FOR THE TECHNOLOGY INSIDER

NOVEMBER 2022

IEEE Spectrum

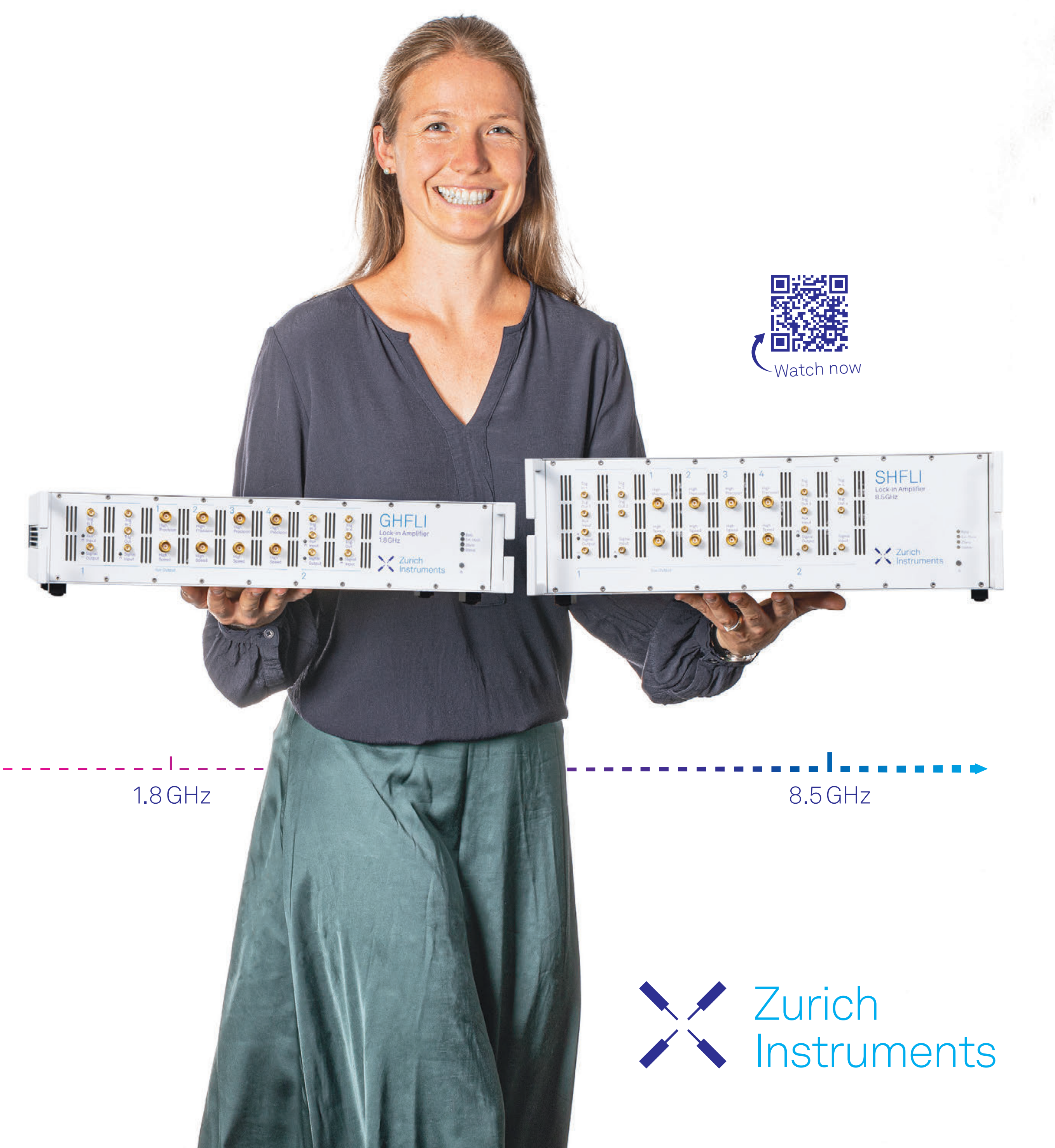
Rewiring the Brain to Smell Again

IEEE

Trailblazers.

New

Meet the Lock-in Amplifiers that measure microwaves.



VOLUME 59 / ISSUE 11 NOVEMBER 2022

Contents

ON THE COVER:
Photo-illustration by Gluekit

COVER PHOTO: TIM MACPHERSON/GETTY IMAGES

Illustration by Edmon de Haro

BY HARRY GOLDSTEIN

Magnus Carlsen of Norway makes a move against Fabiano Caruana of the United States during a game played last January.

AI Helps Humans Level Up

Chess players and programmers now have access to tools that can improve skills and unleash creativity

Back in the mid-1970s, *IEEE Spectrum* senior editor Phil Ross played one of the first chess programs capable of vanquishing humans. He capitulated quickly—too quickly, it turned out: Although the program that beat him was good at openings and the middle game, it was terrible at the end game. Fast forward 50 years and the highest-ranked chess players in the world are AIs, with humans trailing far behind.

As Ross points out in his online piece "AI's Grandmaster Status Overshadows Chess Scandal," the recent brouhaha involving world champion Magnus Carlsen and up-and-comer Hans Niemann highlights how chess-playing AIs, referred to as "engines" by the cognoscenti, have overtaken humans in terms of raw game-playing accuracy. The scandal also shows how AIs are being used by players at all levels to get better faster, fostering a boom in the sport.

If you were born a few decades ago, your best shot at playing and learning from grandmasters was either to be lucky enough to know one or to somehow qualify for the high-level tournaments in which they participated. Nowadays, chess newbies can log in and play engines that far exceed their own abilities, learning strategies and moves in days or weeks that in the past might have taken months or years. Engines can also help neophytes and grandmasters alike analyze their own games to give them an edge against human opponents. In fact, if you burrow down the rabbit holes of chess on YouTube or Twitch, you'll find grandmasters giving move-by-move analyses of games that engines played against *each other.* These AI tools, along with the humans who use them, are readily accessible: You can play the cybernetic versions of super grandmasters, like the open-source, currently top-ranked chess engine Stockfish 14.1, not to mention tens of millions of human opponents on sites like Chess.com, a global community of some 93 million players and a central player in the cheating controversy.

"Expert programmers will still be needed to make breakthrough programming innovations, like the next generation of AI-infused tools or post-AI supertools." —Ben Shneiderman

While it is certainly true that bad actors use AI to cheat at chess—potentially posing an existential threat to the game, as Carlsen has suggested—it is equally true that the chess world has openly embraced AI and has been thriving as a result. Similar risk/reward calculations will need to be made in other domains.

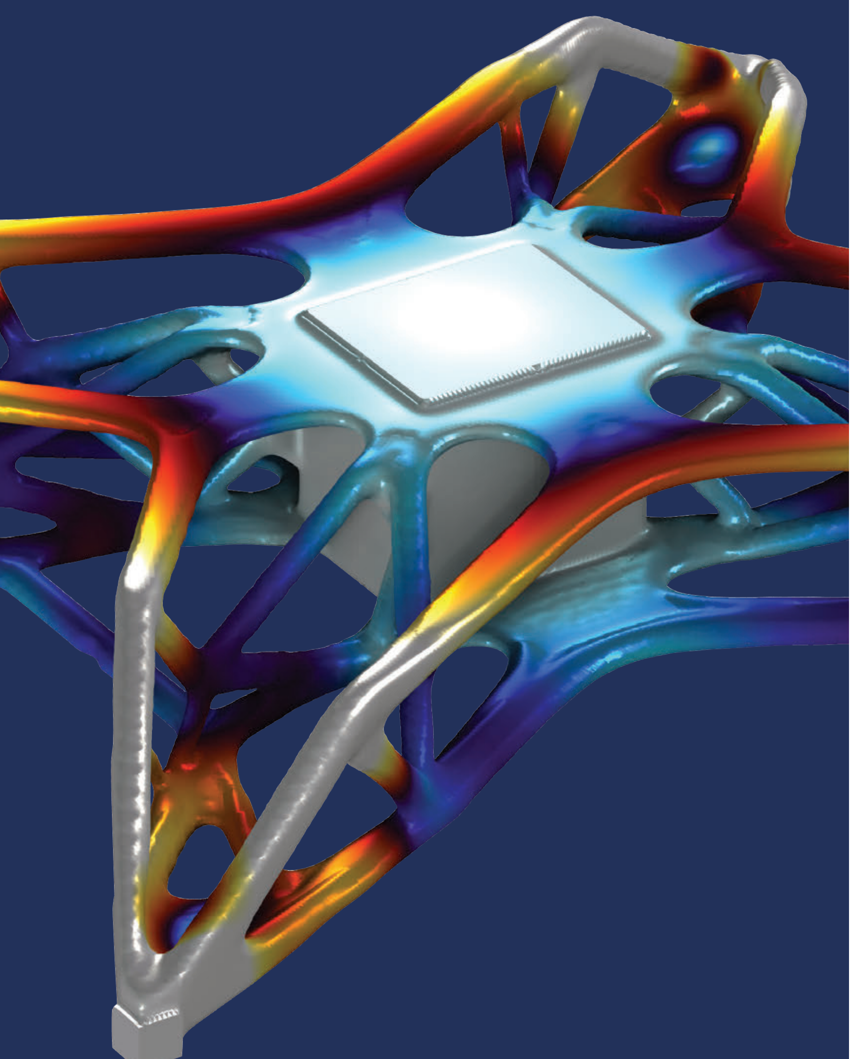
Take software development. On page 5 in this issue, the journalist Craig S. Smith looks at the proliferation of AI-powered code-writing assistants and how they can help programmers compose better programs faster and guide nonprogrammers to instantiate their ideas in software. As for the risk of these AI assistants driving humans out of programming, the experts Smith talked to don't believe that human coders are going to be replaced anytime soon.

Instead, programmers are learning that AI can automate routine tasks such as writing unit tests that verify discrete chunks of code, which can free up time. Amazon's CodeWhisperer, GitHub's Copilot, and Microsoft's TiCoder are all based on large language models trained up on massive code bases. Among other things, these coding assistants can suggest executable instructions and auto-completions for developers as they code.

"Like compilers for high-level languages, code-checkers, and interactive development environments, AI-infused coding tools will speed the programmer's work," says Ben Shneiderman, IEEE Fellow and author of *Human-Centered AI* (Oxford University Press, 2022). "The well-designed user interfaces on these tools lower the threshold of entry but do not raise the ceiling of performance. Expert programmers will still be needed to make breakthrough programming innovations, like the next generation of AI-infused tools or post-AI supertools."

Indeed, while some programmers might be concerned about algorithms edging them out of their jobs, they need only look to chess to see how technology can surpass human abilities and at the same time help people push past their own limits. ■

PORTRAIT BY SERGIO ALBIAC; DEAN MOUHTAROPOULOS/GETTY IMAGES



CONTRIBUTORS

HEATHER L. MACLEAN

MacLean, a professor of chemical engineering and applied chemistry at the University of Toronto, last wrote for us in 2001, in an article on hybrid cars. On page 28, she and her colleagues Alexandre Milovanoff and I. Daniel Posen now turn a critical eye toward pure-electric vehicles. "We still feel that EVs have bright promise,"says MacLean. "It's just that they aren't enough on their own."

MARIOS POULAKIS

After graduating from university, Poulakis spent a year with a team from Greek mobile operator Cosmote, mapping wireless signal strength all over Athens. "This is what triggered my interest in optimization of wireless networks," he explains. Today, as a senior technology planning engineer at Huawei Technologies in Sweden, he is working on one of the most advanced technologies in that field: reconfigurable intelligent surfaces, which he describes on page 40.

YIANNIS PSARAS

Psaras is a research scientist at Protocol Labs, in London. His coauthors, Jorge M. Soares and David Dias, also work at Protocol Labs, where Soares is a technical program manager and Dias is research director. In this issue, they describe a system for peer-to-peer file sharing on the Internet [p. 34]. It's called the InterPlanetary File System because, in theory, it could one day provide an efficient way to pass files between different planets. "For now, though, we're focused on rolling it out for just Earth!" says Psaras.

CRAIG S. SMITH

Smith is a reporter who has covered AI for *The New York Times* and now hosts a podcast on the topic, "Eye on AI." Reviewing the many AI coding applications out there today [p. 5], he says he was struck by how quickly the field is advancing. "In just three years, we've gone from zero to AI writing complete programs," he says. "Imagine what AI will be able to do in the next 10 years."

IEEE Spectrum

ACTING EDITOR IN CHIEF Harry Goldstein, h.goldstein@ieee.org
EXECUTIVE EDITOR Jean Kumagai, j.kumagai@ieee.org
MANAGING EDITOR Elizabeth A. Bretz, e.bretz@ieee.org
CREATIVE DIRECTOR
Mark Montgomery, m.montgomery@ieee.org
PRODUCT MANAGER, DIGITAL Erico Guizzo, e.guizzo@ieee.org
EDITORIAL DIRECTOR, CONTENT DEVELOPMENT
Glenn Zorpette, g.zorpette@ieee.org
SENIOR EDITORS
Evan Ackerman (Digital), ackerman.e@ieee.org
Stephen Cass (Special Projects), cass.s@ieee.org
Samuel K. Moore, s.k.moore@ieee.org
Tekla S. Perry, t.perry@ieee.org
Philip E. Ross, p.ross@ieee.org
David Schneider, d.a.schneider@ieee.org
Eliza Strickland, e.strickland@ieee.org
ART & PRODUCTION
DEPUTY ART DIRECTOR Brandon Palacio, b.palacio@ieee.org
PHOTOGRAPHY DIRECTOR Randi Klett, randi.klett@ieee.org
ONLINE ART DIRECTOR Erik Vrielink, e.vrielink@ieee.org
PRINT PRODUCTION SPECIALIST
Sylvana Meneses, s.meneses@ieee.org
MULTIMEDIA PRODUCTION SPECIALIST
Michael Spector, m.spector@ieee.org
WEB PRODUCTION Michael Novakovic, m.novakovic@ieee.org
NEWS MANAGER Margo Anderson, m.k.anderson@ieee.org
ASSOCIATE EDITORS
Willie D. Jones (Digital), w.jones@ieee.org
Michael Koziol, m.koziol@ieee.org
SENIOR COPY EDITOR Joseph N. Levine, j.levine@ieee.org
COPY EDITOR Michele Kogon, m.kogon@ieee.org
EDITORIAL RESEARCHER Alan Gardner, a.gardner@ieee.org
EDITORIAL INTERN Dina Genkina, d.genkina@ieee.org
CONTRACT SPECIALIST Ramona L. Foster, r.foster@ieee.org
CONTRIBUTING EDITORS Robert N. Charette, Steven Cherry, Charles Q. Choi, Peter Fairley, Edd Gent, W. Wayt Gibbs, Mark Harris, Allison Marsh, Prachi Patel, Julianne Pepitone, Lawrence Ulrich, Emily Waltz
THE INSTITUTE
EDITOR IN CHIEF Kathy Pretz, k.pretz@ieee.org
ASSISTANT EDITOR Joanna Goodrich, j.goodrich@ieee.org
DIRECTOR, PERIODICALS PRODUCTION SERVICES Peter Tuohy
ADVERTISING PRODUCTION MANAGER
Felicia Spagnoli, f.spagnoli@ieee.org
SENIOR ADVERTISING PRODUCTION COORDINATOR
Nicole Evans Gyimah, n.gyimah@ieee.org
ADVERTISING PRODUCTION +1 732 562 6334
EDITOR IN CHIEF EMERITUS Susan Hassler, s.hassler@ieee.org
EDITORIAL ADVISORY BOARD, *IEEE SPECTRUM*
Susan Hassler, *Chair*; Ella M. Atkins, Robert N. Charette, Francis J. Doyle III, Matthew Eisler, Shahin Farshchi, Alissa Fitzgerald, Jonathan Garibaldi, Benjamin Gross, Lawrence O. Hall, Jason K. Hui, Leah H. Jamieson, Mary Lou Jepsen, Michel M. Maharbiz, Somdeb Majumdar, Lisa May, Carmen S. Menoni, Ramune Nagisetty, Paul Nielsen, Sofia Olhede, Christopher Stiller, Wen Tong, Boon-Lock Yeo
EDITORIAL ADVISORY BOARD, *THE INSTITUTE*
Kathy Pretz, *Chair*; Qusi Alqarqaz, Stamatis Dragoumanos, Jonathan Garibaldi, Madeleine Glick, Lawrence O. Hall, Susan Hassler, Francesca Iacopi, Cecilia Metra, Shashi Raj Pandey, John Purvis, Chenyang Xu
MANAGING DIRECTOR, PUBLICATIONS Steven Heffner
DIRECTOR, BUSINESS DEVELOPMENT, MEDIA & ADVERTISING Mark David, m.david@ieee.org
EDITORIAL CORRESPONDENCE
IEEE Spectrum, 3 Park Ave., 17th Floor, New York, NY 10016-5997 **TEL: +1 212 419 7555**
BUREAU Palo Alto, Calif.; Tekla S. Perry **+1 650 752 6661**
ADVERTISING INQUIRIES Naylor Association Solutions, Erik Albin +1 352 333 3371, ealbin@naylor.com
REPRINT SALES +1 212 221 9595, ext. 319

IEEE SPECTRUM (ISSN 0018-9235) is published monthly by The Institute of Electrical and Electronics Engineers, Inc. All rights reserved. © 2022 by The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016-5997, U.S.A. Volume No. 59, Issue No. 11. The editorial content of IEEE Spectrum magazine does not represent official positions of the IEEE or its organizational units. Canadian Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40013087. Return undeliverable Canadian addresses to: Circulation Department, IEEE Spectrum, Box 1051, Fort Erie, ON L2A 6C7. Cable address: ITRIPLEE. Fax: +1 212 419 7570. INTERNET: spectrum@ieee.org. ANNUAL SUBSCRIPTIONS: IEEE Members: $21.40 included in dues. Libraries/institutions: $399. POSTMASTER: Please send address changes to IEEE Spectrum, % Coding Department, IEEE Service Center, 445 Hoes Lane, Box 1331, Piscataway, NJ 08855. Periodicals postage paid at New York, NY, and additional mailing offices. Canadian GST #125634188. Printed at 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 https://www.ieee.org/about/corporate/governance/p9-26.html.

News

CODE EXCERPTS COURTESY DEEPMIND

COMPUTER PROGRAMMING

AI Is Rewriting the Job of Coding › Programming jobs won't disappear, but they'll be increasingly AI-assisted

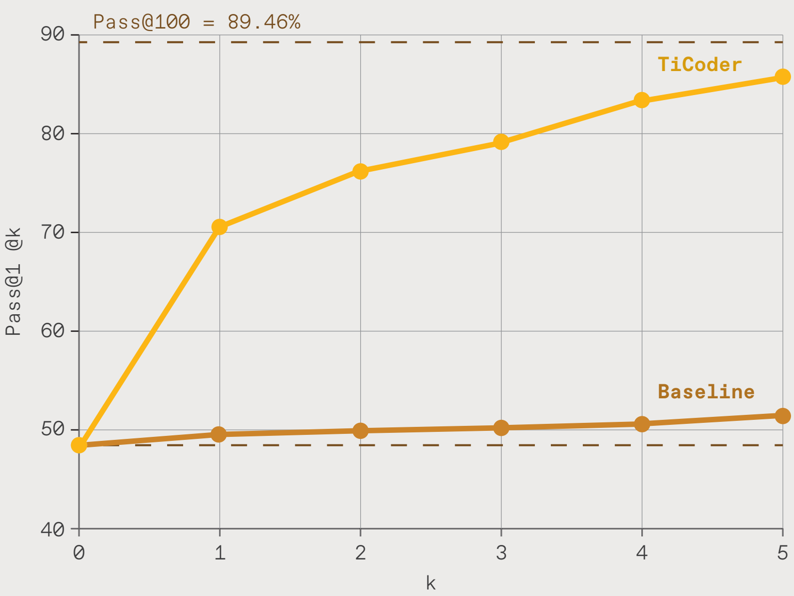
BY CRAIG S. SMITH

Are coders doomed? That question has been bouncing around computer-programming communities ever since OpenAI's large language model, GPT-3, surprised everyone with its ability to create HTML websites from simple written instructions.

In the months since, rapid-fire advances have led to systems that can write complete, albeit simple, computer programs from natural-language descriptions—spoken or written human language—and automated coding assistants that speed the work of computer programmers. How far will artificial intelligence go in replacing or augmenting the work of human coders?

According to the experts *IEEE Spectrum* consulted, the bad news is coding as we know it today may indeed be doomed. But the good news is that computer programming and software development appear poised to remain a very human endeavor for the foreseeable future. In the meantime, AI-powered automated code generation will increasingly speed

Illustration by Edmon de Haro



This graph details coding-accuracy improvements enabled by the program TiCoder. By itself, an AI-coding app reflects the human programmer's intentions around only half the time [dark yellow line]. But with TiCoder improving the system's capabilities [yellow line], code is generated that accurately reflects the human programmer's intentions between 70 and 85 percent of the time.

SOURCE: "INTERACTIVE CODE GENERATION VIA TEST-DRIVEN USER-INTENT FORMALIZATION," LAHIRI ET AL., ARXIV, AUGUST 2022

up software development by allowing more code to be written in less time.

According to these experts, if someone wants to become a software developer, it won't always be the case that they'll need to learn a programming language. The time frame for this switchover to natural-language programming is still an open-ended question. But expect important developments in this direction to be measured in years, not decades. Instead of learning C++ or Python or Ruby, these future coders will instead need to understand the semantics, concepts, and logical sequences of building a computer program. And that should open software development to a wider and more diverse population.

