motors are among the most expensive parts of a robot, Bugeja notes. Using lower-cost PCB motors instead could make robotics more accessible for hobbyists.

Mark Lantz

Lantz joined IBM's research lab in Zurich in 2001 to work on data storage using nanoscale indentations in polymer film. That didn't pan out, but the technologies he helped develop could be applied to magnetic tape. Back then, he says, "10 terabytes of capacity seemed a crazy goal. Now we have 15-TB cartridges and do demos that show we can store hundreds of terabytes," as he writes about in this issue [p. 32]. "I'm always a little bit amazed."

Anthony F.J. Levi

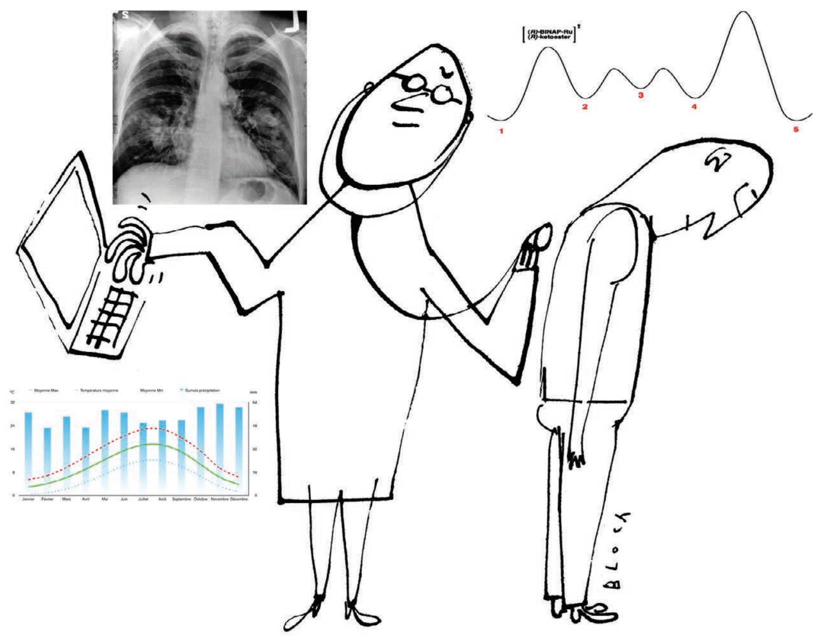
Levi's research into linking computer processors directly with optical fibers was featured 16 years ago in the pages of *IEEE Spectrum*. "Way back then we knew we had to get fiber to the processor," Levi says. In "Silicon Photonics' Last-Meter Problem" [p. 38], the University of Southern California professor explains why this technology still isn't a commercial reality and what it will take to make it one.

Weiyang Li

In 2013, Li heard that lithium-ion battery fires had grounded the Boeing 787 Dreamliner. "I learned that the problem could have been caused by dendritic lithium," she says. Now an assistant professor of engineering at Dartmouth University, Li was a postdoc in Yi Cui's lab at Stanford University at the time. The two researchers turned their attention to the dendrite problem; you can read about their proposed solution in "Less Fire, More Power" [p. 44].

Victor Prado

Prado, a still-life photographer based in New York City, shot the images for the cover and for "Tape Storage Mounts a Comeback" [p. 32]. He initially tried to photograph curls of magnetic tape resting on clear acrylic. "I thought it was going to be easy," says Prado, but static charge doomed the attempt. Still, he prefers such struggles to the challenges of photographing people. "In the end, it's only objects you have to deal with," he says.



Training Tomorrow's Doctor-Engineers

The medical profession is changing, and so must medical education

When King Li went to medical school 40 years ago, he sometimes felt like he was trying to memorize the human body. "I'd see these curves in the textbook, and have to memorize: 'If the patient's kidneys are failing, reduce this drug by 30 percent,' " he remembers. "But I didn't understand why." ¶ In many medical schools, little has changed. But Li is heading up a brand new school with a radically different approach. When it comes time for the students to study pharmacology, he says, they'll have more than a chart in a textbook–they'll have a computer program with a simulated patient to learn on. "If the kidneys start failing, they can experiment with changing the dosage of the drug," he says. ¶ Li is dean and chief academic officer for Carle Illinois College of Medicine, created in partnership by the University of Illinois at Urbana-Champaign and the Carle Health System. Li says the school is the first in the world to apply engineering principles to the teaching of medicine, bringing analytics and problem-based learning to every aspect of the curriculum. The school just welcomed its first class of 32 students, who arrived with more than the typical premed coursework under their belts; students are also required to have high-level math, computer science, and statistics. ¶ While plenty of M.D.-Ph.D. programs already exist that enable students to get both medical and advanced engineering degrees, Li says Carle offers something very different than that "layered-on" method. Rather than going to medical school to learn about the human cardiovascular system and engineering school to learn about fluid dynamics, he says, Carle's students will learn about both at once. Every course will have three instructors, to cover the biological science, clinical applications, and engineering aspects of the topic.

These students are being trained to enter a profession that's in the throes of transformation, with advances in sensors, devices, and computing changing how doctors do their jobs. Artificial intelligence (AI) systems will likely play an increasingly important role, ingesting vast quantities of medical data and provide lightning-fast analytics. In the best-case scenario for doctors, AI and other technologies will take on much of the grunt work of medical practice, giving doctors more time to spend with patients and focus on challenging cases. But there are other possible futures in which machines take on so much of the work that human experts are less necessary. How these trends will play out is highly uncertain.

Li hopes that Carle's students won't just be open to change, but will challenge the status quo and help invent the new ways of doing things. As the students go through each clinical rotation, learning about specialties such as internal medicine, pediatrics, and surgery, they'll be required to identify a problem or inefficiency they see in their patient rounds and come up with a technological solution.

Rashid Bashir, executive associate dean of the new school and department head of bioengineering at Urbana-Champaign, says these "physician-innovators" can take on the grand challenge of health care. The United States spends more than $3 trillion annually on health care, Bashir notes, accounting for nearly 18 percent of the gross domestic product, and the figures are projected to keep rising. "In almost every industry, the insertion of technology has helped reduce costs," he says. "Think about electrical engineering: Every generation of chips is lower cost and more powerful. But in health care, it's the other way around: Costs keep going up." When the Carle students see that cost curve in a textbook, they won't just commit it to memory–they'll start experimenting. –ELIZA STRICKLAND

↗ **POST YOUR COMMENTS** at https://spectrum.ieee.org/carle0918

ILLUSTRATION BY Serge Bloch

NEWS

100

NUMBER OF WIND FIELD MEASUREMENTS THAT AEOLUS TAKES EVERY HOUR

ATG MEDIALAB/ESA

SHINING A LIGHT ON GLOBAL WINDS

The Aeolus satellite uses powerful lidar to measure wind speeds

CYCLONE WATCH: The data collected by Aeolus will improve the accuracy of weather forecasts.

It's closing time for one of Earth observation's most stubborn and critical data gaps: global wind speeds. A European Space Agency (ESA) satellite set for launch from French Guiana late last month—after nearly two decades of challenging engineering—will be the first to directly measure wind speed and direction, from Earth's surface to the stratosphere.

Winds are key determinants of weather and climate, yet most wind data still comes from weather balloons. Readings from commercial jets supplement the balloons' twice-daily samplings, along with estimates inferred from satellites that track moving clouds, atmospheric temperatures, and sea-surface roughness. The result is a patchy wind record that adds uncertainty to weather forecasts.

The new satellite, named Aeolus, will improve on these measurements by deploying a wind-sensing lidar in space for the » first time. It is essentially the same light-pulsing object-detection technology that helps police enforce speed limits and autonomous vehicles navigate roads. But instead of detecting motion by pinging a car with laser light and measuring the Doppler shift in the photons reflected back, Aeolus will ping molecules and aerosols in the air from 320 kilometers above Earth.

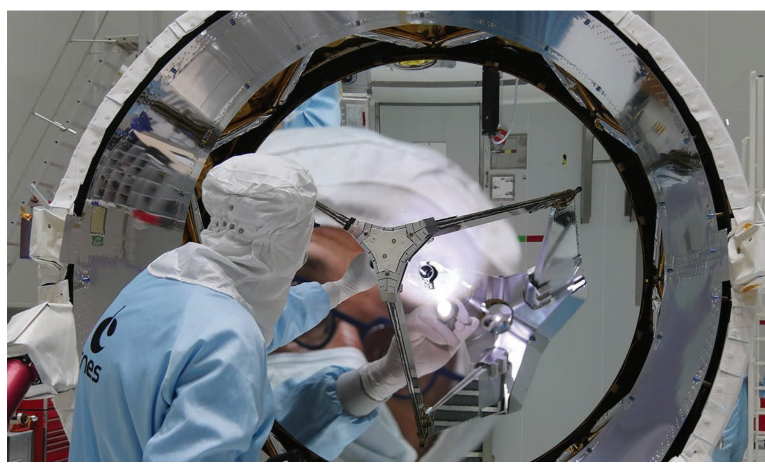
"It's inherently difficult. The backscattered signal from the atmospheric constituents is quite weak, so you need a relatively large laser and a large telescope to gather enough photons," explains Michael Hardesty, a lidar expert and senior research scientist at the Cooperative Institute for Research in Environmental Sciences, based in Boulder, Colo., a partnership of the University of Colorado and the National Oceanic and Atmospheric Administration. Hardesty serves on ESA's mission advisory group for Aeolus.

The scientists working on Aeolus had to assemble their own ultraviolet laser because the existing ones had been built for nuclear weapon simulations and thus were classified. By 2005, they had a suitably powerful and stable laser that would pulse 50 times per second. With a telescope and dual receivers, the team could measure the laser's reflection from atmospheric molecules (mainly nitrogen) or particles and, based on its frequency, discern average wind direction and speed.

Then a potential showstopper emerged when the team added mirrors and lenses and ran the device in a vacuum. The laser's potent pulses heated the optical surfaces to 1,700 °C, damaging their coatings. And volatile organic compounds (VOCs) sucked from the instrument's semiconductors by the vacuum formed a carbon layer on the lenses.

A decade passed as the Aeolus team developed robust coatings and a workaround for the soot buildup—adding an oxygen tank and continuously bleeding in trace amounts to burn away the VOCs in the lidar instrument. Finally, late last year, the scientists vacuum-tested the entire spaceship. In all, the extra work inflated the mission's cost more than 50 percent, to an estimated €481 million (US $560 million).

Last month's launch was supposed to place Aeolus in an inclined polar orbit, sending it up and around Earth about once every 90 minutes. As its lidar looks straight down at Earth, the satellite should produce a continuous ribbon of wind profiles starting 30 km above



SHOWTIME: Aeolus waits on a trolley prior to launch [top]. Its wind lidar, named Aladin, will be the first of its kind in space. The satellite's telescope [bottom] will collect light reflected from atmospheric particles. Here, an engineer checks the telescope for dust and smudges.

ESA (2)

the planet and reaching down to the surface, or to clouds thick enough to attenuate its beam.

Aeolus's data will most dramatically improve weather and climate science for areas with the sparsest wind measurements today, such as over oceans and the tropics. Experts anticipate significant improvement for Atlantic hurricane landfall predictions, and simulations suggest that it could improve the accuracy of typhoon path projections by 9 percent.

The ESA even expects the satellite to reveal some storms that would otherwise surprise forecasters. For example, in March 2014, a weather system flooded Northern Europe without warning because models could not see strong westerly winds blowing above the tropical Pacific.

Gil Lizcano, R&D director for the Barcelona-based meteorologic analytics firm Vortex, says the satellite's data will also help tune algorithms to more accurately predict wind "ramp up" and "ramp down" events, which occur when winds quickly shift from low to high speeds, or vice versa. At present, such surges—and the corresponding rise or fall in output from wind turbines—force power plants to make last-minute adjustments that can be costly for both wind farm owners and grid operators.

Aeolus has enough fuel and oxygen to orbit for only three or four years, but a U.S. mission may follow, according to Gail Skofronick Jackson, a wind detection specialist at NASA's headquarters in Washington, D.C. In January, a National Academies of Sciences report identified winds as a "high priority targeted observable," notes Jackson. And hardware is already taking shape: NASA aircraft recently demonstrated a novel wind lidar developed by Hardesty's team and Colorado-based Ball Aerospace.
—PETER FAIRLEY

↗ **POST YOUR COMMENTS** at https://spectrum.ieee.org/aeolus0918

ROY PEÑA/UNIVERSITY OF TEXAS N

NEWS

HYDROGEN HYBRIDS DEBUT AS ZERO-EMISSION DELIVERY TRUCKS

Fuel cells extend the range of battery-powered trucks now heading for California roads

Austin Mabrey steers the clanging United Parcel Service (UPS) van down a street in Austin, Texas. But he's not driving the boxy brown vehicle to deliver packages. Mabrey is road-testing its zero-emission system—a hybrid of hydrogen fuel cells and lithium iron phosphate batteries.

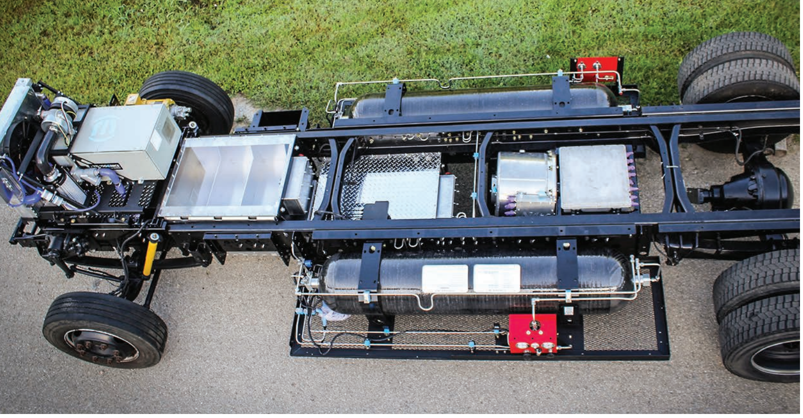
"It's peppier than I would've imagined," he says. Near my perch in the passenger seat, a high-pitched hum emanates from the electric motor that drives the hydraulic power-steering pump. As we approach a narrow turn, Mabrey engages the regenerative braking system, which recharges the batteries, and a whining noise erupts from the back.

We're circling the Center for Electromechanics at the University of Texas (UT), where engineers are almost finished testing the van's power train inside a cavernous research hangar. They began road trials in June after working for more than a year to design and model the concept, though the project first won federal funding in 2013. UPS plans to deploy the prototype in California later this year and, if all goes well, roll out more vehicles just like it.

Logistics companies and automakers worldwide are developing vans and trucks that don't emit any pollution. But it's much more complicated to build a zero-emission cargo truck than it is to produce an emissionless passenger car. New fuel systems can't encroach on cargo space or add more weight to a truck's bulky frame. And trucks must be able to run their normal routes without making extra stops to recharge batteries or refill tanks.

"The driver has to be able to accomplish their mission—it's a work truck," says Joe Ambrosio of Unique Electric Solutions, which is integrating the UPS van's electric components. The New York firm hired six interns from UT to work on the project, including »

ROAD READY: This converted UPS truck features a 32-kilowatt fuel-cell module from Hydrogenics.



Mabrey, who is now an engineer at the company.

From the sidewalk, the van looks like any other delivery vehicle. UPS provided a 2007 diesel van to UT researchers, who converted it into a fuel-cell/battery hybrid. The new system includes a high-power, 99-kilowatt-hour battery pack from Lithium Werks that sits between the chassis frame rails [see above]. Two 10-kilogram hydrogen tanks saddle the rails, while a 32-kilowatt fuel-cell module from Hydrogenics is stored below the hood, where a conventional engine would be.

Engineers designed the van for a range of up to 200 kilometers, which it can achieve thanks to its "range extender"–the fuel-cell module, Ambrosio says. Using hydrogen, the vehicle can travel longer distances and make more stops than a purely battery-powered van, he says.

Michael Lewis, a senior engineering scientist and the project lead at UT, says the first challenge in building the system was "right-sizing" its components. The battery pack, fuel-cell module, and hydrogen tanks needed to be big enough to support the van's operations but still fit within its existing dimensions.

The team refined its early designs based on real-world duty-cycle measurements gathered from UPS vans in California and Texas, which revealed how far the vehicles typically travel and how hilly or strenuous their routes are. UPS's telematics technology can gather 1,700 data points per second, helping engineers troubleshoot problems and spot inefficiencies in fuel or battery use.

Once the road tests are completed this year, UT will transfer the technology to Unique Electric Solutions to retrofit potentially 15 "phase two" vans, which will feature the final fuel-cell/battery system developed in Texas. UPS aims to then deploy those trucks across California.

The UPS van reflects a broader push by state and federal agencies to accelerate clean energy technologies, including hydrogen. The California Energy Commission, the South Coast Air Quality Management District, and the U.S. Department of Energy are funding the project, and the nonprofit Center for Transportation and the Environment is serving as program manager.

Thousands of hydrogen-powered forklifts and passenger cars, dozens of buses, and at least one other delivery van are now on U.S. roads. Hydrogen refueling stations are also beginning to pop up in California and a few other states, as well as in China, Japan, South Korea, and Germany.

"There's a real viable market beginning to blossom in certain areas of the world," says Andy Marsh, CEO of Plug Power.

Marsh's fuel-cell company is working with FedEx and Workhorse Group to build 20 zero-emission delivery vans by next year. As of May, the first of the fuel-cell/battery vehicles had begun to haul packages at a FedEx distribution facility in New York.

"This is the year that hydrogen fuel-cell vehicles began to really show some momentum," Marsh says. —MARIA GALLUCCI

↗ **POST YOUR COMMENTS** at https://spectrum.ieee.org/hybridtrucks0918

AI-HUMAN PARTNERSHIPS TACKLE "FAKE NEWS"

Machine learning can get you only so far—then human judgment is required

During the 2016 U.S. presidential election, inaccurate and misleading articles burned through social networks. Since then, tech companies–from behemoths like Facebook and Google to scrappy startups–have built tools to fight misinformation (including what many call "fake news," though that term is highly politicized). Most companies have turned to artificial intelligence (AI) in hopes that fast and automated computer systems can deal with a problem that's seemingly as big as the Internet.

"They're all using AI because they need to scale," says Claire Wardle, who leads the misinformation-fighting project First Draft, based in Harvard University's John F. Kennedy School of Government. AI can speed up time-consuming steps, she says, such as going through the vast amount of content published online every day and flagging material that might be false.

But Wardle says AI can't make the final judgment calls. "For machines, how do you code for 'misleading'?" she says. "Even humans struggle with defining it. Life is messy and complicated and nuanced, and AI is still a long way from understanding that."

Facebook, which was widely criticized for failing to take action against false content in 2016, says it will use AI to do better in the U.S. midterm elec-

ROY PEÑA/UNIVERSITY OF TEXAS

tions this November—and in other elections around the world. Jim Kleban, a Facebook product manager who works on reducing misinformation in the site's news feeds, explains that Facebook now uses AI to augment human intelligence. The AI goes through the millions of links shared on Facebook every day to identify suspect content, which is then sent to human fact-checkers. "For the foreseeable future, all these systems will require hybrid solutions," he says.

When fact-checkers rate a piece of content as false, Facebook places it lower in users' news feeds. Kleban says this method reduces future views of that content by 80 percent.

