FOR THE TECHNOLOGY INSIDER
MAY 2023

IEEE Spectrum

The Internet Evangelist

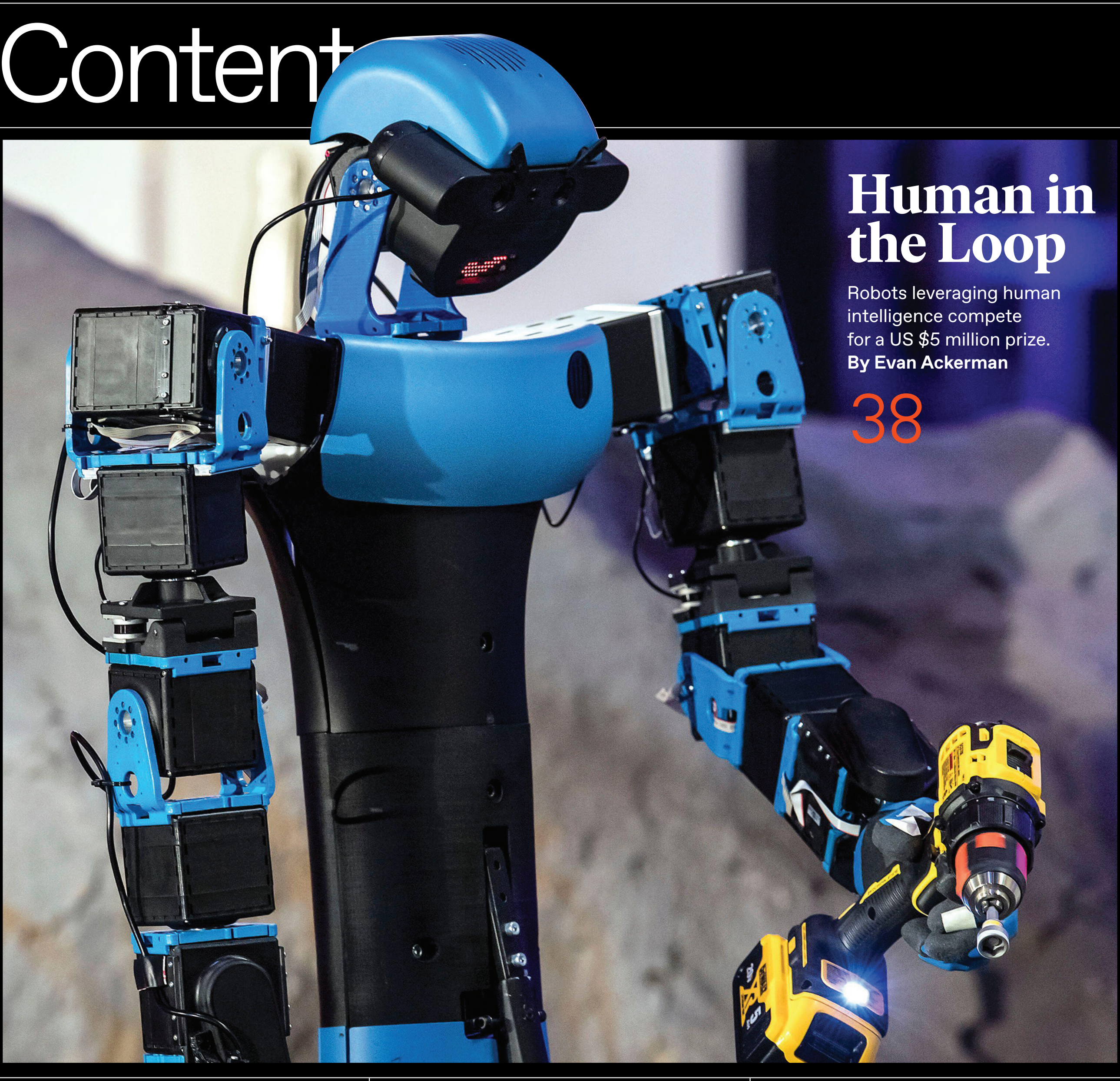
Vint Cerf's Crusade to Network the World

2023 IEEE Medal of Honor Recipient

IEEE

VOLUME 60 / ISSUE 5 MAY 2023

Contents



ON THE COVER:
Photo by Peter Adams

ZUMA/ALAMY

BY HARRY GOLDSTEIN

The Portrait Artist

How Tekla S. Perry chronicles consequential careers

Senior Editor Tekla S. Perry with 2017 IEEE Medal of Honor recipient Kees Immink in Singapore.

Only a handful of engineering luminaries know what it feels like to get "The Call" from *IEEE Spectrum* senior editor Tekla S. Perry. It's a rite of passage, a vital part of the process of being selected by IEEE as a Medal of Honor recipient. This year it was Mr. Internet himself, the inimitable and always dapper Vint Cerf, and you can read Perry's profile of him on page 24.

Perry wrote her first article for *Spectrum* in 1980. She was first assigned the task of profiling some of the most influential and consequential engineers of the last quarter century in 1995, starting with MOH recipient Lotfi A. Zadeh, the father of fuzzy logic.

"I remember being relieved that he could explain fuzzy logic in a way I could easily understand, because I didn't really understand it previously," Perry recalls."That made future MOH interviews less stressful going in, because I had faith that the medalists would be able to make me understand their technologies."

There's an art to writing a good profile, and it all starts with the interview. Perry prefers to meet in person to establish a human connection. That rapport lets her subjects open up, giving readers a glimpse of their personalities, motivations, and attitudes.

"So much of journalism is a quick hit—getting a specific piece of information about a specific topic—it's a real treat to be able to have the time and opportunity to let the conversation go down a very long and meandering path, with these little lightbulbs that go on along the way when I hear a great anecdote or a revealing comment," Perry says. "I generally leave feeling I actually connected with the person, that it was more than just a Q&A."

Perry usually starts her interviews with questions that range from things that happened in the distant past to things that happened last night. The warm-up is aimed at opening drawers in the memory that perhaps haven't been opened in a while. "Then I start the more chronological part of the interview by asking about their parents and their careers, which makes me feel like I am starting to get to know them beyond their achievements," she says. "It's also likely something they haven't talked about in a long time, so I don't get a rehearsed response. But that's why the interview usually takes at least three hours."

While Perry usually gets MOH recipients to reveal things about themselves and their contributions, she occasionally learns something about herself, too. Nick Holonyak, MOH recipient in 2003 for the development of the red LED, actually changed the way Perry thinks about her ethnic roots.

"My family identified as 'Russian' even though my grandparents emigrated under the label of 'Austrian' in the early 1900s, and the town they came from now is in Poland, near Ukraine and Slovakia," Perry says.

"Holonyak started asking me about my origins because of my name, and quickly updated my ethnic understanding: He and I are both Carpatho-Rusyn—along with Andy Warhol and Sandra Dee, he told me—and we were likely more closely related to each other than either of us was to anyone else in southern Illinois at the time. He recommended a few books, I followed up on those, and my somewhat confusing family history fell into place."

And sometimes she gets a memorable meal out of the deal. Last year's recipient, Asad Madni, told Perry she had to try his wife's cooking. "She ended up making lunch for us," Perry recounts. "It was a revelation. I've always liked Indian food, but nothing I've ever tasted before or since comes anywhere close to that meal."

If you're fortunate enough to get The Call next year, gird yourself for a journey of discovery that will take you back to places you haven't thought about in years, with a companion who will bring the rest of us along for the ride. ■

VITAL STATISTICS

Name: Tekla S. Perry

Residence: Palo Alto, Calif.

Education: B.A., journalism, Michigan State University

Family: Eric Nee (husband); Alex, Nadya, and Mischa Nee (children)

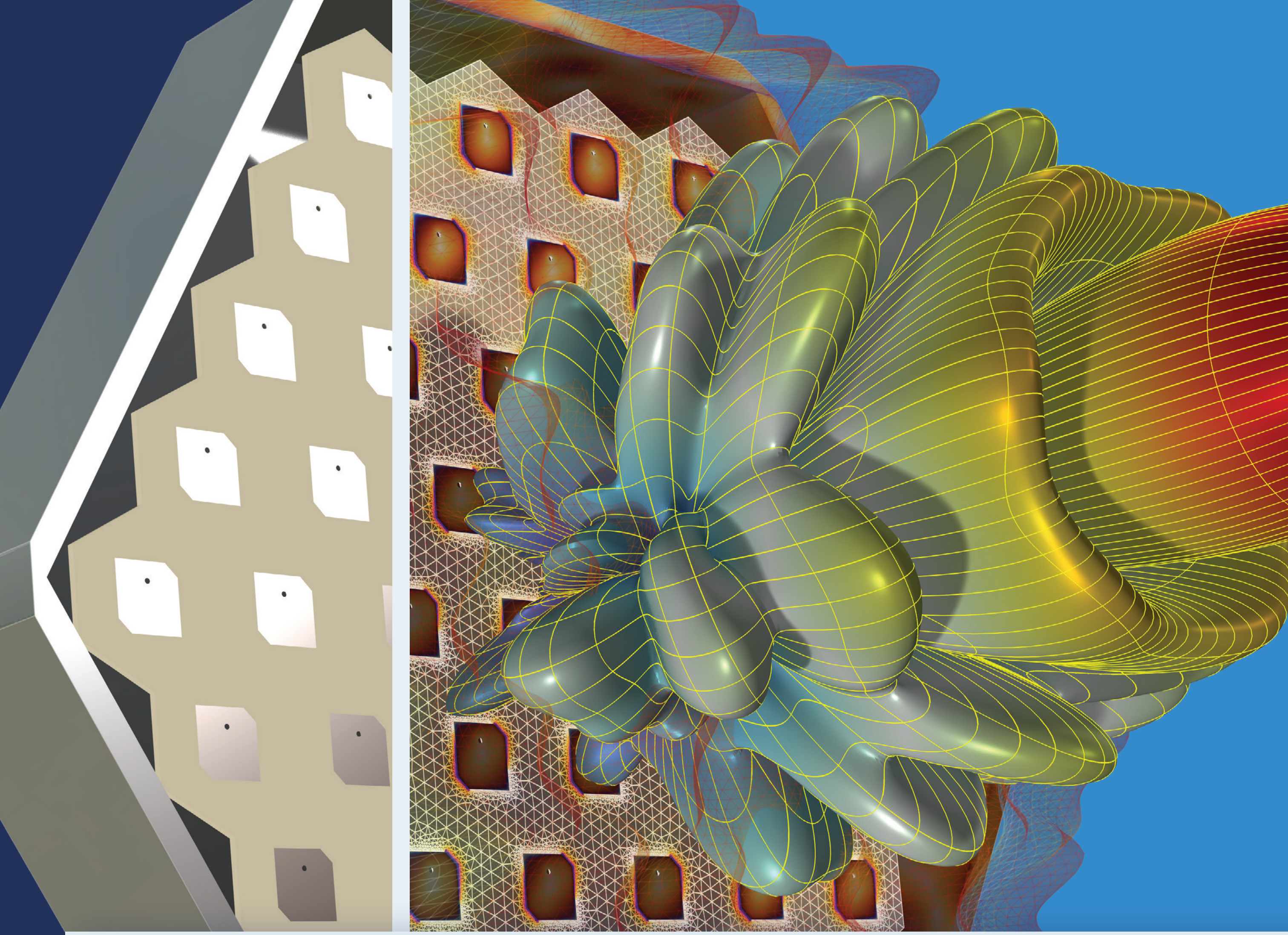
First job: Community datebook writer at WBRW Radio in Somerville, N.J., at age 16

Oddest job: Boardwalk barker

Favorite composer: Bruce Springsteen

Leisure activities: Urban hiking, acrylic painting

PORTRAIT BY SERGIO ALBIAC; TEKLA S. PERRY



PETER ADAMS

Adams is a tech entrepreneur turned photographer who specializes in capturing the tech industry. For this issue, he photographed IEEE Medal of Honor recipient Vint Cerf [p. 24]. Turns out Cerf's fans occasionally send him copies of a portrait Adams took of him at a 2018 shoot, with a request that Cerf sign them. The autographed photos often end up on eBay, he says. After this photo shoot, Cerf handed Adams a business card—with his signature. Adams promised to keep it.

ROEL BAETS

Baets, an IEEE Fellow, is a professor at Ghent University, where he leads the photonics research group and is affiliated with Imec. He and his coauthors—Joris Van Campenhout, program director for optical input/output at Imec; Bernardette Kunert, a principal scientist at Imec; and Gunther Roelkens, also a professor at Ghent and affiliated with Imec—are researching ways to put compound semiconductor lasers on silicon chips. The group describes four options on page 32 of this issue.

MATTHEW HUTSON

The recent "deepfake" video of Volodymyr Zelenskyy purportedly asking his troops to surrender reinforces the continuing need for tools that can detect false and misleading content. In this issue, Hutson, a New York City–based journalist, surveys a variety of the counteralgorithms now being developed by computer scientists [p. 7]. Some of these have shown success in detecting the "signatures" of generative AI, but Hutson is skeptical. "Although I'm impressed that deepfake detection performs so well in many tests," he says, "I think eventually generation will win out."

JAMES L. PELKEY

In this issue, Pelkey and coauthors Loring G. Robbins and Andrew L. Russell describe five events that shaped today's Internet [p. 46]. Pelkey, creator of The History of Computer Communications website, died shortly before this article was published. In 2017, he teamed up with Russell (a history professor and officer-in-charge of SUNY Polytechnic Institute in Albany and Utica, N.Y.) and Robbins (a Hawaii-based technology writer) to transform that website into a book, on which their article is based.

IEEE SPECTRUM (ISSN 0018-9235) is published monthly by The Institute of Electrical and Electronics Engineers, Inc. All rights reserved. © 2023 by The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016-5997, U.S.A. Volume No. 60, Issue No. 5. The editorial content of IEEE Spectrum magazine does not represent official positions of the IEEE or its organizational units. Canadian Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40013087. Return undeliverable Canadian addresses to: Circulation Department, IEEE Spectrum, Box 1051, Fort Erie, ON L2A 6C7. Cable address: ITRIPLEE. Fax: +1 212 419 7570. INTERNET: spectrum@ieee.org. ANNUAL SUBSCRIPTIONS: IEEE Members: $21.40 included in dues. Libraries/institutions: $399. POSTMASTER: Please send address changes to IEEE Spectrum, % Coding Department, IEEE Service Center, 445 Hoes Lane, Box 1331, Piscataway, NJ 08855. Periodicals postage paid at New York, NY, and additional mailing offices. Canadian GST #125634188. Printed at 3401 N. Heartland Dr, Liberty, MO 64068, U.S.A. Printed at Spenta Multimedia Pvt. Ltd., Plot 15, 16 & 21/1, Village Chikhloli, Morivali, MIDC, Ambernath (West), Dist. Thane. IEEE Spectrum circulation is audited by BPA Worldwide. IEEE Spectrum is a member of the Association of Business Information & Media Companies, the Association of Magazine Media, and Association Media & Publishing. IEEE prohibits discrimination, harassment, and bullying. For more information, visit https://www.ieee.org/about/corporate/governance/p9-26.html.