"I don't believe AI is anywhere near replacing human developers," said Vasi Philomin, Amazon's vice president and general manager for machine learning and AI, adding that AI tools will free coders from routine tasks, but the creative work of computer programming will remain.

When the programming of electronic computers began in the 1940s, programmers wrote in numerical machine code. It wasn't until the mid-1950s that Grace Hopper and her team at the computer company Remington Rand developed Flow-Matic, which allowed programmers to use a limited English vocabulary to write programs.

Since then, programming has climbed a ladder of increasingly efficient languages that allow programmers to be more productive.

AI-written code is the cutting edge of a broader movement to allow people to write software without having to code at all. Already, with platforms like Akkio, people can build machine-learning models with simple drag, drop, and button-click features. Users of Microsoft's Power Platform, which includes a family of low-code products, can generate simple applications by just describing them.

In June, Amazon released CodeWhisperer, a coding assistant for programmers, like GitHub's Copilot, which was first released in limited preview in June 2021. Both tools are based on large language models (LLMs) that have been trained on massive code repositories. Both offer auto-complete suggestions as a programmer writes code or suggest executable instructions from simple natural-language phrases.

But to move beyond auto-completion, the problem is teaching the intent to the computer. Software requirements themselves are not always hallmarks of exactitude, while natural language is itself notoriously imprecise.

"To resolve all these ambiguities in English written specification, there needs to be some incremental refinement, some conversation between the human and the machine," said Peter Schrammel, cofounder of Diffblue, which automates the writing of unit tests for Java.

To address these problems, researchers at Microsoft have recently proposed adding a feedback mechanism to LLM-based code generation so that the computer asks the programmer for clarification of any ambiguities before generating code.

The interactive system, called TiCoder, attempts to use iterative feedback to divine the programmer's algorithmic intent and then generate code that is consistent with the expressed intentions.

According to the researchers' paper, TiCoder improves the accuracy of automatically generated code to as much as 85 percent from 48 percent, when evaluated on the Mostly Basic Programming Problems benchmark. MBPP, meant to evaluate machine-generated code, consists of Python programming problems, designed to be solvable by entry-level programmers.

A unit of code, which can be hundreds of lines long, is the smallest part of a program that can be maintained and executed independently. A suite of unit tests, typically consisting of dozens of unit tests, checks that the unit executes as intended, so that when you stack the units together, the program works.

According to a survey by Diffblue, developers spend roughly 35 percent of their time writing quality-control tests (as opposed to writing code destined for production use), so there are significant productivity gains to be made just by automating a part of this.

Unit tests are useful for debugging individual functions and for detecting errors when code is manually changed. But a unit test can also be used as the specification for a unit of code and to guide programmers. While not many programmers pursue true test-driven development, in which the unit tests are written first, unit tests and units are sometimes written together.

GitHub's Copilot, Amazon CodeWhisperer, and AI-programming-assistant packages can be used as interactive auto-completion tools for writing unit tests. The programmer is given suggestions and picks the one they think will work best. Diffblue's system, called Diffblue Cover, uses reinforcement learning to write unit tests automatically, with no human intervention.

Earlier this year, Google's artificial-intelligence lab, DeepMind, based in the United Kingdom, went further in fully automatic code generation with AlphaCode, a large language model that can write simple computer programs from natural-language instructions.

The model was first trained on GitHub's online code repository until the model was able to produce reasonable-looking code. To fine-tune the model, DeepMind used 15,000 pairs of natural-language problem descriptions and successful code solutions from past coding competitions to create a specialized data set of input-output examples.

The final step was to generate many solutions and then use a filtering algorithm to select the best one. "We created many different program possibilities by essentially sampling the language model almost a million times," said Oriol Vinyals, who leads DeepMind's deep-learning team.

To optimize the sample-selection process, Vinyals says, DeepMind uses a clustering algorithm to group the working solutions together, to find candidates that are likely to work as well as those written by human programmers.

To test the system, DeepMind submitted 10 AlphaCode-written programs to a human coding competition on the popular Codeforces platform, where its solutions ranked among the top 54 percent.

"To generate a program, will you just write it in natural language, no coding required, and then the solution comes out at the other end?" Vinyals asked rhetorically in a recent interview. "I believe so."

Vinyals and others caution that it will take time, possibly decades, to reach that goal. "We are still very far away from when a person would be able to tell a computer about the requirements for an arbitrary complex computer program, and have that automatically get coded," said Andrew Ng, a founder and CEO of Landing AI who is an AI pioneer and one of the founders of Google Brain.

But given the speed at which AI-code generation has advanced in a few short years, it seems inevitable that AI systems will eventually be able to write code from natural-language instructions. Hand-coding software programs will increasingly be like hand-knitting sweaters.

To give natural-language instructions to a computer, developers will still need to understand some concepts of logic and functions and how to structure things. They will still need to study the foundations of programming, even if they don't learn specific programming languages or write in computer code. That will, in turn, enable a wider range of programmers to create more and more varied kinds of software.

"I don't believe AI is anywhere near replacing human developers," Amazon's Philomin said. "It will remove the mundane, boilerplate stuff that people have to do, and they can focus on higher-value things."

Diffblue's Schrammel agrees that AI-automated code generation will allow software developers to focus on more difficult and creative tasks. But, he adds, there will need to be at least one interaction with a human to confirm that what the machine has understood is what the human intended.

"Software developers will not lose their jobs because an automation tool replaces them," he said. "There always will be more software that needs to be written." ■

››COMPUTING

"Quantum-Safe" Crypto Cipher Hacked by a 10-Year-Old PC › Supposedly impregnable algorithm felled in 4 minutes

BY CHARLES Q. CHOI

Future quantum computers may rapidly break modern cryptography. Now, researchers find that a promising algorithm designed to protect computers from these advanced attacks could be broken in just 4 minutes. Surprisingly, that 4-minute break-in was not achieved by a cutting-edge machine but by a conventional 10-year-old desktop computer. This latest, unexpected defeat highlights the many hurdles postquantum cryptography will need to clear before adoption, researchers say.

In theory, quantum computers can quickly solve problems that might take classical computers eons to solve. That difference serves as the foundation upon which much of modern cryptography rests. The strength of cryptography's encryption schemes comes from the extreme difficulty that classical computers face when it comes to mathematical problems such as factoring huge numbers. However, quantum computers can, in principle, run algorithms that can rapidly crack such encryption.

To stay ahead of this quantum threat, cryptographers around the world have spent the past two decades designing postquantum cryptography (PQC) algorithms. These are based on new mathematical problems that both quantum and classical computers find difficult to solve.

For years, researchers at organizations such as the National Institute of Standards and Technology (NIST) have been investigating which PQC algorithms should become the new standards the world should adopt. NIST announced it was seeking candidate PQC algorithms in 2016 and received 82 submissions in 2017. In July 2022, after three rounds of review, NIST announced four algorithms that would become standards, and four more would enter another round of review as possible additional contenders.

Now, a new study has revealed a way to completely break SIKE, one of these contenders under review (which Amazon, Cloudflare, Microsoft, and others have investigated). "The attack came from out of the blue, and was a silver bullet," says Christopher Peikert, a cryptographer at the University of Michigan at Ann Arbor who did not take part in this new work.

SIKE (supersingular isogeny key encapsulation) is a family of PQC algorithms involving elliptic curves. "Elliptic curves have long been studied in mathematics," says NIST mathematician Dustin Moody, who was not part of the study. "They are described by an equation in the form of $y^2 = x^3 + Ax + B$, where A and B are numbers. So, for example, an elliptic curve could be $y^2 = x^3 + 3x + 2$."

In 1985, Moody says, "mathematicians figured out a way to make cryptosystems involving elliptic curves, and these systems have been widely deployed. However, these elliptic curve cryptosystems turn out to be vulnerable to attacks from a quantum computer." Around 2010, researchers found a new way to use elliptic curves in cryptography. "It was believed that this new idea wasn't susceptible to attacks from quantum computers," he says.

This new approach is based on how two points can be added on an elliptic curve to derive another point on the elliptic curve, Moody says. An "isogeny" is a map from one elliptic curve to another elliptic curve that preserves this addition law.

"If you make this map complex enough—the conjectured hard problem, which allows encryption of data, is given two elliptic curves—it's hard to find an isogeny between them," says Thomas Decru, a mathematical cryptographer at KU Leuven in Belgium who coauthored the paper describing the study that broke SIKE.

SIKE is a form of isogeny-based cryptography based on the supersingular isogeny Diffie-Hellman (SIDH) key exchange protocol. "SIDH/SIKE was one of the first practical isogeny-based cryptographic protocols," Decru says.

However, one of SIKE's vulnerabilities was that in order for it to work, it needed to provide extra information to the public known as auxiliary torsion points. "Attackers have tried to exploit this extra information for a while but had not been successful in using it to break SIKE," Moody says. "However, this new paper [describes] a way to do it, using some pretty advanced mathematics."

To explain this new attack, Decru says that although elliptic curves are two-dimensional objects, in mathematics elliptic curves can be visualized as objects of any number of dimensions. One can also create isogenies between these generalized objects.

By applying a 25-year-old theorem, the new attack uses the extra information that SIKE makes public to construct an isogeny in two dimensions. This isogeny can then reconstruct the secret key that SIKE uses to encrypt a message. Decru and Wouter Castryck, the team's lead researcher, detailed their findings on 5 August in the Cryptology ePrint Archive.

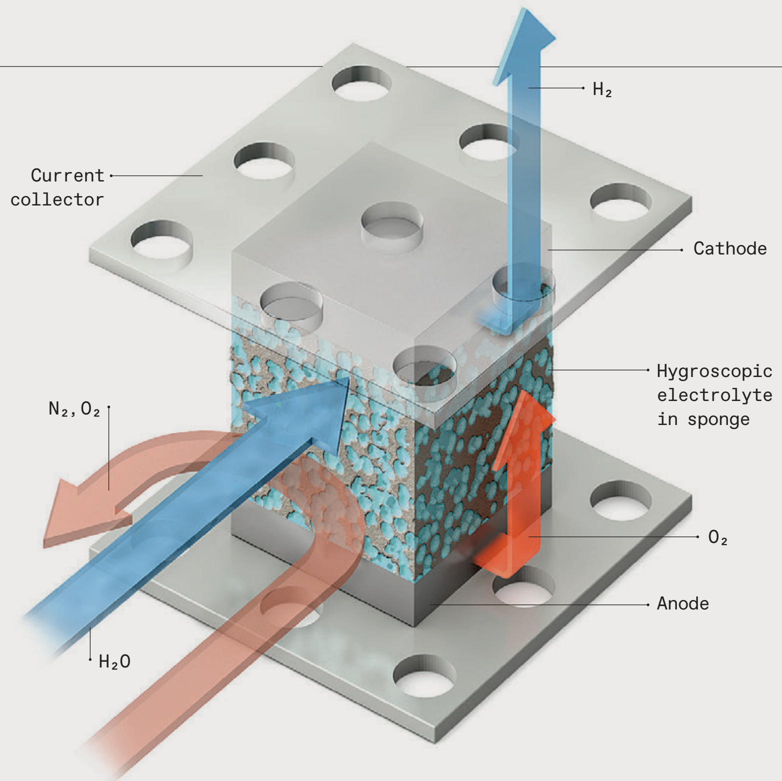
"To me what is most surprising is that the attack seemingly came out of nowhere," says cryptographer Jonathan Katz at the University of Maryland at College Park, who did not take part in this new work. "There were very few prior results showing any weaknesses in SIKE, and then suddenly this result appeared with a completely devastating attack—namely, it finds the entire secret key, and does so relatively quickly without any quantum computation."

Using an algorithm based on this new attack, a 10-year-old Intel desktop took 4 minutes to find a secret key secured by SIKE.

"Attacks on SIDH/SIKE went from essentially no progress for 11 to 12 years, since SIDH was first proposed, to a total break," says Peikert.

One reason SIKE's vulnerability had not been detected until now is that the new attack "applies very advanced mathematics—I can't think of another situation where an attack has used such deep mathematics compared with the system being broken," says mathematician Steven Galbraith at the University of Auckland, in New Zealand, who was not part of this research. Katz agrees, saying, "I suspect that fewer than 50 people in the world understand both the underlying mathematics and the necessary cryptography." ■

Illustration by Dan Page



This schematic of the direct-air electrolyzer demonstrates the basics of the process: moist air and electricity in and, after the water in the air is split, hydrogen and oxygen out.

ENERGY

Pulling Fuel Out of Thin Air › Direct-air electrolyzers can produce hydrogen even in arid regions

BY PRACHI PATEL

The most sustainable way to make hydrogen fuel is to split water using renewable electricity. But heretofore that has required access to freshwater—a resource already in limited supply. Now, researchers have reported a way to make hydrogen fuel from just humidity in the air.

Their electrolyzer extracts moisture from air and splits it via renewably powered electrolysis to create hydrogen. It is the first such electrolyzer to produce high purity (99 percent) hydrogen from air that has as little as 4 percent humidity, says Gang Kevin Li, a professor of chemical engineering at the University of Melbourne, in Australia. The Australian team's breakthrough could open up the possibility of producing hydrogen in semi-arid regions, which also have some of the highest solar- and wind-power potential.

Tests of the prototype direct-air electrolyzer over 12 consecutive days showed that it could produce almost 750 liters of hydrogen a day on average per square meter of electrolyzer. Li and his colleagues reported the details in the journal *Nature Communications*.

Hydrogen offers the prospect of clean, emission-free energy, and the hydrogen economy has gathered steam in the past few years due to increases in funding and improvements in technology. But most of the hydrogen around the world today is still produced from natural gas or coal. Green hydrogen from electrolysis is still a nascent technology because of the need for electrolyzers on a large scale.

Many teams are working on alternative ways to make green hydrogen. Solar-powered water-splitting devices, for example, use photocatalysts, which absorb sunlight to split water into hydrogen and oxygen. But these have low solar-to-hydrogen efficiency—commonly no more than 1 percent. To overcome the need for freshwater, there have been attempts to produce hydrogen from saline and brackish waters, but the devices then must deal with contamination and chlorine as a by-product.

Li and his colleagues decided to use moisture in the air as the water source. Globally, there are nearly 13 trillion tonnes of water in the air at any moment, they say, and even dry environments such as the expansive Sahel region in Africa have an average relative humidity of 20 percent.

To tap into that humidity, the researchers soaked a sponge or foam with a water-absorbing electrolyte liquid and sandwiched it between two electrodes. "Water extracted by the electrolyte is spontaneously transported to the electrodes by capillary force and electrolyzed into hydrogen at the cathode and oxygen at the anode," Li explains. "The whole process is passive, and no moving parts or mechanics are involved."

The researchers demonstrated the use of solar panels or a small wind turbine to power the module. They tested the prototype both indoors and outdoors in the hot, dry Melbourne summer. The solar-to-hydrogen efficiency of the device is over 15 percent, Li says.

For the outdoor tests, they connected five electrolyzers in parallel, which, powered by the sun, produced 745 L of hydrogen per square meter a day, enough fuel to heat a home. They also let the prototype run by itself for eight months to showcase its durability.

The prototype is only a few square centimeters in area right now. But over the next year, with funding from investors, the team plans to make larger electrolyzers, with electrode areas of 10 square meters, Li says. They are also improving the electrolyte recipe to further bolster energy efficiency and output, he says.

Both efficiency and output should not be affected when the device is scaled up. But the main challenge the team faces is to find the right materials for the electrolyzer, Li says. "How can we make it cheaper and better?" ■

Illustration by Emily Cooper

TRANSPORTATION

Blockchain Ride-Hailing App Takes on Uber in India › The Drife app already has 10,000 drivers in Bangalore

BY EDD GENT

Ride-hailing in India is dominated by Uber and local rival Ola, but startup Drife thinks blockchain technology could be the key to breaking up the duopoly.

CEO and cofounder Firdosh Sheikh used to be a power user of Uber, racking up more than 5,000 rides in total. But after speaking to drivers, she discovered that many were unhappy with the commissions the platform charged and the often-opaque rules that governed which rides they get assigned. Riders also get a raw deal, she says, with little control over what they pay for their rides or whom they travel with.

The problem, Sheikh decided, boiled down to having a middleman controlling the relationship between riders and drivers. So she set out to build a decentralized ride-hailing platform that puts control back into the hands of users via blockchain technology. Last November, after three years of development, she and her cofounders launched the Drife app in the southern city of Bangalore. Today, they have more than 10,000 drivers and 100,000 riders signed up on the platform.

One of the main things that differentiates Drife from companies like Uber, says Sheikh, is that it doesn't charge any commission. Drivers get to keep the entire fare and will instead pay a monthly subscription to use the platform, although the company is currently waiving this to encourage sign-ups.

"We don't charge anything from the fare that you pay as a rider," she says. "A centralized entity who has a profit motive in the fare that I pay as a rider will start manipulating the fare and exploiting it for their own profit motive, and that's where both drivers and riders struggle."

A base fare is set based on the class of vehicle and the distance of the trip, but after that riders can boost the amount they are willing to pay to attract more drivers. Drivers can also make counteroffers. The rider then chooses who to go with based either on price or driver ratings. This means pricing is purely market driven, and because Drife isn't taking a cut, fares should be cheaper than alternatives, says Sheikh.

The company also has its own cryptocurrency, called DRF. At present it can't be used for much, but the token will play an important role in Drife's expansion plans, says Sheikh. The company plans to operate on a franchise model, with local entrepreneurs bidding to run Drife operations in new cities in return for a share of subscription fees. But to apply for a franchise, they will need to purchase a large chunk of DRF tokens and lock them up for the duration of their contract. So far, the company has received about 60 franchise requests for cities across the globe, says Sheikh.

About 30 percent of DRF tokens have also been reserved for an "ecosystem fund," which will be used for incentives and rewards for drivers and riders. Besides being tradable for real money, these tokens will also confer the right to vote in a decentralized autonomous organization (DAO)—a kind of member-run organization whose internal rules are encoded into a blockchain. Each city will have its own DAO, which will be responsible for choosing franchisees.

"Nothing we see today can work at the scale that we want to grow," says Sheikh. "That's why we've started our own side project where we're working on a blockchain customized to our own needs."

DRIFE

Drife's cofounders are chief operating officer Surya Ranjith [left], chief executive officer Firdosh Sheikh [middle], and chief technology officer Mudit Marda [right].

Crucially, users won't need to take Drife's word for any of this, as the entire system will be governed by smart contracts. These are software programs that live inside a blockchain and automate transactions according to predefined rules that are visible to everyone. "Nobody's going to trust me if I say I don't manipulate it unless I show them that I don't have any power to manipulate that data," says Sheikh. "That's only possible through blockchain."

In reality, however, Drife is building the plane while flying it. The firm's smart-contract system is built on Polygon, which is an extension of the Ethereum blockchain. But because the app's creators are constantly tweaking features and functionality, Sheikh admits that most of the time, operations are actually running on a back-end server that mimics the processing the blockchain is supposed to do.

And many of the details about how the platform will work still need to be ironed out. How to ensure the DAO voting system isn't dominated by those who hold the most tokens is a work in progress, says Sheikh. Most blockchains are also geared toward settling financial transactions, and it's not clear if they can cope with the volume of real-world data involved in running a large-scale ride-hailing business.

It might still be some time before that becomes a problem. Despite a promising number of sign-ups, the company is currently doing only around 7,000 rides a week, compared with the millions done by Uber and Ola every day.

Mobility consultant Vinay Piparsania, founder of MillenStrat Advisory & Research, says that blockchain technology holds considerable promise for disrupting the ride-hailing industry and disrupting the market positions of the big players. The biggest challenge for startups like Drife, though, is matching the incumbents' financial and operational capabilities.

"Unfortunately, at this time, such driver-focused startups are much too small and fragmented to make the difference to the duopolic might of Uber and Ola," he says. It's a David and Goliath situation, Piparsania adds, so for the time being, these companies should focus on "nibbling away in some key towns and categories by attracting and holding onto drivers, and actually demonstrating to riders that they can compete on delivery." ■

JOURNAL WATCH

This Drone Transforms Itself on the Fly

A couple of years ago, we wrote about Dragon (short for "Dual-rotor embedded multilink Robot with the Ability of multi-deGree-of-freedom aerial transformatiON"). The aerial robot consists of four pairs of gimballed, ducted fans, with each pair linked together through a two-axis actuated joint. This arrangement makes it physically flexible in flight to a crazy degree.

Dragon has more degrees of freedom than it knows what to do with. In other words, the hardware necessary for it to pull off an array of tricks is all there. The hard part will be getting it to use that hardware to do things that are truly useful, and reliably so. In 2018, Dragon was just learning how to shape-shift to fit through small spaces. Now it's able to adapt its entire structure

in order to grasp and manipulate objects.