Facebook's AI is trained via machine learning, a technique in which an AI system takes in a huge data set of labeled material and independently finds patterns. For example, an image-sorting AI might look at millions of photos labeled either "cat" or "dog" and learn the distinguishing characteristics of felines and canines. But training an AI to recognize false content is much trickier.

Kleban says the Facebook AI uses a variety of signals to pick out articles that contain misinformation, starting with the source of the content: "Knowing that a page or a website has shared false content in the past is a good predictor that it will happen again," he says. There may also be a discernible pattern in how false content propagates across the Web; Kleban says that's an active area of research. As for the text itself, the AI isn't equipped to evaluate statements for their truthfulness, but it can find signals, such as expressions of disbelief in the comment section.

The London-based startup Factmata, whose high-profile investors include Twitter cofounder Biz Stone and Craigslist founder Craig Newmark, is developing an AI system with a different approach. It specifically does not look at the content's publishers, says Factmata founder Dhruv Ghulati. "We want to judge content based on content itself," he says.

Factmata's system is also a hybrid of human and machine, though in a different configuration: The humans are experts who label content used for the AI's training. "Things like fake news and propaganda are inherently nuanced and subjective," Ghulati says. "It does require expertise to understand the nature of the content and tag it appropriately." With that proprietary data set, Factmata is training its AI to recognize politically biased content, false content, and hate speech.

Some companies that began with other journalistic purposes have joined the fray.

Krzana, a London-based company, helps journalists find breaking news with a customized real-time news feed. Reporters use Krzana's AI-enabled tool to discover content based on keywords and search terms they've selected. In recent elections in Mexico, a media coalition used Krzana to quickly find stories that might contain misinformation.

"Rather than waiting for these stories to be shared by a lot of people, journalists were among the first to read them," says Krzana cofounder Toby Abel. "If they were fake, they could be countered very quickly."

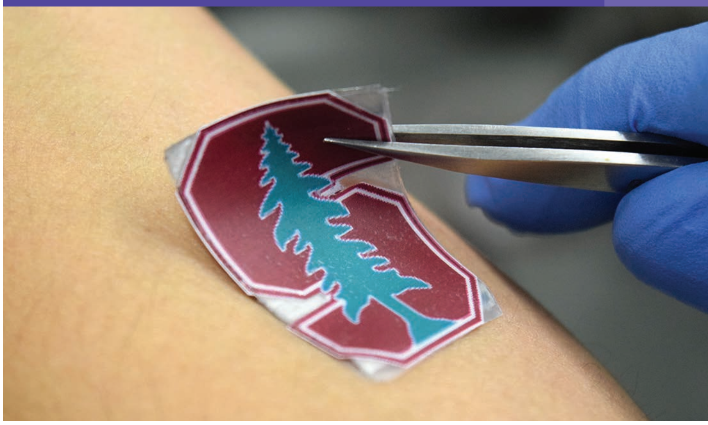
Abel says an AI misinformation detector can't yet be reliable on its own, and he agrees that there needs to be a "human in the loop." He cites an example from the 2018 Mexican election, in which a political candidate responded to accusations of Russian ties with a playful stunt: He went down to the docks and pronounced that he was waiting for his Russian submarine. "If someone had read that without outside context and understanding, it would sound like fake news—but it's not," Abel says.

Satire is one of the toughest problems for AI systems that try to identify false content. Companies are also grappling with misinformation in images, videos, graphs, and other nontextual content. The possibilities for deception seem endless: A photograph might be legitimate, for example, but its caption may be misleading.

Today's AI systems may not be ready to parse complicated claims independently or to make sophisticated decisions about truth, says Ghulati of Factmata, but that doesn't mean they shouldn't be deployed now. "The risk is that you try to get the perfect definition of fake news and never reach an answer," he says. "The important thing is to build something." —ELIZA STRICKLAND

POST YOUR COMMENTS at https://spectrum.ieee.org/misinformation0918

ISTOCKPHOTO



SIGNS OF STRESS: If you're sweating enough for your skin to glisten, a new sensor can measure your cortisol level within seconds.

ONUR PARLAK/STANFORD UNIVERSITY

WEARABLE SENSOR DETECTS STRESS IN SWEAT

Cortisol is key to tracking stress, but it's tough to measure in an instant

Stress—we all know it can be bad for us. It affects blood pressure, metabolism, immune response, and memory. Over time, it can contribute to the development of chronic diseases. So scientists and health professionals are putting a lot of effort into finding ways to measure it.

Wearable devices that detect stress typically track markers such as temperature, heart rate, and perspiration levels. But all of those markers can be affected by factors unrelated to stress.

A far better indicator that someone is under stress is a change in the amount of cortisol produced by the body. Cortisol, a steroid hormone, goes up when a person is under physical or emotional strain. The hormone can be measured by testing blood, saliva, or hair. But none of these options provide quick results, so they're not particularly useful for detecting short-term stressors.

That's why a wearable that can provide instant, accurate, and ongoing measurements of cortisol levels is something of a holy grail for researchers working on these types of sensors.

In July, a team at Stanford said that achievement is within reach. The group, led by materials science and engineering associate professor Alberto Salleo and postdoctoral research fellow Onur Parlak, announced in *Science Advances* that they've developed a wearable patch that can determine how much cortisol someone is producing in seconds, using sweat drawn from the skin under the patch.

The stretchy patch pulls sweat through perforations and into a reservoir. A membrane on top of the reservoir allows charged ions, such as sodium and potassium, to pass through. Cortisol, which has no charge, can't pass this barrier and instead blocks the charged ions. Signals sent from an electrical sensor in the patch can reveal these backups and indicate how much cortisol is in the sweat.

Parlak tested the prototype on several runners. The cortisol levels detected by the wearable sensor patch matched those obtained by putting samples of the runners' sweat through an ELISA (enzyme-linked immunosorbent assay) test, which takes several hours to produce results.

This announcement could be positive news on the sweat research front, but it won't represent a real breakthrough until the technology is proven through more rigorous on-body testing, points out Jason Heikenfeld, a professor at the University of Cincinnati and cofounder of Eccrine Systems, a company that is also developing sweat sensors.

"Sweat sensing is way tougher than conventional assays, where you can control pH or salinity," said Heikenfeld. "Both salinity and pH vary significantly in sweat—so much that any sensor that is sensitive to ions or pH will likely change signal more due to changes in ions or in pH than to cortisol." He says the Stanford team will need to run hours-long tests on many subjects to gauge the sensor's reliability.

Cortisol is a difficult target, agrees John Rogers, a professor at Northwestern University who focuses on wearable devices, "but it's also highly interesting for our group and others," he says. "Right now, the big challenges are thermal stability and temporal stability. The next big step [for the Stanford group] is showing robust performance in field studies over significant time periods and under varying thermal conditions."

The Stanford researchers are now working to miniaturize the device, evaluate the data, develop a user interface, and adapt it to be powered by energy harvesting rather than a battery. They suggest in their paper that the sensor could also be made to detect other noncharged biomarkers and hormones in sweat.
—TEKLA S. PERRY

A version of this article appears in our View From the Valley blog.

POST YOUR COMMENTS at https://spectrum.ieee.org/sweatsensor0918

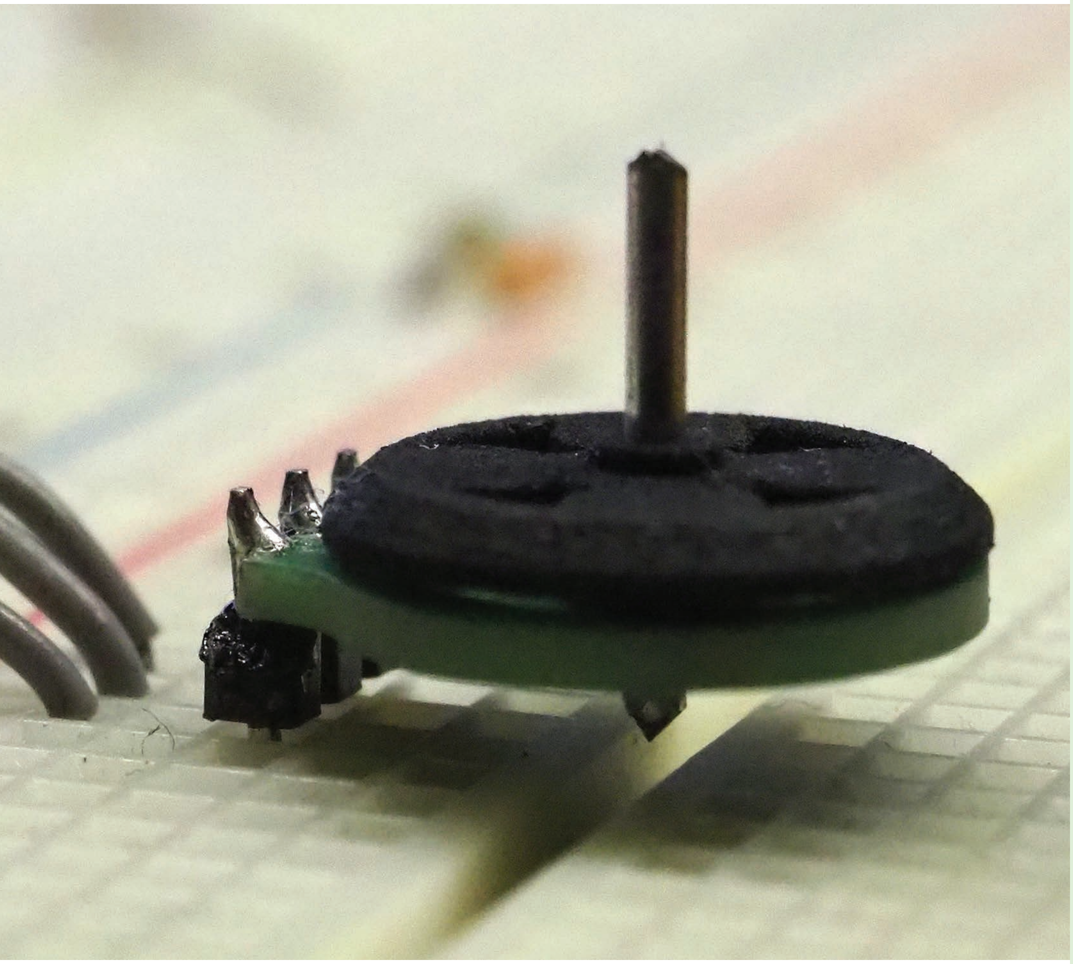
PHOTOGRAPH BY Erhan Elaldi/Anadolu Agency/Getty Images

TOUR DE SMART-PHONE

HOW MANY TIMES have you been on a business trip, deep in serious conversation on your cellphone, when you suddenly notice that the phone's battery is nearly drained? In just about any of the world's transit hubs, you'll find travelers scrambling to claim one of the precious few wall outlets for the electricity that someone else provides. But at least a few airports and railway stations have started changing the game, setting up public charging stations that require you to provide your own energy. At this bicycle-powered charging kiosk, installed in a railway station in Istanbul, a generator delivers charge to your electronic device as you pedal—turning your muscle power into the ability to tweet and send emails.

ONE NANOMETER

THE SIZE OF THE WORLD'S SMALLEST ELECTRIC MOTOR— JUST ONE MOLECULE WIDE— CREATED AT TUFTS UNIVERSITY IN 2011



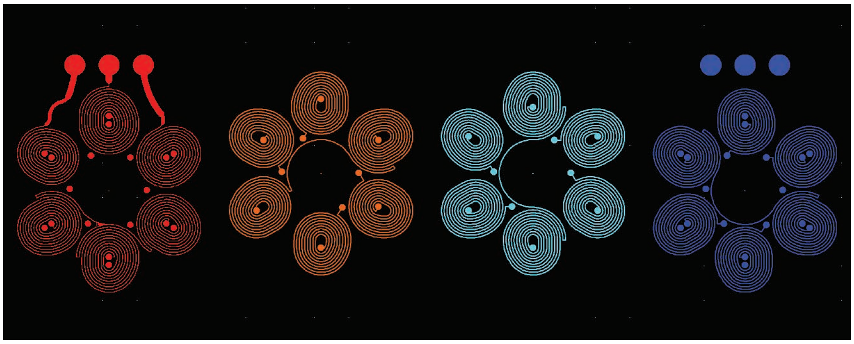
CARL BUGEJA (5)

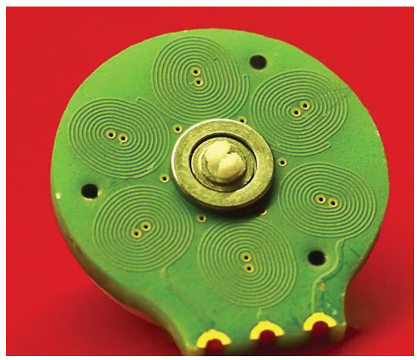
THE PRINTABLE MOTOR

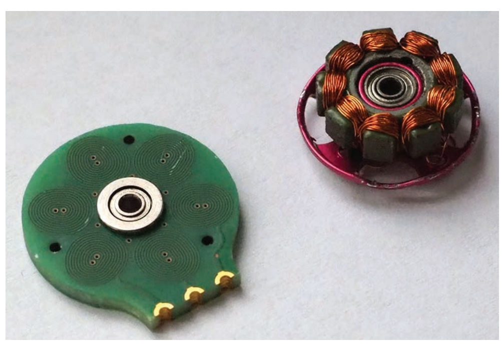
THE TRICK IS TO USE CIRCUIT BOARD TRACES FOR COILS

RESOURCES_HANDS ON

I started out by just wanting to make a very small drone. But I quickly realized that there was a limiting factor in just how small and light I could make any design: the motors. Even small motors are still discrete packages that have to be attached to all the other electronic and structural elements. So I began wondering if there was a way to merge these elements and save some mass. ● I drew inspiration from how some radio systems antennas made from the copper traces on a printed circuit board (PCB). Could I use something similar to create a magnetic field strong enough to drive a motor? I decided to see if I could build a motor of the axial flux type using electromagnetic coils fashioned from a PCB's traces. In an axial flux motor, the electromagnetic coils forming the motor's stator are mounted parallel to a disk-shaped rotor. Permanent magnets are embedded in the disk of the rotor. Driving the stator coils with alternating current causes the rotor to spin. ● The first challenge was making sure I could create enough magnetic flux to turn the rotor. It's simple enough to pattern a flat spiral coil trace and run current through it, but I limited my motor to a diameter of 16 millimeters, so that the overall motor diameter was comparable to that of the smallest off-the-shelf brushless motors. Sixteen millimeters meant I could fit only about 10 turns per spiral and 6 coils in total, arranged under the disk of the rotor. Ten turns just isn't enough to produce a sufficient magnetic field. But







TURN, TURN, TURN: I designed four sets of coils out of copper traces [top] and stacked them vertically on a four-layer printed circuit board [above right]. Pulsing these coils drives a 3D printer rotor with embedded permanent magnets [above]. Although not as strong as a traditional brushless motor, the PCB is cheaper and lighter [right].

the nice thing about PCBs is that it's pretty easy today to make one with multiple layers. By printing stacks of coils, with coils on each of four layers, I was able to get 40 turns per coil, enough to turn a rotor.

A bigger problem emerged as the design progressed. In order to keep a motor spinning, the dynamically changing magnetic field between the rotor and stator must be synchronized. In a typical motor that's driven by alternating current, this synchronization arises naturally due to the arrangement of the brushes that electrically bridge the stator and rotor. In a brushless motor, control electronics implementing a feedback system are required.

In a previous brushless motor driver that I'd built, I measured the back electromotive force as feedback to control the speed. Back EMF is produced because a spinning motor acts like a little generator, inducing a voltage in the stator coils that opposes the voltage used to drive the motor. Sensing the back EMF gives feedback about how the rotor is spinning, and lets the control electronics synchronize the coils. But in my PCB motor, the back EMF was too weak to use. So instead I mounted a Hall-effect sensor, which can measure the change in a magnetic field directly, to gauge how rapidly the rotor and its permanent magnets were spinning above the sensor. This information was then fed into the motor control electronics.

To make the rotor itself, I turned to 3D printing. Initially, I made a rotor that I fitted onto a separate metal shaft, but then I simply began printing the snap-fit shaft as an integral part of the rotor. This reduced the physical components to just the rotor, four permanent magnets, a bearing, and the PCB that provides both the coils and structural support.

I soon had my first motor up and spinning. Testing showed it could deliver a static torque of 0.9 gram-centimeters. This wasn't enough torque to meet my original goal of building an integrated drone motor, but I realized that the motor could still be used for propelling small and inexpensive robots along the ground on wheels, so I persisted (motors are typically among the most expensive parts of robots). The printed motor can operate with voltages from 3.5 to 7 volts, although it does heat up noticeably at higher voltages. At 5 V, its operating temperature is 70 °C, which is still manageable. It draws about 250 milliamperes.

Currently, I've been focusing my efforts on increasing the torque of the motor (you can follow my ongoing efforts on Hackaday). I've been able to almost double it by adding a ferrite sheet to the back side of the stator coils to contain the coils' magnetic field lines. I'm also looking into designing other prototypes with different winding configurations and more stator coils. In addition, I've been working on using the same techniques to build a PCB linear actuator that can drive a 3D-printed slider down a row of 12 coils. And I'm testing a flexible PCB prototype that uses the same printed coils to perform electromagnetic actuation. My goal is—even if I can't take to the sky yet—to start making new robots with smaller and simpler mechanisms than is currently possible. **—CARL BUGEJA**

↗ POST YOUR COMMENTS at https://spectrum.ieee.org/pcb0918

MUSIC FOR ALIENS
JONATHON KEATS CREATES INSTRUMENTS FOR NONHUMAN SENSES

Sure, humans have tried before to communicate with any extraterrestrials who might be out there. Most notably, the twin Voyager probes each carry a copy of the Golden Record, intended to communicate "a story of our world" with images, nature sounds, music, and spoken greetings. But experimental philosopher Jonathon Keats says that if we *really* want to persuade aliens to make contact, we should do more than just tell them our story.

"If you're in a bar and hear someone who just keeps talking about themselves, it gets annoying," he says. "I'm trying to make something that's more universal and more inclusive."

So Keats has founded Intergalactic Omniphonics—which he calls a startup and others might call conceptual art—to bring sentient beings together through music. He has created several new instruments that he hopes will be accessible to aliens, even if their senses are considerably different from ours.

The simplest of the instruments, which Keats created during a residency at the University of North Carolina Asheville's STEAM Studio in collaboration with the sculptor Amelia Pate, is the dog-whistle organ. Recognizing that the human hearing range is limited—spanning the audio frequencies of about 20 to 20,000 hertz—Keats and Pate built an organ with a foot pump that sends a flow of air through various ultrasonic whistles.

But what if the aliens aren't sensitive to the changes in air pressure that we call sound? At its core, "music is the modulation of frequency and amplitude over time," Keats says. "So anything that allows for that can carry a tune." He turned to the electromagnetic spectrum, but he didn't want to limit himself to the portion that's visible to the human eye. He decided to work with gamma rays, which are emitted by radioactive materials. Keats scrounged his samples from eBay.

"I got radium in an old watch dial and uranium in an old marble," he says. To make the gamma rays ring out gaily, a player rhythmically lifts lead casings that cover the materials.