IEEE

YOU

JOINED THE GROUP

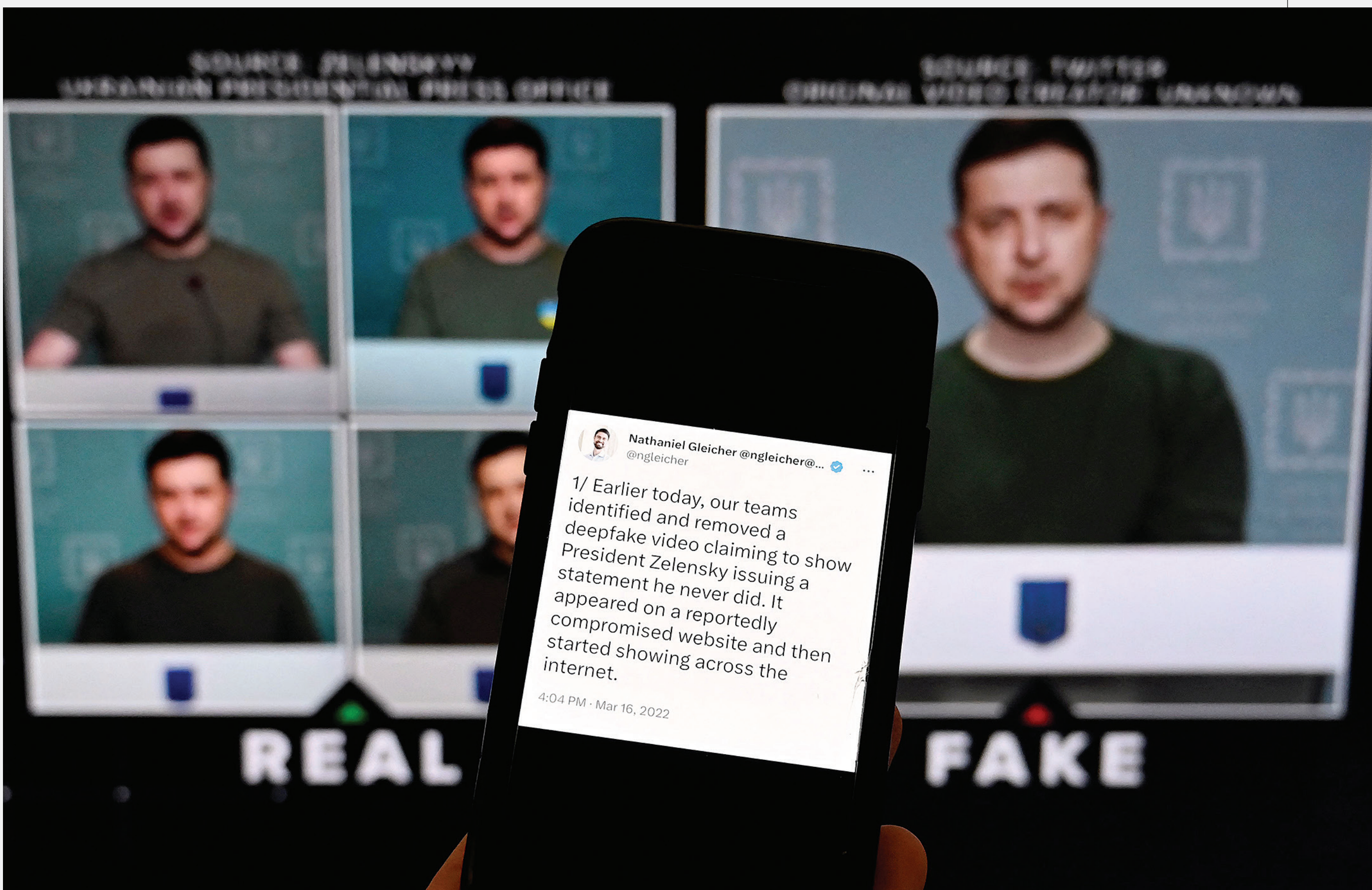
Member Discounts

YOU

HAVE EXCLUSIVE ADVANTAGES

Visit ieee.org/discounts to access your exclusive member savings.

News



Last year, Meta's head of security policy, Nathaniel Gleicher, announced the takedown [see smartphone screen] of a deepfake video of Ukrainian president Volodymyr Zelenskyy calling on his soldiers to lay down their weapons. The real/fake comparison appeared on the website Verify.

OLIVER DOULIERY/AFP/GETTY IMAGES

ARTIFICIAL INTELLIGENCE

The Deepfake Arms Race Heats Up › Detectors mostly work—until deepfakers devise the next workaround

BY MATTHEW HUTSON

In March 2022, a video appeared online that seemed to show Ukraine's president, Volodymyr Zelenskyy, asking his troops to lay down their arms in the face of Russia's invasion. The video—created with the help of artificial intelligence—was poor in quality and the ruse was quickly debunked, but one day a more sophisticated effort could have serious geopolitical consequences.

That's why, as computer scientists continue to devise better algorithms for generating video, audio, images, and text, they're also creating counteralgorithms to detect synthetic content. Today such detection typically looks at the subtle signatures of particular generation tools; in the future it will increasingly rely on underlying physical and biological signals, which are hard for AI to imitate.

It's also entirely possible that both the generators and the detectors of these digital phantoms will become locked in an arms race. "In this respect, it never ends," says Luisa Verdoliva, a computer scientist at the University of Naples Federico II, in Italy.

In November, Intel unveiled a platform for analyzing videos called the Real-Time Deepfake Detector. (The term "deepfake" derives from the use of deep learning—an area of AI that uses many-layered artificial neural networks.) Likely customers include social-media companies, broadcasters, and nongovernmental organizations that can distribute detectors to the general public, says Ilke Demir, a researcher at Intel. According to Intel, the system, consisting of both Intel hardware and software running remotely on an Internet-based platform, can analyze 72 video streams at once per Intel processor. Eventually, the company says, its platform will provide several detection tools. At present it uses one detector, which Demir cocreated with Umur Çiftçi, at Binghamton University, N.Y., called FakeCatcher.

FakeCatcher studies color changes in faces to infer blood flow, a process called photoplethysmography (PPG). The researchers designed the software to focus on certain patterns of color in certain facial regions. "PPG signals are special in the sense that they're everywhere on your skin," Demir says. "It's not just eyes or lips. And changing illumination does not eliminate them. But any generative operation actually eliminates them, because the type of noise that they're adding messes up the spatial, spectral, and temporal correlations." Put another way, FakeCatcher confirms that color fluctuates naturally over time as the heart pumps blood, and that there's coherence across facial regions. In one early test of the technology, originally published in *IEEE Transactions on Pattern Analysis and Machine Intelligence* in July 2020, the detector achieved 91 percent accuracy, nearly nine percentage points better than the next-best system.

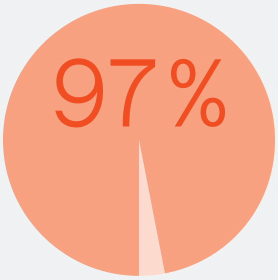
When a new detection method comes along, someone can often train a generative algorithm to fool it. Demir says FakeCatcher can't easily be reverse engineered for the sake of training generators—since fine-grained blood-flow patterns inside a living person are still difficult to simulate artificially.

Michael Goebel, a Ph.D. student in electrical engineering at the University of California at Santa Barbara and a paper coauthor, notes that there's a spectrum of detection methods. "At one extreme, you have very unconstrained methods that are just pure deep learning," meaning they use all the data available. "At the other extreme, you have methods that do things like analyze gaze. Ours is kind of in the middle." Their system, called PhaseForensics, focuses on lips and extracts information about motion at various frequencies before providing this digested data to a neural network. "By using the motion features themselves, we kind of hard-code in some of what we want the neural network to learn," Goebel says.

Audio deepfakes have also become a problem. In January, someone uploaded a fake clip of the actress Emma Watson reading part of Hitler's *Mein Kampf*. Here, too, researchers are on the case. In one approach, scientists at the University of Florida developed a system that models the human vocal tract. The model digitally constructs a sound-producing airway, with the bore of the tube varying along the length, and uses the construction to predict sounds. Given a suspicious audio sample, it can determine if it is biologically plausible. The paper reports accuracy on one data set of around 99 percent.

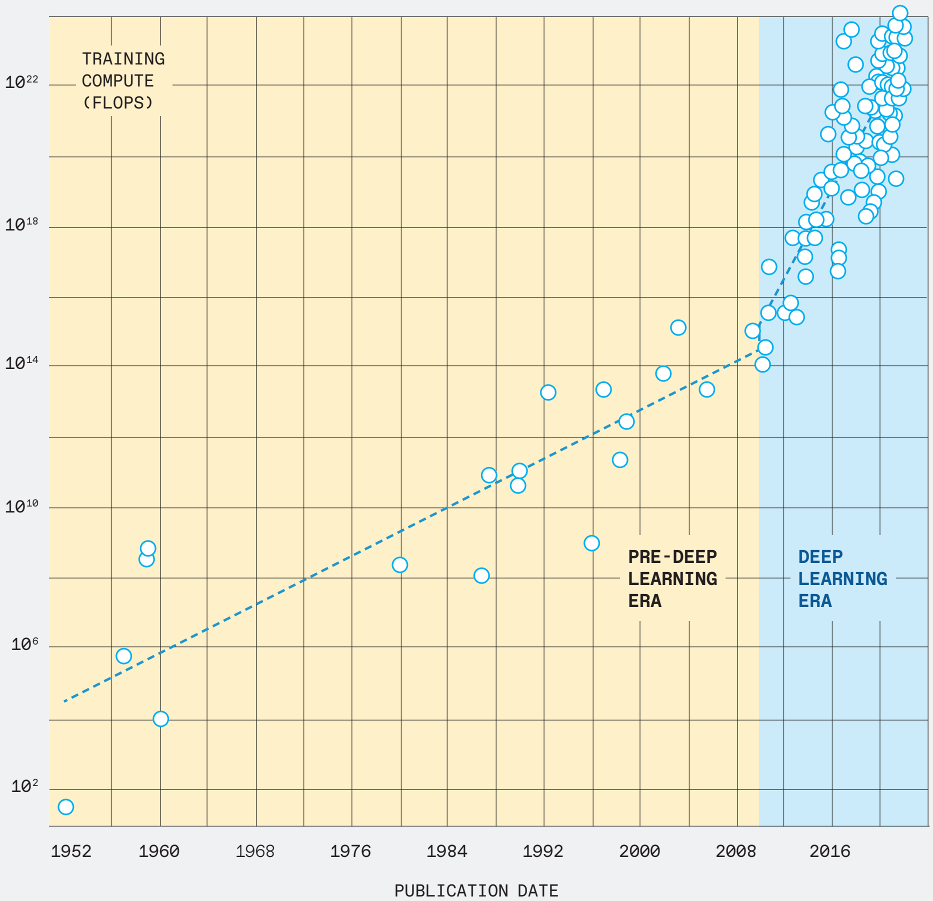
The algorithm doesn't need to have sampled deepfake audio from a particular generation algorithm in order to defend against it. Verdoliva, at the University of Naples, has developed another such method. During training, the algorithm learns to find biometric signatures of speakers. When implemented, it takes recordings of a given speaker, uses its learned techniques to find the biometric signature, then looks for that signature in a questionable recording.

Verdoliva's group has also worked on identifying generated and manipulated images, whether altered by AI or by Adobe Photoshop. The group trained a system called TruFor on photos from 1,475 cameras, and it learned to recog-



Detection accuracy of Intel video FakeCatcher with facial blood flow and facial motion algorithms

It is entirely possible that both the generators and the detectors of these digital phantoms will become locked in an arms race.



THE TREND BEHIND THE TREND—REDOUBLED AI GROWTH

As shown in this graph, drawn from a recent historical review of the field, the AI compute power required to train milestone machine-learning systems has steadily increased over time, starting in the early 1950s and carrying forward another six decades. The pace of progress roughly matched Moore's Law, with a doubling in system capabilities (or the halving of training compute power required) every 17 to 29 months. However, the researchers find that in recent years, the deep-learning revolution—powered in part by the rapid growth of computer centers run by the world's biggest AI companies—has accelerated that progress. Since 2010, the AI doubling time has been gaining in speed, now occurring every 4 to 9 months.

SOURCE: PARAMETER, COMPUTE AND DATA TRENDS IN MACHINE LEARNING BY JAIME SEVILLA, PABLO VILLALOBOS, JUAN FELIPE CERÓN, MATTHEW BURTELL, LENNART HEIM, AMOGH B. NANJAJJAR, ANSON HO, TAMAY BESIROGLU, AND MARIUS HOBBHAHN; 2021

nize the kinds of signatures left by such cameras. Looking at a new image, the system can detect mismatches between parts of the picture.