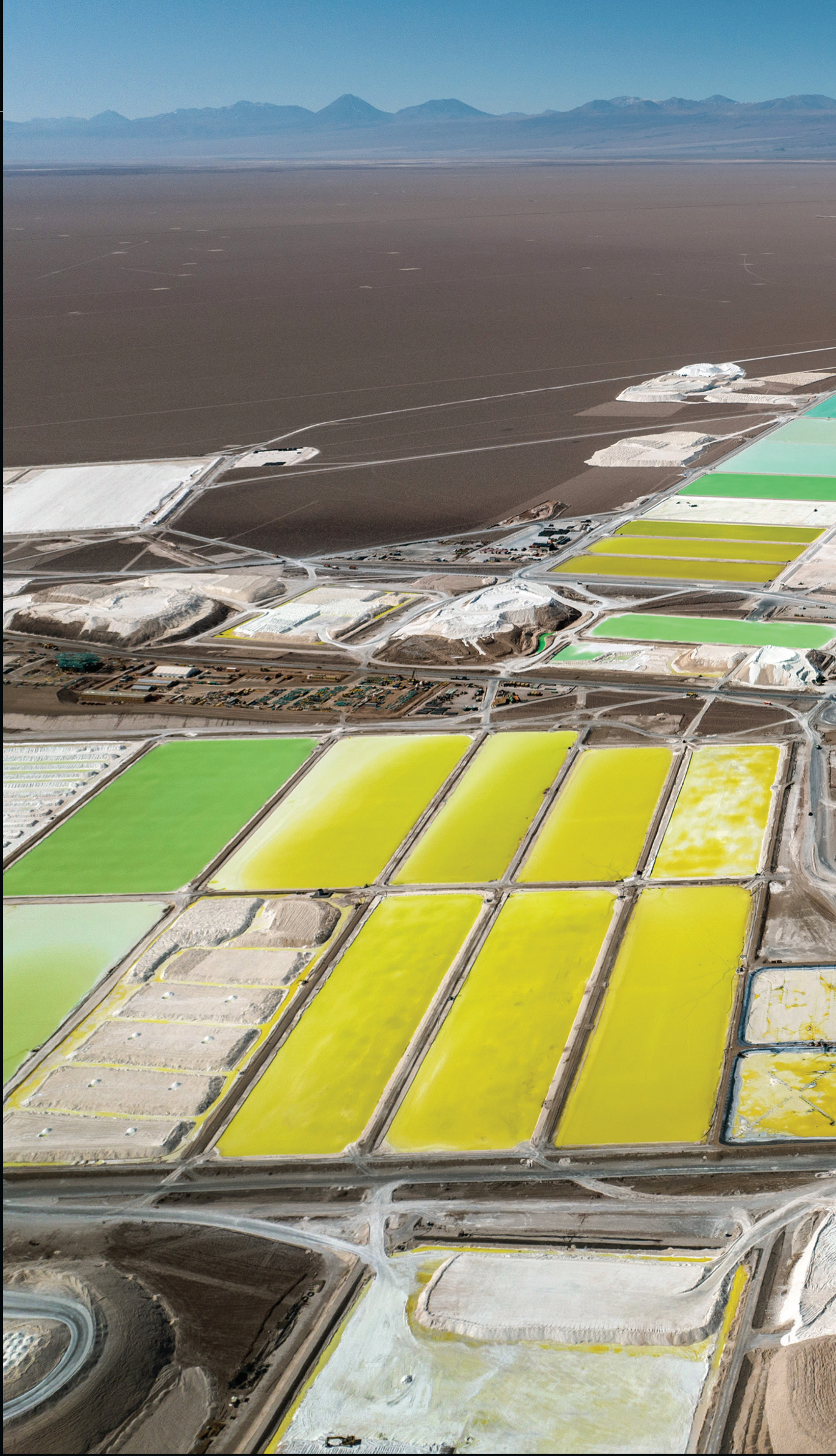
In a couple of recent papers in the *International Journal of Robotics Research* and *IEEE Robotics and Automation Letters*, Moju Zhao and colleagues from the University of Tokyo report some substantially updated capabilities for Dragon. It's much more stable now, they say. The team put the flying mechanical beast to a series of tests, including turning some real industrial valves. They note that the force for the valve turning comes from propeller thrust, not power sent through tacked-on actuators.

This thing is, technically, a mobile manipulator—just not in a form factor we're used to associating with that term, because it's continuously airborne. But it's easy to envision all the jobs Dragon could tackle that a ground-based manipulator simply could not. And making the structure of the drone itself the manipulator is a much more elegant solution than larding on attachments. Or, rather, a potentially elegant solution, considering that Dragon is obviously still very much a research project.

The drone weighs a hefty 7.4 kilograms, and while its payload is a respectable 3.4 kg, the maximum flight time of 3 minutes is a constraint that will need to be solved before it makes it very far beyond the lab environment. To be fair, though, the present focus of the research is about expanding Dragon's capabilities, which is a control problem.

Nevertheless, Zhao reports, the engineers are already considering giving Dragon the ability to walk on the ground in a bid to extend its battery life.

— Evan Ackerman



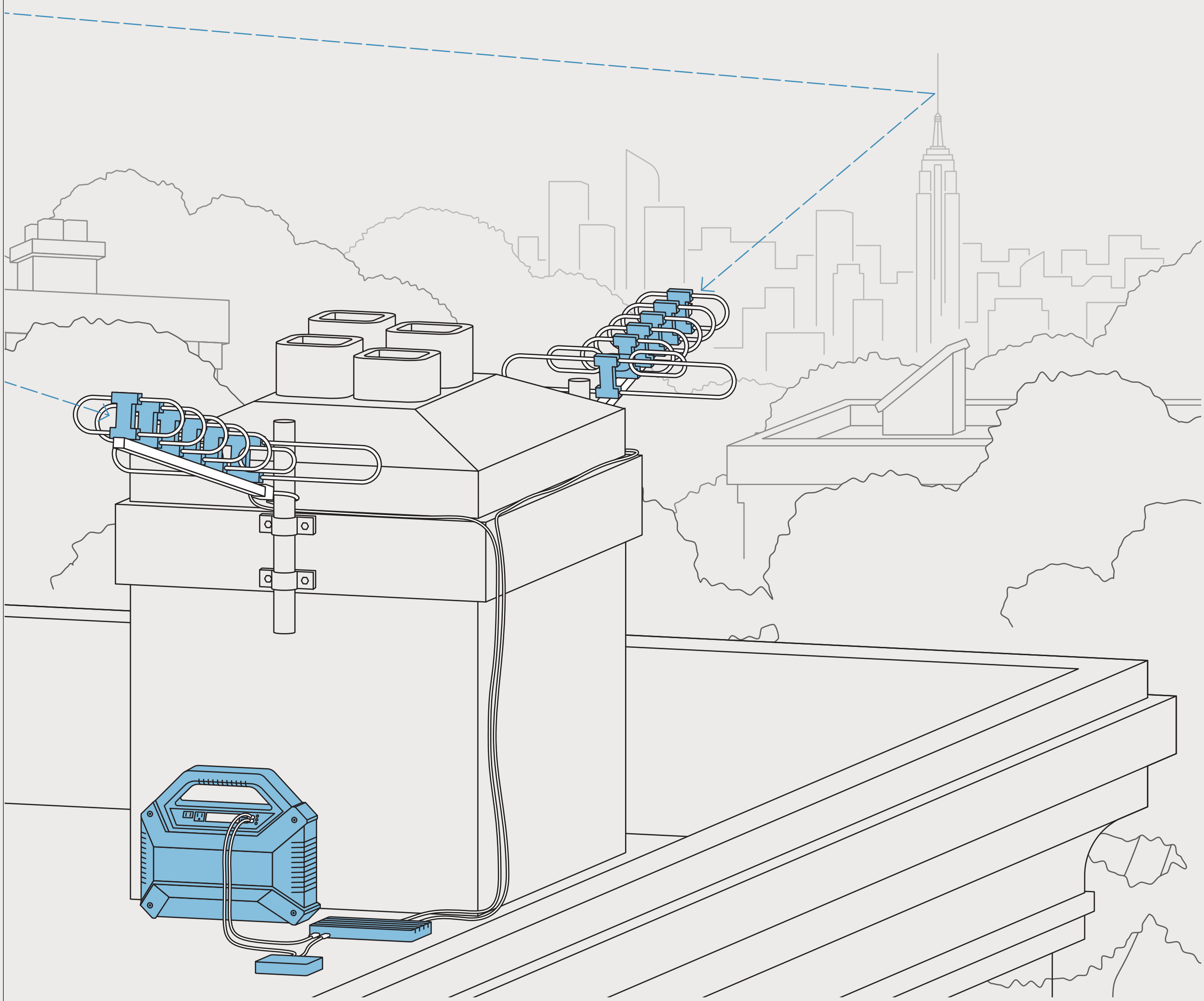
A Rich Harvest in the Desert

By Willie D. Jones

The rise in global demand for lithium—spurred mainly by the manufacturing of lithium-ion batteries for electric vehicles—has triggered expanded mining of the world's known major caches of lithium. One of the richest deposits of lithium ore is in Chile's Atacama Desert. The area, said to be the driest region on earth, has long been a premier source of sodium nitrate, or saltpeter. To extract lithium, miners pump a naturally occurring brine from beneath Atacama's salt flats into giant evaporation pools on the surface. Initially, the liquid is a dark blue color. But over the course of 18 months, the sun and the arid desert air strip away a great deal of the brine's moisture. Left behind are pools of a bright yellow slurry with a roughly 6 percent lithium content. A nearby chemical plant processes the slurry into lithium carbonate powder, well suited to making batteries for automotive propulsion.

PHOTOGRAPH BY JOHN MOORE/GETTY IMAGES

Hands On



Transmissions from a broadcast tower, such as the spire on top of the Empire State Building, can be used with cheap TV antennas and a software-designed radio to track the movements of airplanes.

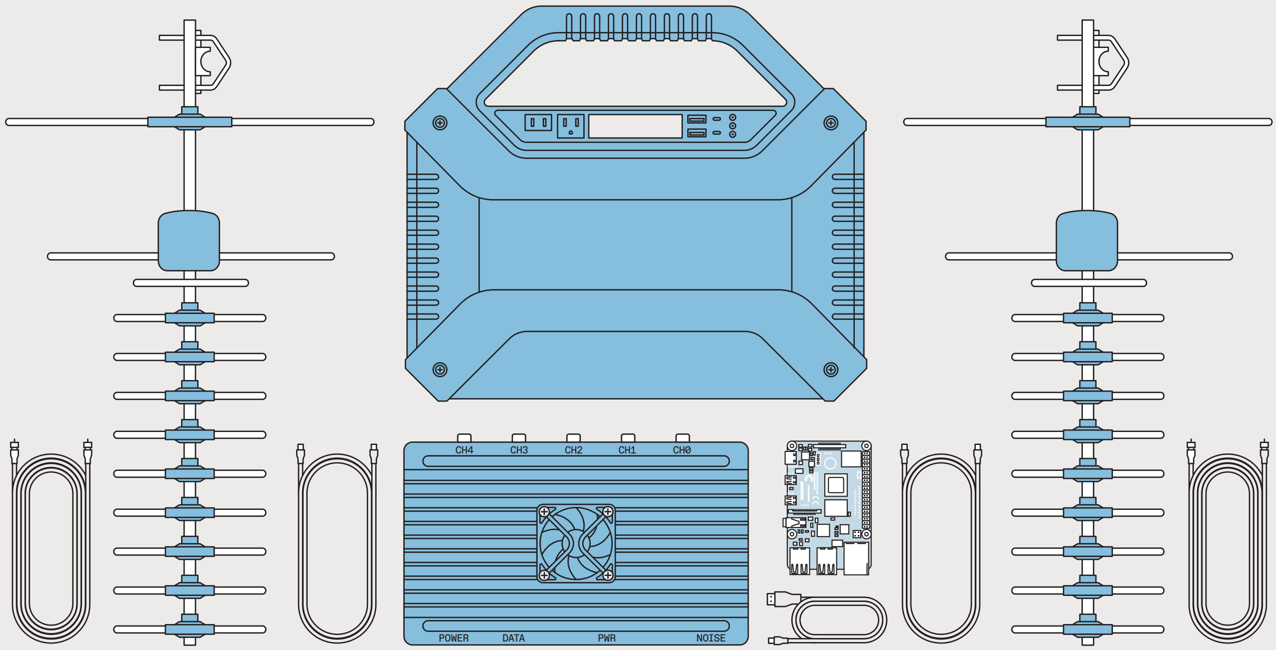
Passive Radar With the KrakenSDR › Spot stuff with TV antennas and a software-defined radio

BY STEPHEN CASS

Normally, when it comes to radio-related projects, my home of New York City is a *terrible* place to be. If we could see and hear radio waves, it would make an EDM rave feel like a sensory deprivation tank. Radio interference plagues the metropolis. But for once, I realized I could use this kaleidoscope of electromagnetism to my advantage—with a passive radar station.

Unlike conventional radar, passive radar doesn't send out pulses of its own

Illustrations by James Provost



Both the KrakenRF SDR and the Raspberry Pi 4 [middle bottom] require a fair amount of power via USB C cables, so a battery pack [top middle] is needed for mobile operation. The Pi is connected to the SDR via a data link, and in turn the SDR is connected via coaxial cables to two directional TV antennas [left and right].

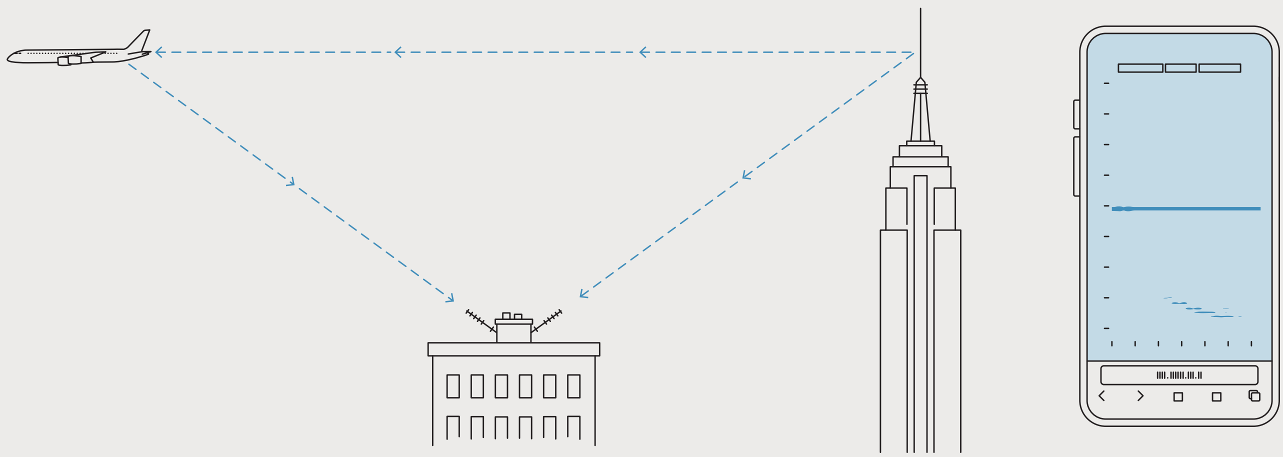
and watch for reflections. Instead, it uses ambient signals. A reference antenna picks up a signal from, say, a cell tower, while a surveillance antenna is tuned to the same frequency. The reference and surveillance signals are compared. If a reflection from an object is detected, then the time it took to arrive at the surveillance antenna gives a range. Frequency shifts indicate the object's speed via the Doppler effect.

I was interested in passive radar because I wanted to put a new software-defined radio (SDR) through its paces. I've checked in with amateur SDR developments for *IEEE Spectrum* since 2006, when SDR become something remotely within a maker's budget. The biggest leap forward happened in 2012 when it was discovered that USB stick TV tuners using the RTL2832U demodulator chip could be tapped to make very cheap but effective SDR receivers. An explosion of interest in SDRs followed. Building off the demand stimulated by this activity, a number of manufacturers have started making premium, but still relatively cheap, SDRs. This includes RTLx-based USB sticks built with better supporting components and designs versus the original TV tuners, and completely new receivers such as the RSPDx. Some of these new SDRs can transmit as well as receive, such as the HackRF One or Lime Mini.

I was researching diving back into SDR with one of these devices when I spotted the CrowdSupply campaign for the US $399 KrakenSDR. It's receive only, but it boasts not one or two tuners, but *five*! The tuners are based on the RTL R820T2/R860 chip, and they are combined with hardware that can automatically do coherence synchronization among them.

What that means is that, for example, you can arrange five omnidirectional antennas in a circle, and do radio direction finding by looking at when a transmission arrives at each antenna. Normally, an amateur looking to do direction finding would have to wave around a directional antenna, something difficult to do while, for example, driving a car.

But it was the KrakenSDR's ability to do passive radar that really caught my eye as a new capability in lowish-cost radio tech , so I plonked down the money. The next step was to get suitable antennas. The radio's manufacturer, KrakenRF, recommends directional Yagi TV antennas for two reasons. First, while the KrakenSDR can work with many signals, including



Comparing the time between the arrival of a signal from a broadcast transmitter and the arrival of a reflection of that signal lets you detect objects such as airplanes and estimate their range. Frequency shifts between the two signals allow you to plot the speed of the object away or toward the antennas along with the range. The trace on the right shows a plane moving away as it increases its speed.

FM radio or cell-tower transmissions, digital TV signals are best to work with because they are fairly evenly distributed across the channel's broadcast band, unlike the narrower and more variable signals from an FM station. (KrakenRF notes that if you *must* use an FM signal, pick a heavy-metal station "since heavy metal is closer to white noise.") The second reason is that pointing a directional antenna away from the reference source means that it's less likely to be swamped by the reference signal.

I ordered two small and light $19 TV antennas. Portability was important because I needed to carry my entire setup to and from my apartment building's roof, where my particular location in an outer borough of the city provided more advantages. First, the sky above has a regular supply of aircraft landing and taking off from NYC's airports—and large metal assemblies moving against an empty background are perfect radar test objects. Second, my roof has a line of sight to the Empire State Building, giving me the ability to choose as a reference signal any one of more than half a dozen TV channels transmitted from its spire.

I deployed my rig: a heavy-duty battery pack, the KrakenSDR, cables, and antennas, along with a Raspberry Pi 4 to process data from the SDR. KrakenRF offers an SD card image for the Pi that bundles an operating system configured to work with its preinstalled open-source software. It also sets up the Pi as a Wi-Fi access point with a Web interface. I really wish more companies would adopt this approach, as installing open-source software is often a frustrating exercise in trying to replicate the precise system environment it was developed in. Even if you want to ultimately install the KrakenSDR software somewhere other than a Pi, having a known-good setup is useful as a reference, and allows you to test the hardware.

I pointed the reference antenna toward the Empire State Building and retreated with the surveillance antenna behind the superstructure of my building's stairwell. This was in a bid to shield the antenna from the reference signal and myself from the wind. Checking the feed from the antennas using the Web interface's built-in spectrum analyzer, I discovered I was almost *too* successful in choosing the Empire State's transmitter tower as a source of radio illumination: The reference signal was saturating the receiver with the default gain setting of 27 decibels, so I dropped it down to just 2.7 dB.

But intense illumination means bright reflections. With one hand I pointed the surveillance antenna at the overcast skies and held my phone in the other. Gratifyingly, I almost instantly started seeing a blip on the speed-versus-range radar plot, matched a few moments later by the rumble of an approaching jet. (The plot updates about once every 3 seconds.) Because of the strength of the echoes, I was able to raise the signal-cutoff threshold significantly, giving me radar returns uncluttered with noise, and often with multiple aircraft. A win for SDR!

Admittedly, my passive radar setup doesn't have much everyday value. But as a demonstration of how far and fast inexpensive SDR technology is advancing, it's a clear signal. ■

I was almost *too* successful in choosing the Empire State's transmitter tower as a source of radio illumination.

Careers

Computer pioneer Craig Partridge is chair of the department of computer science at Colorado State University.

COLORADO STATE UNIVERSITY

Craig Partridge › In the midst of Internet traffic for 40 years and loving It

BY DANIEL P. DERN

For computer pioneer Craig Partridge, pushing the envelope on interesting challenges has been his modus operandi. It's what led him to make key contributions to the early Internet.

The IEEE Fellow worked with TCP/IP developer and advocate Phil Karn on TCP message round-trip time estimation. Partridge also led the team that designed and built the world's first high-speed router. Today, he's working on getting a better understanding of traffic flows in VPNs. In addition, he is trying to figure out what causes high rates of packet errors, which often result in bad file transfers, and how to fix them.

Although his bachelor's degree from Harvard was in history, the computer electives he took convinced him he had both the aptitude and interest for a career in computing. And he was right. He developed networking concepts and technologies while working at the ARPAnet and Internet pioneering firm Bolt Beranek & Newman (BBN), now Raytheon BBN Technologies. And he holds a master's and Ph.D. in computer science, also from Harvard.

"It was 1983, and BBN needed an additional person who could write TCP/IP code for the BSD [Berkeley Software Distribution] Unix operating system," Partridge says. "BBN had folks who wrote most of the original code, and they taught me what networking and TCP/IP was. That was something no grad school at the time could teach me.

"A lot of the Internet tech used today is entirely or almost entirely recognizable from what was created back then," he says.

By 2018 Partridge was chief scientist for networking research at BBN, but he decided he didn't want his tenure there to be the last chapter of his career. Later that year he joined Colorado State University in Fort Collins, as chair of its department of computer science.

Partridge says it's not uncommon for engineers to move from industry to academia later in their careers. They often want to "give back" by becoming professors, deans, or department chairs.

One of his assignments at CSU was to reexamine the school's computer science curriculum. Among other improvements, the faculty now teaches courses in machine learning earlier. Virtual reality and virtual spaces courses were created, and the math requirements were revamped.

He also is working to increase the number of women in computing. One way was to revise the curriculum to follow the BRAID (Building, Recruiting, And Inclusion for Diversity) principles.

"This effort has included figuring out how to make the introductory courses speak to a wider variety of students by covering the same material but using different examples and different presentations that relate to how they could use computing to solve problems that interest them," he says.