Another elegant instrument is a "cello" that sends out gravitational waves—something that Keats says has huge potential for interstellar communication because the waves move through the universe at the speed of light. Thus, an instrument that controls the frequency and amplitude of faint gravitational waves might send a musical message: The cellist plays the instrument by swinging ball bearings of different masses.

As for what to play with these instruments (beyond obvious songs like David Bowie's "Starman"), Keats has created an anthem that he hopes will be "cognitively meaningful" to any being in the universe. "I'm sure I've failed," he says ruefully. "I'm sure there are some beings that are feeling left out. But I tried my best."

The anthem takes as its theme the second law of thermodynamics, which holds that closed systems inevitably become more disordered over time, increasing their entropy. Against this backdrop, we living beings extract useful energy from our environments, to make ourselves less disordered. But eventually we die and decay, and our energy goes back into the mix.

"The anthem communicates what it is to be alive, since that's what we have in common with every organism on this planet and any organism elsewhere in the universe," Keats says. To represent this concept musically, the anthem's score calls for a soloist, representing beings everywhere, to make music that goes from a state of greater to lesser entropy, as the larger orchestra tends toward greater entropy in the background. Ultimately, the soloist's voice becomes more entropic to represent death, and the orchestra becomes marginally less entropic as the being's energy is added back in. Then it repeats. That's the cycle of life.

Keats would love to send the anthem beaming out into space, perhaps via the powerful Arecibo radio telescope, in Puerto Rico. Until then, he says, "I'm hoping we can get it played at some ballparks." **—ELIZA STRICKLAND**

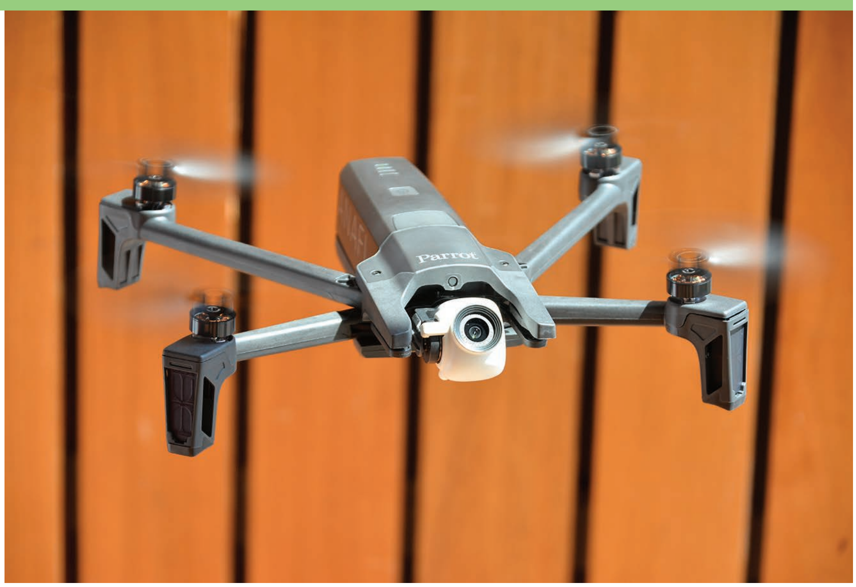
An extended version of this article appears in our Tech Talk blog.

POST YOUR COMMENTS at https://spectrum.ieee.org/music0918

UNIVERSITY OF NORTH CAROLINA ASHEVILLE

PARROT'S NEW DRONE RECLAIMS A NICHE

THE ANAFI MARKS THE COMPANY'S RETURN TO THE CONSUMER SPACE



EVAN ACKERMAN

Parrot was one of the first (if not the absolute first) companies to take a crack at the consumer drone space. The AR Drone came out in 2010, and Parrot followed it up with a solid upgrade in the AR Drone 2.0 a few years later. Since then, we've seen Parrot's Bebop and some clever flying toys and had a bunch of fun with its fixed-wing Disco. But at this point, most consumers probably think of DJI when they think of camera drones because of how pervasive Phantoms and Mavics are. Two years ago, Parrot saw how the market was trending and started working on a new aerial camera drone designed to be exceptionally easy to use and transport.

The result is the Anafi, a US $700 drone with the sort of thoughtful usability and clever design that we've come to expect from Parrot. I got to try one out for a few weeks this summer. [Disclosure: Parrot covered my expenses to attend the Anafi launch event in New York City.]

Anafi's biggest selling point is its camera system. The 21-megapixel sensor shoots 4K video with room to spare. At standard HD, you can shoot 60 frames per second, and it'll also shoot 12-megapixel rectilinear photos that are distortion-free.

Anafi's camera features a 2.8x zoom in HD and 1.4x in 4K. The zoom is digital, not optical, but it doesn't introduce blockiness or blurriness into images. This is possible because the sensor is large enough that the whole thing isn't being used all at once when you're recording video. When "zoomed out," the video is actually downsampled from across the entire sensor. "Zooming in" simply reduces that downsampling over a smaller area until you hit a pixel ratio of 1:1, so the quality stays the same.

Unusually, the Anafi's camera can rotate vertically, from straight down to straight up. This is really a novelty: Drones are generally best at flying above things and looking down. Parrot demonstrates this feature by following someone on a ropes course. Yeah, it's great for that, but how often are you shooting someone on a ropes course?

The Anafi's upward-looking ability is a by-product of the fact that its camera's gimbal has only two axes of physical stabilization: pitch and roll, but not yaw. Yaw movement is stabilized digitally. Most other camera drones use mechanical stabilization in all three axes, which requires another joint in the gimbal that constrains the camera from looking up.

It seems like three-axis mechanical stabilization should be way better than two-axis, but Parrot's testing showed that most of the uncommanded motion of the drone (like fighting gusts of wind) happens through some combination of pitch and roll. Drones simply do not need nearly as much stabilization in the yaw axis, so Parrot has concluded that doing it digitally is totally fine, and it saves the cost of the third axis. (The one downside I did notice is that it's a bit more difficult to do silky-smooth panning shots.)

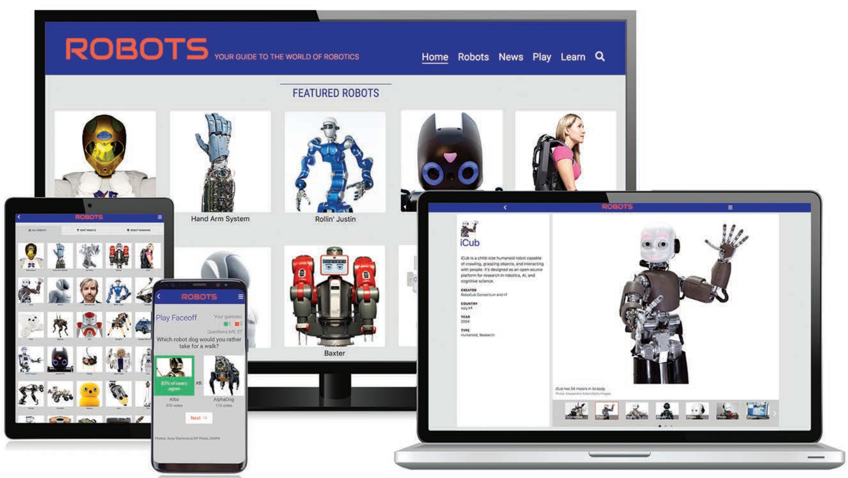
Anafi is as easy to fly as you'd expect from Parrot. Plug phone into controller, turn on drone, push takeoff button, and you're airborne. Anafi has no trouble in light winds; according to the specs, you can fly it in winds of up to 50 kilometers per hour (31 miles per hour). There's no automatic obstacle avoidance, but I didn't find that a big deal.

My flights averaged between 16 and 18 minutes, bringing the battery down to 25 percent—this is very much in line with Parrot's estimate of 25 minutes maximum flight time (at which point the drone initiates its emergency landing mode). And the footage I've been able to take with this machine? Gorgeous. Rich, smooth, full of detail, with well-balanced color. **—EVAN ACKERMAN**

An extended version of this review appears in our Automaton blog.

POST YOUR COMMENTS at https://spectrum.ieee.org/parrotdrone0918

CALLING ALL ROBOTS
WE ARE BUILDING THE WORLD'S LARGEST, COOLEST ROBOTICS CATALOG



Do you like robots? Of course you do! You're reading *IEEE Spectrum*, so you almost certainly love robots. Robots capture our imagination. Robots are the future. Now let us ask you this: What's your favorite robot?

We bet some of you said R2-D2. Or maybe Rosie. Or Robby. Or Johnny 5. Or Data.

These are all cool robots. We like them too! But here's the problem: Those are not *real* robots. We see this happen often, especially with kids. When you ask them which robots they find inspiring, the answers typically come from science fiction.

And why is that a problem? Because we've reached the point where it's clear that robots are going to affect every aspect of our lives. It won't happen overnight, but most of us will likely see the day when robotics will be everywhere—in our homes, offices, schools, factories, hospitals, streets, and even skies.

And that's why it's important to focus on real-world robots, not just sci-fi ones. We want more engineers to pursue careers in robotics. We want more kids to dream of becoming roboticists and technologists, or at least be sufficiently familiar with the details of the technology to make informed, thoughtful, and ethical decisions in the future. So we need to make real robots just as inspiring as their fictional counterparts, and here at *Spectrum,* we have a plan to do just that.

Over the past year, we've been creating a massive portal for everything robotics, built around a fun and unique dynamic catalog. You can see it right now at Robots.ieee.org. There you'll find a vast zoo of humanoids, drones, exoskeletons, quadrupeds, and other kinds of automatons, each with its own profile, with photos, videos, curious facts, and technical specifications.

Many profiles also have special interactives that you won't find anywhere else: You can spin robots 360 degrees or make them move. Take joy in making the robot baby iCub crawl across the screen, wiggling the fingers of NASA's space humanoid Robonaut, or swapping the facial expressions of lifelike android Geminoid DK.

We also want to know how you feel about all these different robots. You can rate them based on their capabilities and appearance, and then you can see how each robot ranks against the others. Based on users' votes, we've created rankings to see which robots are the Top Rated, Creepiest, and Most Wanted.

The site also has a robotics news section, a game called Faceoff, and an educational section that will feature lesson plans and other STEM (science, technology, engineering, math) materials for schools interested in learning about real robots from industry, research, and startups.

If our robot guide happens to look familiar, that's because it's an expansion of *Spectrum's* popular Robots for iPad app. If you know the app, thank you for being a user—an update is coming soon for iOS and Android. In the meantime, we invite you to check out Robots.ieee.org on your desktop, tablet, or phone for the latest content.

Our collection currently has 157 robots, and we're going to be adding more soon. Our goal is to have robots of all types and sizes and from as many countries as possible.

So we ask again: What is your favorite robot? Go to the site and check out the robots we already have. Send us your suggestions of new robots to add by emailing hellorobots@ieee.org. Get them in by 15 October, and you may win an exclusive robot T-shirt. Go robots!

—ERICO GUIZZO & RANDI KLETT

↗ VISIT OUR ROBOTICS SITE at https://robots.ieee.org

TOP: IEEE SPECTRUM; DEVICES: ISTOCKPHOTO; BOTTOM: CARLTON SOOHOO

FIXING SPECTRUM AUCTIONS

THIS FALL, THE U.S. FEDERAL Communications Commission has a chance to set a precedent for how one of the most precious of resources in modern society–radio spectrum–is valued. That precedent could in turn help jump-start the next era of wireless innovation. ● As part of its rulemaking around the Citizens Broadband Radio Service (CBRS), the FCC will make available up to 150 megahertz of spectrum for wireless communications. Historically, the agency and spectrum buyers have arrived at a price for spectrum based on how many people it could cover. Thus, prices are measured relative to "MHz pops" (megahertz passing one person). So if you have 3 MHz capable of reaching 2 million people, you designate that as 6 million MHz pops. ● With the CBRS band, this equation should change. The FCC needs to start thinking beyond cellphones and people: We are now building a world of connected devices, where sensors, appliances, drones, and tractors need cellular connections too. ● CBRS is a vast swath of spectrum in the 3,550-to-3,700-MHz range currently used by the U.S. Navy and some satellite providers. It's relatively lightly used, so the FCC proposed in 2014 that this band be opened, for example, to companies seeking to enhance and deploy LTE networks in a way that's similar to Wi-Fi. ● So far, the band is attracting a lot of interest beyond the traditional cellular carriers and cable companies. For example, the National Football League and GE Digital, the company's IoT arm, have asked the FCC for permission to build trial versions of these new types of Wi-Fi-like networks. The biggest sticking point is how large, geographically, the license grants should be.

These Priority Access Licenses, or PALs, dictate the scope of coverage of a particular chunk of spectrum. In the past, these grants were parceled out based on metropolitan areas–the more populous the metro area, the more expensive the license. But for the CBRS PALs, groups such as the NFL and building owners would like something smaller. For example, the NFL might like a piece of spectrum to cover just a stadium.

That's why the FCC should focus on devices or uses rather than number of people covered. It would be a more complex valuation equation, but if done right, it would more accurately reflect the ways wireless connectivity can add value beyond ordinary telephony. The NFL might offer fans access to a wireless sensor network that tracks players' statistics live on the field, or, taking stock of inventory, sends them to the nearest place to get a hot dog.

The FCC is now accepting comments on a variety of fundamental rules associated with apportioning the PALs. These include the geographic size. Dave Wright, the president of the CBRS Alliance, says that the renewability rules could be a good place to tackle the people-over-devices bias.

In particular, he'd like to see the FCC institute a new auction scheme, with two different kinds of PAL. "One might be based on population coverage, while [another] one, based on IoT, might look at the number of connected clients served per access point," he says. This way, a rural area without many people, but with a huge connected farming operation built on hundreds of thousands of networked sensors and equipment, could be weighed equally when it came time to determine how much use the spectrum has.

Such a scheme would more effectively capture the value of radio spectrum. It could be the start of a new era in terms of how the FCC thinks about spectrum use, and how it fits into our lives. That could lead to better outcomes, whether you're a football fan or not. ■

↗ **POST YOUR COMMENTS** at https://spectrum.ieee.org/spectrumauctions0918

ILLUSTRATION BY J.D. King

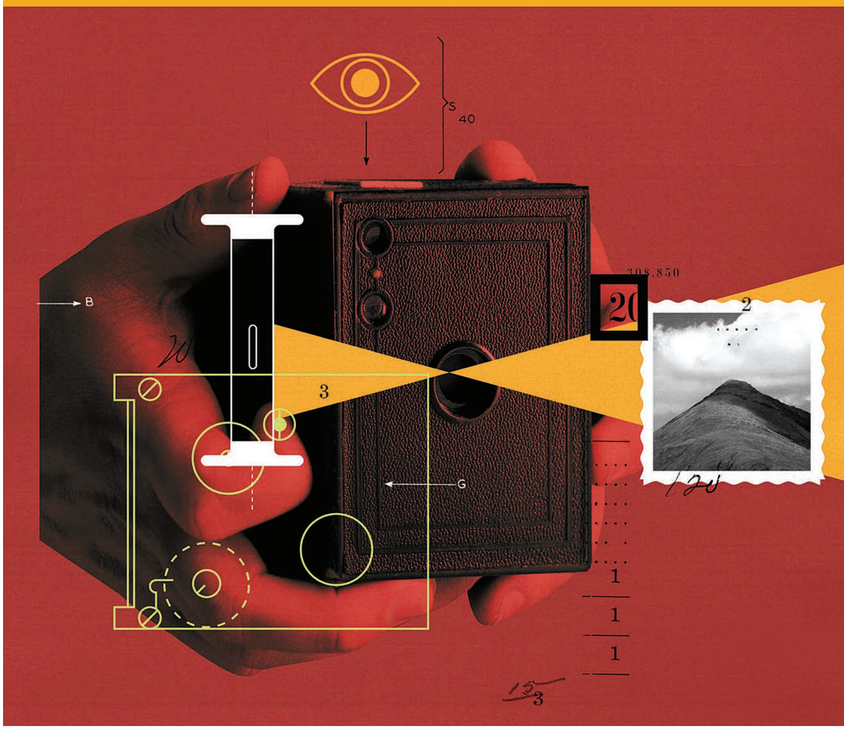


PHOTO-ILLUSTRATION BY Stuart Bradford

SEPTEMBER 1888: KODAK CAMERA IS PATENTED

BY 1888, PHOTOGRAPHERS HAD BEEN fixing images on plates for more than six decades, and some of them produced impressive portraits, photojournalism, and landscapes. But no one could do such things easily. • The first prerequisite for effortless picture taking came in 1871, when Richard Maddox invented highly sensitive dry plates—glass coated with gelatin emulsion. That step eliminated the awkward coating-exposure-processing sequence that had to be done on-site when using the wet-plate process. • But an entire suite of improvements still needed to be made, and they came from an unlikely innovator: George Eastman, a bank clerk in Rochester, N.Y. In 1877, Eastman bought a camera and wet-plate gear to use on a trip to Santo Domingo. The trip fell through, but it prompted Eastman to experiment with new emulsion coatings, and in 1879 he patented a coating machine. In 1884 he replaced the glass support with a negative stripping film made of three layers—paper, soluble gelatin, and gelatin emulsion. In 1885 he added a convenient roll holder. • The final step came 130 years ago this month, on 4 September 1888, when Eastman was awarded U.S. Patent No. 388,850 for a small, handheld, easy-to-use camera. His company had already begun making it three months earlier. • Eastman called the camera a Kodak because he liked the ring of it. It was a wooden rectangular prism 9.5 by 8.3 by 16.5 centimeters (3.75 by 3.25 by 6.5 inches) covered in smooth black leather. Its 57-millimeter lens had good close-focusing capabilities, allowing the photographer to focus on objects as close as 1.2 meters. • It wasn't really as easy to use as its advertisements claimed—"You press the button, we do the rest"—but it was easy enough. You put the film in and advanced it, and after exposing all 50 or 100 frames (there was no exposure counter), you rewound it. Then, you had to ship the entire camera to Eastman's factory in Rochester to be developed and printed. The first model cost US $25 (more than $600 in 2018 terms).