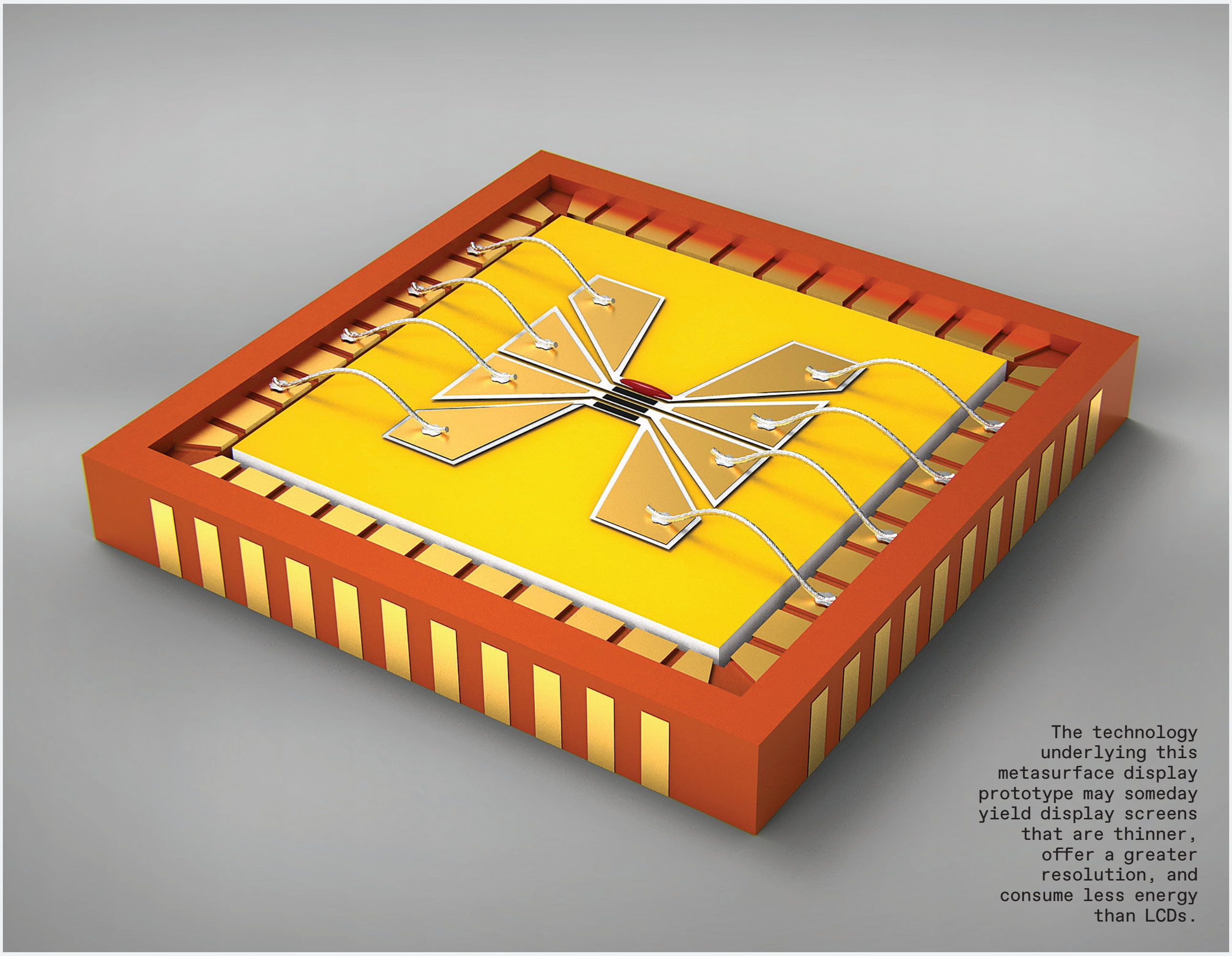
High-school students are now regularly in the game of using AI to generate content, prompting the text-generating system ChatGPT to write essays. One solution is to ask the creators of such systems, called large language models, to watermark the generated text. Researchers at the University of Maryland recently proposed a method that randomly creates a set of green-listed vocabulary words; it then gives a slight preference to those words in its own compositions. If you know this (secret) list of green-listed words, you can look for a predominance of them in a piece of text to build a case that it came from the algorithm. One problem is that there is an increasing number of powerful language models, and we can't expect all of them to watermark their output.

One Princeton student, Edward Tian, created a tool called GPTZero that looks for signs that a text was written by ChatGPT even without watermarking. Humans tend to make more surprising word choices and fluctuate more in sentence length. But GPTZero appears to have limits. One user putting GPTZero to a small test found that it correctly flagged 10 out of 10 AI-authored texts as synthetic, but that it also falsely flagged 8 of 10 human-written ones.

Synthetic-text detection will likely lag far behind detection in other mediums. According to Tom Goldstein, an associate professor of computer science at the University of Maryland who coauthored the watermarking paper, that's because people use language in so many different ways and because there isn't usually a lot of data. An essay might have a few hundred words, versus the million pixels of a picture, and words are discrete, unlike subtle variations in pixel color.

There's a lot at stake in detecting synthetic content. It can be used to sway teachers, courts, or electorates. It can produce humiliating or intimidating adult content. The mere idea of deepfakes can erode trust in mediated reality. Demir calls this future "dystopian." Short-term, she says, we need detection algorithms. Long-term, we also need protocols that establish provenance, perhaps involving watermarks or blockchains.

"People would like to have a magic tool that is able to do everything perfectly and even explain it," Verdoliva says of detection methods. Nothing like that exists—and it probably never will. "You need multiple tools." Even if a quiver of detectors can take down deepfakes, the content will have at least a brief life online before it disappears. It will have an impact. So, Verdoliva says, technology alone can't save us. Instead, people need to find ways to adapt to the new, nonreality-filled reality. ■



The technology underlying this metasurface display prototype may someday yield display screens that are thinner, offer a greater resolution, and consume less energy than LCDs.

ANDREY MIROSHNICHENKO/UNSW CANBERRA

CONSUMER ELECTRONICS

Metasurface Displays Target the LCD

> "Invisibility cloak" physics powers these superhigh-speed prototypes

BY CHARLES Q. CHOI

The physics that makes microscopic "invisibility cloaks" possible may lead to ultrathin displays with 10 times the resolution and half the energy consumption of an LCD screen, which is today's dominant display technology.

The research, detailed online on 22 February in the journal *Light: Science & Applications*, describes a prototype four-pixel device operating at less than 5 volts. It can switch how much visible and near-infrared light it transmits by a factor of nine in just 625 microseconds, a speed that is compatible with 1,600 frames per second—nearly 30 times as fast as present-day video.

LCD technology depends on liquid-crystal cells that are illuminated by a backlight. Polarizers in front and behind the pixels filter light waves based on their polarity, or the direction in which they vibrate. The liquid-crystal cells can rotate in the direction in which these filters are oriented to switch light transmissions on and off. "However, improvements on LCD technologies are now mostly just incremental," says Eric Virey, senior industry analyst at market-research firm Yole Intelligence, in Lyon, France.

The new technology exploits metasurfaces, engineered materials with features not generally found in nature, such as the ability to bend rays of light in unexpected ways. Bend them far enough and they will wrap around an object, creating the illusion that the object does not exist. This is the trick behind proposed invisibility cloaks, which might hide objects from light, sound, heat, and other types of waves.

Optical metamaterials manipulate light with structures that have repeating patterns at scales smaller than the wavelengths they influence. However, their structures are typically static. This is an obstacle to many applications, such as displays, that require changeable optical properties.

In the new study, researchers experimented with electrically tunable metasurfaces compatible with standard CMOS production techniques. These rely on the significant change in the optical properties of silicon that comes with a change in temperature.

The core of the new device consists of a 155-nanometer-thick film in which a precise array of holes has been made, each one from 78 to 101 nm in width. This metasurface is encapsulated by transparent strips of indium tin oxide, which serve as electrically driven heaters. Varying the heat changes how those arrays transmit light.

This new work "enables a fast, efficient, and compact way to tune the response of metasurfaces, which advances the field," says electrical engineer Andrea Alù at the City University of New York (CUNY) Graduate Center, who did not take part in this research.

"I find this work quite remarkable, especially the ability to dynamically modify the state of the metasurface, and do it at such high speed," says Virey, who also did not participate in the new study. "It shows that there's a lot of still unexplored potential applications and properties of metasurfaces, and we're probably just scratching the surface."

"Display makers might be happy to find a new technology that could give their aging LCD fabs a second life."

—ERIC VIREY, SENIOR DISPLAY ANALYST AT FRENCH MARKET-RESEARCH FIRM YOLE INTELLIGENCE

The scientists note that a key advantage of this new approach is stability. "Silicon nanostructures are known for durability, which is one reason they are still the most popular material in the microchip industry," says cosenior study author Dragomir Neshev, a professor of physics at Australian National University, in Canberra. "We have run our prototype samples many times over a few months and did not experience any degradation."

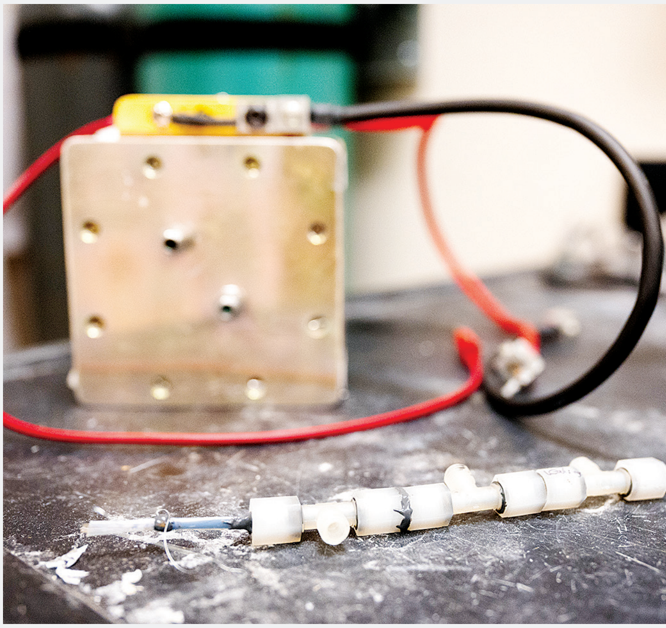
The new metasurfaces can do the job that liquid crystals now do without any need for polarizers. This can lead to major improvements because the polarizers cause the loss of half the light emitted by a display's backlight before it reaches the viewer's eyes.

With minimal modifications, current production lines for LCD displays can be updated to replace liquid-crystal pixels with metasurface versions, says study coauthor Mohsen Rahmani, a professor of engineering at Nottingham Trent University, in England. "Our metasurface pixels are compatible with current silicon-chipmaking technologies, which keep the production costs low," he says.

Promising as metasurface displays are, Virey cautions that organic LED (OLED) displays, which are currently the main rivals of LCDs, remain competitive because they do not need liquid-crystal layers. But that advantage is undercut by the fact that OLEDs are "expensive and have a short life-span," says Rahmani.

The researchers now hope to optimize their metasurface device by tinkering with heating, cooling, and the electrical input. Machine-learning techniques could also help design smaller, thinner, and more efficient metasurface displays, they add. Alù, of CUNY, notes that smaller pixel sizes are also desirable.

The scientists aim to build a prototype big enough to generate images within the next five years. They hope to integrate their technology into flat screens available to the public within 10 years. Given how LCD manufacturers have spent more than US $100 billion on current fabs, "display makers might be happy to find a new technology that could give their aging LCD fabs a second life," Virey says. ■



ENERGY

Building a Better Flow Battery for Grid Storage ›

New designs will make wind and solar power more practical

BY PRACHI PATEL

By substituting tubular shapes for the traditionally flat electrodes and membranes in a flow-battery cell, researchers have shrunk the cell size by 75 percent. This could cut the cost of an important front-runner in the effort to store renewable energy on the grid.

Unlike lithium-ion batteries, which store energy in solid electrodes, flow batteries store chemical energy in liquid electrolytes that sit in tanks. This stored charge is converted into an electric current (and vice versa) in a power module that is made up of a stack of electrochemical cells.

In each cell, the positive and negative electrodes are separated by a membrane that allows certain ions to move through when the battery is charging and discharging. To operate the battery, two electrolytes are pumped through either side of the electrochemical cell.

In theory, the amount of charge that flow batteries can store can be easily scaled up to megawatt-hours by increasing the size of the electrolyte tanks. Flow batteries are also safer and longer lasting than lithium-ion batteries. They remain costly, though, with a capital cost of around US $800 per kilowatt-hour, more than twice that of lithium-ion batteries. “But they can be much cheaper, and our work accelerates this process,” says Nian Liu, a chemical and biomolecular engineering professor at Georgia Tech.

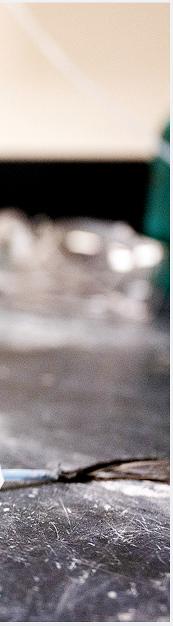
Liu and colleagues focused on redesigning the power module, which makes up 40 percent of the cost of a flow battery. Besides the flat components, these modules now contain inactive parts, such as frames, rubber gaskets, bipolar plates, and flow distributors, which take up space and thus bring down the cell’s power density. “They also make it expensive,” Liu says, “and they are not necessary.”

The team’s design eliminates these components. A 0.65-millimeter-wide tube, made of the ion-transporting material Nafion, acts as the membrane. The researchers thread a carbon-fiber cathode through the tube, place four of these tubes inside a larger, 2-millimeter plastic tube, and then push a zinc wire through to serve as the anode. During operation, the team pumps one electrolyte through the outer plastic tube and the other electrolyte through the four inner tubes.

The distance between the electrodes and membranes decreased from the few millimeters common in a planar cell to less than a millimeter in this design, boosting power density. “So we use less material to achieve the same performance,” Liu says. “Less material means it is cheaper.”

A proof-of-concept test using zinc-iodine chemistry—one of the common chemical makeups used in flow-battery technology—showed that the battery had a charge density of about 1,322 watts per liter of electrolyte and a discharge density of about 306

GEORGIA INSTITUTE OF TECHNOLOGY



A team of engineers at Georgia Tech [from far left, Nian Liu, Ryan Lively, Xing Xie, and Alex Filippas] has invented a new flow-battery cell configuration. The result [left] will be smaller, cheaper flow batteries.

W/L. For a conventional planar cell, the charge density and discharge density are under 60 W/L and 45 W/L, respectively. The researchers reported on 3 January in *Proceedings of the National Academy of Sciences* that the battery lasted through over 2,500 charge cycles and is compatible with other typical flow-battery chemistries.

Liu estimates that the tubular design should cut the cost of flow-battery power modules by roughly half. Furthermore, all the components in the cell can be purchased off the shelf, and scaling up the reactor cell design should be easy, since it is based on a commonly used design in the chemical industry.

Liu expects that the overall cost of flow batteries should go down in coming years. There is a lot of ongoing research on improving electrolytes already, he says, "and also people are starting to think about recycling electrolytes. If this is demonstrated, then the electrolyte cost can be significantly reduced."

A flow-battery company is in talks to sponsor further research on scaling up the tubular flow cell. ■

A rider tests the HydroBlade eFoil prototype in the waters off South Florida.