Partridge says the number of women majoring in computing at CSU has nearly doubled, from 12 percent to more than 20 percent.

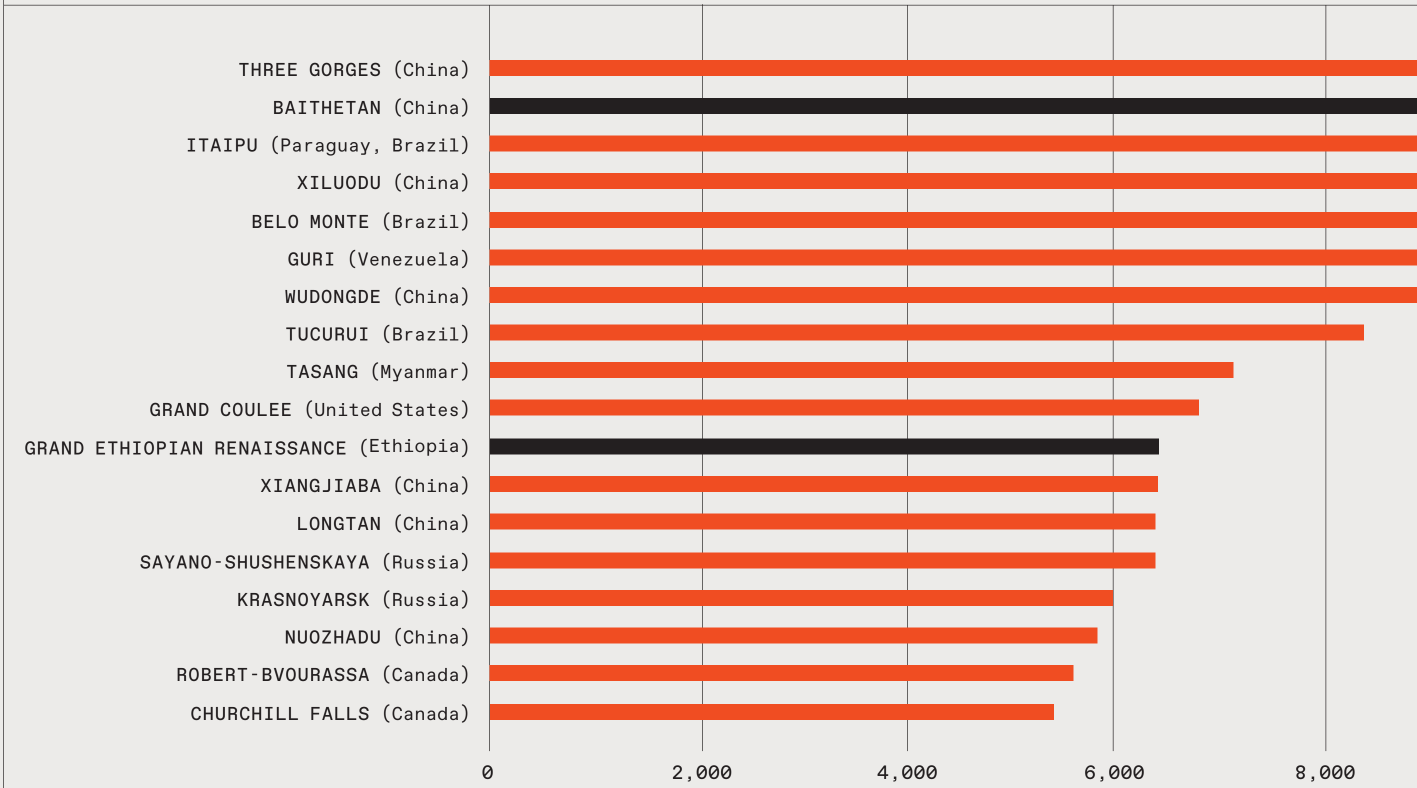
He also has continued to conduct research, which he finds very different from how he did research at BBN.

"You are trying to do research while trying to teach the students," he says. "You could often do the research faster yourself, but then the student wouldn't learn. By comparison, in industry I knew I had a team I could give a broad description to and get the right thing the first time. With students, by the time they get good at it, they are graduating."

Partridge would like today's engineers to pay more attention to edge computing, smart home infrastructure, vehicle safety systems, and walled-garden social networks. He says walled gardens are "social networks where people don't want to connect to people who don't want to connect with those who think differently."

One challenge in preparing students for the world they will be working in, he says, "is getting them to understand that just because you can create something new doesn't mean that society wants it." ■

Numbers Don’t Lie



Hydropower, the Forgotten Renewable

This still-mighty source of electricity has been unfairly relegated to the scrap heap

I live in Manitoba, a province of Canada where all but a tiny fraction of electricity is generated from the potential energy of water. Unlike in British Columbia and Quebec, where generation relies on huge dams, our dams on the Nelson River are low, with hydraulic heads of no more than 30 meters, which creates only small reservoirs. Of course, the potential is the product of mass, the gravitational constant, and height, but the dams’ modest height is readily compensated for by a large mass, as the mighty river flowing out of Lake Winnipeg continues its course to Hudson Bay.

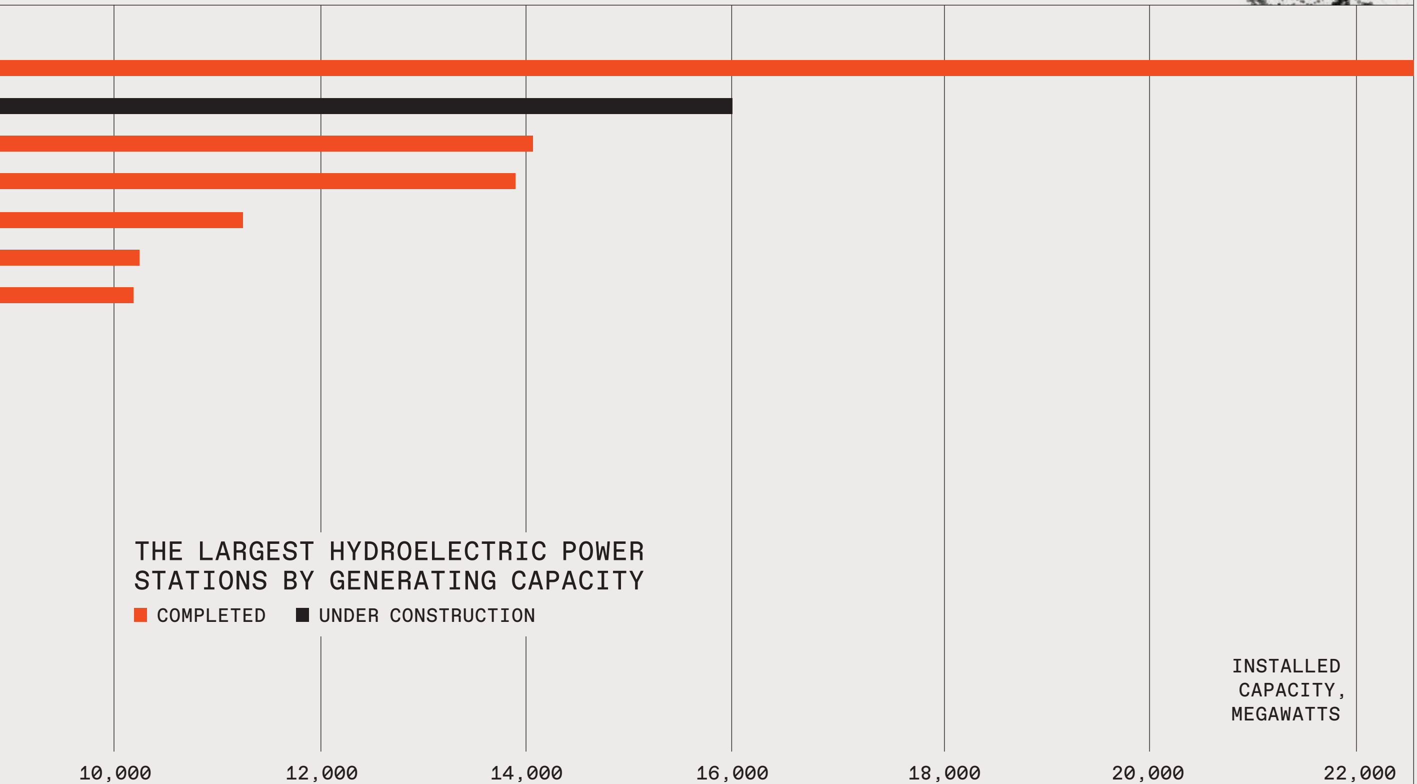
You would think this would be lauded for being about as “green” as it could get, but in 2022 that would be a mistake. There is no end of gushing about China’s cheap solar panels—but when was the last time you saw a paean to hydroelectricity?

Construction of large dams began before World War II. The United States got the Grand Coulee on the Columbia River, the Hoover Dam on the Colorado, and the dams of the Tennessee Valley Authority. After the war, construction of large dams moved to the Soviet Union, Africa, South America (Brazil’s Itaipu, at its completion in 1984 the world’s largest dam, with 14 gigawatts capacity), and Asia, where it culminated in China’s unprecedented effort. China now has three of the world’s six largest hydroelectric stations: Three Gorges, 22.5 GW (the world’s largest); Xiluodu, 13.86 GW; and Wudongde, 10.2 GW. Baihetan on the Jinsha River should soon become the world’s second-largest station (16 GW).

But China’s outsize drive for hydroelectricity is unique. By the 1990s, large hydro stations had lost

DATA SOURCES: WIKIPEDIA, HYDROPOWER.ORG

BY VACLAV SMIL



their green halo in the West and come to be seen as environmentally undesirable. They are blamed for displacing populations, disrupting the flow of sediments and the migration of fish, destroying natural habitat and biodiversity, degrading water quality, and for the decay of submerged vegetation and the consequent release of methane, a greenhouse gas. There is thus no longer a place for Big Hydro in the pantheon of electric greenery. Instead, that pure status is now reserved above all for wind and solar. This ennoblement is strange, given that wind projects require enormous quantities of embodied energy in the form of steel for towers, plastics for blades, and concrete for foundations. The manufacture of solar panels involves the environmental costs from mining, waste disposal, and carbon emissions.

And hydro still matters more than any other form of renewable generation. In 2020, the world's hydro stations produced 75 percent more electricity than wind and solar combined (4,297 versus 2,447 terawatt-hours) and accounted for 16 percent of all global generation (compared with nuclear electricity's 10 percent). The share rises to about 60 percent in Canada and 97 percent in Manitoba. And some less affluent countries in Africa and Asia are still determined to build more such stations. The largest projects now under construction outside China are the Grand Ethiopian Renaissance Dam on the White Nile (6.45 GW) and Pakistan's Diamer-Bhasha (4.5 GW) and Dasu (4.3 GW) on the Indus.

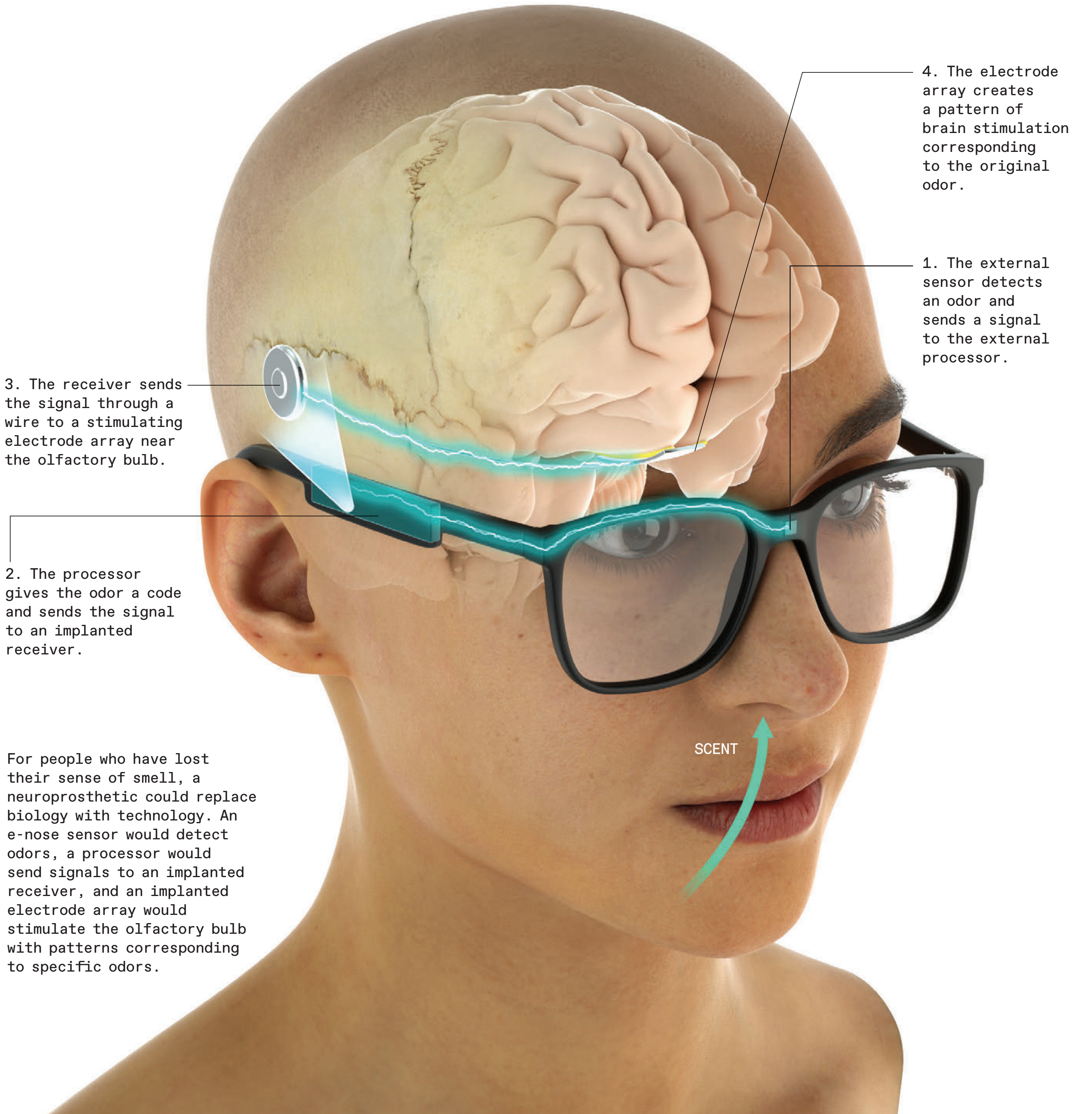
In 2020, the world's hydro stations produced 75 percent more electricity than wind and solar combined and accounted for 16 percent of all global generation.

I never understood why dams have suffered such a reversal of fortune. There is no need to build megastructures, with their inevitable undesirable effects. And everywhere in the world there are still plenty of opportunities to develop modest projects whose combined capacities could provide not only excellent sources of clean electricity but also serve as long-term stores of energy, as reservoirs for drinking water and irrigation, and for recreation and aquaculture.

I am glad to live in a place that is reliably supplied by electricity generated by low-head turbines powered by flowing water. Manitoba's six stations on the Nelson River have a combined capacity slightly above 4 GW. Just try to get the equivalent here from solar in January, when the snow is falling and the sun barely rises above the horizon! ■

A Bionic Nose to Smell

COVID SURVIVORS DRIVE DEMAND FOR



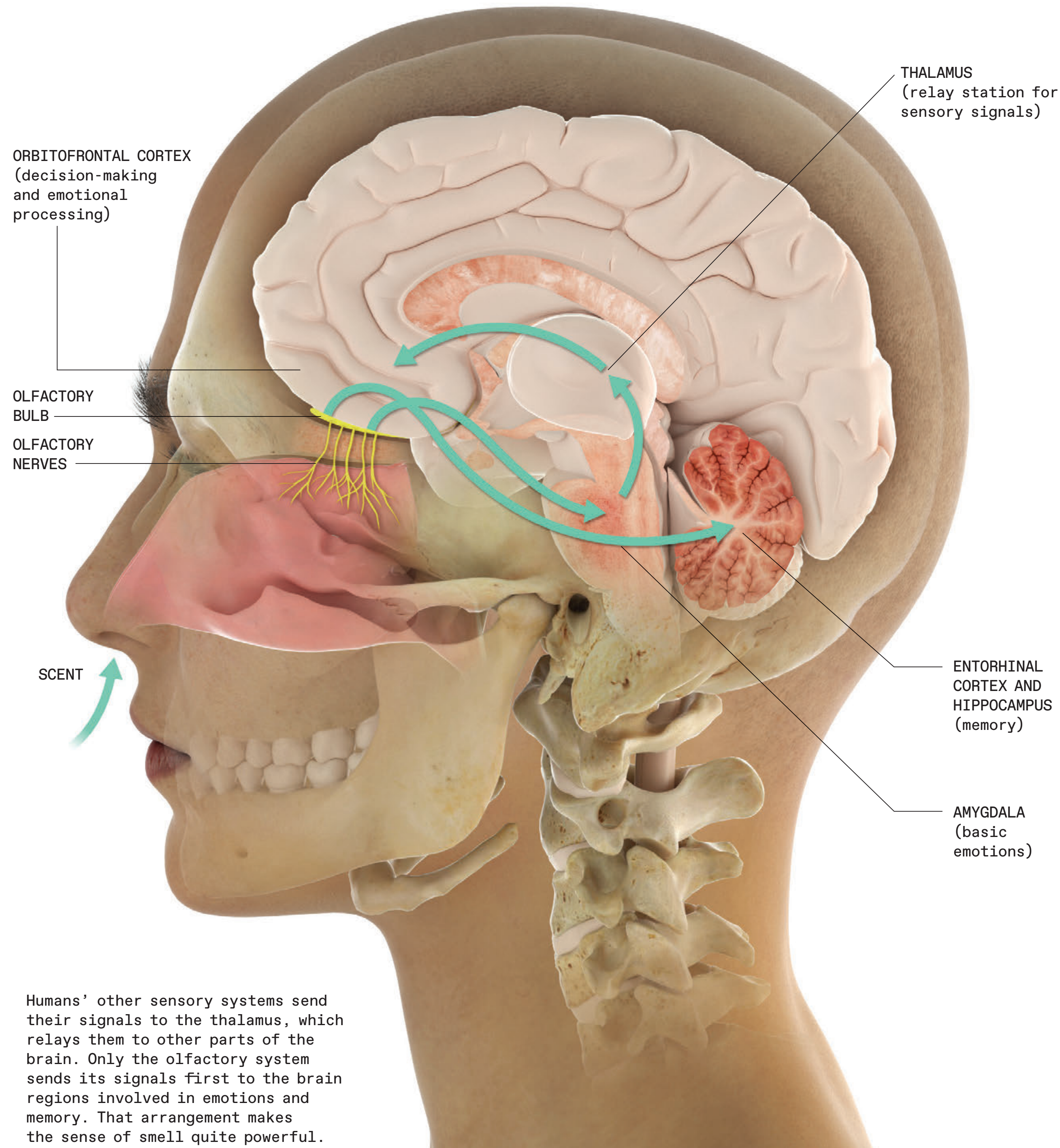
For people who have lost their sense of smell, a neuroprosthetic could replace biology with technology. An e-nose sensor would detect odors, a processor would send signals to an implanted receiver, and an implanted electrode array would stimulate the olfactory bulb with patterns corresponding to specific odors.

the Roses Again

A NEUROPROSTHETIC NOSE

BY ELIZA STRICKLAND

Illustrations by Anatomy Blue



Humans' other sensory systems send their signals to the thalamus, which relays them to other parts of the brain. Only the olfactory system sends its signals first to the brain regions involved in emotions and memory. That arrangement makes the sense of smell quite powerful.

DEAUDREA 'SHA' AGUADO

RICHARD COSTANZO STANDS BESIDE A MANNEQUIN HEAD sporting spectacles decked with electronics and holds a vial of blue liquid up to a tiny sensor. An LED glows blue, and Costanzo's phone displays the word "Windex." Then he waves a vial of purple liquid and gets a purple light along with the message "Listerine." • "There won't be Scotch tape on the final model," says Costanzo, as he rearranges the gear in his lab at Virginia Commonwealth University (VCU), in Richmond. The prototype is a partial demonstration of a concept that he's been working on for decades: a neuroprosthetic for smell. The mannequin represents someone who has lost their sense of smell to COVID-19, brain injury, or some other medical condition. It is also intended to show off the sensor, which is the same type used for commercial electronic noses, or e-noses. In the final product, the sensor won't light up an LED but will instead send a signal to the user's brain.

In the lab's back room, another model shows the second half of the concept: There, the e-nose sensor transmits its signal to a small array of electrodes taken from a cochlear implant. For people with hearing loss, such implants feed information about sound to the inner ear and then to the brain. The implant is also about the right size for the olfactory bulb on the edge of the brain. Why not use it to convey information about odor?

This project could be a career-capping achievement for Costanzo, a professor emeritus of physiology and biophysics who in the 1980s cofounded VCU's Smell and Taste Disorders Center, one of the first such clinics in the country. After years of research on olfactory loss and investigations into the possibility of biological regeneration, he began working on a hardware solution in the 1990s.