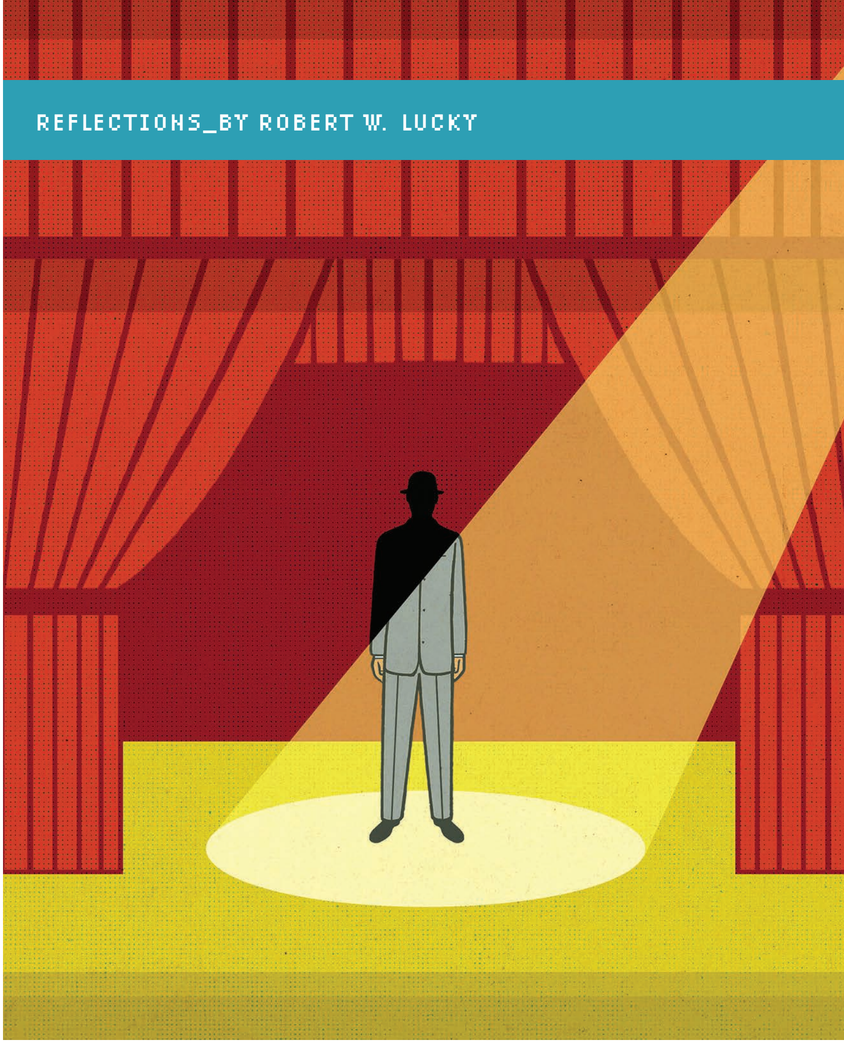
Before the century's end, the product line included, for the first time, film for a non-Kodak camera: the miniature (5 by 4 by 4 cm) Kombi pocket camera, with 25 exposures and a price of about $3.00. Also added were the Bull's Eye camera, which allowed users to remove the exposed film in daylight and develop it at home, selling for $7.50 or so, and (in 1897) the Folding Pocket Kodak camera, the prototype of the roll-film designs that dominated the market over the next six decades.

Kodak kept its technical leadership by introducing the first safety film (made of cellulose acetate, rather than of highly flammable cellulose nitrate) in 1908. Amateur movie cameras followed in 1923, the first 35-mm precision Kodak Retina cameras in 1934, the first slide projector in 1937, Kodacolor film in 1942, and the Kodak Instamatic in 1963. That line of cameras sold more than 50 million units by 1970 making it the most successful tool of mass-scale amateur photography. I owned one.

Impressively, it was Kodak—the master of film—that in 1975 invented the world's first digital camera, a toaster-size contraption with a resolution of 10,000 pixels. Yet although it kept a competitive position in this new filmless business for a while, it eventually lost the contest.

In 2012, Kodak entered Chapter 11 bankruptcy and sold all of its consumer and digital imaging patents for a mere pittance. The reorganized company, focusing on diverse imaging solutions (including instant-print cameras and 3D printing), is still around, but its stock has recently traded about 75 percent below its 2014 postbankruptcy peak. A brand that embodied generations of innovative American technical leadership has become yet another entry in the country's manufacturing retreat. ■

↗ **POST YOUR COMMENTS** at https://spectrum.ieee.org/kodak0918



THE AGE OF THE HEROIC INVENTOR IS OVER

EVERYONE KNOWS THE GREAT ELECTRICAL inventors of the past, such as Alexander Graham Bell of the telephone, Marconi of the radio, and Edison of the incandescent light. But few know who invented the transistor, laser, integrated circuit, optical fiber, stored-program computer, GPS, or Internet. Yet these are probably the most significant inventions of the last century, and I wonder: Why this disparity? • Perhaps it is because the world of the past was a simpler place and there was magic in the air. I remember, when I was a child, my mother telling me about the thrill she felt in first hearing a voice emanate from a radio, first speaking on a telephone, and the first time she saw an airplane fly over her farm in Virginia. They were seen as miracles, and accepted as such. In a relatively empty, largely agrarian world, these magical inventions had enormous impact, and even before they were widely adopted, it was understood almost immediately what they did and, to some degree, how they would be used. • In contrast, the world today is a busy, connected place. Magic has been replaced by science and technology. Inventions are numerous, seemingly incremental, and mostly taken for granted. My grandchildren are not going to ask me what it was like to use the first transistor. It's a meaningless question. Even engineers initially viewed the transistor as a simple replacement for the vacuum tube, while the laser was generally believed to be a glorified flashlight. No one remembers the first time they used the Internet or GPS; they just evolved. There was no initial flash of discovery.

And unlike Bell, Marconi, and Edison, none of the inventors of these new technologies started out an amateur. They were all highly trained engineers and scientists, often working in teams supported by government research grants or corporate research labs and churned out all over the world in universities running degree programs that didn't even exist in the early 20th century.

Moreover, today's inventions are rarely credited to single individuals. All of the recent inventions that I cited above involved multiple creators. Three people were given the Nobel Prize for the transistor (Shockley, Bardeen, and Brattain) and three for the laser (Townes, Basov, and Prokhorov). Four were given the Draper Prize for the creation of the Internet (Cerf, Kahn, Kleinrock, and Roberts). Of course, even in the past, the provenance of inventions was a tangled affair, but it was a question of which of several competing individual inventors should receive credit.

While our current inventors have received much well-deserved recognition and awards within their industries, they have not become celebrities. I suspect that even many engineers would not know their names. On the other hand, we all know the names of Gates, Jobs, Zuckerberg, Musk, and Bezos. Their accomplishments are not in the mold of yesterday's famous inventors, but their social and commercial innovations are the hallmark inventions we now celebrate.

On reflection, I think that this lack of celebrity for our esteemed electrical pioneers of today may not be a bad thing. Instead of concentrating all the acclaim on a few individuals, we engineers and scientists can all feel a shared pride in the accomplishments of our professions. We have, indeed, changed the world. ■

↗ **POST YOUR COMMENTS** at https://spectrum.ieee.org/reflections0918

ILLUSTRATION BY Dan Page

THE QUEST TO SAVE GENOMICS

Unless researchers solve the looming data compression problem, biomedical science could stagnate

By DMITRI PAVLICHIN & TSACHY WEISSMAN
Illustration by GREG MABLY

HAVE YOU HAD YOUR GENOME SEQUENCED YET? Millions of people around the world already have, and by 2025 that number could reach a billion.

The more genomics data that researchers acquire, the better the prospects for personal and public health. Already, prenatal DNA tests screen for developmental abnormalities. Soon, patients will have their blood sequenced to spot any nonhuman DNA that might signal an infectious disease. In the future, someone dealing with cancer will be able to track the progression of the disease by having the DNA and RNA of single cells from multiple tissues sequenced daily.

And DNA sequencing of entire populations will give us a more complete picture of society-wide health. That's the ambition of the United Kingdom's Biobank, which aims to sequence the genomes of 500,000 volunteers and follow them for decades. Already, population-wide genome studies are routinely used to identify mutations that correlate with specific diseases. And regular sequencing of organisms in the air, soil, and water will help track epidemics, food pathogens, toxins, and much more.

This vision will require an almost unimaginable amount of data to be stored and analyzed. Typically, a DNA sequencing machine that's processing the entire genome of a human will generate tens to hundreds of gigabytes of data. When stored, the cumulative data of millions of genomes will occupy dozens of exabytes.

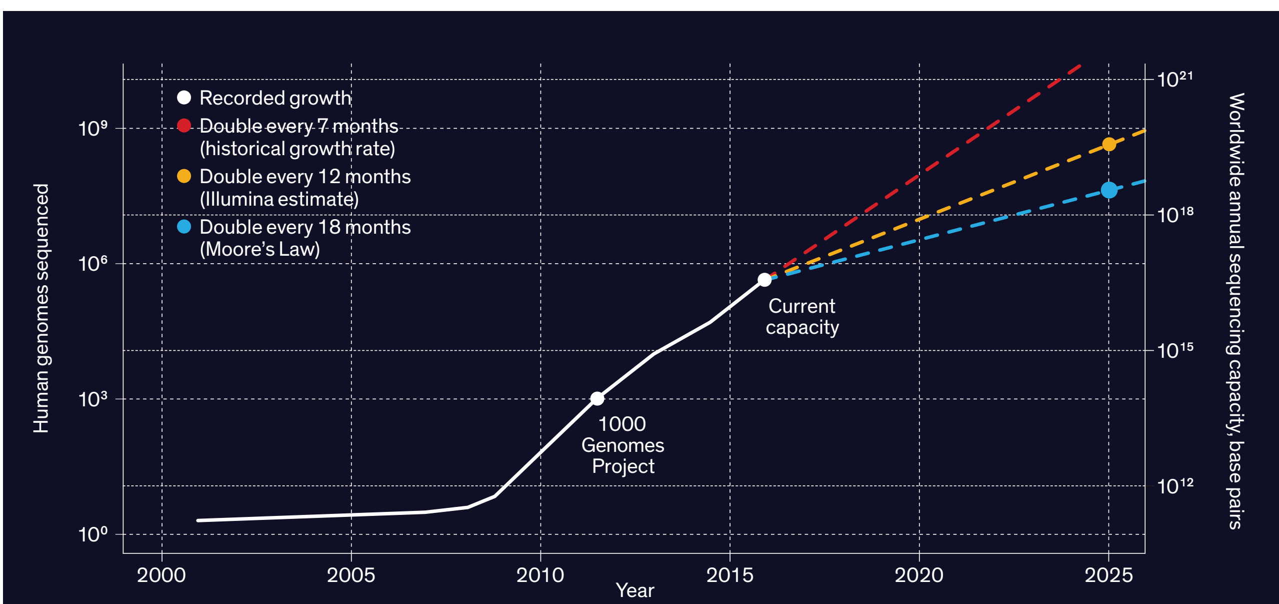
And that's just the beginning. Scientists, physicians, and others who find genomic data useful aren't going to stop at sequencing each individual just once—in the same individual, they'll want to sequence multiple cells in multiple tissues repeatedly over time. They'll also want to sequence the DNA of other animals, plants, microorganisms, and entire ecosystems as the speed of sequencing increases and its cost falls—it's just US $1,000 per human genome now and rapidly dropping. And the emergence of new applications—and even new industries—will compel even more sequencing.

While it's hard to anticipate all the future benefits of genomic data, we can already see one unavoidable challenge: the nearly inconceivable amount of digital storage involved. At present the cost of storing genomic data is still just a small part of a lab's overall budget. But that cost is growing dramatically, far outpacing the decline in the price of storage hardware. Within the next five years, the cost of storing the genomes of billions of humans, animals, plants, and microorganisms will easily hit billions of dollars per year. And this data will need to be retained for decades, if not longer.

Compressing the data obviously helps. Bioinformatics experts already use standard compression tools like gzip to shrink the size of a file by up to a factor of 20. Some researchers also use more specialized compression tools that are optimized for genomic data, but none of these tools have seen wide adoption. The two of us do research on data compression algorithms, and we think it's time to come up with a new compression scheme—one that's vastly more efficient, faster, and better tailored to work with the unique characteristics of genomic data. Just as special-purpose video and audio compression is essential to streaming services like YouTube and Netflix, so will targeted genomic data compression be necessary to reap the benefits of the genomic data explosion.

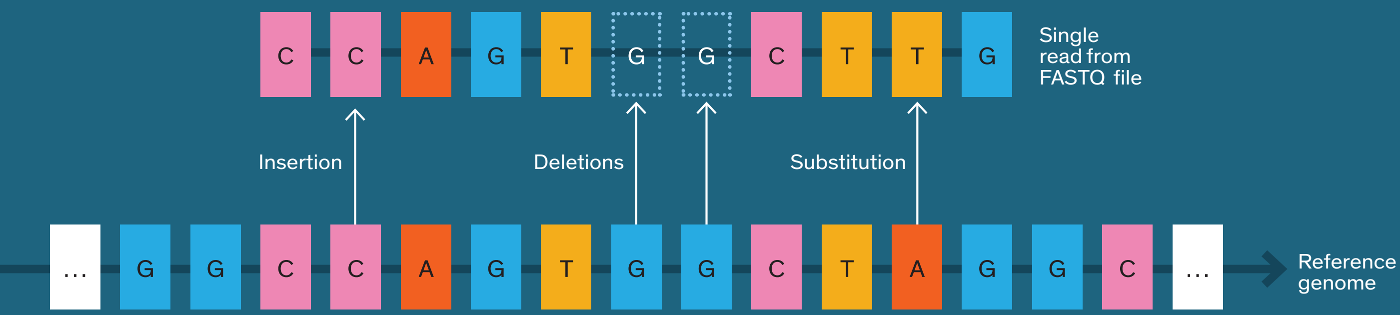
BEFORE WE EXPLAIN how genomic data could be better compressed, let's take a closer look at the data itself. "Genome" here refers to the sequence of four base nucleotides—adenine, cytosine, guanine, and thymine—that compose the familiar A, C, G, T alphabet of DNA. These nucleotides occur in the chains of A-T and C-G pairs that make up the 23 pairs of chromosomes in a human genome. These chromosomes encompass some 6 billion nucleotides in most human cells and include coding genes, noncoding elements (such as the telomeres at the ends of chromosomes), regulatory elements, and mitochondrial DNA. DNA sequencing machines like those from Illumina, Oxford Nanopore Technologies, and Pacific Biosciences are able to automatically sequence a human genome from a DNA sample in hours.

These commercial DNA sequencers don't produce a single genome-long string of ACGTs but rather a large collection of substrings, or "reads." The reads partially overlap each other, requiring sequence-assembly software to reconstruct the full genome from them. Typically, when whole-genome sequencing is performed, each piece of the genome appears in no more than about 100 reads.



GROWTH OF HUMAN GENOME SEQUENCING: Since the first publication of a draft human genome sequence in 2001, there's been a dramatic increase in the pace of growth of both the number of genomes sequenced and the sequencing capacity. The numbers after 2015 represent three possible projected growth curves.

SOURCE: STEPHENS ZD, LEE SY, FAGHRI F, CAMPBELL RH, ZHAI C, EFRON MJ, ET AL., 2015, PLOS BIOL 13(7)



READ AND REFERENCE: A DNA "read" [top string] approximately matches a portion of the reference human genome [bottom]. Insertions, deletions, and substitutions (due either to mutations or to noise in the DNA sequencing process) result in an imperfect match. To encode a read, we can state its starting position in the reference genome and describe any variations.

Depending on the sequencing technology used, a read can vary in length from about 100 to 100,000 base pairs, and the total number of reads varies from millions to tens of billions. Short reads can turn up single base-pair mutations, while longer reads are better for detecting complicated variations like deletions or insertions of thousands of base pairs.

DNA sequencing is a noisy process, and it's common for reads to contain errors. And so, besides the string of ACGT nucleotides, each read includes a quality score indicating the sequencing machine's confidence in each DNA nucleotide. Sequencers express their quality scores as logarithms of error probabilities. The algorithms they use to do so are proprietary but can be checked after the fact. If a quality score is 20—corresponding to an error probability of 1 percent—a user can confirm that about 1 percent of the base pairs were incorrect in a known DNA sequence. Programs that use these files rely on quality scores to distinguish a sequencing error from, say, a mutation. A true mutation would show a higher average quality score—that is, a lower probability of error—than a sequencing error would.

The sequencer pastes together the strings and the quality scores, along with some other metadata, read by read, to form what is called a FASTQ file. A FASTQ file for an entire genome typically contains dozens to hundreds of gigabytes.

The files are also very redundant, which stems from the fact that any two human genomes are nearly identical. On average, they differ in about one nucleotide per 1,000, and it's typically these genetic differences that are of interest. Some DNA sequencing targets specific areas of difference—for example, DNA-genotyping applications like 23andMe look only for specific variations, while DNA profiling in criminal investigations looks for variations in the number of repetitions of certain markers.

But you need to sequence the whole genome if you don't know where the interesting stuff lies—when you're trying to diagnose a disease of unknown genetic origin, say—and that means acquiring much larger quantities of sequencing data.

The repetition in sequencing data also comes from reading the same portions of the genome multiple times to weed out errors. Sometimes a single sample contains multiple variations of a sequence, so you'll want to sequence it repeatedly to catch those variations. Let's say you're trying to detect a few cancer cells in a tissue sample or traces of fetal DNA in a pregnant woman's blood. That may mean sequencing each DNA base pair many times, often more than 100, to distinguish the rare variations from the more common ones and also the real differences from the sequencing errors.

BY NOW, YOU SHOULD HAVE a better appreciation of why DNA sequencing generates so much redundant data. This redundancy, it turns out, is ideal for data compression. Rather than storing multiple copies of the same chunk of genomic data, you can store just one copy.

To compress genomic data, you could first divide each DNA sequence read into smaller chunks, and then assign each chunk a numerical index. Eventually, the sum total of indexes constitutes a dictionary, in which each entry isn't a word but a short sequence of DNA base pairs.

Text compressors work this way. For example, GitHub hosts a widely used list of words that people can use to assign each word its own numerical index. So to encode a passage of text into binary, you'd replace each word with its numerical index—the list on GitHub assigns the number 64,872 to the word *compression*—which you'd then render in binary format. To compress the binary representation, you could sort the dictionary by word usage frequency instead of alphabetical order, so that more common words get smaller numbers and therefore take fewer bits to encode.

Another common strategy— the Lempel-Ziv family of algorithms—builds up a dictionary of progressively longer phrases rather than single words. For example, if your text often contains the word *genomic* followed by *data*, a single numerical index would be assigned to the phrase *genomic data*.

Many general-purpose compression tools such as gzip, bzip2, Facebook's Zstandard, and Google's Brotli use both of these approaches. But while these tools are good for compressing generic text, special-purpose compressors built to exploit patterns in certain kinds of data can dramatically outperform them.

Consider the case of streaming video. A single frame of a video and the direction of its motion enable video compression software to predict the next frame, so the compressed file won't include the data for every pixel of every frame. Moreover, the viewer can tolerate some barely perceptible loss of video information or distortion, which isn't the case with text-based data. To take advantage of that fact, an international consortium spent years developing the H.264 video compression standard (now used by Blu-ray Disc, YouTube, the iTunes store, Adobe Flash Player, and Microsoft's Silverlight, among many others).

Researchers have likewise been devising special-purpose tools for compressing genomic data, with new ones popping up in the academic literature about once a month. Many use what's called reference-based compression, which starts with one human genome sequence as its reference. Any short human DNA sequence—that is, one made up of 100 base pairs or less—is likely to appear somewhere in that reference, albeit with sequencing errors and mutations. So instead of listing all the base pairs in a string of 100, a specialized compressor notes only where the string starts within the reference (for example, "1,000th base pair in chromosome 5") and describes any deviations from the reference sequence (for example, "delete the 10th base pair"). The reference-based approach requires the user to have a copy of the reference human genome, about 1 gigabyte in size, in addition to the compressor software.