PELAGION

TRANSPORTATION

Reimagined Jet Ski Brings EVs to the Beach

› The battery-powered HydroBlade nixes noise and noxious fumes

BY LAWRENCE ULRICH

Cars aren't the only conveyances being transformed by electricity. Along with electric motorcycles and snowmobiles, personal watercraft are floating better ways to coexist with nature and neighbors. This new breed of machines brings requisite thrills to the Great Outdoors, but without fouling the atmosphere or disturbing the peace with an internal-combustion racket.

The latest comes from Florida-based Pelagion, whose founder and chief executive, engineer Jamie Schlinkmann, was inspired by childhood adventures on a watersports icon: A 1973 Kawasaki Jet Ski. Schlinkmann's machine, just the 213th ever built, is still one of his prized possessions. His company's Pelagion HydroBlade is an ingenious mash-up of classic stand-up Jet Skis and modern surfboard-style "eFoils." Those electric-powered boards had a real breakthrough in 2020 when Facebook founder Mark Zuckerberg was spotted sailing over Hawaiian waters on one model.

The metaverse may spring to mind the first time you see an eFoil, with its rider seeming to fly above the waves on a magic carpet. It's the magic of hydrodynamics: A hydrofoil works like an airplane wing, creating high and low pressure areas as it slices through water. It thus generates lift with precious little drag. Add an electric motor and propeller to create thrust and you've got a hydrofoil that doesn't require surf waves, a kite, or a tow boat to generate power.

"They're kind of like flying an airplane with no rudder and no ailerons," Schlinkmann says. "You shift your weight to fly that plane."

There's only one problem: Powered or not, a hydrofoil takes some practice and patience to learn to ride in a standing position, especially for people with no surfing or wakeboarding experience, or so-so balance skills. So Schlinkmann added a boom-mounted canard and rudder up front to keep the craft airborne and steady without the rider having to constantly expend energy and adjust body position. Add a trusty set of handlebars, says Schlinkmann, and the HydroBlade handles more like a bicycle. That makes it much less intimidating.

"There's a much lower learning curve," Schlinkmann says. "Like a bicycle, you lean into a curve and add another control element."

The design began to take shape around 2020. Schlinkmann pulled the engine and other internal-combustion engine guts from his old Jet Ski and studied how he could make it electric. He realized that a conventional electric Jet Ski might only have a 15- or 20-minute runtime on a single charge, which wasn't good enough. But after he rode a few eFoils, the idea came together.

The foil improves efficiency so much, he thought at the time. If he could make something that combined the two worlds, he'd get a range and usage that was practical.›

For the HydroBlade, a pair of permanent-magnet radial-flux motors drive dual propellers at a peak 16 kilowatts (21 horsepower). They're fed by two battery packs, with a combined 600 cylindrical 2170 NCM cells and a total 11 kilowatt-hours of energy—about eight to 10 times the capacity onboard a typical eFoil. A separate 1.6-kW charger can refill batteries in about 4 hours.

Schlinkmann says the HydroBlade should cruise for about 4 hours at speeds of up to a hair-whipping 70 kilometers per hour (44 miles per hour). The company is targeting a roughly 100-kilometer range, enough to get from Fort Lauderdale to Bimini. The twin battery packs weigh 58 kilograms, a bit more than half the craft's total 104-kg weight. The removable, swappable packs act as stressed members for a hull that's skinned in ABS composite. A smartphone app doubles as an electronic dashboard. Put the phone in the craft's handlebar mount and get readings, including speed, the state of charge, and telemetry.

Putting that much juice underwater requires special attention to safety—the first concern many people have about electric watercraft.

"If you're in a Tesla and catch fire, you just get out," says Schlinkmann, a longtime Model S owner. "If you're on the water and catch on fire, you're swimming."

Schlinkmann says the company's battery-management system monitors the health and state of individual cells and communicates along a typical CAN bus. The in-house design also fuses cells in parallel at a notably low current for cell balancing, keeping power current on a separate path.

"We're fusing the cells at a much lower level, so if there's any cell failure, you won't have a fire," he says.

The company plans to tool up production in coming months and to bring the HydroBlade to customers by around year's end. Initial capacity, according to the company, will be 1,000 units a year. Next time you're lolling on a beach, keep your eyes peeled for a HydroBlade—because you certainly won't hear it shrieking past. ■

> **"There's a much lower learning curve [than other airfoils]. Like a bicycle, you lean into a curve and add another control element."**
>
> –JAMIE SCHLINKMANN, PELAGION FOUNDER AND CEO

JOURNAL WATCH

Wireless Power Boosted for Big Gadgets

Despite the success in wirelessly charging small devices like phones, challenges remain when it comes to doing so with power-hungry devices like e-bikes. Shu-Yuen Ron Hui and his colleagues at Nanyang Technological University outlined ways to overcome these barriers in a study published 23 January in the *IEEE Journal of Emerging and Selected Topics in Power Electronics*.

One major hurdle is the thermal limit of batteries. To address this issue, Hui and his colleagues developed a new temperature-regulated current-control technique that decreases the charging time without overheating the battery.

A second challenge is transferring more power, which requires that gate drivers work at exceptionally high frequencies. Existing gate drivers have a latency of about 100 nanoseconds, but the researchers have developed one with a latency of just 6 ns—enough to speed up power inverters by an order of magnitude, to tens of megahertz.

The team's technology also keeps track of the system's maximum-efficiency operating point as the battery is being charged, increasing its effectiveness. Today's standards cover the charging of devices drawing 15 watts or less, but plans are underway to create standards for devices that require about 200 W, such as e-bikes.

The groundwork still must be laid for even bigger electronics, so Hui and his colleagues plan to keep forging ahead. They have already filed a patent for a resonator that could transfer hundreds of watts.

—Michelle Hampson

Autodidacts Build Brain-Controlled Prosthetic

By Willie D. Jones

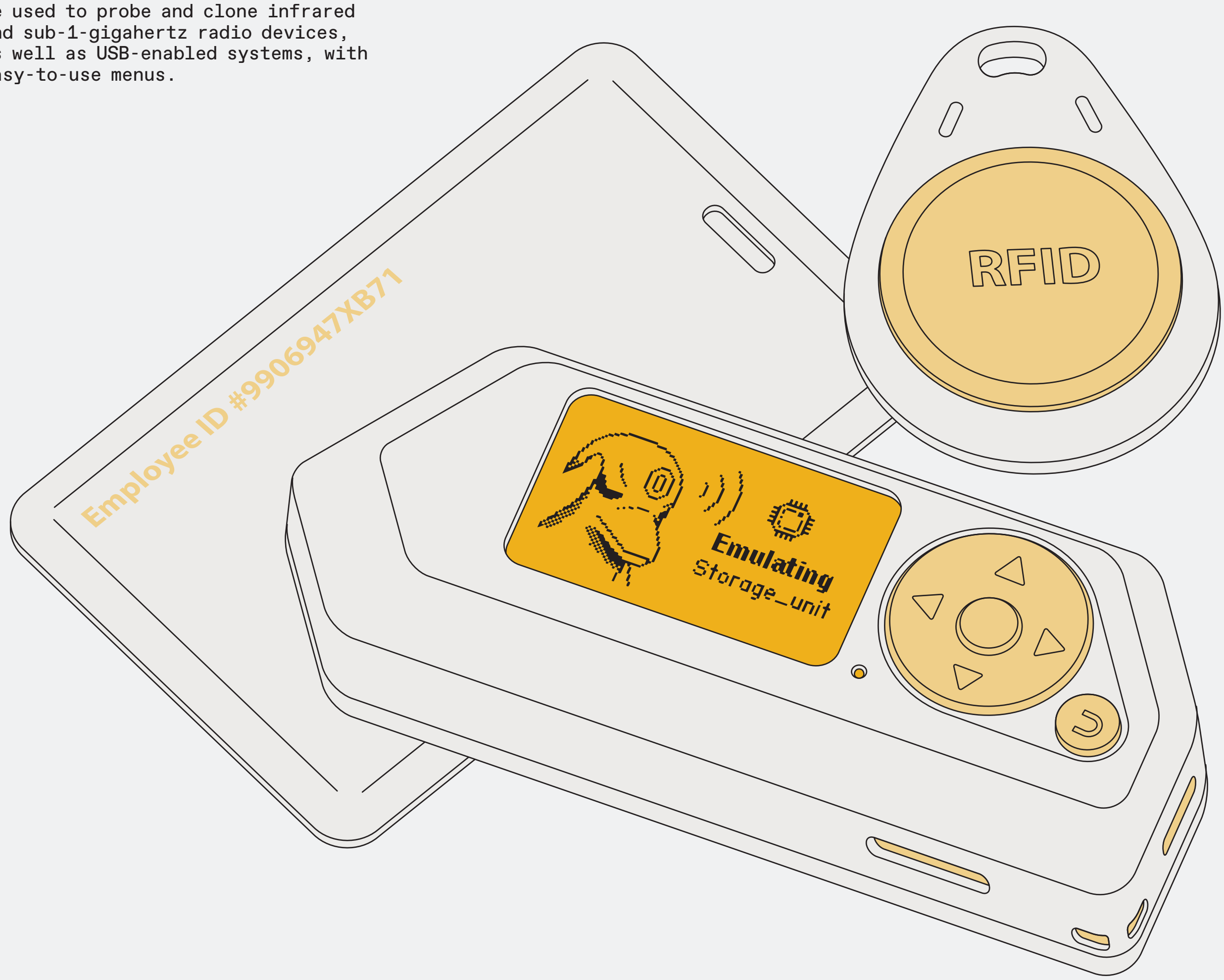
Self-taught Kenyan engineer David Gathu has turned his ingenuity and interest in electronics into innovations that help the disabled. Here he's wearing the brain-controlled prosthetic arm he and his auto-didact cousin Moses Kiuna cobbled together from salvaged parts. The headset picks up brain signals and converts them to electric current that is directed to a transmitter. The transmitter wirelessly relays the signals in the form of commands to the arm, which responds with movements that match the wearer's intentions.

PHOTOGRAPH BY TONY KARUMBA/ AFP/GETTY IMAGES


G

Hands On

Out of the box, the Flipper Zero can be used to probe and clone infrared and sub-1-gigahertz radio devices, as well as USB-enabled systems, with easy-to-use menus.



A Hacker's Delight

› You'll either love or hate the Flipper Zero

BY STEPHEN CASS

Readers of this Hands On are likely to fall into one of two camps: those who'll view the Flipper Zero with fascination, and those who'll view it with loathing. Among the former are security researchers and hardware developers trying to debug a wireless setup. Among the latter are IT folks charged with defending their realm from physical or network attacks. But whatever camp you

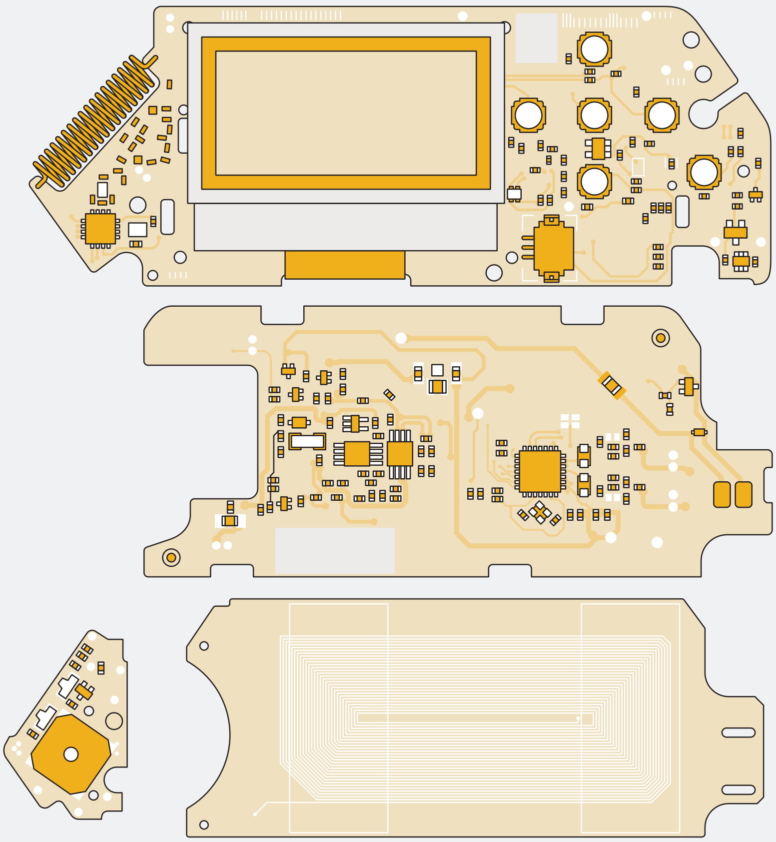
Illustrations by James Provost

fall into, the Flipper is something you'll need to know about.

The Flipper is an open-source hacking tool of exceptional polish and functionality. Its official price is US $169, but it sells out as fast as it can be manufactured and so can often only be found at a hefty markup—I paid $250 from one reseller.

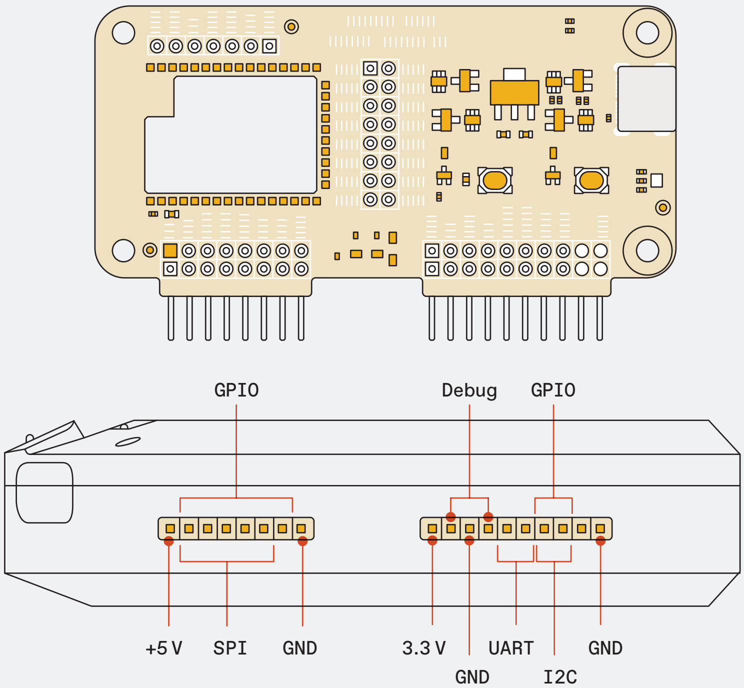
Hacking software and hardware tends to adopt, either consciously or unconsciously, design aesthetics that wouldn't be out of place in a William Gibson cyberpunk novel. Hardware is utilitarian, with boxy enclosures painted or printed black. Software often relies on opaque commands. They are *serious tools* for *serious people*. The Flipper stands this schema on its head. Its line of aesthetic descent is more Tamagotchi than tech dystopia, with a brightly colored white-and-orange case molded to fit your palm. An onscreen animated anthropomorphic dolphin pops up to guide you through setup menus. It looks and feels like a child's toy. It isn't.