A self-described electronics buff, Costanzo enjoyed his experiments with sensors and electrodes. But the project really took off in 2011 when he began talking with his colleague Daniel Coelho, a professor of otolaryngology at VCU and an expert in cochlear implants. They recognized at once that a smell prosthetic could be similar to a cochlear implant: "It's taking something from the physical world and translating it into electrical signals that strategically target the brain," Coelho says. In 2016

Richard Costanzo [left] and Daniel Coelho [right] demonstrate the external components of their olfactory prosthetic. In a complete system, after the sensor detects an odor, the transmitter would send a signal to a stimulator implanted in the brain.

the two researchers were awarded a U.S. patent for their olfactory-implant system.

Costanzo's quest became abruptly more relevant in early 2020, when many patients with a new illness called COVID-19 realized they had lost their senses of smell and taste. Three years into the pandemic, some of those patients have still not recovered those faculties. When you also consider people who have lost their sense of smell due to other diseases, brain injury, and aging, this niche technology starts to look like a viable product. Add in Costanzo and Coelho's other collaborators—including an electronic nose expert in England, several clinicians in Boston, and a businessman in Indiana—and you have a dream team who just might make it happen.

Costanzo says he's wary of hype and doesn't want to give people the impression that a commercial device will be available any day now. But he does want to offer hope. Right now, the team is focused on getting the sensors to detect more than a few odors and figuring out how best to interface with the brain. "I think we're several years away from cracking those nuts," Costanzo says, "but I think it's doable."

Scott Moorehead just wanted to teach his 6-year-old son how to skateboard. On a Sunday in 2012 he was demonstrating some moves in the driveway of his Indiana home when the skateboard hit a crack and flipped him off. "The back of my skull bore the brunt of the fall," he says. He spent three days in the intensive care unit, where doctors treated him for multiple skull fractures, massive internal bleeding, and damage to his brain's frontal lobe.

Over weeks and months his hearing came back, his headaches went away, and his irritability and confusion faded. But he never regained his sense of smell.

Moorehead's accident permanently disconnected the nerves that run from the nose to the olfactory bulb at the base of the brain. Along with his sense of smell, he lost all but a rudimentary sense of taste. "Flavor comes mostly from smell," he explains. "My tongue on its own can only do sweet, salty, spicy, and bitter. You can blindfold me and put 10 flavors of ice cream in front of me, and I won't know the difference: They'll all taste slightly sweet, except chocolate that's a bit bitter."

Moorehead grew depressed: Even more than the flavors of food, he missed the unique smells of the people he loved. And on one occasion he was oblivious to a gas leak, only realizing the danger when his wife came home and raised the alarm.

Anosmia, or the inability to smell, can be caused not only by head injuries but also by exposure to certain toxins and by a variety of medical problems—including tumors, Alzheimer's, and viral diseases, such as COVID. The sense of smell also commonly atrophies with age; in a 2012 study in which more than 1,200 adults were given olfactory exams, 39 percent of participants age 80 and above had olfactory dysfunction.

The loss of smell and taste have been dominant symptoms of COVID since the beginning of the pandemic. People with COVID-induced anosmia currently have only three options: Wait and see if the sense comes back on its own, ask for a steroid medication that reduces inflammation and may speed recovery, or begin smell rehab, in which they expose themselves to a few familiar scents each day to encourage the restoration of the nose-brain nerves. Patients typically do best if they seek out medication and rehab within a few weeks of experiencing symptoms, before scar tissue builds up. But even then, these interventions don't work for everyone.

In April 2020, researchers at VCU's smell and taste clinic launched a nationwide survey of adults who had been diagnosed with COVID to determine the prevalence and duration of smell-related symptoms. They've followed up with those people at regular intervals, and this past August they published results from people who were two years past their initial diagnosis. The findings were striking: Thirty-eight percent reported a full recovery of smell and taste, 54 percent reported a partial recovery, and 7.5 percent reported no recovery at all. "It's a serious quality of life issue," says Evan Reiter, director of the VCU clinic.

While other researchers are investigating biological approaches, such as using stem cells to regenerate odor receptors and nerves, Costanzo believes the hardware approach is the only solution for people with total loss of smell. "When the pathways are really out of commission, you have to replace them with technology," he says.

Unlike most anosmics, Scott Moorehead didn't give up when his doctors told him there was nothing he could do to recover his sense of smell. As the CEO of a cellphone retail company with stores in 43 states, he had the resources to invest in long-shot research. And when a colleague told him about the work at VCU, he got in touch and offered to help. Since 2015, Moorehead has put almost US $1 million into the research. He also licensed the technology from VCU and launched a startup called Sensory Restoration Technologies.

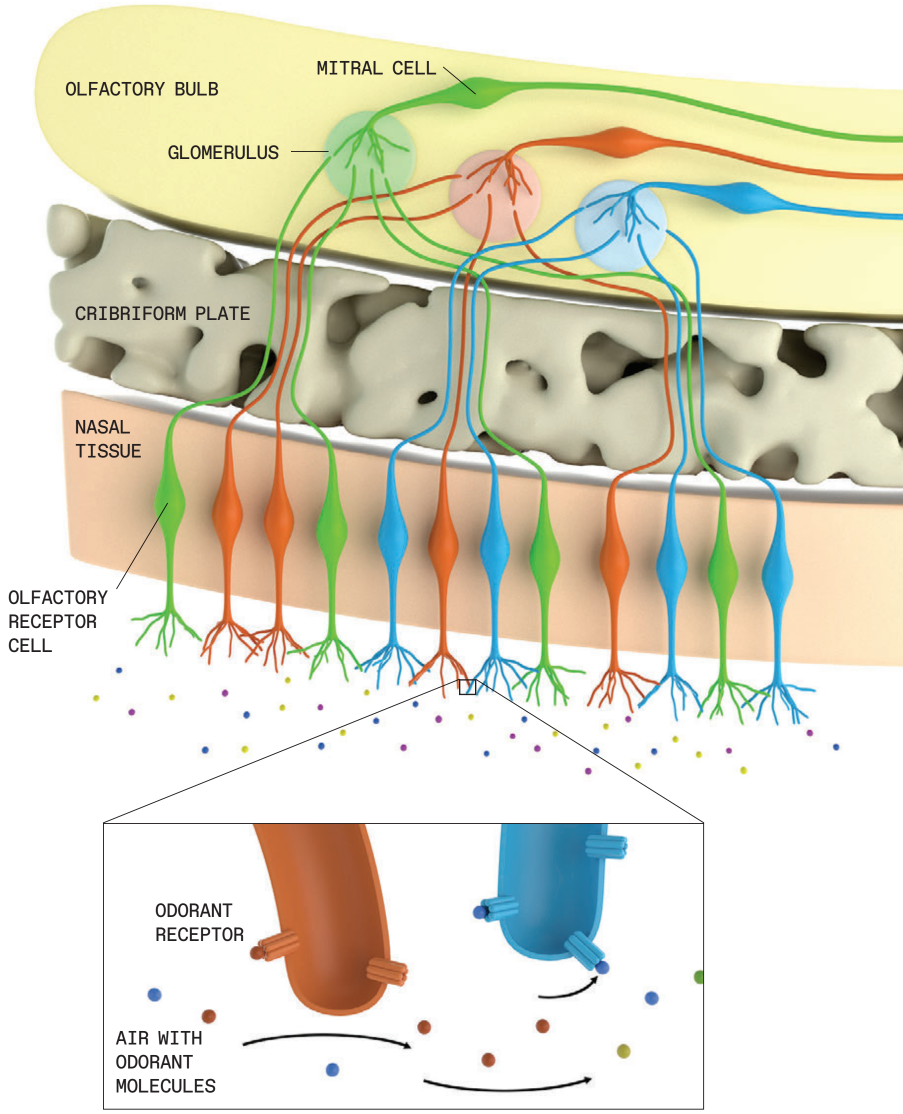
When COVID struck, Moorehead saw an opportunity. Although they were far from having a product to advertise, he scrambled to put up a website for the startup. He remembers saying: "People are losing their sense of smell. People need to know we exist!"

Equivalent neuroprosthetics exist for other senses. Cochlear implants are the most successful neurotechnology to date, with more than 700,000 devices implanted in ears around the world. Retina implants have been developed for blind people (though some bionic-vision systems have had commercial trouble), and researchers are even working on restoring the sense of touch to people with prosthetic limbs and paralysis. But smell and taste have long been considered too hard a challenge.

To understand why, you need to understand the marvelous complexity of the human olfactory system. When the smell of a rose wafts up into your nasal cavity, the odor molecules bind to

7.5% OF PEOPLE WHO WERE DIAGNOSED WITH COVID-19 MORE THAN TWO YEARS AGO HAVE NOT RECOVERED THEIR SENSE OF SMELL.

Odor molecules that enter the nose bind to olfactory receptor cells, which send signals through the bone of the cribriform plate to reach the olfactory bulb. From there, the signals are sent to the brain.



receptor neurons that send electrical signals up the olfactory nerves. Those nerves pass through a bony plate to reach the olfactory bulb, a small neural structure in the forebrain. From there, information goes to the amygdala, a part of the brain that governs emotional responses; the hippocampus, a structure involved in memory; and the frontal cortex, which handles cognitive processing.

Those branching neural connections are the reason that smells can sometimes hit with such force, conjuring up a happy memory or a traumatizing event. "The olfactory system has access to parts of the brain that other senses don't," Costanzo says. The diversity of brain connections, Coelho says, also suggests that stimulating the olfactory system could have other applications, going well beyond appreciating food or noticing a gas leak: "It could affect mood, memory, and cognition."

The biological system is difficult to replicate for a few reasons. A human nose has around 400 different types of receptors that detect odor molecules. Working together, those receptors enable humans to distinguish between a staggering number of smells: A 2014 study estimated the number at 1 trillion. Until now, it hasn't been practical to put 400 sensors on a chip that would be attached to a user's eyeglasses. What's more, researchers don't yet fully understand the olfactory code by which stimulating certain combinations of receptors leads to perceptions of odor in the brain. Luckily, Costanzo and Coelho know people working on both of those problems.

E-**NOSES ARE ALREADY** used today in a variety of industrial, office, and residential settings—if you have a typical carbon-monoxide detector in your home, you have a very simple e-nose.

"Traditional gas sensors are based on semiconductors like metal oxides," explains Krishna Persaud, a leading e-nose researcher and a professor of chemoreception at the University of Manchester, in England. He's also an advisor to Costanzo and Coelho. In the most typical e-nose setup, he says, "when a molecule interacts with the semiconductor material, a change in resistance occurs that you can measure." Such sensors have been shrinking over the last two decades, Persaud says, and they're now the size of a microchip. "That makes them very convenient to put in a small package," he says. In the VCU team's early experiments, they used an off-the-shelf sensor from a Japanese company called Figaro.

The problem with such commercially available sensors, Persaud says, is that they can't distinguish between very many different odors. That's why he's been working with new materials, such as conductive polymers that are cheap to manufacture, low power, and can be grouped together in an array to provide sensitivity to dozens of odors. For the neuroprosthetic, "in principle, several hundred [sensors] could be feasible," Persaud says.

A first-generation product wouldn't allow users to smell hundreds of different odors. Instead, the VCU team imagines initially including receptors for a few safety-related smells, such as smoke and natural gas, as well as a few pleasurable ones. They could even customize the prosthetic to give users smells that are meaningful to them: the smell of bread for a home baker, for example, or the smell of a pine forest for an avid hiker.

Pairing this e-nose technology with the latest neurotechnology is Costanzo and Coelho's current challenge. While working with Persaud to test new sensors, they're also partnering with clinicians in Boston to investigate the best method of sending signals to the brain.

The VCU team laid the groundwork with animal experiments. In experiments with rats in 2016 and 2018, the team showed that using electrodes to directly stimulate spots on the

“THE OLFACTORY SYSTEM HAS ACCESS TO PARTS OF THE BRAIN THAT OTHER SENSES DON’T.” —Richard Costanzo

surface of the olfactory bulb generated patterns of neural activity deep in the bulb, in the neurons that passed messages on to other parts of the brain. The researchers called these patterns odor maps. But while the neural activity indicated that the rats were perceiving something, the rats couldn’t tell the researchers what they smelled.

Their next step was to recruit collaborators who could perform similar trials with human volunteers. They started with one of Costanzo’s former students, Eric Holbrook, an associate professor of otolaryngology at Harvard Medical School and director of rhinology at Massachusetts Eye and Ear. Holbrook spends much of his time operating on people’s sinus cavities, including the ethmoid sinus cavities, which are positioned just below the cribriform plate, a bony structure that separates the olfactory receptors from the olfactory bulb.

Holbrook discovered, in 2018, that placing electrodes on the bone transmitted an electrical pulse to the olfactory bulb. In a trial with awake patients, three of the five volunteers reported smell perception during this stimulation, with the reported odors including “an onionlike smell,” “antiseptic-like and sour,” and “fruity but bad.” While Holbrook sees the trial as a good proof of concept for an olfactory-implant system, he says that poor conductance through the bone was an important limiting factor. “If we are to provide discrete, separate areas of stimulation,” he says, “it can’t be through bone and will need to be on the olfactory bulb itself.”

Placing electrodes on the olfactory bulb would be new territory. “Theoretically,” says Coelho, “there are many different ways to get there.” Surgeons could go down through the brain, sideways through the eye socket, or up through the nasal cavity, breaking through the cribriform plate to reach the bulb. Coelho explains that rhinology surgeons often perform low-risk surgeries that involve breaking through the cribriform plate. “What’s new isn’t how to get there or clean up afterward,” he says, “it’s how do you keep an indwelling foreign body in there without causing problems.”

Another tactic entirely would be to skip over the olfactory bulb and instead stimulate “downstream” parts of the brain that receive signals from the olfactory bulb. Championing that approach is another of Costanzo’s former students, Mark Richardson, director of functional neurosurgery at Massachusetts General Hospital. Richardson often has epilepsy patients spend several days in the hospital with electrodes in their brains, so that doctors can determine which brain regions are involved in their seizures and plan surgical treatments. While such patients are waiting around, however, they’re often recruited for neuroscience studies.

To contribute to Costanzo and Coelho’s research, Richardson’s team asked epilepsy patients in the monitoring unit to take a sniff of a wand imbued with a smell such as peppermint, fish, or banana. The electrodes in their brains showed the pattern of resulting neural activity “in areas where we expected, but also in areas where we didn’t expect,” Richardson says. To better understand the brain responses, his team has just begun another round of experiments with a tool called an olfactometer that will release more precisely timed bursts of smell.

Once the researchers know where the brain lights up with activity in response to, say, the smell of peppermint, they can try stimulating those areas with electricity alone in hopes of creating the same sensation. “With the existing technology, I think we’re closer to inducing the [smell perceptions] with brain stimulation than with olfactory-bulb stimulation,” Richardson says. He notes that there are already approved implants for brain stimulation and says using such a device would make the regulatory path easier. However, the distributed nature of smell perception within the brain poses a new complication: A user would likely need multiple implants to stimulate different areas. “We might need to hit different sites in quick succession or all at once,” he says.

ACROSS THE ATLANTIC, the European Union is funding its own olfactory-implant project, called ROSE (Restoring Odorant detection and recognition in Smell dEficits). It launched in 2021 and involves seven institutions across Europe.

Thomas Hummel, head of the Smell & Taste Clinic at the Technical University of Dresden and a member of the consortium, says the ROSE researchers are partnering with Aryballe, a French company that makes a tiny sensor for odor analytics. The partners are currently experimenting with stimulating both the olfactory bulb and the prefrontal cortex. “All the parts that are needed for the device, they already exist,” he says. “The difficulty is to bring them together.” Hummel estimates that the consortium’s research could lead to a commercial product in 5 to 10 years. “It’s a question of effort and a question of funding,” he says.

Persaud, the e-nose expert, says the jury is out on whether a neuroprosthetic could be commercially viable. “Some people with anosmia would do anything to have that sense back to them,” he says. “It’s a question of whether there are enough of those people out there to make a market for this device,” he says, given that surgery and implants always carry some amount of risk.

The VCU researchers have already had an informal meeting with regulators from the U.S. Food and Drug Administration, and they’ve started the early steps of the process for approving an implanted medical device. But Moorehead, the investor who tends to focus on practical matters, says this dream team might not take the technology all the way to the finish line of an FDA-approved commercial system. He notes that there are plenty of existing medical-implant companies that have that expertise, such as the Australian company Cochlear, which dominates the cochlear-implant market. “If I can get [the project] to the stage where it’s attractive to one of those companies, if I can take some of the risk out of it for them, that will be my best effort,” Moorehead says.

Restoring people’s ability to smell and taste is the ultimate goal, Costanzo says. But until then, there’s something else he can give them. He often gets calls from desperate people with anosmia who have found out about his work. “They’re so appreciative that someone is working on a solution,” Costanzo says. “My goal is to provide hope for these people.” ■

To decarbonize road transport we need to complement EVs with bikes, rail, city planning, and alternative energy

THE ELECTRIC VEHICLE IS NOT ENOUGH

BY HEATHER L. MACLEAN, ALEXANDRE MILOVANOFF & I. DANIEL POSEN

EVS HAVE FINALLY COME OF AGE. The total cost of purchasing and driving one—the cost of ownership—has fallen nearly to parity with a typical gasoline-fueled car. Scientists and engineers have extended the range of EVs by cramming ever more energy into their batteries, and vehicle charging networks have expanded in many countries. In the United States, for example, there are more than 49,000 public charging stations, and it is now possible to drive an EV from New York to California using public charging networks.

With all this, consumers and policymakers alike are hopeful that society will soon greatly reduce its carbon emissions by replacing today's cars with electric vehicles. Indeed, adopting electric vehicles will go a long way in helping to improve environmental outcomes. But EVs come with important weaknesses, and so people shouldn't count on them alone to do the job, even for the transportation sector.

Why not? EVs lack tailpipe emissions, sure, but producing, operating, and disposing of these vehicles creates greenhouse-gas emissions and other environmental burdens. Driving an EV pushes these problems upstream, to the factory where the vehicle is made and beyond, as well as to the power plant where the electricity is generated. The entire life cycle of the vehicle must be considered, from cradle to grave. When you do that, the promise of electric vehicles doesn't shine quite as brightly. Here we'll show you in greater detail why that is.

Illustration by MCKIBILLO

THE LIFE CYCLE to which we refer has two parts: The vehicle cycle begins with mining the raw materials, refining them, turning them into components, and assembling them. It ends years later with salvaging what can be saved and disposing of what remains. Then there is the fuel cycle—the activities associated with producing and using the fuel or electricity to power the vehicle through its working life.

For EVs, much of the environmental burden centers on the production of batteries, the most energy- and resource-intensive component of the vehicle. Each stage in production matters—mining, refining, and producing the raw materials, manufacturing the components, and finally assembling them into cells and battery packs.

Where all this happens matters, too, because a battery factory uses a lot of electricity, and the source for that electricity varies from one region to the next. Manufacturing an EV battery using coal-based electricity results in more than three times the greenhouse-gas emissions of manufacturing a battery with electricity from renewable sources. And about 70 percent of lithium-ion batteries are produced in China, which derived 64 percent of its electricity from coal in 2020.

Most automotive manufacturers say they plan to use renewable energy in the future, but for now, most battery production relies on electric grids largely powered by fossil fuels. Our 2020 study, published in *Nature Climate Change*, found that manufacturing a typical EV sold in the United States in 2018 caused about 7 to 12 tonnes of carbon dioxide emissions, compared with about 5 to 6 tonnes for a gasoline-fueled vehicle.

You also must consider the electricity that charges the vehicle. In 2019, 63 percent of global electricity was produced from fossil-fuel sources, the exact nature of which varies substantially among regions. China, using largely coal-based electricity, had 6 million EVs in 2021, constituting the largest total stock of EVs in the world.

But coal use varies, even within China. The southwest province of Yunnan derives about 70 percent of its electricity from hydropower, slightly more than the percentage in Washington state, while Shandong, a coastal province in the east, derives about 90 percent of its electricity from coal, similar to West Virginia.

Norway has the highest per capita number of EVs, which represented more than 86 percent of vehicle sales in that country in 2021. And it produces almost all its electricity from hydro and solar. Therefore, an EV operated in Shandong imposes a much bigger environmental burden than that same EV would in Yunnan or Norway.

The U.S. falls somewhere in the middle, deriving about 60 percent of its electricity from fossil fuels, primarily natural gas, which produces less carbon than coal does. In our model, using electricity from the 2019 U.S. grid to charge a typical 2018 EV would produce between 80 and 120 grams of carbon dioxide per kilometer traveled, compared with about 240 to 320 g/km for a gasoline vehicle. Credit the EV's advantage to its greater efficiency in the conversion of chemical energy to motion—77 percent, compared with 12 to 30 percent for a gasoline car—along with the potential to generate electricity using low-carbon sources. That's why operating EVs typically releases less carbon than operating gasoline vehicles of similar size, even in coal-heavy grids like Shandong or West Virginia.

But when you factor in the greenhouse-gas emissions associated with vehicle manufacture, the calculus changes. As an illustration, an EV operated in Shandong or West Virginia causes about 6 percent *more* greenhouse gas emissions over its lifetime than does a conventional gasoline vehicle of the same size. An EV operated in Yunnan causes about 60 percent less.

An EV operated in Shandong or West Virginia causes about 6 percent *more* greenhouse gas emissions over its lifetime than does a conventional gasoline vehicle of the same size. An EV operated in Yunnan causes about

60 PERCENT LESS.

CAN EVS BE good enough—and can manufacturers roll them out fast enough—to meet the goals set in 2021 by the 26th United Nations Climate Change Conference (COP26)? The 197 signatory nations agreed to hold the increase in the average global temperature to no more than 2 °C above preindustrial levels and to pursue efforts to limit the increase to 1.5 °C.