As mentioned, FASTQ files contain not just DNA sequences but also quality scores indicating potential errors. Unfortunately, reference-based compression can't be used to compress FASTQ quality scores because there is no reference sequence for quality scores. Instead, these tools look at patterns in the quality scores—that a low-quality score is likely to be followed by another low-quality score, for example, or that quality scores tend to be higher at the beginning of a DNA read than toward the end. Just as numbering all words in order of decreasing usage frequency lets us compress text, numbering the set of possible quality scores in the order of their predicted likelihood lets us compress this data. Instead of storing and compressing low-quality data, researchers sometimes discard it, but the data compression program might not be able to decide exactly which data to discard or what the threshold for "low quality" is.

THESE NEW COMPRESSORS are a good start, but they're far from perfect. As our understanding of the data evolves, so will our ability to compress the data. Data compression forces us to look for nonobvious patterns and redundancies in the data; when we reach the point of compressing the data deeply, we'll know that we finally understand it. A genomic data compressor that factors in subtle patterns in the data will result in smaller file sizes and reduced storage costs.

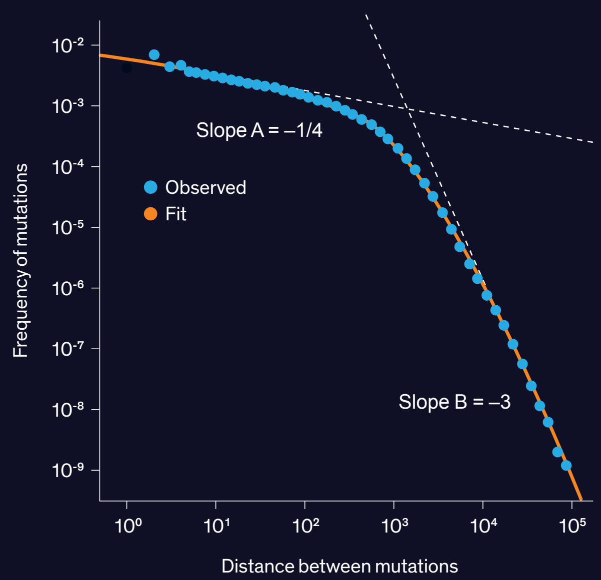
In our own research at Stanford University, we've made one potentially useful observation: The distance along the genome between consecutive DNA variations follows a "double power law" distribution. You may be familiar with the concept of the "power law" distribution, in which the probability of an outcome is proportional to the inverse magnitude of that outcome, possibly raised to some power. City populations typically follow this distribution: There are about half as many cities with 2 million people as there are cities with 1 million people. This law can also apply to a country's distribution of wealth, where 20 percent of the population holds 80 percent of the wealth.

A double power law consists of two different power laws operating on the same type of data but covering different ranges. For example, an 80/20 rule could apply to the lower half of a population by wealth, while a 90/10 rule applies to the upper half. Double power laws can be used to describe the distribution of the number of friends on Facebook, durations of phone calls, and file sizes on a hard drive.

And it turns out that a histogram plot of the distance between adjacent genetic variations, measured in DNA base pairs, looks like a double power law, with the crossover point between the two power laws happening at around 1,000 DNA base pairs [see graph, "Double Power Law"]. It's an open question as to what evolutionary process created this distribution, but its existence could potentially enable improved compression. One of Claude Shannon's foundational achievements in information theory states that data can't be compressed below the Shannon entropy—a measure of randomness—of its distribution. A double power law distribution turns out to be less random—that is, it has lower entropy—than a model that assumes each location in the genome is equally likely to contain a variation. We are excited by this discovery—both as an intriguing biological phenomenon and as a hint that greater compression savings lie untapped.

THE GENOMIC DATA compressors in use today are lossless—that is, they allow you to recover the uncompressed file bit for bit, exactly as it was before compression. But there's a case for allowing some amount of loss, not in the DNA sequences but in the quality scores denoting the sequencer's confidence in the data. While there are only four DNA nucleotides (A,C,G,T), there are typically about 40 possible quality scores, so most of the bits in a lossless compressed FASTQ file make up the quality score rather than the DNA sequences. This amount of precision is wasteful, as applications that use genomic data tend to ignore small variations in quality scores or may discard the quality scores entirely. Indeed, performance on some tasks, like finding variations between two genomes, actually improves when the quality scores are compressed in a lossy way, because lossy compression smooths out irrelevant variations among the quality scores, effectively removing noise from the data.

We can also save on storage space by discarding other pieces of genomic sequencing information. The exact order in which



DOUBLE POWER LAW: Distances between adjacent DNA mutations within a single genome follow a "double power law" probability distribution. At around 1,000 DNA base pairs, the slope changes from about –1/4 to about –3 on a plot that uses logarithmic scales on both axes. Understanding this distribution could lead to better compression algorithms.

SORTING: Arranging DNA substrings, or "reads," in alphabetical order results in similar reads being placed near one another, reducing the number of differences between adjacent reads, and thus reducing the space to compress the reads.

SOURCE, TOP: DMITRI S. PAVLICHIN, TSACHY WEISSMAN, AND GOLAN YONA, 2013, BIOINFORMATICS

DNA reads appear in a FASTQ file is often unimportant for subsequent analysis: You could, in many cases, such as when identifying genetic variations, shuffle the reads in a random order and expect nearly the same output. So you could sort the DNA reads alphabetically, and then exploit the fact that sorted lists can be compressed more than unsorted ones. The analogous case in text compression would be to sort a list of words and state the distance between adjacent words. The words *decompressed* and *decompresses*, for example, are adjacent in the dictionary, and their last letters (*d* and *s*) are 15 letters apart in the alphabet, so you can encode the entire second word with just the integer 15.

As an example of how this method works on DNA, let's sort the sequences ACGAAA, ACGAAG, and ACGAAT alphabetically. The first five letters are all the same, so we're interested only in the differences between the sixth letters. The second sequence is then encoded as the integer 2 (because the last letter, G, is two letters after A in the nucleotide alphabet ACGT), and the third sequence is encoded as a 1 (because its last letter, T, is one letter after G). This approach could result in a twofold or greater savings relative to storing the DNA reads in their original order.

Of course, the compression ratio is just one measure of a compression tool's capabilities. Speed is another factor. Some special-purpose FASTQ compressors run in parallel, saving time over a single-CPU implementation; others exploit GPU and field-programmable gate array processors, hardware that's more commonly used to accelerate video processing and machine learning. Another useful feature is being able to search compressed data. You don't want to have to decompress an entire file in order to do a quick search for occurrences of a particular DNA sequence within it.

While a lot of genomic compression options are emerging, what's needed now is standardization. Just as video compression technology couldn't take off until a good portion of the industry agreed to a standard, genomic compression technology will have to move toward one standard—or at least a small set of standards.

Fortunately, work on a compression standard for genomic sequencing data has begun. The Moving Picture Experts Group (MPEG)—the same body that developed the MP3 audio format and several popular video formats—has for several years been developing a standard for compressed genomic data, an effort named MPEG-G. The specification is expected to be completed late this year. This standard will evolve as the technology improves, in the same way that video compression standards have been doing.

The pace at which we develop efficient, sound, and standardized genomic data compression is simply a matter of economics. As the amount of stored data skyrockets, and the cost of storage becomes onerous, reducing the cost will propel the industry toward better compression methods.

Right now, genomic research may be on the cusp of reaping unexpected benefits as the total amount of sequence data accumulates; today, the field is approximately where artificial intelligence was a mere decade ago. The recent dramatic advances in AI have been driven in large part by the availability of vast data sets; deep-learning algorithms that performed poorly on moderate amounts of data became powerful when used with massive sets. Genomic researchers have started applying deep learning to their data, but they'll likely have to wait for a critical mass of genomic information to accumulate before realizing similar gains. One thing is clear, though: They won't get there without major advances in genomic data compression technology. ■

↗ **POST YOUR COMMENTS** at http://spectrum.ieee.org/genome0918

Tape Storage Mounts a Comeback

Disk drives are reaching their limits, but magnetic tape just gets better and better

By **Mark Lantz** • Photography by **Victor Prado**

It should come as no surprise

that recent advances in big-data analytics and artificial intelligence have created strong incentives for enterprises to amass information about every measurable aspect of their businesses. And financial regulations now require organizations to keep records for much longer periods than they had to in the past. So companies and institutions of all stripes are holding onto more and more.

Studies show that the amount of data being recorded is increasing at 30 to 40 percent per year. At the same time, the capacity of modern hard drives, which are used to store most of this, is increasing at less than half that rate. Fortunately, much of this information doesn't need to be accessed instantly. And for such things, magnetic tape is the perfect solution.

Seriously? Tape? The very idea may evoke images of reels rotating fitfully next to a bulky mainframe in an old movie like *Desk Set* or *Dr. Strangelove*. So, a quick reality check: Tape has never gone away!

Indeed, much of the world's data is still kept on tape, including data for basic science, such as particle physics and radio astronomy, human heritage and national archives, major motion pictures, banking, insurance, oil exploration, and more. There is even a cadre of people (including me, trained in materials science, engineering, or physics) whose job it is to keep improving tape storage.

Tape has been around for a long while, yes, but the technology hasn't been frozen in time. Quite the contrary. Like the hard disk and the transistor, magnetic tape has advanced enormously over the decades.

The first commercial digital-tape storage system, IBM's Model 726, could store about 1.1 megabytes on one reel of tape. Today, a modern tape cartridge can hold 15 terabytes. And a single robotic tape library can contain up to 278 *petabytes* of data. Storing that much data on compact discs would require more than 397 million of them, which if stacked would form a tower more than 476 kilometers high.

It's true that tape doesn't offer the fast access speeds of hard disks or semiconductor memories. Still, the medium's advantages are many. To begin with, tape storage is more energy efficient: Once all the data has been recorded, a tape cartridge simply sits quietly in a slot in a robotic library and doesn't consume any power at all. Tape is also exceedingly reliable, with error rates that are four to five orders of magnitude lower than those of hard drives. And tape is very secure, with built-in, on-the-fly encryption and additional security provided by the nature of the medium itself. After all, if a cartridge isn't mounted in a drive, the data cannot be accessed or modified. This "air gap" is particularly attractive in light of the growing rate of data theft through cyberattacks.

The offline nature of tape also provides an additional line of defense against buggy software. For example, in 2011, a flaw in a software update caused Google to accidentally delete the saved email messages in about 40,000 Gmail accounts. That loss occurred despite there being several copies of the data stored on hard drives across multiple data centers. Fortunately, the data was also recorded on tape, and Google could eventually restore all the lost data from that backup.

The 2011 Gmail incident was one of the first disclosures that a cloud-service provider was using tape for its operations. More recently, Microsoft let it be known that its Azure Archive Storage uses IBM tape storage equipment.

1951
Magnetic tape first used to record data on a computer (Univac)

1952
Introduction of commercial computer tape storage (IBM 726)

1964
Introduction of 9-track tape storage

1968
Introduction of the self-threading drive (IBM 2420)

1974
First automated (robotic) tape library (IBM 3850)

IBM (3)

INSIDE AND OUT: **A modern Linear Tape-Open (LTO) tape cartridge consists of a single reel. After the cartridge is inserted, the tape is fed automatically to a reel built into the drive mechanism.**

All these pluses notwithstanding, the main reason why companies use tape is usually simple economics. Tape storage costs one-sixth the amount you'd have to pay to keep the same amount of data on disks, which is why you find tape systems almost anyplace where massive amounts of data are being stored. But because tape has now disappeared completely from consumer-level products, most people are unaware of its existence, let alone of the tremendous advances that tape recording technology has made in recent years and will continue to make for the foreseeable future.

All this is to say that tape has been with us for decades and will be here for decades to come. How can I be so sure? Read on.

Tape has survived for as long as it has for one fundamental reason: It's cheap. And it's getting cheaper all the time. But will that always be the case?

You might expect that if the ability to cram ever more data onto magnetic disks is diminishing, so too must this be true for tape, which uses the same basic technology but is even older. The surprising reality is that for tape, this scaling up in capacity is showing no signs of slowing. Indeed, it should continue for many more years at its historical rate of about 33 percent per year, meaning that you can expect a doubling in capacity roughly every two to three years. Think of it as a Moore's Law for magnetic tape.

That's great news for anyone who has to deal with the explosion in data on a storage budget that remains flat. To understand why tape still has so much potential relative to hard drives, consider the way tape and hard drives evolved.

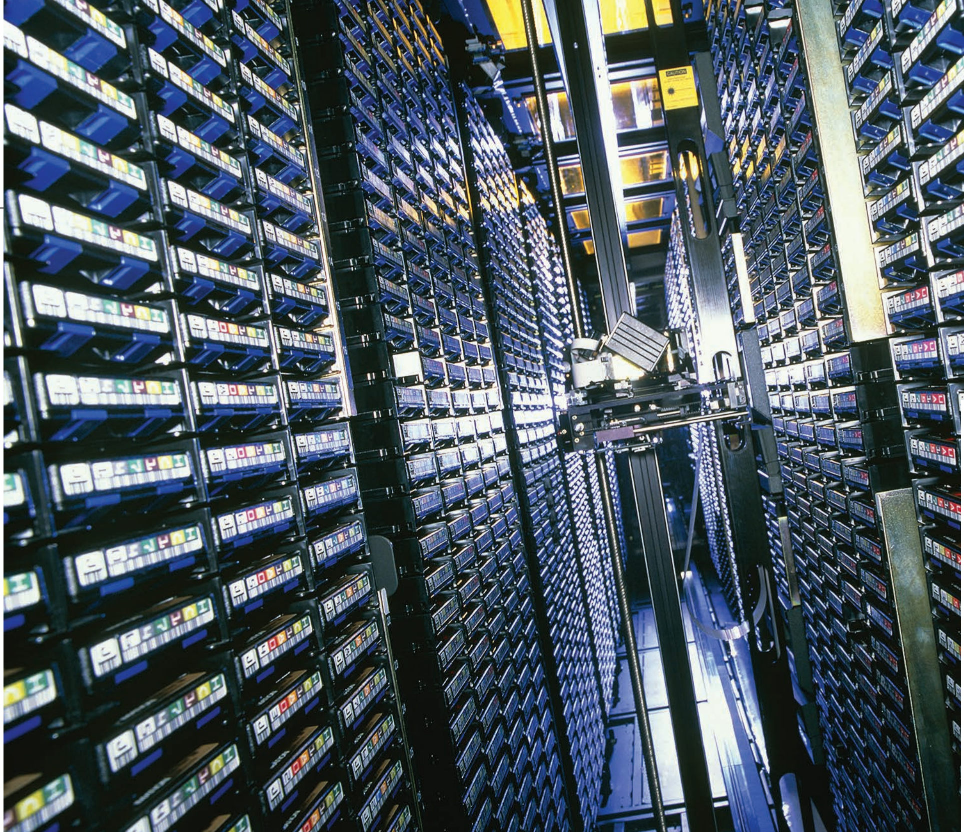
Both rely on the same basic physical mechanisms to store digital data. They do so in the form of narrow tracks in a thin film of magnetic material in which the magnetism switches between two states of polarity. The information is encoded as a series of bits, represented by the presence or absence of a magnetic-polarity transition at specific points along a track. Since the introduction of tape and hard drives in the 1950s, the manufacturers of both have been driven by the mantra "denser, faster, cheaper." As a result, the cost of both, in terms of dollars per gigabyte of capacity, has fallen by many orders of magnitude.

These cost reductions are the result of exponential increases in the density of information that can be recorded on each square millimeter of the magnetic substrate. That areal density is the product of the recording density along the data tracks and the density of those tracks in the perpendicular direction.

Early on, the areal densities of tapes and hard drives were similar. But the much greater market size and revenue from the sale of hard drives provided funding for a much larger R&D effort, which enabled their makers to scale up more aggressively. As a result, the current areal density of high-capacity hard drives is about 100 times that of the most recent tape drives.

Nevertheless, because they have a much larger surface area available for recording, state-of-the-art tape systems

PHOTOGRAPH BY Victor Prado



provide a native cartridge capacity of up to 15 TB—greater than the highest-capacity hard drives on the market. That's true even though both kinds of equipment take up about the same amount of space.

With the exception of capacity, the performance characteristics of tape and hard drives are, of course, very different. The long length of the tape held in a cartridge—normally hundreds of meters—results in average data-access times of 50 to 60 seconds compared with just 5 to 10 milliseconds for hard drives. But the rate at which data can be written to tape is, surprisingly enough, more than twice the rate of writing to disk.

Over the past few years, the areal density scaling of data on hard disks has slowed from its historical average of around 40 percent a year to between 10 and 15 percent. The reason has to do with some fundamental physics: To record more data in a given area, you need to allot a smaller region to each bit. That in turn reduces the signal you can get when you read it. And if you reduce the signal too much, it gets lost in the noise that arises from the granular nature of the magnetic grains coating the disk.

It's possible to reduce that background noise by making those grains smaller. But it's difficult to shrink the magnetic grains beyond a certain size without compromising their ability to maintain a magnetic state in a stable way. The smallest size that's practical to use for magnetic recording is known in this business as the superparamagnetic limit. And disk manufacturers have reached it.

Until recently, this slowdown was not obvious to consumers, because disk-drive manufacturers were able to compensate by adding more heads and platters to each unit, enabling a higher capacity in the same size package.

A DATA DELUGE: **Modern tape libraries can hold hundreds of petabytes, whereas the IBM 726 [right], introduced in 1952, could store just a couple of megabytes.**

But now both the available space and the cost of adding more heads and platters are limiting the gains that drive manufacturers can make, and the plateau is starting to become apparent.

There are a few technologies under development that could enable hard-drive scaling beyond today's superparamagnetic limit. These include heat-assisted magnetic recording (HAMR) and microwave-assisted magnetic recording (MAMR), techniques that enable the use of smaller grains and hence allow smaller regions of the disk to be magnetized. But these approaches add cost and introduce vexing engineering challenges. And even if they are successful, the scaling they provide is, according to manufacturers, likely to remain limited. Western Digital Corp., for example, which recently announced that it will probably begin shipping MAMR hard drives in 2019, expects that this technology will enable areal density scaling of only about 15 percent per year.

In contrast, tape storage equipment currently operates at areal densities that are well below the superparamagnetic limit. So tape's Moore's Law can go on for a decade or more without running into such roadblocks from fundamental physics.

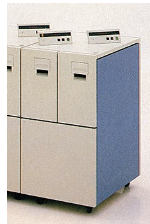
LEFT: DAVID PARKER/SCIENCE SOURCE; RIGHT: IBM

Still, tape is a tricky technology. Its removable nature, the use of a thin polymer substrate rather than a rigid disk, and the simultaneous recording of up to 32 tracks in parallel create significant hurdles for designers. That's why my research team at the IBM Research–Zurich lab has been working hard to find ways to enable the continued scaling of tape, either by adapting hard-drive technologies or by inventing completely new approaches.

In 2015, we and our collaborators at FujiFilm Corp. showed that by using ultrasmall barium ferrite particles oriented perpendicular to the tape, it's possible to record data at more than 12 times the density achievable with today's commercial technology. And more recently, in collaboration with Sony Storage Media Solutions, we demonstrated the possibility of recording data at an areal density that is about 20 times the current figure for state-of-the-art tape drives. To put this in perspective, if this technology were to be commercialized, a movie studio, which now might need a dozen tape cartridges to archive all the digital components of a big-budget feature, would be able to fit all of them on a single tape.