The Flipper is powered by a 32-bit Arm processor core with a top speed of 64 megahertz. That's hardly anything to get excited about, but the core is paired with a bunch of analog and serial peripheral interfaces and, most importantly, a sophisticated radio transceiver. Out of the box, the Flipper can do quite a few interesting things. For starters, it lets you read many common types of RFID key cards by holding them up to the Flipper, which stores the data on its SD card—and then you can use the Flipper to emulate those cards, opening, say, an office door as easily as if you had the original card. (It's flexible enough that I was even able to read the microchips implanted into my cats, which operate at a different frequency than RFID key cards.) This can be handy if, like me, you sometimes find yourself digging out a small sheaf of blank white cards from your wallet and trying each in turn to open a seldom-used door, but the implications for physical security are obvious. (It's true that you can also easily clone key cards at, for example, the kind of kiosk that also cuts keys, but you need to bring the card to the kiosk, something harder to do surreptitiously than simply cloning the card in situ.)



An open-source design, the Flipper is composed of a main board with an Arm-based CPU and transceiver chip [top], a board for NFC and RFID communications [middle], a PCB antenna [bottom], and an extra board to handle IR and iButton interfacing [bottom left].

In a similar fashion, you can clone an iButton key fob, and many radio remotes that work in the popular 433-MHz band. It can also read near-field communication (NFC) devices, such as MIFARE key cards, and even pick up the signals your contactless credit card's EMV chip puts out (although this is of very limited use, since the Flipper can't duplicate the output of the algorithms used to create a per-transaction one-time code). There's even an infrared port that allows you to clone, say, a TV remote, à la a universal remote control, and if you don't have access to the original remote control, it can cycle through a set of common infrared protocols for



The Flipper's capabilities can be enhanced with add-on boards, such as a Wi-Fi board or other hardware that can communicate via a number of serial protocols. In addition, the Flipper can help debug hardware by generating various test signals, including, for example, PWM signals, which can be used to test a servo.

basic commands like on/off and changing the channel, making it a superpowered version of the TV-B-Gone. This last capability is popular for petty mischief, judging by the numerous videos online showing people turning off lecture hall projectors, or fast-food-restaurant menu displays.

The Flipper also has the ability to emulate a USB keyboard and mouse, and can execute scripts that can control a graphical user interface—a boon for those seeking to automate tasks, another security headache for others. You can also use the Flipper as a UART-to-USB bridge, allowing connectivity with many devices that don't support USB.

A $45 plug-in board adds Wi-Fi capabilities, although as shipped this merely enables over-the-air updates. But there are plenty of hacks of the Flipper and the Wi-Fi board themselves. It's possible to replace their firmware with alternatives that allow low-level control of Wi-Fi signals, albeit with a return to the land of cryptic commands. This opens up the ability to do a number of attacks on a Wi-Fi network. In one built-in example, you can spoof a series of access-point names, so the list of Wi-Fi networks in your area is spammed, line by line, by the lyrics to Rick Astley's "Never Going to Give You Up," in an unusual version of Rickrolling. (To test this without provoking the ire of either the IEEE's IT department or my neighbors, I took the Flipper to the end of a pier jutting 110 meters out from the aptly named Transmitter Park into New York City's East River). Another point-and-shoot attack lets you target an access point with a flood of deauthorization packets, killing active connections.

All this functionality is, in itself, legal. Just as with a set of lockpicks, the line is not the possession or even use of the tools, but how the tools are used. Attacking your own network or cloning your own key card is one thing, going after someone else's is another. But the modified firmware can also allow the Flipper to perform actions that are inherently illegal, such as removing regional restrictions on which ISM bands the device can transmit on: The moment you pump out a single photon on a band that's a no-no in your region, you are operating illegally.

Just like the smartphone, the Flipper is a demonstration of technological convergence, merging many different devices into one slick package. One wonders what the next-generation version will bring—and whether or not it'll look like a Furby. ■

The Flipper Zero looks and feels like a child's toy. It isn't.

BY DANIEL P. DERN

Careers: Jonathan L. Zittrain

The Harvard law professor is an expert on digital technology

Harvard professor Jonathan L. Zittrain is an expert on the Internet, digital technology, law, and public policy.

JON CHASE/HARVARD UNIVERSITY

Jonathan L. Zittrain wears many hats. An expert on the Internet, digital technology, law, and public policy, he regularly contributes to public discussions about what digital tech is doing to us and what we should do about it—most recently around the governance of AI and the incentives that shape major social media platforms.

He holds several positions, all at Harvard, reflecting his many converging interests. He is a professor of international law at Harvard Law School, a professor of public policy at its Kennedy School, and a professor of computer science at the university's John A. Paulson School of Engineering and Applied Sciences. He's also cofounder and faculty director of Harvard's Berkman Klein Center for Internet & Society.

In his various capacities, he has been tackling many sticky cyberpolicy issues over the past 25 years.

"Lately, I've been working on the question of how to regulate and govern generative technologies—that is, technologies like the Internet, the Web, and generative AI that allow contributions and development from nearly anyone or anywhere," Zittrain says.

He's also curious about what role public interest plays in tech's evolution, which these days is largely a product of market forces.

"I'm deeply interested in whether and how fast-growing and rapidly deployed technologies such as AI large language models and new distributed activity networks like those of Web3 should be governed or regulated."

> "Remarkably little is settled around the wise use of technology."

Zittrain first got involved with computers in 1983, when he was in high school, and his parents gave him a US $99 Texas Instruments TI-99/4A, the first 16-bit home computer. Using a hand-me-down television set as a monitor, and a dial-up modem, he connected the computer to CompuServe, one of the first commercial online service providers. "It was a pre-Internet online community-of-communities," he recalls. "There were no graphics on the screen back then, just text."

CompuServe charged a flat fee per minute and the user paid any additional phone charges. Zittrain ran up a $300 bill on his parent's credit card the first month, he says, and his parents threatened to disconnect him.

But when he posted a farewell message on CompuServe, one of the company's chief system operators offered him free connection time in exchange for becoming an assistant sysop, a person who runs a computer server. "I gave an enthusiastic Yes! before he could change his mind," Zittrain says.

In that role, he helped resolve disputes among users. Later, he became chief administrator for CompuServe's private Sysop Forum, where members discussed how to manage their own forums.

Zittrain also learned to code, writing a host program for a bulletin board system in TI Extended Basic.

These early experiences led him to pursue a bachelor's degree in cognitive science and artificial intelligence at Yale. He continued to code, mostly in Lisp.

His interest in how online communities might govern themselves led him to study law and public policy. In 1995, he earned a law degree from Harvard Law School and a master's degree in public administration from the Kennedy School.

"Helping people resolve the disputes that arise in rolling conversations—when intensity of emotion does not necessarily track to the gravity of the issue—is still a central concern for me today," says Zittrain.

For Zittrain, the Berkman Klein Center remains the hub for most of his activities.

It has spawned programs such as the Institute for Rebooting Social Media, a research initiative aimed at addressing misinformation, privacy breaches, harassment, and content governance on social media platforms.

Over the years, the center has hosted hundreds of faculty, staff, fellows, and affiliates from over 40 countries, Zittrain says. "Whether it's producing code for GitHub, publications, or podcasts and videos to the world at large, or bringing together people who would never have otherwise met—and who might even be skeptical of one another—we've tried to bring new perspective and energy to the cultivation of a digital world in the public interest."

This year he is teaching courses that address the ethical implications of artificial intelligence systems and whether and how to govern digital platforms.

"Students explore different frameworks for understanding the evolution and use of technology in society, such as what this technology is doing to us, and how we might together affect how it works," Zittrain says.

Zittrain conducts research on what's happening across the sweep of digital technology and writes about his findings in blogs, magazines, and books. In his 2009 book, *The Future of the Internet—and How to Stop It* (Yale University Press), he examined many of the problems that still affect us today, such as massive security breaches, ubiquitous surveillance, and social media that foments harassment and spreads lies. He predicted, correctly, that those problems would only get worse.

Much of Zittrain's work looks at ways to make information more widely available.

In that vein, he has helped launch several organizations that provide free access to information. These include the nonprofit Creative Commons, a global nonprofit that enables sharing and reuse of content and knowledge. Another is the Open Casebook series, a joint project of MIT Press and the Harvard Law School Library that offers open-source digital legal textbooks for free.

Zittrain oversees the Harvard Law School Library, whose open source tools built by its Library Innovation Lab have helped preserve the pioneering URL naming convention commonly known as permalink. Perma Links, administered by Perma.cc, archive Web pages and create permanent links to them, ensuring the information will remain available to courts, researchers, libraries, and others. The library's CaseLaw Access Project has digitized more than 6.5 million U.S. state and federal court decisions and made them available for free.

"We need more lawyers and public policy experts who have a grounding in computer science, so they understand the complexities and challenges of monitoring and moderating large-scale social media forums."

What are some career possibilities for people looking to enter tech policy and law? "Remarkably little is settled around the wise use of technology," Zittrain says. "For example, for products that can be controlled by their manufacturer—like smart home devices and new cars—it's unclear what the manufacturer's responsibilities are. Is it the responsibility of the vendors, the insurance companies, or the regulatory sector to ensure privacy? And whose responsibility is it when something goes wrong? Every one of these companies will need their own thoughtful experts to help research and write, debate, decide, and implement policies."

These gaps between new technologies and their fair, safe use create job and career opportunities—just as CompuServe's online forums created the need for sysops and sysop forums.

"We need more lawyers and public policy experts who have a grounding in computer science," says Zittrain, "so they understand the complexities and challenges of monitoring and moderating large-scale social media forums like Facebook or Twitter. Or of not doing so."

Given the complicated and profound problems these new technologies create, he says, those who design and build technology should consider more than just the engineering aspects of their work. They should also study the philosophy and ethics of what they are creating as well as their responsibilities to the consumer.

Having coding experience is helpful for those interested in getting involved in technology policy, Zittrain adds.

"Knowing how software works and doesn't work, both abstractly and in reality, is hugely helpful in thinking about what kinds of digital futures are possible—and what means of technical, political, and social intervention are available to get there." ■

Employer:
Harvard

Titles:
Professor of international law,
professor of public policy,
professor of computer science

Education:
Yale, Harvard

Book:
The Future of the Internet—and How to Stop It
(Yale University Press, 2009)

5 Questions for Anton Troynikov

His company's creation identifies the art behind AI-generated images

Anton Troynikov is the cofounder of Chroma, an AI company focused on understanding the behavior of AI through data. Previously, Troynikov worked on robotics with a focus on 3D computer vision. He does not believe AI is going to kill us all.

Shortly after their first releases to the public, text-to-image artificial intelligence models like Stable Diffusion and Midjourney also became the focal points in debates around the ethics of their usage. Anton Troynikov is a cofounder of Chroma, a startup working to improve AI interpretability. With AI art generators, Troynikov and others at Chroma saw an opportunity to build a tool that would try to attribute the source images used in a particular image generation. Troynikov answered five quick questions on the project, Stable Attribution, and how artists and engineers can stop "talking past each other."

What were your first impressions of AI art generators when they were released?

Anton Troynikov: I started paying attention to the AI art discourse after Stable Diffusion was released and a lot more people got access to the model. And I started to realize that people on both sides of the conversation were talking past each other. I wanted to see if there was a technical solution to the problem of making sure that technologists and creatives were not antagonists to one another.

What's your goal with Stable Attribution?

Troynikov: I wanted to demonstrate that this problem is not technically infeasible to tackle. We felt it was the right thing to just go ahead and see what kind of reaction we'd get when we launched it.

Very quickly, how does Stable Attribution work?

Troynikov: Stable Diffusion is in a class of models called latent diffusion models. Latent diffusion models encode images and their text captions into vectors (unique numerical representations). During training, the model tries to reproduce the original numerical representation of every image in its training set, based on that image's accompanying text caption.

Because these vectors come from these pretrained models that turn images into vectors and back, the idea is basically, "Okay, it's trying to reproduce images as similarly as possible." So a generated image wants to be similar to the images that most influenced it, by having a similar numerical representation. That's the very short explanation.

How do you make that final step and determine who the artists and creators are?

Troynikov: We would really like to be able to attribute directly back to the human who created the source images. What we have—and what's available in the public training data set of Stable Diffusion—are URLs for images, and those URLs often come from a CDN [content delivery network]. The owners of the sites where those images appear and the owners and operators of those CDNs could make that connection.

We do have a little submission form on the site. If people recognize who the creator is, they can submit it to us, and we'll try to link it back.

How do you see generative AI and source attribution affecting artistic creation?

Troynikov: One is, by being able to do attribution, you can then proportionately compensate the contributors to your training set. The other really interesting thing is it turns them from just a generator into a search engine. You can iteratively find that aesthetic that you like and then link back to the things that are contributing to the generation of that image. ■

Photo-illustration by Stuart Bradford

By Tekla S. Perry

Photography by **Peter Adams**

Mr. Internet

Vint Cerf's 1973 sketch kicked off five decades of improving and evangelizing what we now know as the Internet

It was June 1973. For the past three months, Vint Cerf and Bob Kahn had been working together on a problem Kahn had been pondering for some time: how to connect ground-based military computers seamlessly to communications satellites and mobile radios.

The ARPANET and the way it handled communications was already well established. But extending it to handle multiple networks—whose reliability couldn't be taken for granted—was a different story.

The two had been exchanging ideas in person and via email and reviewing the work of others who were trying to solve similar issues. But now, Cerf sat alone in the lobby of San Francisco's Jack Tar Hotel, on a break from a computing conference. And the problem was on his mind.