Our analysis shows that to bring the United States into line with even the more modest 2 °C goal would require electrifying about 90 percent of the U.S. passenger-vehicle fleet by 2050—some 350 million vehicles.

To arrive at this number, we first had to decide on an appropriate carbon budget for the U.S. fleet. Increases in global average temperature are largely proportional to cumulative global emissions of carbon dioxide and other greenhouse gases. Climate scientists use this fact to set a limit on the total amount of carbon dioxide that can be emitted before the world surpasses the 2 °C goal: This amount constitutes the global carbon budget.

We then used results from a model of the global economy to allocate a portion of this global budget specifically to the U.S. passenger-vehicle fleet over the period between 2015 and 2050. This portion came out to around 45 billion tonnes

Cobalt mining for batteries in the Democratic Republic of Congo has been linked to water-quality problems, armed conflicts, child labor, respiratory disease, and birth defects.

SEBASTIAN MEYER/CORBIS/GETTY IMAGES

of carbon dioxide, roughly equivalent to a single year of global greenhouse-gas emissions.

This is a generous allowance, but that's reasonable because transportation is harder to decarbonize than many other sectors. Even so, working within that budget would require a 30 percent reduction in the projected cumulative emissions from 2015 to 2050 and a 70 percent reduction in annual emissions in 2050, compared with the business-as-usual emissions expected in a world without EVs.

Next, we turned to our model of the U.S fleet of light vehicles. Our model simulates for each year from 2015 to 2050 how many new vehicles are manufactured and sold, how many are scrapped, and the associated greenhouse-gas emissions. We also keep track of how many vehicles are on the road, when they were made, and how far they are likely to drive. We used this information to estimate annual greenhouse-gas emissions from the fuel cycle, which depend partly on the average vehicle size and partly on how much vehicle efficiency improves over time.

Finally, we compared the carbon budget with our model of total cumulative emissions (that is, both vehicle-cycle and fuel-cycle emissions). We then systematically increased the share of EVs among new vehicle sales until the cumulative fleet emissions fell within the budget. The result: EVs had to make up the vast majority of vehicles on the road by 2050, which means they must make up the vast majority of vehicle sales a decade or more earlier.

That would require a dramatic increase in EV sales: In the United States in 2021, just over 1 million vehicles—less than 1 percent of those on the road—were fully electric. And only 3 percent of the new vehicles sold were fully electric. Considering the long lifetime of a vehicle, about 12 years in the United States, we would need to ramp up sales of EVs dramatically starting now to meet the 2 °C target. In our model, over 10 percent of all new vehicles sold by 2020 would have had to be electric, rising above half by 2030, and essentially all by 2035. Studies conducted in other countries, such as China and Singapore, have arrived at similar results.

The good news is that 2035 is the year suggested at the COP26 for all new car and vans in leading markets to be zero-emissions vehicles, and many manufacturers and governments have committed to it. The bad news is that some major automotive markets, such as China and the United States, have not yet made that pledge, and the United States has already missed the 10 percent sales share for 2020 that our study recommended. Of course, meeting the more ambitious 1.5 °C climate target would require even larger-scale deployment of EVs and therefore earlier deadlines for meeting these targets.

IT'S A TALL order, and a costly one, to make and sell so many EVs so soon. Even if that were possible, there would also have to be an enormous increase in charging infrastructure and in material supply chains. And that much more vehicle charging would then put great pressure on our electricity grids.

Charging matters because one of the commonly cited obstacles to EV adoption is range anxiety. Shorter-range EVs, like the Nissan Leaf, have a manufacturer's reported range of just 240 km, although a 360-km model is also available. Longer-range EVs, like the Tesla Model 3 Long Range, have a manufacturer's reported range of 600 km. The shorter driving ranges of most EVs are no problem for daily commutes, but range anxiety is real for longer trips, especially in cold weather, which can cut driving ranges substantially due to the energy demand of heating the cabin and the lower capacity of cold batteries.

Most EV owners recharge their cars at home or at work, meaning that chargers need to be available in garages, driveways, on-street parking, apartment building parking areas, and commercial parking lots. A couple of hours at home is sufficient to recharge from a typical daily commute, while overnight charging is needed for longer trips. In contrast, public charging stations that use fast charging can add several hundred kilometers of range in 15 to 30 minutes. This is an impressive feat, but it still takes longer than refilling a gas tank.

Another barrier to the adoption of EVs is the price, which is largely a function of the cost of the batteries, which make the purchase price 25 to 70 percent higher than that of an equivalent conventional vehicle. Governments have offered subsidies or tax rebates to make EVs more appealing, but such measures, while easy enough to implement in the early days of a new technology, would become prohibitively expensive as EV sales mount.

Although EV battery costs have fallen dramatically over the past decade, the International Energy Agency is projecting a sudden reversal of that trend in 2022 due to increases in prices of critical metals and a surge in demand for EVs. While projections of future prices vary, highly cited long-term projections from BloombergNEF suggest the cost of new EVs will reach price parity with conventional vehicles by 2026, even without government subsidies. In the meantime, EV buyers' sticker shock could be alleviated by the knowledge that fuel and maintenance costs are far lower for EVs and that total ownership costs are about the same.

But what drivers gain, governments might lose. The International Energy Agency estimates that by 2030 the deployment of EVs could cut global receipts from fossil-fuel taxes by around US $55 billion. Those tax revenues are necessary for the maintenance of roads. To make up for their loss, governments will need some other source of revenue, such as vehicle registration fees.

T

THE GROWTH IN the number of EVs introduces various other challenges, too, not the least of which are the greater demands placed on material supply chains for EV batteries and electricity grids. Batteries require raw materials such as lithium, copper, nickel, cobalt, manganese, and graphite. Some of these materials are highly concentrated in a few countries.

For example, the Democratic Republic of Congo (DRC) holds about 50 percent of the world's cobalt reserves. Just two countries—Chile and Australia—account for over two-thirds of global lithium reserves, and South Africa, Brazil, Ukraine, and Australia have almost all the manganese reserves. This concentration is problematic because it can lead to volatile markets and supply disruptions.

The COVID pandemic has shown just what supply-chain disruptions can do to other products dependent on scarce materials, notably semiconductors, the shortage of which has forced several automotive manufacturers to

AFP/GETTY IMAGES

The manufacture of lithium batteries for EVs, like those shown here, is energy intensive, as is the mining and refining of the raw materials.

China has more EVs than any other country—but also gets most of its electricity from coal.

VCG/GETTY IMAGES

stop producing vehicles. It is unclear whether suppliers will be able to meet the future demand for some critical raw materials for electric batteries. Market forces may lead to innovations that will increase the supplies of these materials or reduce the need for them. But for now, the implications for the future are not at all obvious.

The scarcity of these materials reflects not only the irregular endowment of various countries but also the social and environmental consequences of extraction and production. The presence of cobalt mines in the DRC, for example, reduced water quality and expanded armed conflicts, child labor, respiratory disease, and birth defects. International regulatory frameworks must therefore not only protect supply chains from disruption but also protect human rights and the environment.

Some of the problems in securing raw material could be mitigated by new battery chemistries—several manufacturers have announced plans to switch to lithium iron phosphate batteries, which are cobalt free—or battery-recycling programs. But neither option totally removes supply-chain or socio-environmental concerns.

That leaves the electricity grid. We estimate that electrifying 90 percent of the U.S. light-duty passenger fleet by 2050 would raise demand for electricity by up to 1,700 terawatt-hours per year—41 percent of U.S. electricity generation in 2021. This additional new demand would greatly change the shape of the consumption curve over daily and weekly periods, which means the grid and its supply would have to be remodeled accordingly.

And because the entire point of EVs is to replace fossil fuels, the grid would need more renewable sources of energy, which typically generate energy intermittently. To smooth out the supply and ensure reliability, the grid will need to add energy-storage capacity, perhaps in the form of vehicle-to-grid technologies that exploit the installed base of EV batteries. Varying the price of electricity throughout the day could also help to flatten the demand curve.

A ALL SAID, EVS present both a challenge and an opportunity. The challenge could be hard to manage if EVs are deployed too rapidly—but rapid deployment is exactly what is needed to meet climate targets. These hurdles can be overcome, but they cannot be ignored: In the end, the climate crisis will require us to electrify road transport. But this step alone cannot solve our environmental woes. We need to pursue other strategies.

We should try as much as possible, for example, to avoid motorized travel by cutting the frequency and length of car trips through better urban planning. Promoting mixed-use neighborhoods—areas that put work and residence in proximity—would allow more bicycling and walking.

Between 2007 and 2011, the city of Seville built an extensive cycling network, increasing the number of daily bike trips from about 13,000 to more than 70,000—or 6 percent of all trips. In Copenhagen, cycling accounts for 16 percent of all trips. Cities around the world are experimenting with a wide range of other supporting initiatives, such as Barcelona's superblocks, regions smaller than a neighborhood that are designed to be hospitable to walking and cycling. Congestion charges have been levied in Stockholm and London to limit car traffic. Paris has gone further, with a forthcoming private-vehicle ban. Taken together, changes in urban form can reduce transport energy demand by 25 percent, according to a recent installment of the Sixth Assessment Report from the Intergovernmental Panel on Climate Change.

We should also shift from using cars, which often have just one person inside, to less energy-intensive modes of travel, such as public transit. Ridership on buses and trains can be increased by improving connectivity, frequency, and reliability. Regional rail could supplant much intercity driving. At high occupancy, buses and trains can typically keep their emissions to below 50 grams of carbon dioxide per person per kilometer, even when powered by fossil fuels. In electrified modes, these emissions can drop to a fifth as much.

Between 2009 and 2019, Singapore's investment in mass rapid transit helped reduce the share of private vehicle transport from 45 percent to 36 percent. From 1990 to 2015, Paris slashed vehicle travel by 45 percent through sustained investment in both public transit and in infrastructure to encourage walking and other human-powered transportation.

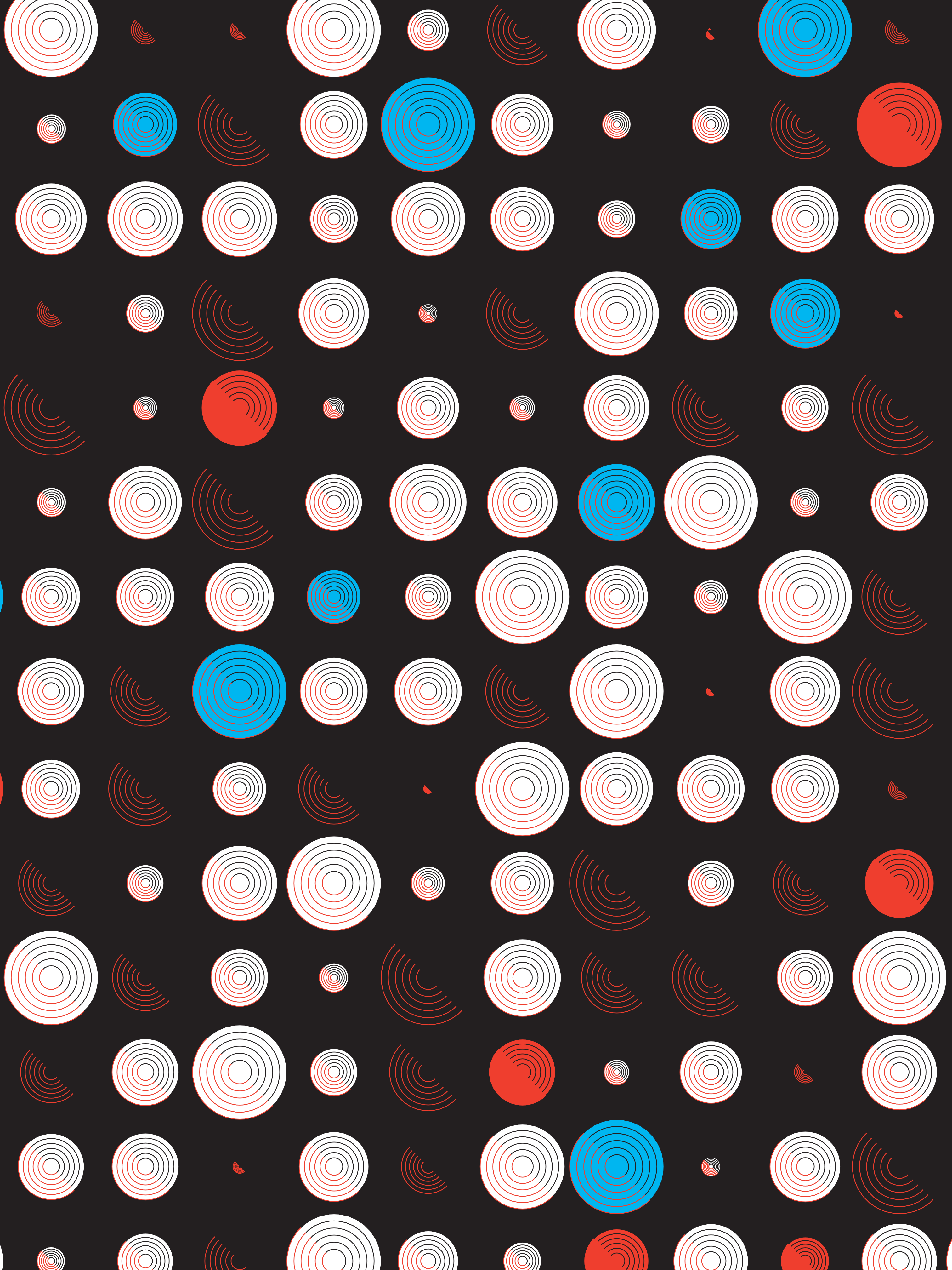
Implementing these complementary strategies could ease the transition to EVs considerably. We shouldn't forget that addressing the climate crisis requires more than just technology fixes. It also demands individual and collective action. EVs will be a huge help, but we shouldn't expect them to do the job alone. ■

To the → → → InterPlanetary File System —and Beyond!

Peer-to-peer file sharing would make the Internet far more efficient

By
→ Yiannis Psaras
→ Jorge M. Soares
→ David Dias

Illustrations by Carl De Torres



When the COVID-19 pandemic erupted in early 2020, the world made an unprecedented shift to remote work. As a precaution, some Internet providers scaled back service levels temporarily, although that probably wasn't necessary for countries in Asia, Europe, and North America, which were generally able to cope with the surge in demand caused by people teleworking (and binge-watching Netflix). That's because most of their networks were overprovisioned, with more capacity than they usually need. But in countries without the same level of investment in network infrastructure, the picture was less rosy: Internet service providers (ISPs) in South Africa and Venezuela, for instance, reported significant strain.

But is overprovisioning the only way to ensure resilience? We don't think so. To understand the alternative approach we're championing, though, you first need to recall how the Internet works.

The core protocol of the Internet, aptly named the Internet Protocol (IP), defines an addressing scheme that computers use to communicate with one another. This scheme assigns addresses to specific devices—people's computers as well as servers—and uses those addresses to send data between them as needed.

It's a model that works well for sending unique information from one point to another, say, your bank statement or a letter from a loved one. This approach made sense when the Internet was used mainly to deliver different content to different people. But this design is not well suited for the mass consumption of static content, such as movies or TV shows.

The reality today is that the Internet is more often used to send exactly the same thing to many people, and it's doing a huge amount of that now, much of which is in the form of video. The demands grow even higher as our screens obtain ever-increasing resolutions, with 4K video already in widespread use and 8K on the horizon.

The content delivery networks (CDNs) used by streaming services such as Netflix help address the problem by temporarily storing content close to, or even inside, many ISPs. But this strategy relies on ISPs and CDNs being able to make deals and deploy the required infrastructure. And it can still leave the edges of the network having to handle more traffic than actually needs to flow.

The real problem is not so much the volume of content being passed around—it's how it is being delivered, from a central source to many different far-away users, even when those users are located right next to one another.

Instead of asking a particular provider, "Please send me this file," your machine asks the network, "Who can send me this file?"

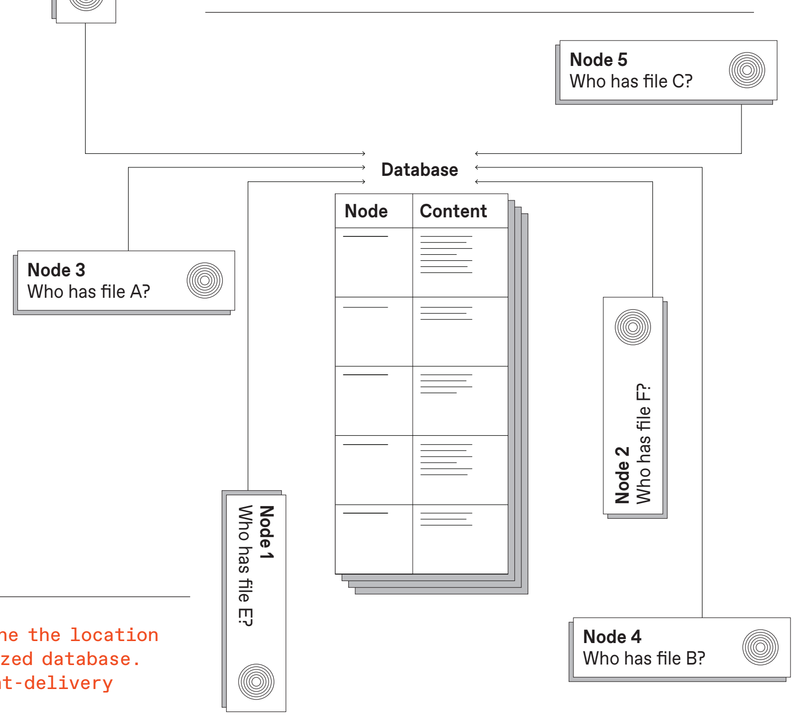
A more efficient distribution scheme in that case would be for the data to be served to your device from your neighbor's device in a direct peer-to-peer manner. But how would your device even know whom to ask? Welcome to the InterPlanetary File System (IPFS).

The InterPlanetary File System gets its name because, in theory, it could be extended to share data even between computers on different planets of the solar system. For now, though, we're focused on rolling it out for just Earth!

The key to IPFS is what's called content addressing. Instead of asking a particular provider, "Please send me this file," your machine asks the network, "Who can send me this file?" It starts by querying peers: other computers in the user's vicinity, others in the same house or office, others in the same neighborhood, others in the same city—expanding progressively outward to globally distant locations, if need be, until the system finds a copy of what you're looking for.

These queries are made using IPFS, an alternative to the Hypertext Transfer Protocol (HTTP), which powers the World Wide Web. Building on the principles of peer-to-peer networking and content-based addressing, IPFS allows for a decentralized and distributed network for data storage and delivery.

The benefits of IPFS include faster and more-efficient distribution of content. But they don't stop there. IPFS can also improve security with content-integrity checking so that data cannot be tampered with by intermediary actors. And with IPFS, the network can continue operating even if the connection to the originating server is cut or if the service that



One scheme used by peer-to-peer systems to determine the location of a file is to keep that information in a centralized database. Napster, the first large-scale peer-to-peer content-delivery system, used this approach.

initially provided the content is experiencing an outage—particularly important in places with networks that work only intermittently. IPFS also offers resistance to censorship.

To understand more fully how IPFS differs from most of what takes place online today, let's take a quick look at the Internet's architecture and some earlier peer-to-peer approaches.

As mentioned above, with today's Internet architecture, you request content based on a server's address. This comes from the protocol that underlies the Internet and governs how data flows from point to point, a scheme first described by Vint Cerf and Bob Kahn in a 1974 paper in the *IEEE Transactions on Communications* and now known as the Internet Protocol. The World Wide Web is built on top of the Internet Protocol. Browsing the Web consists of asking a specific machine, identified by an IP address, for a given piece of data.

The process starts when a user types a URL into the address bar of the browser, which takes the hostname portion and sends it to a Domain Name System (DNS) server. That DNS server returns a corresponding numerical IP address. The user's browser will then connect to the IP address and ask for the Web page located at that URL.

In other words, even if a computer in the same building has a copy of the desired data, it will neither see the request, nor would it be able to match it to the copy it holds because the content does not have an intrinsic identifier—it is not content-addressed.

A content-addressing model for the Internet would give data, not devices, the leading role. Requesters would ask for the content explicitly, using a unique identifier (akin to the DOI number of a journal article or the ISBN of a book), and the Internet would handle forwarding the request to an available peer that has a copy.

The major challenge in doing so is that it would require changes to the core Internet infrastructure, which is owned and operated by thousands of ISPs worldwide, with no central authority able to control what they all do. While this distributed architecture is one of the Internet's greatest strengths, it makes it nearly impossible to make fundamental changes to the system, which would then break things for many of the people using it. It's often very hard even to implement incremental improvements. A good example of the difficulty encountered when introducing change is IPv6, which expands the number of possible IP addresses. Today, almost 25 years after its introduction, it still hasn't reached 50 percent adoption.

A way around this inertia is to implement changes at a higher layer of abstraction, on top of existing Internet protocols, requiring no modification to the underlying networking software stacks or intermediate devices.