To enable this degree of scaling, we had to make a bunch of technical advances. For one, we improved the ability of the read and write heads to follow the slender tracks on the tape, which were just 100 or so nanometers wide in our latest demo.

We also had to reduce the width of the data reader—a magnetoresistive sensor used to read back the recorded data tracks—from its current micrometer size to less than 50 nm. As a result, the signal we could pick up with such a tiny reader got very noisy. We compensated by increasing the signal-to-noise ratio inherent to the media, which is a function of the size and orientation of the magnetic particles as well as their composition and the smoothness and slickness of the tape surface. To help further, we improved the signal processing and error-correction schemes our equipment employed.



1984
Introduction of thin-film head technology (IBM 3480) and of the 4-by-5-inch cartridge

1989
Introduction of Digital Data Storage (helical scan)

1993
Introduction of Digital Linear Tape

2000
Introduction of Linear Tape-Open (LTO)

2009
IBM introduces the Linear Tape File System

2017
Release of the latest generation of LTO (LTO-8)

TOP AND BOTTOM: IBM; MIDDLE: CHRISTIAN TAUBE/WIKIPEDIA

To ensure that our new prototype media can retain recorded data for decades, we changed the nature of the magnetic particles in the recording layer, making them more stable. But that change made it harder to record the data in the first place, to the extent that a normal tape transducer could not reliably write to the new media. So we used a special write head that produces magnetic fields much stronger than a conventional head could provide.

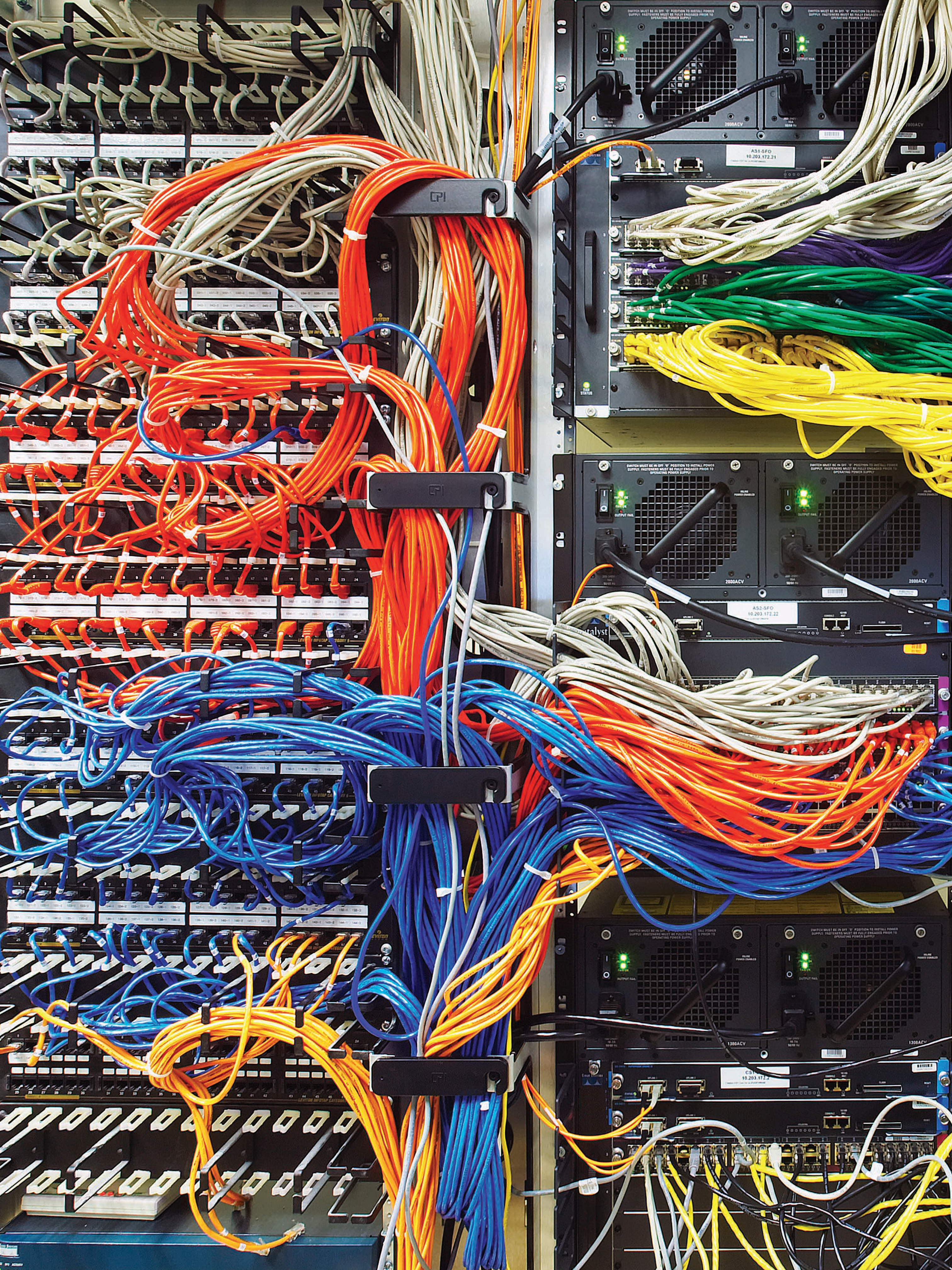
Combining these technologies, we were able to read and write data in our laboratory system at a linear density of 818,000 bits per inch. (For historical reasons, tape engineers around the world measure data density in inches.) In combination with the 246,200 tracks per inch that the new technology can handle, our prototype unit achieved an areal density of 201 gigabits per square inch. Assuming that one cartridge can hold 1,140 meters of tape—a reasonable assumption, based on the reduced thickness of the new tape media we used—this areal density corresponds to a cartridge capacity of a whopping 330 TB. That means that a single tape cartridge could record as much data as a wheelbarrow full of hard drives.

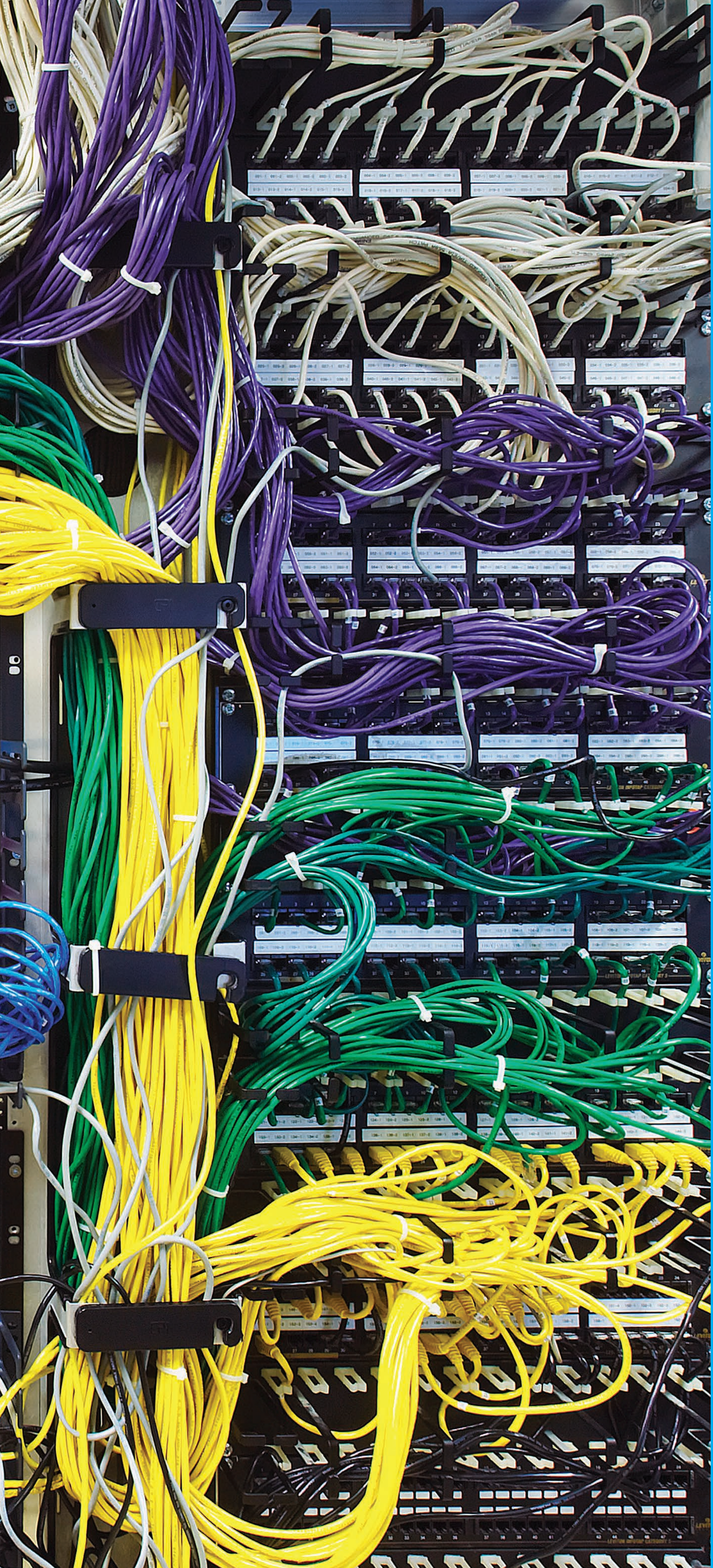
In 2015, the Information Storage Industry Consortium, an organization that includes HP Enterprise, IBM, Oracle, and Quantum, along with a slew of academic research groups, released what it called the "International Magnetic Tape Storage Roadmap." That forecast predicted that the areal density of tape storage would reach 91 Gb per square inch by 2025. Extrapolating the trend suggests that it will surpass 200 Gb per square inch by 2028.

The authors of that road map each had an interest in the future of tape storage. But you needn't worry that they were being too optimistic. The laboratory experiments that my colleagues and I have recently carried out demonstrate that 200 Gb per square inch is perfectly possible. So the feasibility of keeping tape on the growth path it's had for at least another decade is, to my mind, well assured.

Indeed, tape may be one of the last information technologies to follow a Moore's Law–like scaling, maintaining that for the next decade, if not beyond. And that streak in turn will only increase the cost advantage of tape over hard drives and other storage technologies. So even though you may rarely see it outside of a black-and-white movie, magnetic tape, old as it is, will be here for years to come. ■

↗ **POST YOUR COMMENTS** at https://spectrum.ieee.org/tapestorage0918





Silicon Photonics' Last-Meter Problem

Economics and physics still pose challenges to "fiber to the processor" tech

By **ANTHONY F.J. LEVI**

If you think we're on the cusp of a technological revolution today, imagine what it felt like in the mid-1980s. Silicon chips used transistors with micrometer-size features. Fiber-optic systems were zipping trillions of bits per second around the world.

With the combined might of silicon digital logic, optoelectronics, and optical-fiber communication, anything seemed possible.

Engineers envisioned all of these advances continuing and converging to the point where photonics would merge with electronics and eventually replace it. Photonics would move bits not just across countries but inside data centers, even inside computers themselves. Fiber optics would move data from chip to chip, they thought. And even those chips would be photonic: Many expected that someday blazingly fast logic chips would operate using photons rather than electrons.

It never got that far, of course. Companies and governments plowed hundreds of millions of dollars into developing new photonic components and systems that link together racks of computer servers inside data centers using optical fibers. And indeed, today, those photonic devices link racks in many modern data centers. But that is where the photons stop. Within a rack, individual server boards are still connected to each other with inexpensive copper wires and high-speed electronics. And, of course, on the boards themselves, it's metal conductors all the way to the processor.

Attempts to push the technology into the servers themselves, to directly feed the processors with fiber optics, have foundered on the rocks of economics. Admittedly, there is an Ethernet optical transceiver market of close to US $4 billion per year that's set to grow to nearly $4.5 billion and 50 million components by 2020, according to market research firm LightCounting. But photonics has never cracked those last few meters between the data-center computer rack and the processor chip.

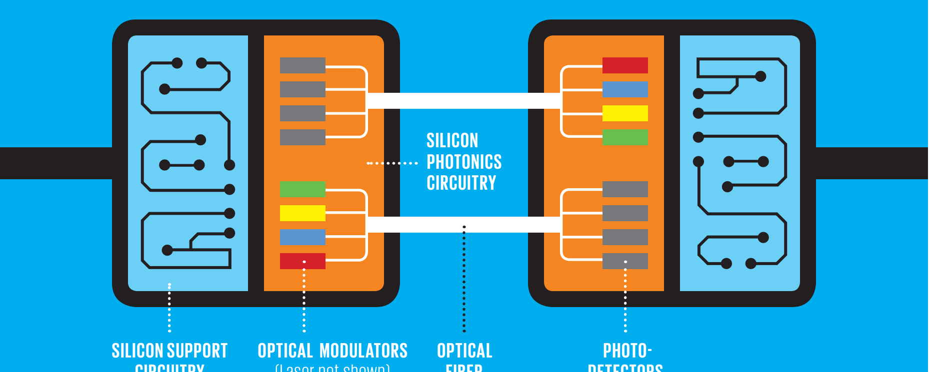
Nevertheless, the stupendous potential of the technology has kept the dream alive. The technical challenges are still formidable. But new ideas about how data centers could be designed have, at last, offered a plausible path to a photonic revolution that could help tame the tides of big data.

ANYTIME YOU ACCESS the Web, stream television, or do nearly anything in today's digital world, you are using data that has flowed through photonic transceiver modules. The job of these transceivers is to convert signals back and forth between electrical and optical. These devices live at each end of the optical fibers that speed data within the data centers of every major cloud service and social media company. The devices plug into switchgear at the top of each server rack, where they convert optical signals to electrical ones for delivery to the group of servers in that rack. The transceivers also convert data from those servers to optical signals for transport to other racks or up through a network of switches and out to the Internet.

Each photonics transceiver module has three main kinds of components: a transmitter containing one or more optical modulators, a receiver containing one or more photodiodes, and CMOS logic chips to encode and decode data. Because ordinary silicon is actually lousy at emitting light, the photons come from a laser that's separate from the silicon chips (though it can be housed in the same package with them). Rather than switch the laser on and off to represent bits, the laser is kept on, and electronic bits are encoded onto the laser light by an optical modulator.

This modulator, the heart of the transmitter, can take a few forms. A particularly nice and simple one is called the Mach-Zehnder modulator. Here, a narrow silicon waveguide channels the laser's light. The guide then splits in two, only to rejoin a few millimeters later. Ordinarily, this diverging and converg-

Inside a Photonics Module

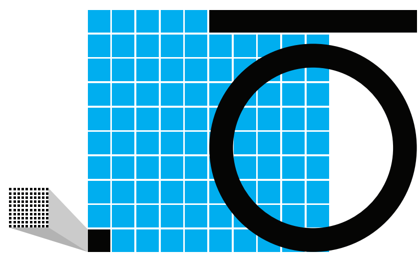


PLUG AND PLAY: A silicon photonics module converts electronic data to photons and back again. Silicon circuits [light blue] help optical modulators [bottom row, left] encode electronic data into pulses of several colors of light. The light travels through an optical fiber to another module, where photodetectors [gray] turn the light back into electronic bits. These are processed by the silicon circuits and sent on to the appropriate servers.

PREVIOUS PAGES: GETTY IMAGES; THIS PAGE: LUXTERA (PHOTO) ing wouldn't affect the light output, because both branches of the waveguide are the same length. When they join up, the light waves are still in phase with each other. However, voltage applied to one of the branches has the effect of changing that branch's index of refraction, effectively slowing down or speeding up the light's wave. Consequently, when light waves from the two branches meet up again, they destructively interfere with each other and the signal is suppressed. So, if you vary a voltage on that branch, what you're actually doing is using an electrical signal to modulate an optical one.

The receiver is much simpler; it's basically a photodiode and some supporting circuitry. After traveling through an optical fiber, light signals reach the receiver's germanium or silicon germanium photodiode, which produces a voltage with each pulse of light.

Both the transmitter and receiver are backed up by circuitry that does amplification, packet processing, error correction, buffering, and other tasks to comply with the Gigabit Ethernet standard for optical fiber. How much of this is on the same chip as the photonics, or even in the same package, varies according to the vendor, but most of the electronic logic is separate from the photonics.



PHOTONICS FAIL: Photonics will never be a real option to transport data from one part of a silicon chip to another. A single optical switch, a ring oscillator in this case, performs the same function as an individual transistor, but it takes up 10,000 times as much area.

WITH OPTICAL COMPONENTS on silicon integrated circuits becoming increasingly available, you might be tempted to think that the integration of photonics directly into processor chips was inevitable. And indeed, for a time it seemed so. [See "Linking With Light," *IEEE Spectrum*, October 2001.]

You see, what had been entirely underestimated, or even ignored, was the growing mismatch between how quickly the minimum size of features on electronic logic chips was shrinking and how limited photonics was in its ability to keep pace. Transistors today are made up of features only a few nanometers in dimension. In 7-nanometer CMOS technology, more than 100 transistors for general-purpose logic can be packed onto every square micrometer of a chip. And that's to say nothing of the maze of complex copper wiring above the transistors. In addition to the billions of transistors on each chip, there are also a dozen or so levels of metal interconnect needed to wire up all those transistors into the registers, multipliers, arithmetic logic units, and more complicated things that make up processor cores and other crucial circuits.

The trouble is that a typical photonic component, such as a modulator, can't be made much smaller than the wavelength of the light it's going to carry, limiting it to about 1 micrometer wide. There is no Moore's Law that can overcome this. It's not a matter of using more and more advanced lithography. It's simply that electrons—having a wavelength on the order of few nanometers—are skinny, and photons are fat.

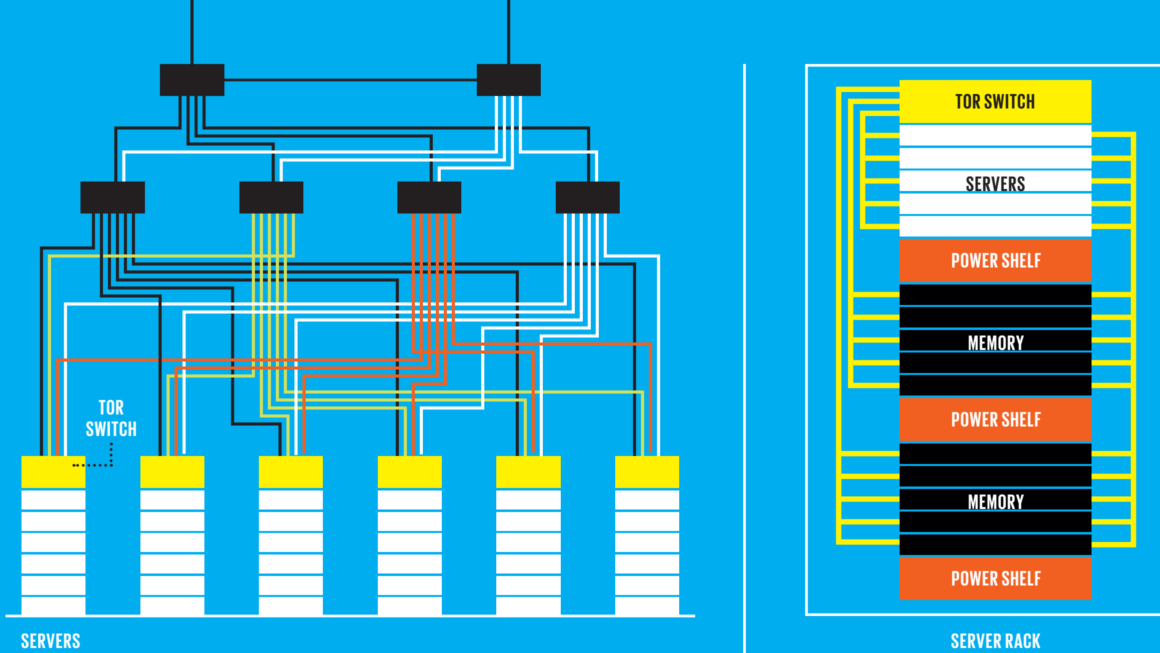
But, still, couldn't chipmakers just integrate the modulator and accept that the chip will have fewer transistors? After all, a chip can now have billions of them. Nope. The massive amount of system function that each square micrometer of a silicon electronic chip area can deliver makes it very expensive to replace even relatively few transistors with lower-functioning components such as photonics.