Cerf pulled out an envelope. Recalling what the two had figured out so far, he began to sketch the main components and key interfaces. He scrawled clouds representing three different packet-switched networks—the ARPANET, packet radio, and packet satellite—and boxes representing the computers hanging off those networks. These would be the host computers, running applications that needed to use the network.

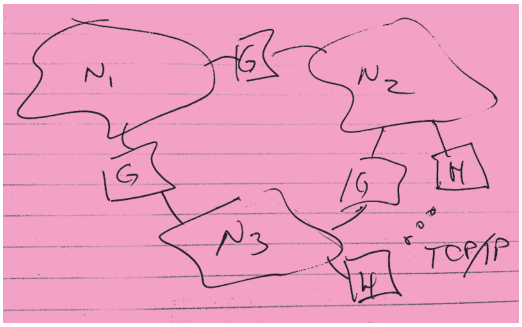
"The networks couldn't be changed and couldn't know that they were part of the Internet, because they already existed," Cerf recalls recently in an interview at his office at Google, in Reston, Va.

So he sketched in another set of computers—gateways—that would know about other networks.

"Those were the constraints of the problem," he says. "Sometimes, if you can constrain a problem enough, you can see the solution pop out in front of you. The diagram helped me to see where protocols would need to be standardized."

Cerf describes the communication protocols that he and Kahn came up with as comparable to a set of postcards and envelopes: The postcard has a message and an address for the intended destination. The address on the envelope is either that of the destination host in the local network or of a gateway that leads toward the next network along the route to the final destination.

When a message arrives at that next gateway, the gateway opens the envelope and checks the address on the postcard. If the message is intended for a destination inside the gateway's home network, it gets delivered in an appropriate envelope; if not, it goes in an envelope addressed to the next gateway en route to the destination network, where the process repeats.



Vint Cerf re-created his original sketch, with clouds representing three packet-switched networks and boxes representing gateway and host computers.

That, essentially, is how the Internet works today.

For the past five decades, Cerf, now 79, has been perfecting, extending, and evangelizing the Internet. It is for this—his contributions in cocreating the Internet architecture and his leadership in its growth to date—that Cerf is the recipient of the 2023 IEEE Medal of Honor.

Cerf came to computers early; his first encounter was in 1958. When he was 15, a family friend working at System Development Corp. in Santa Monica, Calif., arranged a visit to a Semi-Automatic Ground Environment computer center. The SAGE system analyzed radar data, looking for Soviet bombers heading toward the United States.

"You literally walked inside the computer, a room with glowing red tubes on the walls," Cerf recalls. "It was weird, but I was mesmerized."

During high school in Van Nuys, Calif., his best friend, Steve Crocker, wrangled permission for the two to occasionally use a Bendix G-15 computer at the University of California, Los Angeles.

"That's when I realized that you could create your own artificial world with software, and it would do what you told it to do," Cerf says. "And there was something utterly beguiling about this idea."

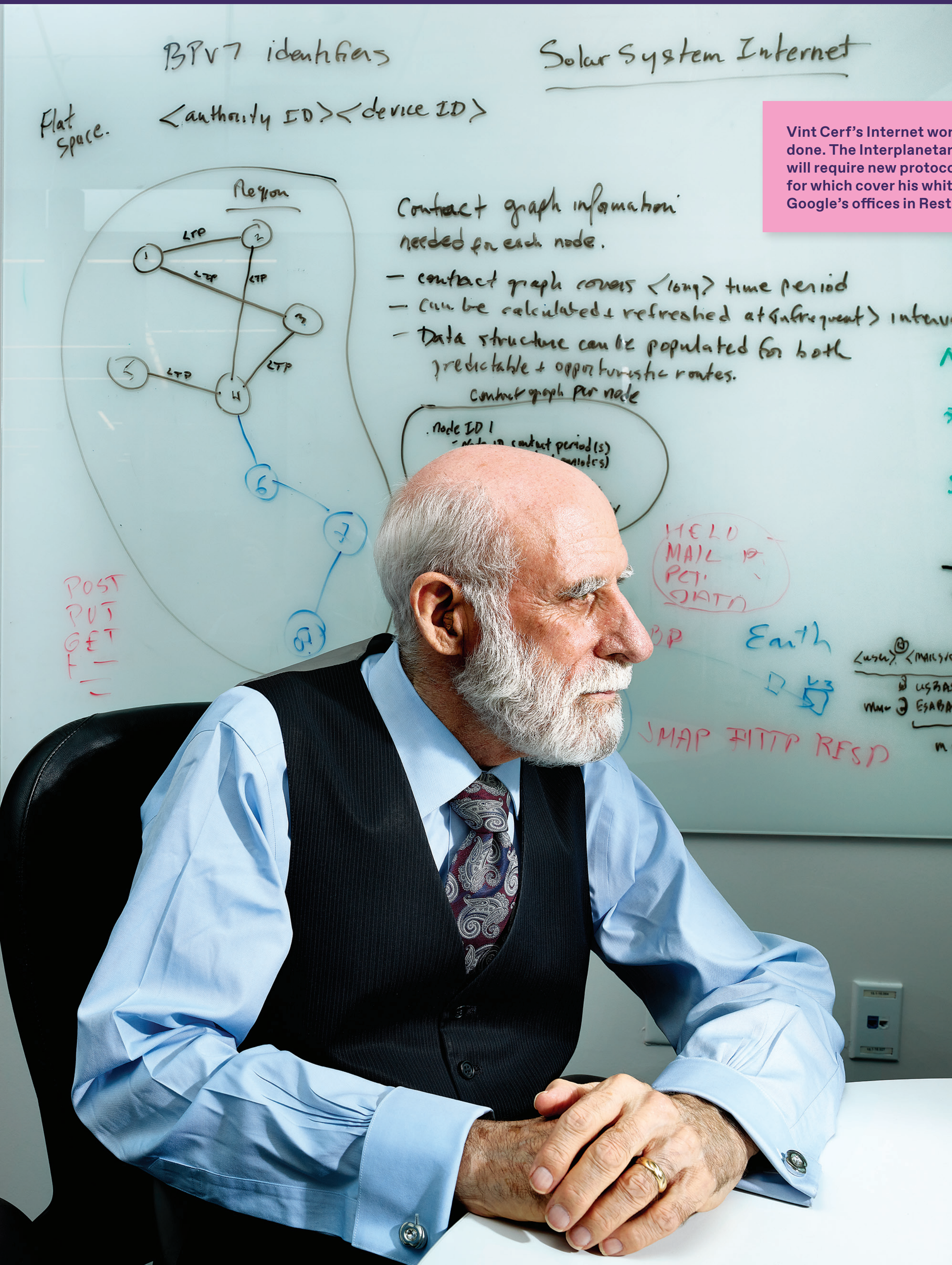
The two found it hard to stay away from UCLA's G-15. One Saturday, Crocker recalls, the two were working on some mathematical equations they wanted the computer to solve. They went over to Engineering Building 1, where the computer was housed, and found the building locked.

"I was crestfallen," Crocker says. "Then Vint observed that a second-floor window was open. I'm thinking, 'Nooo...,' but he was already on my shoulders climbing in. He went through and opened the door, and then we taped over the door latch so we could go in and out during the day."

Cerf was just 17 when he first got paid for developing code in 1961—testing software for the Rocketdyne F-1 engine destined for NASA's Apollo program. But when he entered Stanford University later that year, he envisioned a career in mathematics, not computer science. Then he encountered Riemannian geometry, a mathematical way of describing multidimensional surfaces.

"I broke my pick on that, realizing that I was probably not going to be a professional mathematician," Cerf recalls. "I was relieved that I could program so at least there was a job I could do." He took all the computer classes he could fit into his schedule.

After graduating in 1965 with a B.S. in mathe-



Vint Cerf's Internet work is not yet done. The Interplanetary Internet will require new protocols, notes for which cover his whiteboard in Google's offices in Reston, Va.

Vinton G. Cerf

CURRENT JOB:
Vice president and chief Internet evangelist, Google

DATE OF BIRTH:
23 June 1943

BIRTHPLACE:
New Haven, Conn.

FAMILY:
Sigrid, his wife; sons David and Bennett

EDUCATION:
B.S. 1965, mathematics, Stanford; M.S. 1970 and Ph.D. 1972, computer science, University of California, Los Angeles

FIRST JOB:
Cleaning and refilling automatic coffee machines, at age 14

FIRST JOB IN SOFTWARE:
Rocketdyne, working on test software for the F-1 engines used in the Apollo program's Saturn V rocket

BIGGEST SURPRISE IN CAREER:
"The explosive use of the Internet"

PATENTS:
Two, for an "Internet Radio Communication System" and for a "System of Distributed Task Execution"

CORPORATE BOARD MEMBERSHIPS:
The Marconi Society, the Folger Shakespeare Library Board of Governors, the National Science and Technology Medals Foundation, the American Institute of Physics Foundation

HEROES:
Bob Kahn, Steve Crocker, Gerald Estrin

MOST RECENT BOOK READ:
The Music of the Bees, by Eileen Garvin

FAVORITE BOOKS:
Lord of the Rings by J.R.R. Tolkien, Isaac Asimov's Foundation series, *Mission of Gravity* by Hal Clement

FAVORITE KIND OF MUSIC:
Classical before 1900, particularly Wagner

FAVORITE WEBSITES:
Google, Wordle, *USA Today*'s Sudoku and crossword puzzles, OnlineJigsawpuzzles.net

FAVORITE FOOD:
Häagen Daz coffee ice cream

FAVORITE RESTAURANTS:
Dante Ristorante in Great Falls, Va., 2941 Restaurant in Falls Church, Va.

FAVORITE MOVIES:
Lord of the Rings, the Harry Potter series

MOTTO:
Patience and persistence count

WHAT ADJECTIVE WOULD YOUR SPOUSE USE TO DESCRIBE YOU?
Driven

HOBBIES:
Reading science fiction, doing jigsaw puzzles, collecting stamps and coins (sporadically), wine collecting and drinking

LANGUAGES SPOKEN:
English and some German

KEY ORGANIZATIONAL MEMBERSHIPS:
IEEE, Association for Computing Machinery (ACM), American Association for the Advancement of Science (AAAS), American Academy for the Arts and Sciences, American Philosophical Society, British Computer Society, the Royal Society, the U.S. National Academies of Science and Engineering

MAJOR AWARDS:
IEEE Medal of Honor "for cocreating the Internet architecture and providing sustained leadership in its phenomenal growth in becoming society's critical infrastructure," ACM Turing Award, Queen Elizabeth Prize for Engineering, VinFuture Prize, U.S. National Medal of Technology and Innovation, Presidential Medal of Freedom, the Japan Prize

matics, Cerf joined IBM as a systems engineer and was soon assigned to work with the Quiktran time-sharing system, which ran on a disappointingly old IBM 7044. After two years of tangling with Quiktran, Cerf realized there were fundamental things about computing he didn't know, so he went back to school at UCLA.

There, he joined Crocker in a research group run by Len Kleinrock, with Gerald Estrin as their thesis advisor. Mainly, the group aimed to build a way to model the performance of the ARPANET and its gateways under different traffic conditions. But Cerf, Crocker, and a few others also thought about the computers that would attach to the network, considering what they would do and how they would do it.

It wasn't easy, Cerf recalls. The computers had different operating systems; some even represented the characters of the alphabet differently.

Working with their counterparts at other ARPANET sites, Crocker, Cerf, and others in Kleinrock's group eventually figured it out. And then they set about breaking the network by overloading it with artificial traffic. That mission came from Bob Kahn, then at Bolt Beranek and Newman (BBN), the company contracted by the U.S. Advanced Research Projects Agency (ARPA) to build the switches for the nascent network. [For more on Kahn's ARPANET efforts, see "The Early Internet's Do-or-Die Moments," p. 46.]

"We shot the ARPANET down repeatedly," Cerf says. "I was tempted to get a rubber stamp with a kind of network pattern on it to stamp the side of the computer, the way guys that shoot down airplanes stamp the sides of their planes."

Cerf finished his Ph.D. in 1972, and in October he returned to Stanford as an assistant professor of computer science and electrical engineering, after turning down the job a few times. (Not because he didn't like Stanford, he says, but because he knew how smart the students were and didn't think he had anything to teach them.)

That same month Kahn moved from BBN to what is now the Defense Advanced Research Projects Agency (DARPA). And in March 1973, he contacted Cerf.

"'I've started this program called internetting,'" Cerf recalls Kahn telling him. "'It's intended to find ways of using computers in command and control... in mobile vehicles, in ships at sea, and in airplanes. But what we have with the ARPANET is computers sitting in air-conditioned buildings connected by dedicated telephone circuits. How are we going to hook those all together?'"

And that was the problem that Cerf was contemplating that day in San Francisco in 1973, when he put pen to envelope.

It took him and Kahn six months to flesh out what they called the Transmission Control Protocol (TCP). There had to be Internet addresses, for example, to direct messages to the various networks. There had to be error correction, but performed by the computers attached to the network rather than within the network itself. The two also worried that some packets might not fit with the next network, so they included an elaborate mechanism for fragmenting packets when they needed to travel across networks that could only carry shorter packets.

Then Cerf and Kahn wrote a paper laying it all out, briefed other computer science researchers on the details, and submitted it to the *IEEE Transactions on Communications*. A draft of a formal TCP standard came out in December 1973.

The first real-world test came in 1975, connecting Stanford to BBN in Cambridge, Mass., and University College London. And it didn't work as expected.

"It turned out that we needed to do a three-way handshake to synchronize the packet streams," Cerf says. "The first host computer sends a synchronization request with a sequence number to the other one, and that one sends back a request with its sequence number and an acknowledgment of the first request. Then the first one sends back an acknowledgment."

Eventually, after four iterations of the standard, Cerf says, things were finally looking stable. In 1976 he left Stanford for DARPA, taking over management of its Internetting program. Cerf stayed there for six years as a program manager, helping to work out the details that would make the Internet more reliable and secure. He resigned in late 1982, just before January 1983, when the Internet was to be made operational for the U.S. Defense Department, cooperating non-U.S. defense departments, and research and development contractors.

His concerns at that point, Cerf says, were purely financial. College expenses for his sons loomed on the horizon, and he worried that a government salary wouldn't cover the costs. It was time to go commercial.

Cerf joined MCI as vice president of engineering to build what the company was calling a digital post office.

"We got what we eventually called MCI Mail up and running in nine months," Cerf says.

It wasn't the first commercially available electronic mail service—CompuServe, Telenet, Tymnet, and others allowed subscribers to send email to other subscribers using dial-up modems. But those were isolated islands; someone using one system could not email someone on another.