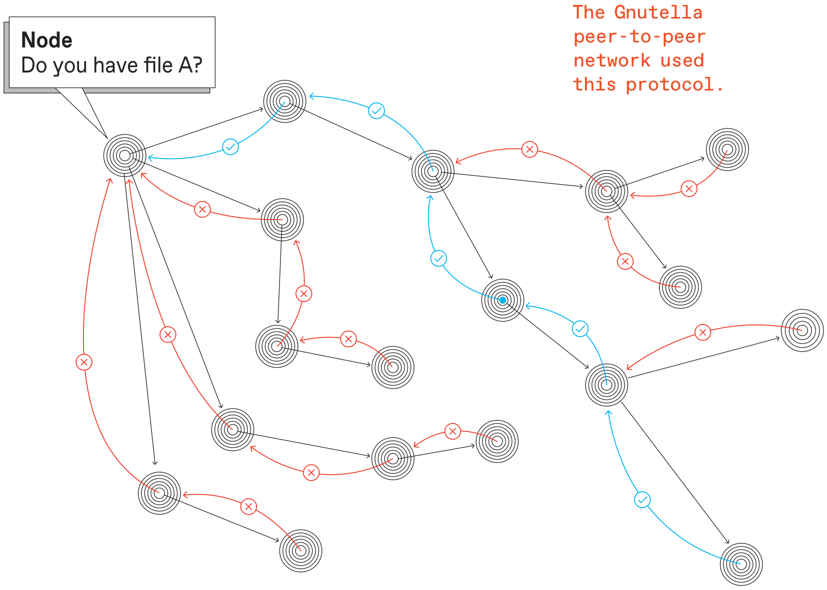
Other peer-to-peer systems besides IPFS, such as BitTorrent and Freenet, have tried to do this by introducing systems that can operate in parallel with the World Wide Web, albeit often with Web interfaces. For example, you can click on a Web link for the BitTorrent tracker associated with a file, but this process typically requires that the tracker data be passed off to a separate application from your Web browser to handle the transfers. And if you can't find a tracker link, you can't find the data.

Freenet also uses a distributed peer-to-peer system to store content, which can be requested via an identifier and can even be accessed using the Web's HTTP protocol. But Freenet and IPFS have different aims: Freenet has a strong focus on anonymity and manages the replication of data in ways that serve that goal but lessen performance and user control. IPFS provides flexible, high-performance sharing and retrieval mechanisms but keeps control over data in the hands of the users.

We designed IPFS as a protocol to upgrade the Web and not to create an alternative version. It is designed to make the Web better, to allow people to work offline, to make links permanent, to be faster and more secure, and to make it as easy as possible to use.

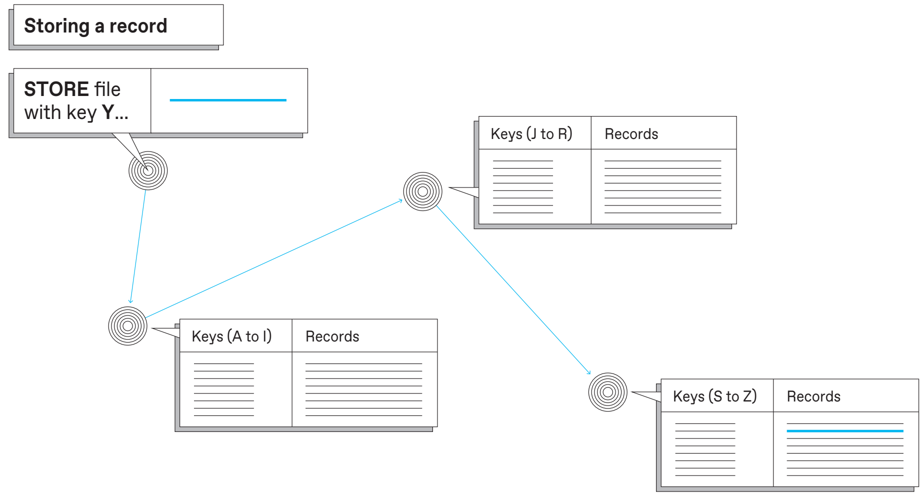
IPFS started in 2013 as an open-source project supported by Protocol Labs and built by a vibrant community and ecosystem with hundreds of organizations and thousands of developers. IPFS is built on a strong foundation of previous work in peer-to-peer (P2P) networking and content-based addressing.

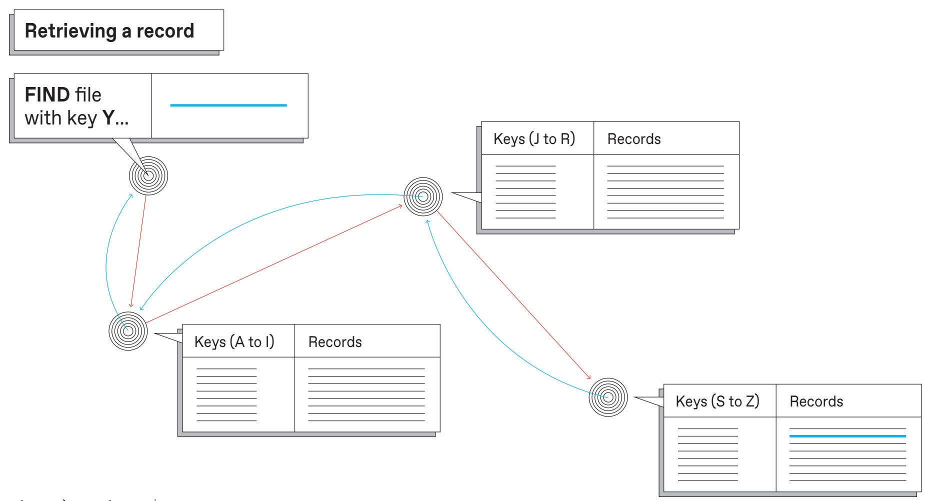
The core tenet of all P2P systems is that users simultaneously participate as clients



Another approach to finding a file in a peer-to-peer network is called query flooding. The node seeking a file broadcasts a request for it to all nodes to which it is attached. If the node receiving the request does not have the file [red], it forwards the request to all the nodes to which it is attached until finally a node with the file passes a copy back to the requester [blue]. The Gnutella peer-to-peer network used this protocol.

To keep track of which nodes hold which files, the InterPlanetary File System uses what's called a distributed hash table. In this simplified view, three nodes hold different parts of a table that has two columns: One column (Keys) contains hashes of the stored files; the other column (Records) contains the files themselves. Depending on what its hashed key is, a file gets stored in the appropriate place [top]—depicted here as though the system checked the first letter of hashes and stored different parts of the alphabet in different places. The actual algorithm for distributing files is more complex, but the concept is similar. Retrieving a file is efficient because it's possible to locate the file according to what its hash is [bottom].





(which request and receive files from others) and as servers (which store and send files to others). The combination of content addressing and P2P provides the right ingredients for fetching data from the closest peer that holds a copy of what's desired—or more correctly, the closest one in terms of network topology, though not necessarily in physical distance.

To make this happen, IPFS produces a fingerprint of the content it holds (called a hash) that no other item can have. That hash can be thought of as a unique address for that piece of content. Changing a single bit in that content will yield an entirely different address. Computers wanting to fetch this piece of content broadcast a request for a file with this particular hash.

Because identifiers are unique and do not change, people often refer to IPFS as the "Permanent Web." And with identifiers that never change, the network will be able to find a specific file as long as some computer on the network stores it.

Name persistence and immutability inherently provide another significant property: verifiability. Having the content and its identifier, a user can verify that what was received is what was asked for and has not been tampered with, either in transit or by the provider. This not only improves security but also helps safeguard the public record and prevent history from being rewritten.

You might wonder what would happen with content that needs to be updated to include fresh information, such as a Web page. This is a valid concern and IPFS does have a suite of mechanisms that would point users to the most up-to-date content.

The world had a chance to observe how content addressing worked in April 2017 when the government of Turkey blocked access to Wikipedia because an article on the platform described Turkey as a state that sponsored terrorism. Within a week, a full copy of the Turkish version of Wikipedia was added to IPFS, and it remained accessible to people in the country for the nearly three years that the ban continued.

A similar demonstration took place half a year later, when the Spanish government tried to suppress an independence referendum in Catalonia, ordering ISPs to block related websites. Once again, the information remained available via IPFS.

IPFS is an open, permissionless network: Any user can join and fetch or provide content. Despite numerous open-source success stories, the current Internet is heavily based on closed platforms, many of which adopt lock-in tactics but also offer users great convenience. While IPFS can provide improved efficiency, privacy, and security, giving this decentralized platform the level of usability that people are accustomed to remains a challenge.

You see, the peer-to-peer, unstructured nature of IPFS is both a strength and a weakness. While CDNs have built sprawling infrastructure and advanced techniques to provide high-quality service, IPFS nodes are operated by end users. The network therefore relies on their behavior—how long their computers are online, how good their connectivity is, and what data they decide to cache. And often those things are not optimal.

One of the key research questions for the folks working at Protocol Labs is how to keep the IPFS network resilient despite shortcomings in the nodes that make it up—or even when those nodes exhibit selfish or malicious behavior. We'll need to overcome such issues if we're to keep the performance of IPFS competitive with conventional distribution channels.

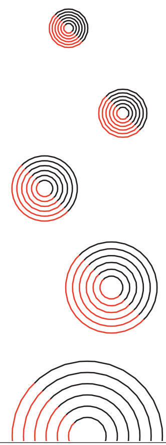
You may have noticed that we haven't yet provided an example of an IPFS address. That's because hash-based addressing results in URLs that aren't easy to spell out or type.

For instance, you can find the Wikipedia logo on IPFS by using the following address in a suitable browser: ipfs://QmRW3V9znzFW9M5FYbitSEvd-5dQrPWGvPvgQD6LM22Tv8D/. That long string can be thought of as a digital fingerprint for the file holding that logo.

There are other content-addressing schemes that use human-readable naming, or hierarchical, URL-style naming, but each comes with its own set of trade-offs. Finding practical ways to use human-readable names with IPFS would go a long way toward improving user-friendliness. It's a goal, but we're not there yet.

Protocol Labs, where we work, has been tackling these and other technical, usability, and societal issues for most of the last decade. Over this time, we have been seeing rapidly increasing adoption of IPFS, with its network size doubling year over year. Scaling up at such speeds brings many challenges. But that's par for the course when your intent is changing the Internet as we know it.

Widespread adoption of content addressing and IPFS should help the whole Internet ecosystem. By empowering users to request exact content and verify that they received it unaltered, IPFS will improve trust and security. Reducing the duplication of data moving through the network and procuring it from nearby sources will let ISPs provide faster service at lower cost. Enabling the network to continue providing service even when it becomes partitioned will make our infrastructure more resilient to natural disasters and other large-scale disruptions.



Reducing the duplication of data moving through the network and procuring it from nearby sources will let ISPs provide faster service at lower cost.

But is there a dark side to decentralization? We often hear concerns about how peer-to-peer networks may be used by bad actors to support illegal activity. These concerns are important but sometimes overstated.

One area where IPFS improves on HTTP is in allowing comprehensive auditing of stored data. For example, thanks to its content-addressing functionality and, in particular, to the use of unique and permanent content identifiers, IPFS makes it easier to determine whether certain content is present on the network, and which nodes are storing it. Moreover, IPFS makes it trivial for users to decide what content they distribute and what content they stop distributing (by merely deleting it from their machines).

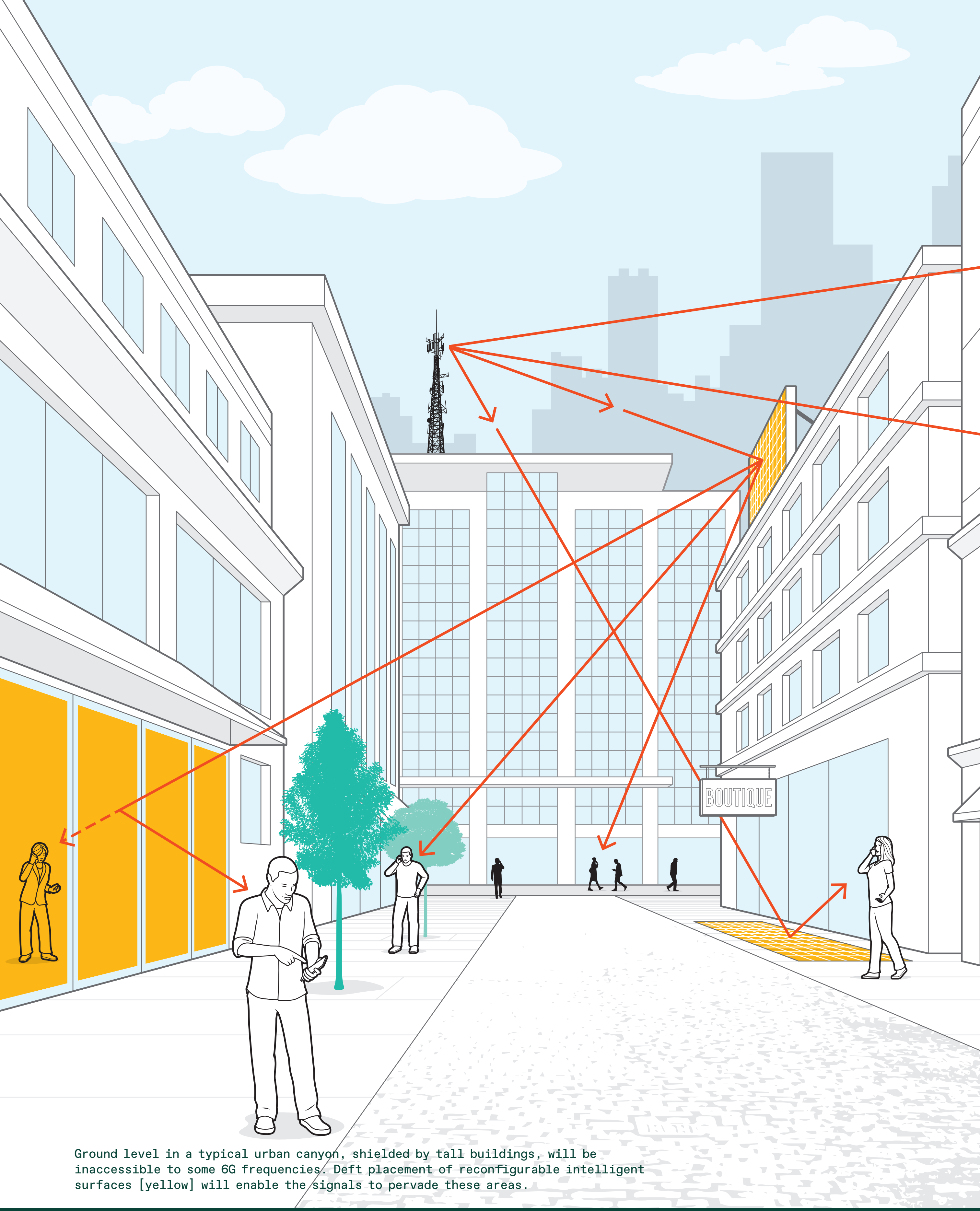
At the same time, IPFS provides no mechanisms to allow for censorship, given that it operates as a distributed P2P file system with no central authority. So there is no actor with the technical means to prohibit the storage and propagation of a file or to delete a file from other peers' storage. Consequently, censorship of unwanted content cannot be technically enforced, which represents a safeguard for users whose freedom of speech is under threat. Lawful requests to take down content are still possible, but they need to be addressed to the users actually storing it, avoiding commonplace abuses (like illegitimate DMCA takedown requests) against which large platforms have difficulties defending.

Ultimately, IPFS is an open network, governed by community rules, and open to everyone. And you can become a part of it today! The Brave browser ships with built-in IPFS support, as does Opera for Android. There are browser extensions available for Chrome and Firefox, and IPFS Desktop makes it easy to run a local node. Several organizations provide IPFS-based hosting services, while others operate public gateways that allow you to fetch data from IPFS through the browser without any special software.

These gateways act as entries to the P2P network and are important to bootstrap adoption. Through some simple DNS magic, a domain can be configured so that a user's access request will result in the corresponding content being retrieved and served by a gateway, in a way that is completely transparent to the user.

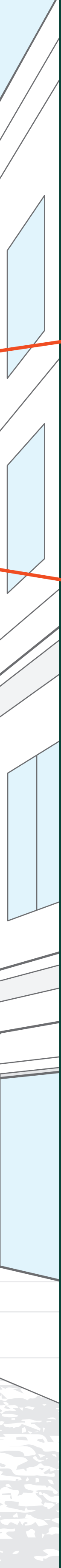
So far, IPFS has been used to build varied applications, including systems for e-commerce, secure distribution of scientific data sets, mirroring Wikipedia, creating new social networks, sharing cancer data, blockchain creation, secure and encrypted personal-file storage and sharing, developer tools, and data analytics.

You may have used this network already: If you've ever visited the Protocol Labs site (Protocol.ai), you've retrieved pages of a website from IPFS without even realizing it! ■



Ground level in a typical urban canyon, shielded by tall buildings, will be inaccessible to some 6G frequencies. Deft placement of reconfigurable intelligent surfaces [yellow] will enable the signals to pervade these areas.

Illustration by Chris Philpot



6G's Metamaterials Solution

There's plenty of bandwidth available if we use reconfigurable intelligent surfaces By Marios Poulakis

For all the tumultuous revolution in wireless technology over the past several decades, there have been a couple of constants. One is the overcrowding of radio bands, and the other is the move to escape that congestion by exploiting higher and higher frequencies. And today, as engineers roll out 5G and plan for 6G wireless, they find themselves at a crossroads: After years of designing superefficient transmitters and receivers, and of compensating for the signal losses at the end points of a radio channel, they're beginning to realize that they are approaching the practical limits of transmitter and receiver efficiency. From now on, to get high performance as we go to higher frequencies, we will need to engineer the *wireless channel itself.* But how can we possibly engineer and control a wireless environment, which is determined by a host of factors, many of them random and therefore unpredictable?

Perhaps the most promising solution, right now, is to use reconfigurable intelligent surfaces. These are planar structures typically ranging in size from about 100 square centimeters to about 5 square meters or more, depending on the frequency and other factors. These surfaces use advanced substances called metamaterials to reflect and refract electromagnetic waves. Thin two-dimensional metamaterials, known as metasurfaces, can be designed to sense the local electromagnetic environment and tune the wave's key properties, such as its amplitude, phase, and polarization, as the wave is reflected or refracted by the surface. In fact, these metasurfaces can be programmed to make these changes dynamically, reconfiguring the signal in real time in response to changes in the wireless channel. As the waves fall on such a surface, it can respond in real time, altering the incident waves' direction so as to strengthen the channel. Think of reconfigurable intelligent surfaces as the next evolution of the repeater concept.

That's important, because as we move to higher frequencies, the propagation characteristics become more "hostile" to the signal. The wireless channel varies constantly depending on surrounding objects. At 5G and 6G frequencies, the wavelength is vanishingly small compared to the size of buildings, vehicles, hills, trees, and rain. Lower-frequency waves diffract around or through such obstacles, but higher-frequency signals are absorbed, reflected, or scattered. Basically, at these frequencies, the line-of-sight signal is about all you can count on.

This article was jointly produced by *IEEE Spectrum* and *Proceedings of the IEEE* with similar versions published in both publications.

Such problems help explain why the topic of reconfigurable intelligent surfaces (RIS) is one of the hottest in wireless research. The hype is justified. A landslide of R&D activity and results has gathered momentum over the last several years, set in motion by the development of the first digitally controlled metamaterials almost 10 years ago.

RIS prototypes are showing great promise at scores of laboratories around the world. And yet one of the first major projects, the European-funded Visorsurf, began just five years ago and ran until 2020. The first public demonstrations of the technology occurred in late 2018, by NTT Docomo in Japan and Metawave, of Carlsbad, Calif.

Today, hundreds of researchers in Europe, Asia, and the United States are working on applying RIS to produce programmable and smart wireless environments. Vendors such as Huawei, Ericsson, NEC, Nokia, Samsung, and ZTE are working alone or in collaboration with universities. And major network operators, such as NTT Docomo, Orange, China Mobile, China Telecom, and BT are all carrying out substantial RIS trials or have plans to do so. This work has repeatedly demonstrated the ability of RIS to greatly strengthen signals in the most problematic bands of 5G and 6G.

To understand how RIS improves a signal, consider the electromagnetic environment. Traditional cellular networks consist of scattered base stations that are deployed on masts or towers, and on top of buildings and utility poles in urban areas. Objects in the path of a signal can block it, a problem that becomes especially bad at 5G's higher frequencies, such as the millimeter-wave bands between 24.25 and 52.6 gigahertz. And it will only get worse if communications companies go ahead with plans to exploit subterahertz bands, between 90 and 300 GHz, in 6G networks. Here's why. With 4G and similar lower-frequency bands, reflections from surfaces can actually strengthen the received signal, as reflected signals combine. However, as we move higher in frequencies, such multipath effects become much weaker or disappear entirely. The reason is that surfaces that appear smooth to a longer-wavelength signal are relatively rough to a shorter-wavelength signal. So rather than reflecting off such a surface, the signal simply scatters.

One solution is to use more powerful base stations or to install more of them throughout an area. But that strategy can double costs, or worse. Repeaters or relays can also improve coverage but here, too, the costs can be prohibitive. RIS, on the other hand, promises greatly improved coverage at just marginally higher cost.

The key feature of RIS that makes it attractive in comparison with these alternatives is its nearly passive nature. The absence of amplifiers to boost the signal means that an RIS node can be powered with just a battery and a small solar panel.

RIS functions like a very sophisticated mirror, whose orientation and curvature can be adjusted in order to focus and redirect a signal in a specific direction [see opening page]. But rather than physically moving or reshaping the mirror, you electronically alter its surface so that it changes key properties of the incoming electromagnetic wave, such as the phase.