Here's the math. Say there are on average 100 transistors per square micrometer. Then a photonic modulator that occupies a relatively small area of 10µm by 10µm is displacing a circuit comprising 10,000 transistors! And recall that a typical photonic modulator acts as a simple switch, turning light on and off. But each individual transistor can act as a switch, turning current on and off. So, roughly speaking, the opportunity cost for this primitive function is 10,000:1 *against* the photonic component because there are at least 10,000 electronic switches available to the system designer for every one photonic modulator. No chipmaker is willing to accept such a high price, even in exchange for the measurable improvements in performance and efficiency you might get by integrating the modulators right onto the processor.

The idea of substituting photonics for electronics on chips encounters other snags, too. For example, there are critical on-chip functions, such as memory, for which photonics has no comparable capability. The upshot is that photons are simply incompatible with basic computer chip functions. And even when they are not, integrating a competing photonic function on the same chip as electronics makes no sense.

THAT'S NOT TO SAY photonics can't get a lot closer to processors, memory, and other key chips than it does now. Today, the market for optical interconnects in the data center focuses on systems called top-of-rack (TOR) switches, into which the photonic transceiver modules are plugged. Here at the top of 2-meter tall racks that house server chips, memory, and other resources, fiber optics link the TORs to each other via a separate layer of switches. These switches, in turn, connect to yet another set of switches that form the data center's gateway to the Internet.

The faceplate of a typical TOR, where transceiver modules are plugged in, gives a good idea of just how much data is in motion. Each TOR is connected to one transceiver module, which is in turn connected to two optical fibers (one to transmit and one to receive). Thirty-two modules, each with 40-gigabit-per-second data rates in each direction, can be plugged into a TOR's 45-millimeter-high faceplate, allowing for as many as 2.56 terabits per second to flow between the two racks.



Data-Center Design

TODAY: Photonics slings data through a multitiered network in the data center. The link to the Internet is at the top (core) level. Switchgear moves data via optical fibers to top-of-rack (TOR) switches, which sit atop each rack of servers.

TOMORROW: Photonics could facilitate a change in data-center architecture. Rack-scale architecture would make data centers more flexible by physically separating computers from their memory resources and connecting them through an optical network.

But the flow of data within the rack and inside the servers themselves is still done using copper wires. That's unfortunate, because they are becoming an obstacle to the goal of building faster, more energy-efficient systems. Photonic solutions for this last meter (or two) of interconnect–either to the server or even to the processor itself–represent possibly the best opportunity to develop a truly high-volume optical component market. But before that can happen, there are some serious challenges to overcome in both price and performance.

So-called fiber-to-the-processor schemes are not new. And there are many lessons from past attempts about cost, reliability, power efficiency, and bandwidth density. About 15 years ago, for example, I contributed to the design and construction of an experimental transceiver that showed very high bandwidth. The demonstration sought to link a parallel fiber-optic ribbon, 12 fibers wide, to a processor. Each fiber carried digital signals generated separately by four vertical-cavity surface-emitting lasers (VCSELs)–a type of laser diode that shines out of the surface of a chip and can be produced in greater density than so-called edge-emitting lasers. The four VCSELs directly encoded bits by turning light output on and off, and they each operated at different wavelengths in the same fiber, quadrupling that fiber's capacity using what's called coarse wavelength-division multiplexing. So, with each VCSEL streaming out data at 25 Gb/s, the total bandwidth of the system would be 1.2 Tb/s. The industry standard today for the spacing between neighboring fibers in a 12-fiber-wide array is 0.25 mm, giving a bandwidth density of about 0.4 Tb/s/mm. In other words, in 100 seconds each millimeter could handle as much data as the U.S. Library of Congress's Web Archive team stores in a month.

Data rates even higher than this are needed for fiber-to-the-processer applications today, but it was a good start. So why wasn't this technology adopted? Part of the answer is that this system was neither sufficiently reliable nor practical to manufacture. At the time, it was very difficult to make the needed 48 VCSELs for the transmitter and guarantee that there would be no failures over the transmitter lifetime. In fact, an important lesson was that one laser using many modulators can be engineered to be much more reliable than 48 lasers.

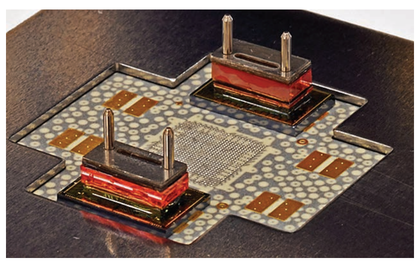
But today, VCSEL performance has improved to the extent that transceivers based on this technology could provide effective short-reach data-center solutions. And those fiber ribbons can be replaced with multicore fiber, which carries the same amount of data by channeling it into several cores embedded within the main fiber. Another recent, positive development is the availability of more complex digital-transmission standards such as PAM4, which boosts data transmission rates because it encodes bits on four intensities of light rather than just two. And research efforts, such as MIT's Shine program, are working toward bandwidth density in fiber-to-the-processor to demonstration systems with about 17 times what we achieved 15 years ago.

These are all major improvements, but even taken together they are not enough to enable photonics to take the next big leap toward the processor. However, I still think this leap can occur, because of a drive, just now gathering momentum, to change data-center system architecture.

Today processors, memory, and storage make up what's called a server blade, which is housed in a chassis in a rack in the data center. But it need not be so. Instead of placing memory with the server chips, memory could sit separately in the rack or even in a separate rack. This rack-scale architecture (RSA) is thought to use computing resources more efficiently, especially for social media companies such as Facebook where the amount of computing and memory required for specific applications grows over time. It also simplifies the task of replacing and managing hardware.

Why would such a configuration help enable greater penetration by photonics? Because exactly that kind of reconfigurability and dynamic allocation of resources could be made possible by a new generation of efficient, inexpensive, multi-terabit-per-second optical switch technology.

THE MAIN OBSTACLE to the emergence of this data-center remake is the price of components and the cost of their manufacture. Silicon photonics already has one cost advantage, which is that it can leverage existing chip manufacturing, taking advantage of silicon's huge infrastructure and reliability. Nevertheless, silicon and light are not a perfect fit: Apart from their crippling inefficiency at emitting light, silicon components suffer from large optical losses as well. As measured by light in to light out, a typical silicon photonic transceiver experiences greater than a 10-decibel (90 percent) optical loss. This inefficiency does not matter much for short-reach optical interconnects between TOR switches because, at least for now, the silicon's potential cost advantage outweighs that problem.



PAST PERFECT: We've had the technology to bring optical fiber directly to the processor for more than a decade. The author helped conceive this 0.4-terabit-per-second-per-millimeter demonstrator more than 15 years ago.

ANTHONY F.J. LEVI

An important cost in a silicon photonics module is the humble, yet critically important, optical connection. This is both the physical link between the optical fiber and the transmitter or receiver chip as well as the link between fibers. Many hundreds of millions of such fiber-to-fiber connectors must be manufactured each year with extreme precision. To understand just how much precision, note that the diameter of a human hair is typically a little less than the 125-μm diameter of a single-mode silica glass fiber used for optical interconnects. The accuracy with which such single-mode fibers must be aligned in a connector is around 100 nm—about one one-thousandth the diameter of a human hair—or the signal will become too degraded. New and innovative ways to manufacture connectors between fibers and from fiber to transceiver are needed to meet growing customer demand for both precision and low component price. However, very few manufacturing techniques are close to being inexpensive enough.

One way to reduce cost is, of course, to make the chips in the optical module cheaper. Though there are other ways to make these chips, a technique called wafer-scale integration could help. Wafer-scale integration means making photonics on one wafer of silicon, electronics on another, and then attaching the wafers. The paired wafers are then diced up into chips designed to be nearly complete modules. (The laser, which is made from a semiconductor other than silicon, remains separate.) This approach cuts manufacturing costs because it allows for assembly and production in parallel.

Another factor in reducing cost is, of course, volume. Suppose the total optical Gigabit Ethernet market is 50 million transceivers per year and each photonic transceiver chip occupies an area of 25 square millimeters. Assuming a foundry uses 200-mm-diameter wafers to make them and that it achieves a 100 percent yield, then the number of wafers needed is 42,000.

That might sound like a lot, but that figure actually represents less than two weeks of production in a typical foundry. In reality, any given transceiver manufacturer might capture 25 percent of the market and still support only a few days of production. There needs to be a path to higher volume if costs are really going to fall. The only way to make that happen is to figure out how to use photonics below the TOR switch, all the way to the processors inside the servers.

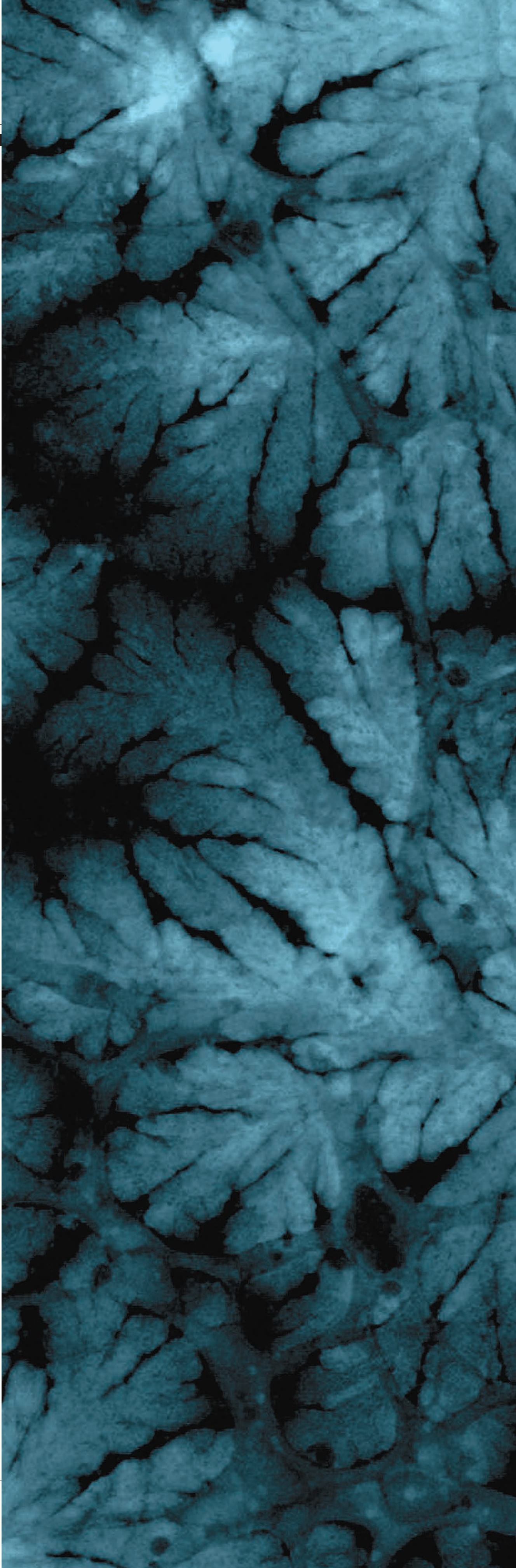
IF SILICON PHOTONICS is ever going to make it big in what are otherwise all-electronic systems, there will have to be compelling technical and business reasons for it. The components must solve an important problem and greatly improve the overall system. They must be small, energy efficient, and super-reliable, and they must move data extraordinarily fast.

Today, there is no solution that meets all these requirements, and so electronics will continue to evolve without becoming intimately integrated with photonics. Without significant breakthroughs, fat photons will continue to be excluded from places where skinny electrons dominate system function. However, if photonic components could be reliably manufactured in very high volume and at very low cost, the decades-old vision of fiber-to-the-processor and related architectures could finally become a reality.

We've made a lot of progress in the past 15 years. We have a better understanding of photonic technology and where it can and can't work in the data center. A sustainable multibillion-dollar-per-year commercial market for photonic components has developed. Photonic interconnects have become a critical part of global information infrastructure. However, the insertion of very large numbers of photonic components into the heart of otherwise electronic systems remains just beyond the edge of practicality.

Must it always be so? I think not. ■

↗ **POST YOUR COMMENTS** at https://spectrum.ieee.org/siliconphotonics0918



LESS FIRE, MORE POWER

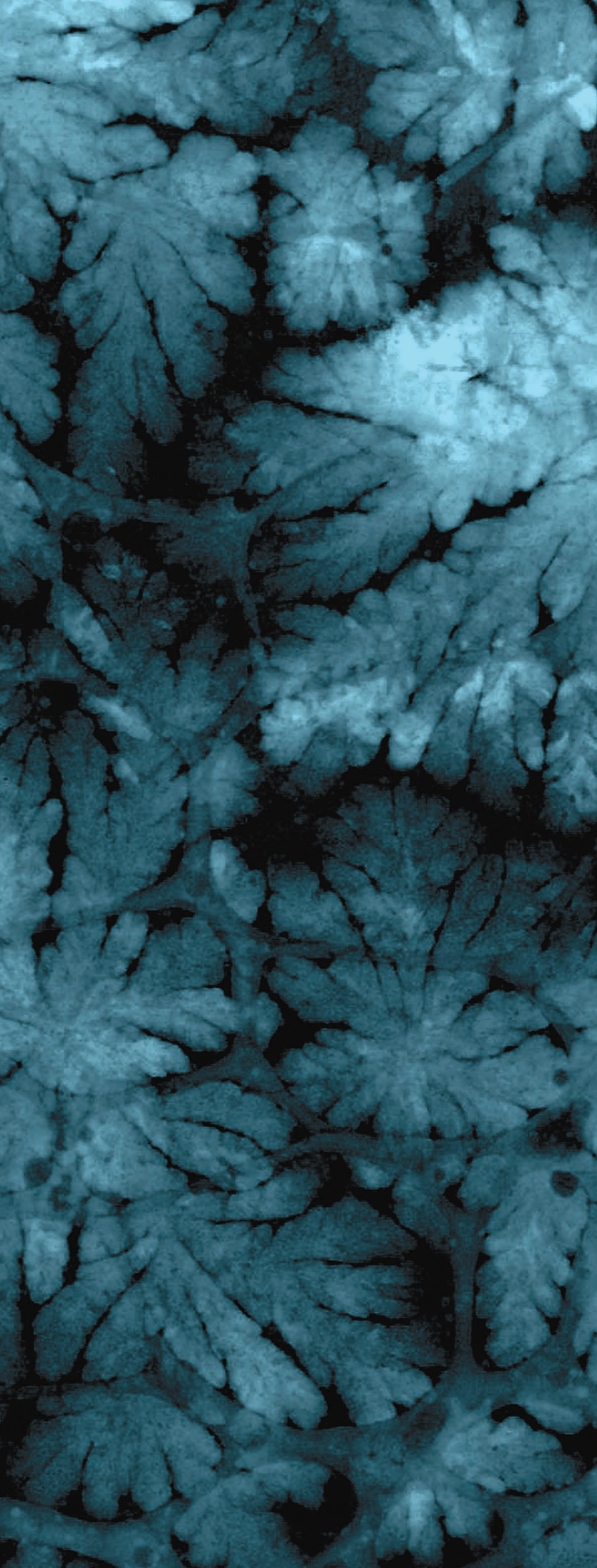
Without the needle-like growths that can short out cells, lithium-ion batteries will be safer

By Weiyang Li & Yi Cui

LITHIUM-ION BATTERIES HAVE MADE HEADLINES for the wrong reason: as a fire hazard. Just this past May, three apparent battery fires in Tesla cars were reported in the United States and Switzerland. In the United States alone, a fire in a lithium-ion battery grounds a flight every 10 days on average, according to the Federal Aviation Administration. And the same problem afflicts electronic cigarettes, which have been blowing up in people's faces sporadically.

No other drawback has so hobbled the advance of what is by far the most promising battery technology to emerge in our lifetimes. Lithium-ion batteries store much more energy than previous chemistries could manage, making them crucial to the future success of phones, drones, cars, even airplanes.

Solving this problem would not only protect lives and property, it would also make it possible to use larger battery packs with more closely packed cells. We'd finally be able to fully exploit the great



ROOT AND BRANCH: Crystalline lithium-metal structures grow out of the anode of a lithium-ion cell in a branching pattern, thus their name, dendrites (from the Greek *dendron* meaning "tree"). If they grow too long, they can short out the cell.

BROOKHAVEN NATIONAL LABORATORY/ SCIENCE SOURCE

energy-to-weight ratio, or specific energy, that this technology allows. What's more, we'd be able to make progress with the next generation of batteries, the ones that use lithium metal.

The problems of today's lithium-ion batteries can be traced largely to dendrites, tiny threadlike structures that form on the surface of an electrode over repeated cycles of charging and discharging. But through our work at Dartmouth and Stanford, the two of us have found that a little chemical tweaking of the electrolyte can head off the pesky growths.

LITHIUM-ION BATTERY PACK is invariably composed of one or more compartments, or cells, each of which has two electrodes covered by an extremely thin polymer film, called a separator, which prevents their coming into direct contact. Permeating the porous separator is the electrolyte, a material—today generally a liquid—that allows lithium ions to move back and forth during charging and discharging.

The slightest damage to the ultrathin separator can put the electrodes into direct contact and create an internal short circuit, which can generate enough heat to make the cell catch fire. The heat of the fire may then overheat adjacent cells, resulting in a chain reaction that can easily cause the whole battery pack to explode.

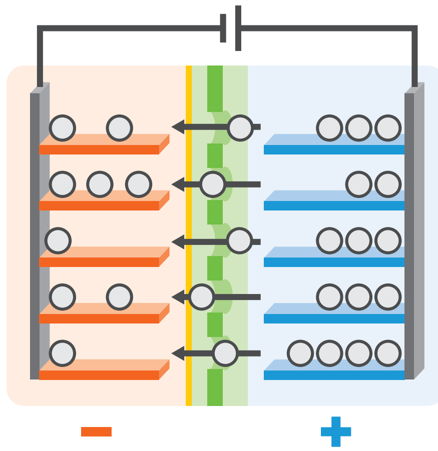
So it's the integrity of the cell's separator that matters most. Of course, every effort must be made during the manufacturing process to prevent damage to the separator, but even a perfectly fabricated separator can fail in operation if dendrites later damage it.