MCI Mail was different. Any user of MCI Mail could communicate with users of other communications services, including telex, fax, and even the U.S. Post Office—the service would print out the message and send it via traditional mail. And it included provisions that would allow it to interconnect with other electronic mail services.

“I haven’t stopped because there’s always more to be done. It doesn’t get boring, ever.”

Cerf oversaw improvements to the system for three years after the launch, then left when former colleague Kahn started the Corporation for National Research Initiatives (CNRI), a nonprofit contract research organization. There, Cerf focused on Internet applications and launched the Internet Society to support the continuing evolution of Internet standards.

His work on MCI Mail would have a big influence on his future work with the Internet. At that point, much of the communications going across the Internet relied on networks that were funded at least in part by the National Science Foundation (NSF), including the NSFNet backbone network, which connected six supercomputer sites and various regional networks to thousands of U.S. universities. Other agencies funded similar networks to support their work. Use of these networks was generally restricted to researchers and academics, with commercial activity, businesses, and the general public banned. Cerf, recalling the challenge of getting MCI Mail to connect to the Telex system and different email providers, wanted to change that policy.

He went to the Federal Networking Council, an organization of the four government agencies—DARPA, the Department of Energy, NASA, and the NSF—that were funding much of the Internet’s rollout. He asked for permission to run a little experiment—to connect MCI Mail to the NSFNet backbone and see if it would work with the email system currently used on the Internet.

He got the okay, and CNRI announced the project in June 1989. Immediately all the other commercial email services clamored to get onto the Internet backbone as well—and got permission, Cerf recalls.

“Then they discovered that because they were all now connected to the Internet, all of their customers could talk to their competitors’ customers—an unexpected consequence,” Cerf says.

In 1994 Cerf went back to MCI. As senior vice president for data architecture, he worked to help the company expand the Internet side of its business. He weathered years of business turmoil—

mergers begun and abandoned, MCI's acquisition by WorldCom, and finally a declaration of bankruptcy in 2003 and sale to Verizon in 2005. Once the dust had settled, Cerf sent an email to his old friend Eric Schmidt, who had been hired as CEO of Google in 2001.

"Hi Eric, would you like some help?" Cerf recalls asking.

Schmidt responded simply: "Yes."

"That," Cerf says, "was my job interview." The toughest detail to work out with Schmidt and Google cofounders Larry Page and Sergey Brin was Cerf's title.

"I said, 'How about "archduke"?'" he recalls.

After some thought, Page and Brin responded, "The previous archduke was Ferdinand. He was assassinated and that started World War I, so maybe that's a bad title. Why don't you be chief Internet evangelist?"

Cerf agreed, and he holds that title to this day, overseeing a small group that deals with Internet policy and standardization issues out of offices in Reston, Va. He meets regularly with members of governing bodies around the world to discuss issues involving Internet regulation, a perennial battle between what freedoms to allow and what activities to regulate. He also keeps a hand in technical work on Internet protocols, including areas in which protocol development might lead to harmful side effects on either Google or the public.

Cerf is often described as the consummate statesman of the Internet world. Judith Estrin, a serial entrepreneur and former chief technology officer of Cisco Systems, has known Cerf as a family friend, a research supervisor, and a fellow networking industry executive. "There are few people in the industry who have the combination of technical understanding, integrity, openness to new ideas, and kindness," she says. "It is rare for someone to be as capable as he is *and* as wonderful a person. He is always professional. He is also infinitely curious; so many people get to a place in their careers where they don't think they need to learn anymore, but his curiosity continues to be fascinating and wonderful."

Cerf embraced the statesman role early on. Testifying before Congress in his early days at DARPA, he wore a three-piece suit. It became a trademark; few today can recall seeing him dressed in anything else. And he knows how to use that image for effect.

"In 1992, when Vint was part of the Internet Architecture Board of the Internet Engineering Task Force, there was a tense time around the future addressing scheme of the Internet," Steve Crocker recalls. "The IETF met, but the head of the architecture board couldn't make it, so Vint was going to run the meeting. As the meeting began, he took off his coat, then his vest, eventually getting down to his T-shirt. It was so out of character, it brought the house down and instantly lowered the temperature of the meeting."

E **Even after 50 years,** Cerf says, the Internet needs work. "I got involved in this and haven't stopped because there's always more to be done. It doesn't get boring, ever."

The to-do list for those involved in Internet protocol development includes domain-name system security—preventing domain names from being hijacked for nefarious purposes and improving resilience, so that a shutdown in one part of the world doesn't cause problems elsewhere.

Cerf doesn't work directly to police problematic uses of the Internet. "When people ask, 'How could you let *that* happen?' my answer is, 'Well, we gave you a bunch of rules to keep trucks, motorcycles, and automobiles from running into each other, but you get to decide what's in the cars and trains and what buildings to put next to the highways,'" he says.

He admits that this answer doesn't always cut it. So he tries to help people who want to make the Internet a safer place "see that some of their solutions have unwanted side effects. You don't want to use a mallet to squash a fly, breaking the network unnecessarily."

T **There are other Internets** to be architected. Cerf points to his office whiteboard, where a scrawled diagram looks something like a complicated version of that first sketch for the Internet. The diagram is part of the design-in-progress for the Interplanetary Internet, an effort to connect a future Internet on the moon, other planets, and traveling space probes to one another and the terrestrial Internet.

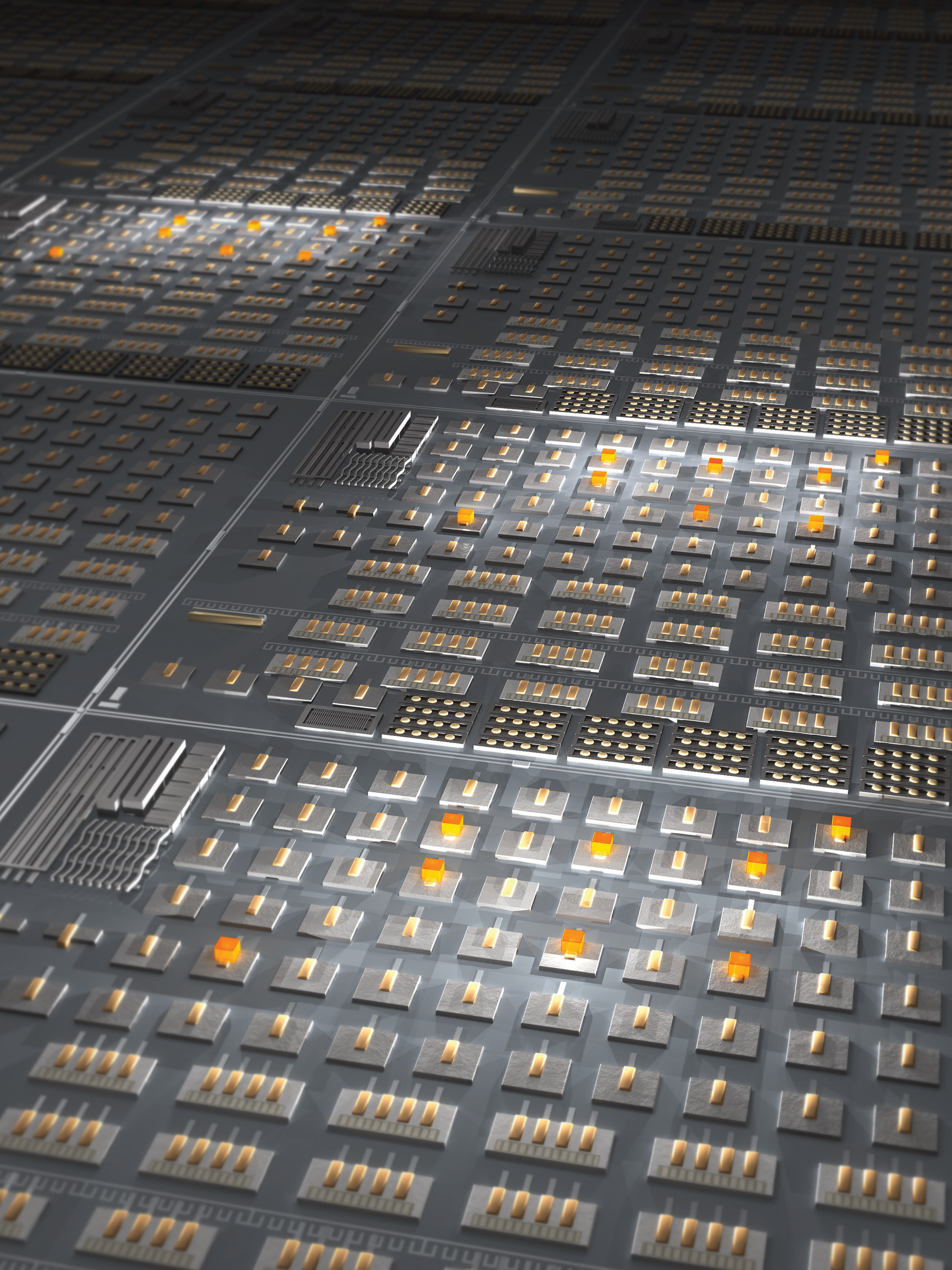
"It requires a different set of protocols," Cerf explains, "because TCP is not designed to do flow control with a 40-minute round trip. The problem gets worse when you go to the outer planets. Instead of minutes, it's hours or even days. And the planets are rotating, disrupting communication. So we had to develop delay-and-disruption-tolerant networking, a protocol we now call the Bundle Protocol Suite." Cerf has been working on the Interplanetary Internet since 1998; the effort has grown from a handful of people to hundreds today.

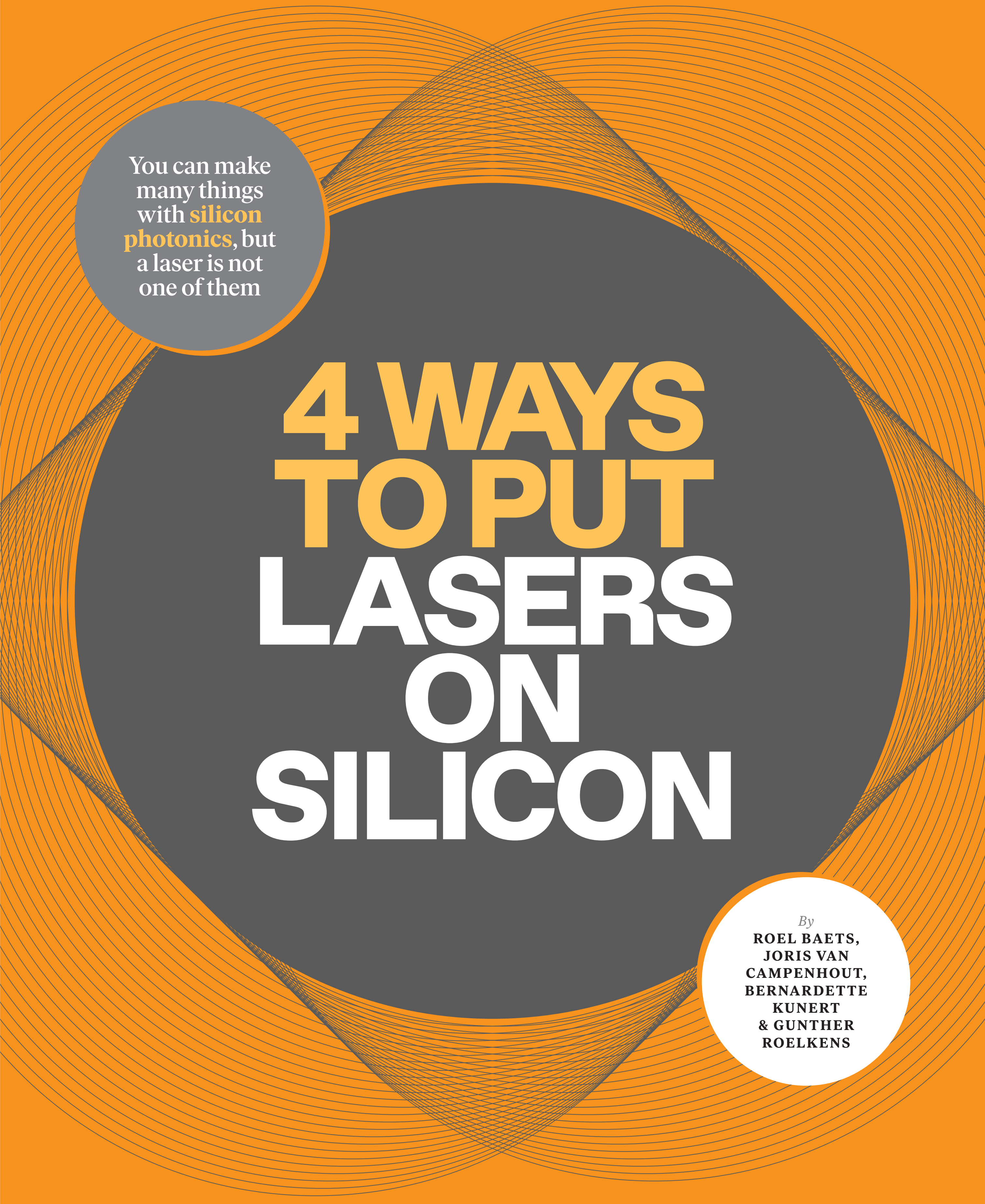
"Patience and persistence," Cerf says. That's been his motto throughout his career. "I'm not going to see the end of this. I feel like I'm in chapter two of what will be a much longer story about the history of interplanetary networking."

Then there is the Interspecies Internet, an effort launched in 2007 by Cerf; Diana Reiss, director of the animal behavior and conservation graduate programs at Hunter College, in New York City; Neil Gershenfeld, director of MIT's Center for Bits and Atoms; and musician Peter Gabriel. This global think tank now has more than 4,500 members and is looking to AI to help translate the signals from one species into those that other species can understand.

"It's been a slow process," Cerf says, "but it's like all my other projects—it might take decades.