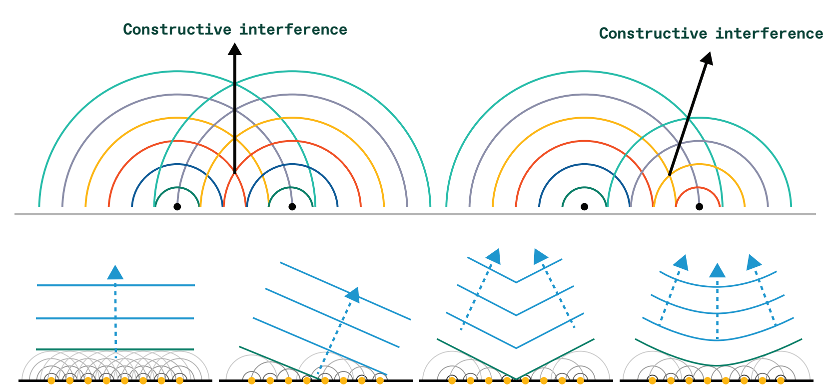
That's what the metamaterials do. This emerging class of materials exhibits properties beyond (from the Greek *meta*) those of natural materials, such as anomalous reflection or refraction. The materials are fabricated using ordinary metals and electrical insulators, or dielectrics. As an electromagnetic wave impinges on a metamaterial, a predetermined gradient in the material alters the phase and other characteristics of the wave, making it possible to bend the wave front and redirect the beam as desired.

An RIS node is made up of hundreds or thousands of metamaterial elements called unit cells. Each cell consists of metallic and dielectric layers along with one or more switches or other tunable components. A typical structure includes an upper metallic patch with switches, a biasing layer, and a metallic ground layer separated by dielectric substrates. By controlling the biasing—the voltage between the metallic patch and the ground layer—you can switch each unit cell on or off and thus control how each cell alters the phase and other characteristics of an incident wave.

To control the direction of the larger wave reflecting off the entire RIS, you synchronize all the unit cells to create patterns of constructive and destructive interference in the larger reflected waves [see illustration below]. This interference pattern reforms the incident beam and sends it in a particular direction determined by the pattern. This basic operating principle, by the way, is the same as that of a phased-array radar.

An RIS has other useful features. Even without an amplifier, an RIS manages to provide substantial gain—about 30 to 40 decibels relative to isotropic (dBi)—depending on the size of the surface and the frequency. That's because the gain of

Beamforming by constructive and destructive interference

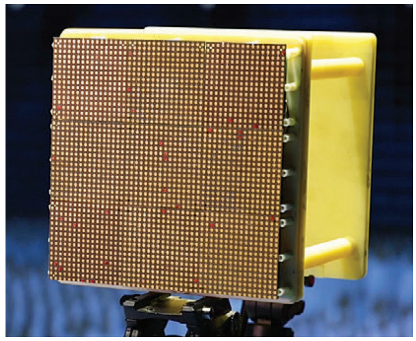


A reconfigurable intelligent surface comprises an array of unit cells. In each unit cell, a metamaterial alters the phase of an incoming radio wave, so that the resulting waves interfere with one another [above, top illustration]. Precisely controlling the patterns of this constructive and destructive interference allows the reflected wave to be redirected [bottom], greatly improving signal coverage.

an antenna is proportional to the antenna's aperture area. An RIS has the equivalent of many antenna elements covering a large aperture area, so it has higher gain than a conventional antenna does.

All the many unit cells in an RIS are controlled by a logic chip, such as a field-programmable gate array with a microcontroller, which also stores the many coding sequences needed to dynamically tune the RIS. The controller gives the appropriate instructions to the individual unit cells, setting their state. The most common coding scheme is simple binary coding, in which the controller toggles the switches of each unit cell on and off. The unit-cell switches are usually semiconductor devices, such as PIN diodes or field-effect transistors.

The important factors here are power consumption, speed, and flexibility, with the control circuit usually being one of the most power-hungry parts of an RIS. Reasonably efficient RIS implementations today have a total power consumption of around a few watts to a dozen watts during the switching state of reconfiguration, and much less in the idle state.



An experimental reconfigurable intelligent surface with 2,304 unit cells was tested at Tsinghua University, in Beijing, last year.

To deploy RIS nodes in a real-world network, researchers must first answer three questions: How many RIS nodes are needed? Where should they be placed? And how big should the surfaces be? As you might expect, there are complicated calculations and trade-offs.

Engineers can identify the best RIS positions by planning for them when the base station is designed. Or it can be done afterward by identifying, in the coverage map, the areas of poor signal strength. As for the size of the surfaces, that will depend on the frequencies (lower frequencies require larger surfaces) as well as the number of surfaces being deployed.

To optimize the network's performance, researchers rely on simulations and measurements. At Huawei Sweden, where I work, we've had a lot of discussions about the best placement of RIS units in urban environments. We're using a proprietary platform, called the Coffee Grinder Simulator, to simulate an RIS installation prior to its construction and deployment. We're partnering with CNRS Research and CentraleSupélec, both in France, among others.

In a recent project, we used simulations to quantify the performance improvement gained when multiple RIS were deployed in a typical urban 5G network. As far as we know, this was the first large-scale, system-level attempt to gauge RIS performance in that setting. We optimized the RIS-augmented wireless coverage through the use of efficient deployment algorithms that we developed. Given the locations of the base stations and the users, the algorithms were designed to help us select the optimal three-dimensional locations and sizes of the RIS nodes from among thousands of possible positions on walls, roofs, corners, and so on. The output of the software is an RIS deployment map that maximizes the number of users able to receive a target signal.

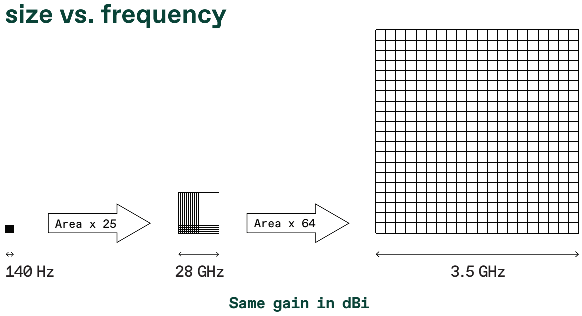
Of course, the users of special interest are those at the edges of the cell-coverage area, who have the worst signal reception. Our results showed big improvements in coverage and data rates at the cell edges, and also for users with better signal reception, especially in the millimeter band.

We also investigated how potential RIS hardware trade-offs affect performance. Simply put, every RIS design requires compromises—such as digitizing the responses of each unit cell into binary phases and amplitudes—in order to construct a less complex and cheaper RIS. But it's important to know whether a design compromise will create additional beams to undesired directions or cause interference to other users. That's why we studied the impact of network interference due to multiple base stations, reradiated waves by the RIS, and other factors.

Not surprisingly, our simulations confirmed that both larger RIS surfaces and larger numbers of them improved overall performance. But which is preferable? When we factored in the costs of the RIS nodes and the base stations, we found that in general a smaller number of larger RIS nodes, deployed further from a base station and its users to provide coverage to a larger area, was a particularly cost-effective solution.

The size and dimensions of the RIS depend on the operating frequency. We found that a small number of rectangular

Reconfigurable intelligent surfaces size vs. frequency



FREQUENCY BAND	NUMBER OF RIS ELEMENTS	RIS SIZE IN SQUARE METERS	RIS ANTENNA GAIN IN DECIBELS RELATIVE TO ISOTROPIC
3.5 gigahertz (sub-6GHz)	2,500 (50 × 50)	4.41	38.8
	10,000 (100 × 100)	17.98	44.9
	22,000 (150 × 150)	40.83	48.4
28 GHz (millimeter wave)	2,500 (50 × 50)	0.068	38.8
	10,000 (100 × 100)	0.28	44.9
	22,000 (150 × 150)	0.64	48.4
140 GHz (subterahertz)	2,500 (50 × 50)	0.0025	38.8
	10,000 (100 × 100)	0.021	44.9
	22,000 (150 × 150)	0.026	48.4

The sizes of typical reconfigurable intelligent surfaces vary from a few square centimeters to several square meters. For a given wavelength, larger area means higher gain. SOURCE: MARIOS POULAKIS

TSINGHUA UNIVERSITY

RIS nodes, each around 4 meters wide for C-band frequencies (3.5 GHz) and around half a meter wide for millimeter-wave band (28 GHz), was a good compromise, and could boost performance significantly in both bands. This was a pleasant surprise: RIS improved signals not only in the millimeter-wave (5G high) band, where coverage problems can be especially acute, but also in the C band (5G mid).

To extend wireless coverage indoors, researchers in Asia are investigating a really intriguing possibility: covering room windows with transparent RIS nodes. Experiments at NTT Docomo and at Southeast and Nanjing universities, both in China, used smart films or smart glass. The films are fabricated from transparent conductive oxides (such as indium tin oxide), graphene, or silver nanowires and do not noticeably reduce light transmission. When the films are placed on windows, signals coming from outside can be refracted and boosted as they pass into a building, enhancing the coverage inside.

Planning and installing the RIS nodes is only part of the challenge. For an RIS node to work optimally, it needs to have a configuration, moment by moment, that is appropriate for the state of the communication channel in the instant the node is being used. The best configuration requires an accurate and instantaneous estimate of the channel. Technicians can come up with such an estimate by measuring the "channel impulse response" between the base station, the RIS, and the users. This response is measured using pilots, which are reference signals known beforehand by both the transmitter and the receiver. It's a standard technique in wireless communications. Based on this estimation of the channel, it's possible to calculate the phase shifts for each unit cell in the RIS.

The current approaches perform these calculations at the base station. However, that requires a huge number of pilots, because every unit cell needs its own phase configuration. There are various ideas for reducing this overhead, but so far none of them are really promising.

The total calculated configuration for all of the unit cells is fed to each RIS node through a wireless control link. So each RIS node needs a wireless receiver to periodically collect the instructions. This of course consumes power, and it also means that the RIS nodes are fully dependent on the base station, with unavoidable—and unaffordable—overhead and the need for continuous control. As a result, the whole system requires a flawless and complex orchestration of base stations and multiple RIS nodes via the wireless-control channels.

We need a better way. Recall that the "I" in RIS stands for intelligent. The word suggests real-time, dynamic control of the surface from within the node itself—the ability to learn, understand, and react to changes. We don't have that now. Today's RIS nodes cannot perceive, reason, or respond; they only execute

Typical Parameters of 5G Network Nodes

	SYSTEM ROLE	RELATIVE COST / IMPLEMENTATION COMPLEXITY	POWER CONSUMPTION	RANGE (COVERAGE)	NETWORK PERFORMANCE IMPROVEMENT	INSTALLATION REQUIREMENTS
RECONFIGURABLE INTELLIGENT SURFACES Controlled radio-frequency signal reflection or refraction	Optional add-on component for system improvement	Low	Low (1-5 watts)	Medium/ high	Medium/ high	Low: • Low space (low-profile, lightweight, potential large 2D surface area) • No power supply • No backhaul
ACTIVE REPEATERS RF signal amplification and retransmission	Optional add-on component for system improvement	Medium	Medium (20-60 W)	Medium	Medium	Medium: • Medium space • Power supply • No backhaul
INTEGRATED ACCESS AND BACKHAUL (IAB) (SMALL-CELL BASE STATION) RF signal regeneration and retransmission (base-station capabilities)	Optional add-on component for system improvement	High	Medium/ high (100-300 W)	Medium/ high	High	Medium: • Medium space • Power supply • Backhaul
MASSIVE MIMO BASE STATION (MACROCELL BASE STATION) RF signal transceiving (full base-station capabilities)	Essential component allowing denser deployment of base stations and system improvement	Very high	Very high (1,000-2,000 W)	Very high	Very high	Medium/high: • Medium/high space • Power supply • Backhaul

remote orders from the base station. That's why my colleagues and I at Huawei have started working on a project we call Autonomous RIS (AutoRIS). The goal is to enable the RIS nodes to autonomously control and configure the phase shifts of their unit cells. That will largely eliminate the base-station-based control and the massive signaling that either limit the data-rate gains from using RIS, or require synchronization and additional power consumption at the nodes. The success of AutoRIS might very well help determine whether RIS will ever be deployed commercially on a large scale.

Of course, it's a rather daunting challenge to integrate into an RIS node the necessary receiving and processing capabilities while keeping the node lightweight and low power. In fact, it will require a huge research effort. For RIS to be commercially competitive, it will have to preserve its low-power nature.

With that in mind, we are now exploring the integration of an ultralow-power AI chip in an RIS, as well as the use of extremely efficient machine-learning models to provide the intelligence. These smart models will be able to produce the output RIS configuration based on the received data about the channel, while at the same time classifying users according to their contracted services and their network operator. Integrating AI into the RIS will also enable other functions, such as dynamically predicting upcoming RIS configurations and grouping users by location or other behavioral characteristics that affect the RIS operation.

Intelligent, autonomous RIS won't be necessary for all situations. For some areas, a static RIS, with occasional reconfiguration—perhaps a couple of times per day or less—will be entirely adequate. In fact, there will undoubtedly be a range of deployments from static to fully intelligent and autonomous. Success will depend on not just efficiency and high performance but also ease of integration into an existing network.

The real test case for RIS will be 6G. The coming generation of wireless is expected to embrace autonomous networks and smart environments with real-time, flexible, software-defined, and adaptive control. Compared with 5G, 6G is expected to provide much higher data rates, greater coverage, lower latency, more intelligence, and sensing services of much higher accuracy. At the same time, a key driver for 6G is sustainability—we'll need more energy-efficient solutions to achieve the "net zero" emission targets that many network operators are striving for. RIS fits all of those imperatives.

Start with massive MIMO, which stands for multiple-input multiple-output. This foundational 5G technique uses multiple antennas packed into an array at both the transmitting and receiving ends of wireless channels, to send and receive many signals at once and thus dramatically boost network capacity. However, the desire for higher data rates in 6G will demand even more massive MIMO, which will require many more radio-frequency chains to work and will be power-hungry and costly to operate. An energy-efficient and less costly alternative will be to place multiple low-power RIS nodes between massive MIMO base stations and users as we have described in this article.

The millimeter-wave and subterahertz 6G bands promise to unleash staggering amounts of bandwidth, but only if we can surmount a potentially ruinous range problem without resorting to costly solutions, such as ultradense deployments of base stations or active repeaters. My opinion is that only RIS will be able to make these frequency bands commercially viable at a reasonable cost.

Only RIS will be able to make the highest 6G-frequency bands commercially viable at a reasonable cost.

The communications industry is already touting sensing—high-accuracy localization services as well as object detection and posture recognition—as an important possible feature for 6G. Sensing would also enhance performance. For example, highly accurate localization of users will help steer wireless beams efficiently. Sensing could also be offered as a new network service to vertical industries such as smart factories and autonomous driving, where detection of people or cars could be used for mapping an environment; the same capability could be used for surveillance in a home-security system. The large aperture of RIS nodes and their resulting high resolution mean that such applications will be not only possible but probably even cost effective.

And the sky is not the limit. RIS could enable the integration of satellites into 6G networks. Typically, a satellite uses a lot of power and has large antennas to compensate for the long-distance propagation losses and for the modest capabilities of mobile devices on Earth. RIS could play a big role in minimizing those limitations and perhaps even allowing direct communication from satellite to 6G users. Such a scheme could lead to more efficient satellite-integrated 6G networks.

As it transitions into new services and vast new frequency regimes, wireless communications will soon enter a period of great promise and sobering challenges. Many technologies will be needed to usher in this next exciting phase. None will be more essential than reconfigurable intelligent surfaces. ■

*The author wishes to acknowledge the help of **Ulrik Imberg** in the writing of this article.*

OTHER CHARACTERISTICS

- Dynamic beamforming capabilities
- Nearly passive (no amplification)
- Full-duplex (no self-interference)
- Energy efficient
- Low TCO
- Control overhead (additional signaling)
- Minimal signal processing
- Not standardized yet

- Signal amplification
- Standardized
- Standard-transparent (no additional signaling)
- Duplex loss or isolation requirement
- Interference amplification
- Static or no beamforming capabilities
- Power supply required

- Dynamic beamforming and other base-station capabilities
- Standardized
- Duplex loss or isolation requirement
- Additional latency
- Power supply required

- Full base-station capabilities
- Many simultaneous signals and users (many transceivers, large bandwidth)
- Standardized
- Power supply required

RIT | Rochester Institute of Technology

Tenure-Track Assistant Professor

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

Apply online at https://apptrkr.com/3474406
Keyword Search: 7166BR or 7167BR.
Inquiries may be sent to Dr. Sohail Dianat (sadeee@rit.edu).

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

NYU | TANDON SCHOOL OF ENGINEERING

Tenure-Track Faculty, Department of Electrical and Computer Engineering, New York University, Brooklyn, NY

The Department of Electrical and Computer Engineering at the NYU Tandon School of Engineering invites applications for three tenure-track Assistant Professor positions, to start on September 1, 2023.

NYU Tandon has nationally renowned research centers that ECE faculty lead and participate in, including the Center for Cybersecurity, NYU Wireless, the Center for Advanced Technology in Telecommunications (CATT), and the Center for Urban Science + Progress (CUSP). ECE faculty are also highly engaged in several school-wide research initiatives (including communications, cybersecurity, data science/AI/robotics, and sustainability).

We are looking to hire in the broad areas of Computer Engineering (including architecture, circuits, hardware devices, and systems), Control Systems, Energy/Power, and Wireless Communications (including antennas/radio propagation, circuits, networking, and theory). The applicant should have a Ph.D. degree in Electrical Engineering, Computer Engineering, or a closely related discipline. The individual should have the potential to develop a strong record of scholarship, leadership, curricular innovation, and an excellent funding record. The applicant should demonstrate excellence in research, teaching and mentoring.

For application instructions and more information about NYU Tandon, see: **https://apply.interfolio.com/112470**.

Applications should be submitted by **December 15, 2022** for full consideration.

JOINT INSTITUTE 交大密西根学院

The University of Michigan-Shanghai Jiao Tong University (UM-SJTU) Joint Institute invites applications for tenure-track or tenured positions at all levels (Assistant, Associate, and Full) in Electrical and Computer Engineering. Candidates should hold a Ph.D. in electrical and computer engineering, computer science, or a closely related field. The Joint Institute particularly seeks candidates in the areas of robotics, image processing and computer vision, micro-electronics, wireless communication theory, computer architecture, and database. The candidates are expected to establish vigorous research programs and contribute to undergraduate and graduate education. Salaries are highly competitive and commensurate with qualifications and experience.

The UM-SJTU Joint Institute receives strong support from both partner universities and the Chinese government. It offers B.S., M.S., and Ph.D. degrees in Electrical and Computer Engineering and related fields. Its ECE program is the first ABET accredited ECE program in the mainland of China, and its students are among China's best. The Joint Institute models itself after the world class U.S. research universities, in terms of its tenure review and promotion system, academic environment, research program, and undergraduate curriculum. Its official language is English.

For full consideration, please send a CV, statement of research interests and teaching goals, copies of three key publications, and the names and contact information of five references, as a single PDF file, to the Search Committee of the Joint Institute at ji-search@sjtu.edu.cn. More information is available at ***http://ji.sjtu.edu.cn/***.

UNIVERSITY OF MINNESOTA — Driven to Discover®

Electrical and Computer Engineering (ECE), University of Minnesota Twin Cities (**https://ece.umn.edu/**) invites applications for two faculty positions. The first is in the area of photonics, electronics, or magnetics. We seek an experimentalist working in micro- and nano-fabrication who will make extensive use of the facilities in the Minnesota Nano Center. The second is in computer engineering, with knowledge and expertise in hardware design and algorithms of emerging computing platforms, including AI/ML, quantum, and quantum-inspired computing.

ECE is committed to fostering a culturally and academically diverse community; candidates who will actively contribute to this commitment – both in identity and professional vision - are particularly encouraged to apply. An earned doctorate in an appropriate discipline is required. Rank and salary will be commensurate with qualifications and experience. Applications will be considered as they are received, and will be accepted until the positions are filled, but for full consideration, please apply by the priority deadline of **December 15, 2022**.

To be considered for a position, candidates must apply online. Application instructions and additional information can be found at **https://z.umn.edu/ecefacultyjobs**

The University of Minnesota is an equal opportunity educator and employer.

BY ALLISON MARSH

Past Forward

This coin-operated *Pong* prototype proved wildly popular when it debuted in November 1972.

MARK RICHARDS/COMPUTER HISTORY MUSEUM

In Praise of Pong

Fifty years ago this month, a new game debuted at Andy Capp's Tavern in Sunnyvale, Calif. *Pong* was the simplest of arcade games—represented on screen as a moving dot, two vertical lines with which to hit the dot, and squared-off digits to keep score. And yet people lined up to play, emptying their pockets of quarters when their turn came. No one was more surprised by the game's success than the company that developed it, Atari. Nolan Bushnell and Ted Dabney had recently started Atari and hired a young engineer named Allan Alcorn as their third employee. Under the guise of a contract for General Electric, Bushnell asked Alcorn to design a game that simulated table tennis and could be played on a television. The result was *Pong*, and with it, Atari kick-started an industry. ■

FOR MORE ON THE HISTORY OF PONG, SEE spectrum.ieee.org/pastforward-nov2022


LIVE
LIFE ON
YOUR
TERMS.
INSURANCE
THAT FITS
YOU.
Group Term Life Insurance
engineered for you.