Dendrites are sharp bits of lithium metal that grow from the anode. These fibers can spread like kudzu vines into the electrolyte, pierce the separator and make their way to the cathode. It's amazing how

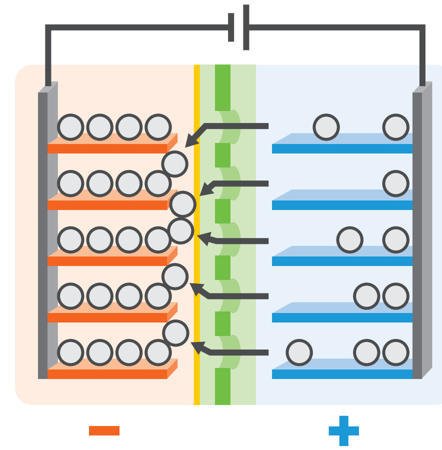
ILLUSTRATIONS BY James Provost

SHARP DENDRITES FROM TINY IONS GROW

Anode
SEI
Separator
Electrolyte
Cathode



CHARGING: Lithium ions move from the cathode [right] through the separator [middle] and on to the anode [left].



OVERCHARGING: When no more room remains for ions, any excess ions will begin to accumulate on the surface.

such tiny little things can cause so much destruction: They were responsible, for example, for the fires that grounded the worldwide fleet of Boeing 787s in 2013.

Dendrites tend to grow when the battery is overcharged, because that's when the lithium ions migrating into the anode can no longer find a berth. Normally, the ions slip between the atomic layers of the anode, a process called intercalation, but when the space between the layers is all filled up—as can happen during overcharging—there's nowhere else for the lithium to go but onto the surface. There they form the seeds of a metallic crystal, which grows with each new charge-discharge cycle.

Solving this problem of dendrite growth matters not just for today's generation of lithium-ion batteries but also for future batteries that will need lithium-metal anodes. That's because lithium metal has a high theoretical specific energy capacity—3,860 milliampere-hours per gram—and a negative electrochemical potential no other anode material can match. A higher potential allows for a higher battery voltage, which is just what's needed in electric cars and in mobile devices.

Both these qualities make lithium anodes critical to battery technologies that are still in the lab, like the highly promising lithium-sulfur and lithium-air batteries, which can store 5 to 10 times as much energy by weight as today's lithium-ion designs. Those future batteries may not be able to incorporate—as lithium-ion batteries do—anodes made of graphite, which has a theoretical capacity of only 372 mAh/g.

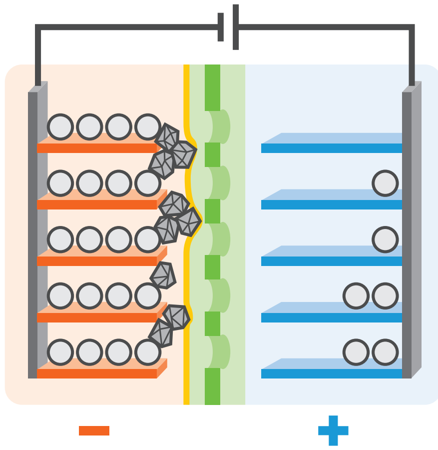
THE FORMATION OF LITHIUM DENDRITES takes place at the meeting point between the anode and the electrolyte, in a layer called the solid electrolyte interphase (SEI). After enough lithium ions move into the anode and accept electrons there, the anode finally expands enough to break the SEI layer. From that point onward, lithium begins to form deposits at the broken part of the SEI. And these deposits seed dendrites.

Later, during discharging, lithium ions are pulled out of the anode, shrinking it again. The SEI layer collapses, generating more cracks and pinholes from which still more dendrites can begin to shoot out the next time the cell charges. Also, by exposing so much metallic lithium to the electrolyte, these cracks enable the two components to react chemically. As lithium disappears into the resulting chemical product, the lithium that remains for use in the cell diminishes. That decline lowers what's called the coulombic efficiency, which can be determined by dividing the amount of lithium removed from use by the amount of lithium still participating in the reactions during each charge-discharge cycle.

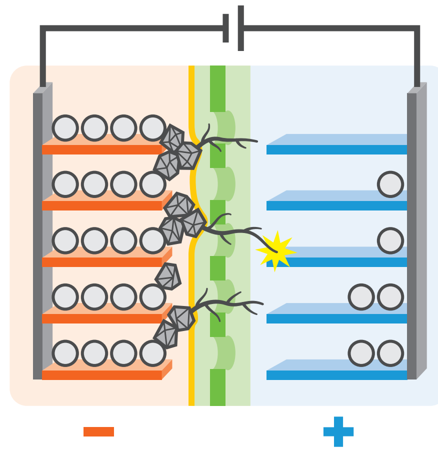
Also, because they are very fragile, dendrites often break off from the anode, generating "dead lithium" that cannot be reused, which further lowers the coulombic efficiency of the cell. To compensate for such losses, today's batteries must include excess lithium, which adds substantially to their weight and cost.

To head off dendrites, we need to shore up the SEI by forming a "super" SEI layer that's uniform and stable. One way to achieve this is to modify the anode's surface by laying down an artificial SEI layer, as it were. We've tried it, and it works. Unfortunately, this approach greatly complicates the fabrication of lithium-ion cells.

Another tactic is to reformulate the electrolyte by including substances that reinforce the SEI layer. The challenge



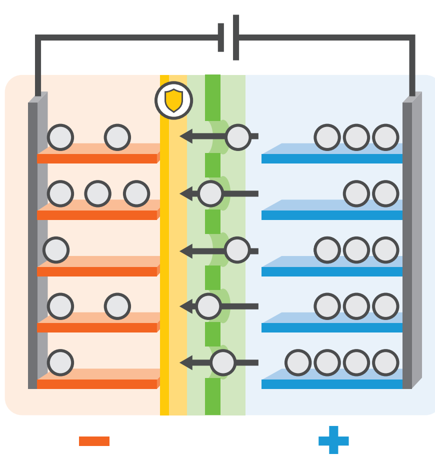
CRYSTALLIZATION: Accumulating ions form metallic crystals, which damage the solid electrolyte interphase (SEI), where the anode meets the electrolyte.



BREAKING AND ENTERING: Dendrites branching out from the crystals pierce first the SEI and then the separator, forming a bridge to the cathode—and thus a short circuit.

HEADING OFF DENDRITES

You can avoid dendrites by shoring up the solid electrolyte interphase layer [yellow] with a "super" layer [light yellow]. A simpler way, now under development, is to add chemicals to the electrolyte.



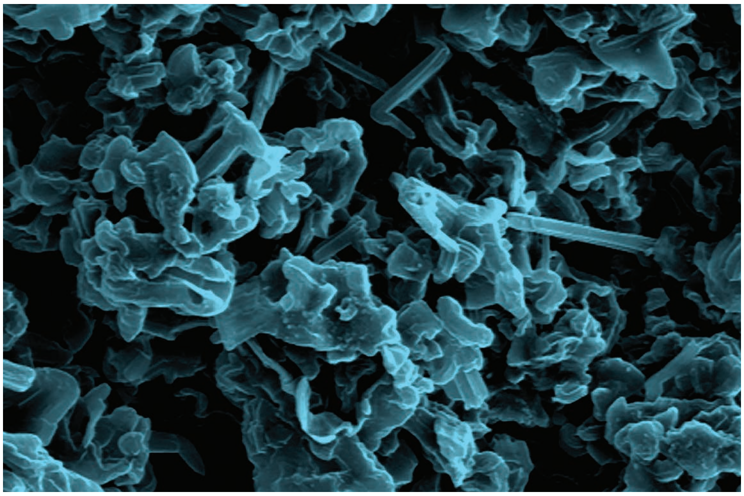
is that such additives must easily dissolve, and most candidate materials don't–a long-standing problem for battery researchers. The additives that do dissolve quickly are consumed during cycling, and as a result the SEI layers fall apart over the long term.

How, then, do you find the right additive? We got our key idea following a roundabout path.

WE'D BEEN CONSIDERING THE PROBLEM of the lithium-sulfur battery, the futuristic technology we mentioned earlier. What makes this combination so attractive, in theory, is its ability to store–in the same amount of mass–more than five times the energy of today's lithium-ion batteries. Such a battery uses metallic lithium for the anode and sulfur for the cathode, and during the reactions that take place while charging or discharging there are a number of steps that create intermediate products at the sulfur cathode. These products, known as polysulfide ions, are highly soluble in the electrolyte, and that means that when the battery is in operation they can travel from the cathode, pass through the separator, and then arrive at the anode. That's not good: Only the lithium ions ought to get that far. When these polysulfide ions hit the anode, they react vigorously with the lithium, accept electrons, and are reduced to a solid. Not only does this process slowly deplete sulfur from the system, it also gradually forms a coating that can wreck the lithium anode. This has been the main difficulty dogging the development of lithium-sulfur batteries.

To avoid this parasitic reaction, researchers have mainly tried to restrict the polysulfide from leaching out of the sulfur cathode in the first place. In one of our brainstorming sessions, we began to think different, as Steve Jobs might say: What if we could actually take advantage of this reaction? By controlling how the polysulfide ions react with lithium, perhaps we could not just form a strong and stable SEI layer but actually nip dendrites in the bud!

Meanwhile, we discovered that lithium nitrate–a very commonly used lithium salt–had long been considered as a potential electrolyte additive because of its ability to restrict–or passivate–the reactivity of the lithium metal. Perhaps by adding both polysulfide and lithium nitrate to the electrolyte



HIGH THREAD COUNT: Dendrites take on a threadlike form as they grow on the surface of the electrode.

we could create complementary actions: Polysulfide reacts with lithium metal, while lithium nitrate can help to prevent the lithium from reacting with polysulfide. By manipulating these two competing reactions, we should be able to turn the sulfur-lithium reaction from a flaw to a feature.

We added lithium polysulfide and lithium nitrate to the electrolyte in various concentrations. We studied the effect on the process of lithium plating and stripping in a two-electrode test cell that used lithium metal as one electrode and a stainless-steel foil as the other.

We assembled coin cells, also called button cells, similar to the ones that power small electronic devices, such as watches, calculators, and hearing aids, and we applied a constant current during charging, allowing the same current to flow during discharge. We deposited a bit of lithium onto the stainless steel by charging the cell; then we stripped it off in discharge, repeating the cycle many times. Finally, we took the cell apart and examined the lithium deposit under a scanning electron microscope.

What we saw was intriguing. Without electrolyte additives, the plated lithium formed structures that were thin, sharp, and fiberlike–dendritic, in a word. But when we added lithium nitrate to the electrolyte, the deposited lithium was thicker, less sharp, and shaped more or less like a noodle. The lithium nitrate had moderated dendrite growth but not prevented it.

Next, we added both lithium polysulfide and lithium nitrate to the electrolyte in various quantities. At just the right balance of additives, the synergistic effect we'd sought came through: No harmful dendritic structures grew. Instead we got flat, pancake-shaped lithium deposits. Even after hundreds of charge-discharge cycles, the surface of the plated lithium was still flat, without any dendritic structures.

Besides heading off dendrites, our two additives together greatly enhanced the coulombic efficiency and the cycling stability. The coulombic efficiency was better than 99 percent over more than 300 charge-discharge cycles. Charging caused plating on only a tiny bit more lithium than was stripped off during discharge. In contrast, with lithium nitrate alone, coulombic efficiency drops to less than 92 percent after just 180 cycles, and with polysulfide alone it's only about 80 percent.

These two additives, working together, bring a huge improvement because of their effect on the SEI layer. To figure out the exact mechanism of that effect, we used a technique called X-ray photoelectron spectroscopy and also conventional electron microscopy to deduce the structure and chemical composition of the SEI layer. In cells using one or the other additive, we found that the SEI layers were marred by lots of cracks and pinholes. When both additives were present, though, we got a flat, uniform SEI. And the chemical breakdown of the SEI layers confirmed that the two additives indeed had competing effects.

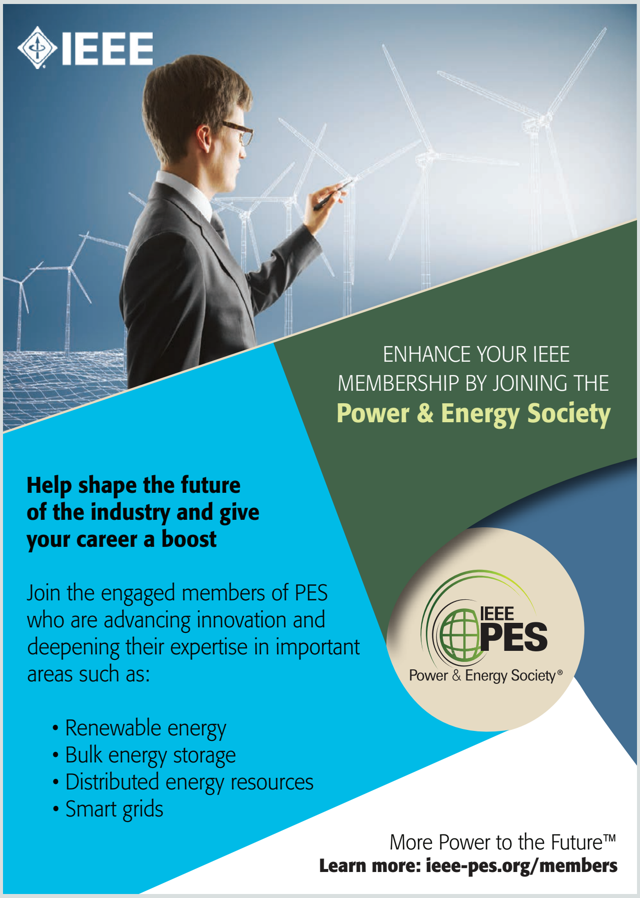
When we added both lithium nitrate and polysulfide, the lithium nitrate was the first to react with the lithium metal, and it did passivate the metal's surface, as expected, drastically reducing the metal's reaction with the polysulfide. The product from the first reaction formed mainly in the upper layer of the SEI, and it effectively suppressed the formation of dendrites.

This technique for preventing the growth of dendrites is still in its early days. We have problems to solve before we can think about commercialization. A particular difficulty is finding the precise formulation of the electrolyte additives for each of the several different kinds of lithium batteries.

But this new strategy holds out the promise not only of creating a safer, higher-energy lithium-ion battery but also of paving the way for next-generation battery chemistries.

With dendrites defeated, a lithium-metal design could store far more energy than today's batteries while lasting through the many charging cycles that consumer products require. We predict that in another 5 to 10 years, our technology will allow the commercialization of safe, superhigh-capacity batteries for phones, laptops, cars, and airplanes. That would make headlines for rechargeable batteries of a much more positive sort. ■

↗ **POST YOUR COMMENTS** online at https://spectrum.ieee.org/dendrites0918



IUPUI

SCHOOL OF ENGINEERING AND TECHNOLOGY

INDIANA UNIVERSITY–PURDUE UNIVERSITY
Indianapolis

Purdue School of Engineering and Technology, IUPUI
Chair, Department of Biomedical Engineering
and
Thomas J. Linnemeier Guidant Foundation Chair

The Department of Biomedical Engineering in the Purdue School of Engineering and Technology at Indiana University-Purdue University Indianapolis (IUPUI) is seeking highly qualified individuals to apply for the position of Department Chair and holder of the Thomas J. Linnemeier Guidant Foundation Chair in Biomedical Engineering, an endowed, tenured position at the rank of Professor.

This senior level leadership position within the School is intended for candidates who are dedicated to the advancement of undergraduate and graduate education, have a distinguished record of research and professional accomplishment, have strong leadership and management abilities, and a steadfast commitment to faculty governance. This is an outstanding opportunity to impact the future direction of the department and its programs with the potential to guide the department's ascension to among the premier departments of its kind in the nation.

Applicants must have a Ph.D. in engineering or a related discipline and a well-documented research career evidenced by significant publications and extramural research funding. The search encourages applications from not only well-established Professors but also from rapidly progressing Associate Professors whose work would merit promotion to Professor. The incumbent's research would have a primary application in an established area of biomedical engineering such as biomaterials, biomechanics, or bioinstrumentation. Other areas in BME will be considered. Besides leading and managing the department, the successful candidate will be expected to establish a state-of-the-art research program in collaboration with the Indiana University School of Medicine, also on the IUPUI campus. Major institutes and centers at IUPUI include the Biomechanics and Biomaterials Research Center, the Indiana Center for Musculoskeletal Health, the Simon Cancer Center, the Krannert Institute of Cardiology, and the Stark Neurosciences Institute. A qualified candidate may be offered joint appointments in both engineering and medicine. A more complete description of BME at IUPUI is available at **www.engr.iupui.edu/bme/**

The Purdue School of Engineering and Technology at IUPUI has 7 academic departments with an extensive undergraduate and graduate degree portfolio, including numerous CAC, EAC, and ETAC of ABET accredited B.S programs. The School has over 3,500 students, including approximately 500 graduate students pursuing either M.S. or Ph.D. programs. More information about the School is available at **engr.iupui.edu/**

The IUPUI campus has approximately 30,000 students and over 200 degree programs with annual research expenditures totaling over $300M. IUPUI has received the Higher Education Excellence in Diversity (HEED) Award for five consecutive years. More information about IUPUI is available at **www.iupui.edu**

*Applications must include a statement of interest, curriculum vitae, and contact information for at least three references. Apply for this position at **https://indiana.peopleadmin.com/postings/4229**. Applications are welcome until the position is filled. IUPUI is an Equal Opportunity/Affirmative Action educator and employer and affords reasonable accommodations to persons with disabilities.*

THE ABCs OF TELEGRAPHY

Compared with Morse code, which required specific expertise to send and receive telegrams, Louis-François Breguet's dial telegraph was, on the face of it, a snap. No complicated code of dots and dashes to memorize—just move the pointer to the appropriate letter and send the signal. On the other end, the needle on the receiver [above] would point to that letter. Breguet's 1842 invention drew on his knowledge of watchmaking—the family business—and its clocklike face was easier to read than William Cooke and Charles Wheatstone's needle telegraph of 1837. Although Breguet's design allowed for 25 letters—the letter *W* was left out—and 25 numbers, there was no room for the diacritical marks that are widely used in French and many other languages. Morse code, by contrast, could accommodate special characters, and for this and other reasons, it eventually prevailed. ■

↗ **For more on the history of telegraphy, go to https://spectrum.ieee.org/pastforward0918**

TECHNICAL UNIVERSITY OF MADRID

WANT TO LEAD GROUNDBREAKING RESEARCH IN SINGAPORE?

The National Research Foundation Singapore (NRF) invites outstanding researchers in their early stage of research careers to apply for the NRF Fellowship.

We welcome researchers in the following disciplines of science and technology:

- Engineering
- Computer Science, including Infocomm Technology and Interactive & Digital Media
- Natural or Physical Sciences
- Life Sciences

Each Fellow is provided with a research grant of up to S$3 million. The research grant can be used to cover personnel, equipment and consumables costs. Application is open from December 2018 to March 2019.

Visit www.nrf.gov.sg/NRFfellowship for more details.

NRF Fellowship

NATIONAL RESEARCH FOUNDATION
PRIME MINISTER'S OFFICE
SINGAPORE

SG
singapore
Passion Made Possible