"I feel like Lewis and Clark, wandering in a landscape full of ideas and endless frontiers. Software, and therefore computer communication, simply has no limits. You never know what you are going to turn up next." ■





You can make many things with silicon photonics, but a laser is not one of them

4 WAYS TO PUT LASERS ON SILICON

By
ROEL BAETS, JORIS VAN CAMPENHOUT, BERNARDETTE KUNERT & GUNTHER ROELKENS

Illustration by Emily Cooper

PHOTONIC INTEGRATED CIRCUITS,

which combine a collection of optoelectronic functions on a single chip, are an increasingly common part of everyday life. They are used in high-speed optical transceivers that link server racks in data centers, including the one used to deliver the *IEEE Spectrum* website, in lidars to keep self-driving cars on track, and in spectrometers to spot chemicals in the atmosphere, among many other applications. All these systems have grown less expensive and, in some cases have become economically feasible, by making most of the IC with silicon fabrication technologies.

Engineers have been able to integrate nearly every important optical function, including the essentials of modulation and detection, on silicon photonic chips, except for one: light emission. Silicon itself doesn't do that efficiently, so semiconductors made of so-called III-V materials, named for the place of their constituents on the periodic table, are typically used to make separately packaged components to produce light.

If you can live with an external laser diode in your design, there's no issue. But several factors have recently been pushing engineers to integrate lasers with silicon photonics. There may be, for example, no space for a separate light source. Tiny devices meant to be implanted in the body to monitor blood-sugar levels might face this problem. Or an application's cost might call for closer integration: When you can fit hundreds or thousands of lasers on a single silicon wafer, you will end up with a lower cost and often higher reliability than when you need to connect separate chips.

There are many ways to achieve this tighter integration of lasers and silicon. Working at Imec, a Belgium-based nanoelectronics R&D center, we are currently pursuing four basic strategies: flip-chip processing, microtransfer printing, wafer bonding, and monolithic integration. What follows is a guide to how these approaches work, their level of scalability and maturity, and their pros and cons.

1 FLIP-CHIP INTEGRATION

A straightforward way of directly integrating lasers on silicon wafers is a chip-packaging technology called flip-chip processing, which is very much what it sounds like.

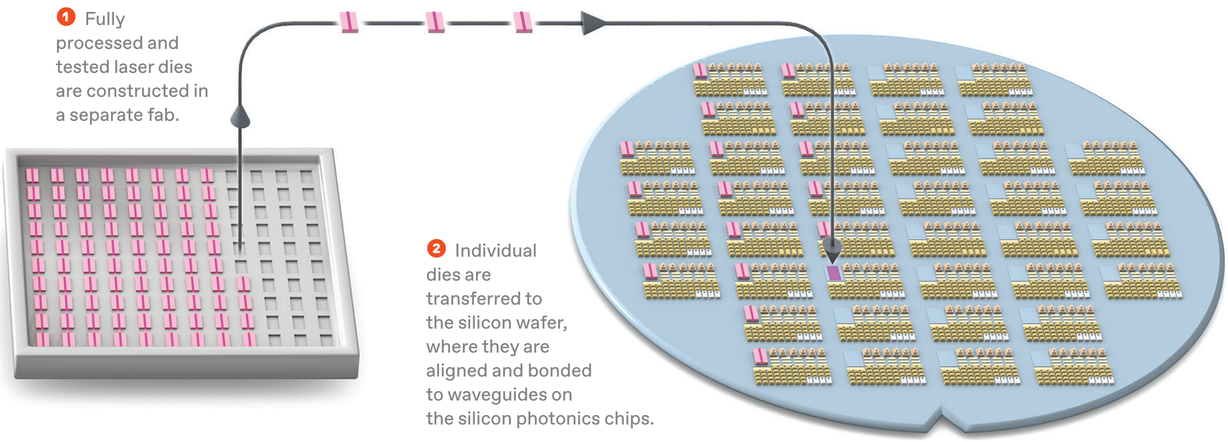
A chip's electrical connections are on top where the uppermost layer of interconnects terminate on metal pads. Flip-chip technology relies on balls of solder attached to those pads. The chip is then flipped over so the solder lines up with corresponding pads on the chip's package (or in our case onto another chip). The solder is then melted, bonding the chip to the package.

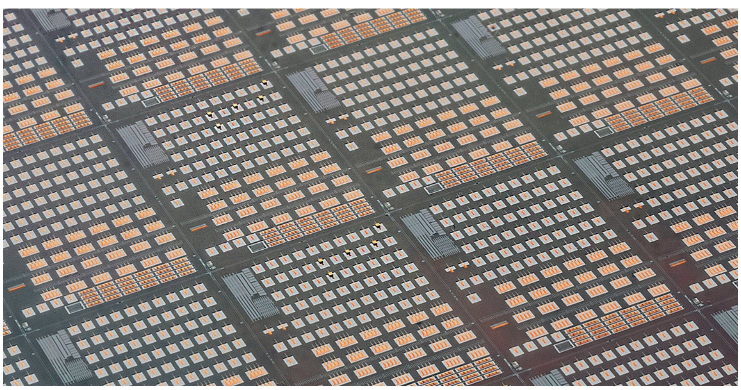
The concept is similar but more exacting when trying to bond a laser chip to a silicon photonics chip. Edge-emitting lasers are fully processed on a wafer, diced into individual chips, and tested by the vendor. The individual laser chips are then bonded to the target silicon photonic wafer, using a high-precision version of the flip-chip process, one laser die at a time. The difficult part is ensuring that the output of the laser, which emits at the edge, lines up with the input of the silicon photonics chip. We use a process called butt-coupling, where the laser is placed in a recessed part of the silicon, so it is laterally abutted to the etched facet of a silicon photonics waveguide.

For this to work, the flip-chip process requires submicrometer alignment precision in all three dimensions. Specialized flip-chip bonding tools have been developed over the past several years to do the job, and we and our collaborators and development partners have used them to optimize the assembly processes. Leveraging an advanced pick-and-place tool that uses machine vision to maintain precise alignment, we can place and bond laser devices with precisions better than 500 nanometers in just a few tens of seconds.

In 2021, we also established a wafer-scale silicon-photonics process that improves on this performance. It adds mechanical alignment pedestals and a more precisely etched butt-coupling interface to the silicon chip to achieve vertical alignment of better than a few hundred nanometers. Using these techniques, we assembled certain laser devices on a 300-millimeter silicon photonics wafer. We were delighted to see that as much as 80 percent of the 50 milliwatts of laser light from each device was coupled into the silicon photonics chip to which it was attached. In the worst cases, the coupling was still around 60 percent across the whole wafer. These results rival the kind of coupling efficiencies achieved with active alignment, a more time-consuming process where light from the laser itself is used to steer the alignment process.

A significant advantage of the flip-chip approach is simplicity and flexibility in the kind of chips that get mated. Because they can be produced in existing fabrication lines with limited additional engineering, they can each be sourced from multiple manufactur-





In flip-chip bonding, laser dies [above pink, small raised shapes at left] are individually transferred and bonded to a silicon photonics wafer.

ers. And, with increasing market demand, flip-chip assembly services are being offered by a growing number of vendors. On the other hand, the sequential nature of the process—each laser die needs to be picked up and placed individually—is a significant drawback. It limits the manufacturing throughput and the potential for deep cost reductions in the long run. That's especially important for cost-sensitive applications, like consumer products, and for systems that require multiple laser devices per chip.

2 MICROTRANSFER PRINTING

Microtransfer printing removes some of the alignment difficulties of butt-coupling, while also making the assembly process faster. Just as in flip-chip processing, the light-emitting devices are grown on III-V semiconductor substrates. But there's a big difference: The III-V wafers are not diced into individual chips. Instead, the lasers on the wafer are undercut so that they are attached to the source wafer only by small tethers. The devices are then picked up all together with a tool that's like an ink stamp, breaking the tethers. The stamp then aligns the lasers with waveguide structures on the silicon photonics wafer and bonds them there.

While flip-chip technology uses metallic solder bumps, microtransfer printing uses an adhesive or can even make do with just molecular bonding, which relies on the van der Waals forces between two flat surfaces, to hold the laser in place. Also, the optical coupling between the light source and the waveguide in the silicon photonics chip happens through a different process. Called evanescent coupling, the process places the laser on top of the silicon waveguide structures and the light "bleeds" into it. Although less power is transferred this way, evanescent coupling demands less precise alignment than does butt-coupling.

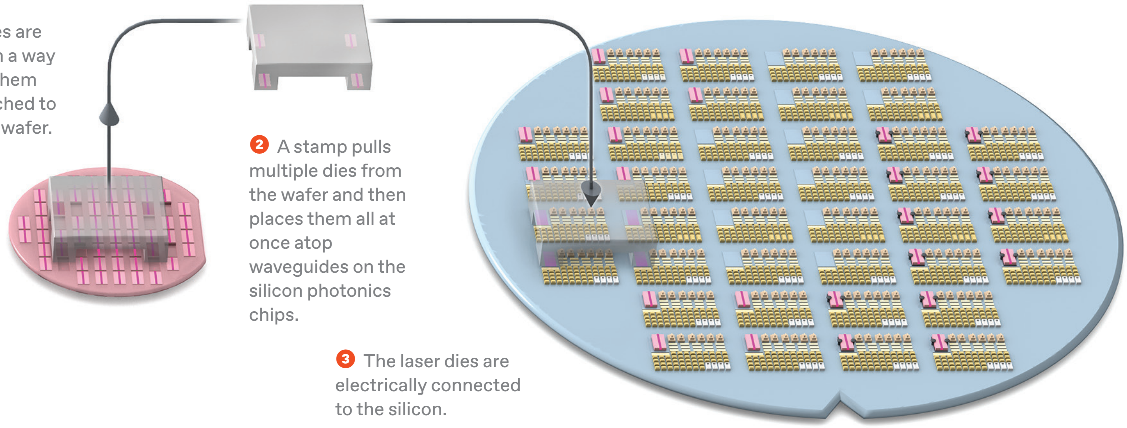
Having greater tolerance in alignment enables this technique to transfer thousands of devices at once. So it should, in principle, allow for higher throughput than flip-chip processing and be ideal for applications that call for the integration of large numbers of III-V components per unit area.

Although transfer printing is an established process for making microLED displays, such as those needed for many augmented reality and virtual reality products, it is not yet ready for printing lasers or optical amplifiers. But we're getting there.

Last year, Imec succeeded in using transfer printing to attach such light sources onto a wafer containing silicon-photonic waveguides, high-speed optical modulators, and photodetectors. We've also printed infrared lasers tunable over 45 nm of wavelength and high-pulse energy devices suitable for chip-based spectroscopy systems. These were made only for demonstration purposes, but we see no fundamental reason that this approach can't achieve good results with high yields. So we expect the technology to be ready for deployment on manufacturing lines within a few years.

OPTOELECTRONICS

Illustration by Emily Cooper



Above: In microtransfer printing, laser dies are weakly held in place on their own wafer. A stamp picks multiple lasers up at once and places them on the silicon photonics wafer. Below: In die-to-wafer bonding, blank pieces of III-V semiconductor are bonded to an already-processed silicon photonics wafer. The III-V material is processed into lasers above the silicon waveguides. The rest of the III-V material is then etched away.

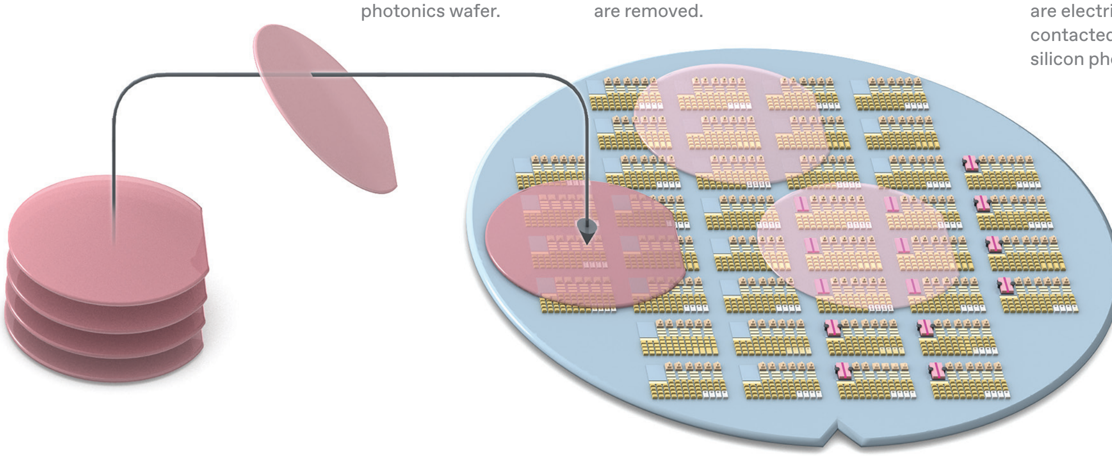
3 DIE-TO-WAFER BONDING

Precisely aligning light-emitting components with their silicon-photonics mates is the critical step in the two technologies we discussed. But one technique, a form of what's called III-V-to-silicon-wafer bonding, finds a way around that. Instead of transferring already-constructed lasers (or other light-emitting components) to a processed silicon wafer, this scheme bonds blank dies (or even small wafers) of a III-V semiconductor to that silicon wafer. You then build the laser devices you need on top of where the corresponding silicon waveguides already are.

Of the transferred material, we're only interested in that thin stratum of crystalline III-V material, called the epitaxial layers. So after bonding with the silicon wafer, the rest of the material is removed. Laser diodes can be fabricated in the epitaxial layers aligned to underlying silicon waveguides using standard lithographic and wafer-scale processing. Any unneeded III-V material is then etched away.

Engineers at Intel developed this approach in the past decade, and the first commercial products built with it, optical transceivers, were launched in 2016. The method allows high throughput integration, because it enables parallel processing of many devices at once. Like transfer printing, it uses evanescent coupling between the III-V and silicon materials, yielding an efficient optical interface.

One drawback of III-V-to-silicon-wafer bonding is that you need substantial investment to establish a manufacturing line that can handle the III-V processing steps using tools meant for fabricating silicon wafers, which are either 200-mm or 300-mm diameter. Such tools are very different from those used in a laser-diode foundry, where the typical wafer diameter is considerably smaller.



Illustrations by Emily Cooper

❹ MONOLITHIC INTEGRATION

The ideal approach to mating the two different materials involved would be to grow III-V semiconductors directly on silicon, an approach called monolithic integration. This would do away with any need for bonding or alignment, and it would reduce the amount of III-V material that is wasted. But many technological hurdles will have to be overcome for this tactic to be practical. So research toward this goal continues at Imec and elsewhere.

The main aim of that research is to create crystalline III-V materials with a low density of defects. The fundamental problem is that there is quite a mismatch—more than 4 percent—between the lattice spacing of atoms in silicon and that of the atoms in the III-V semiconductors of interest.

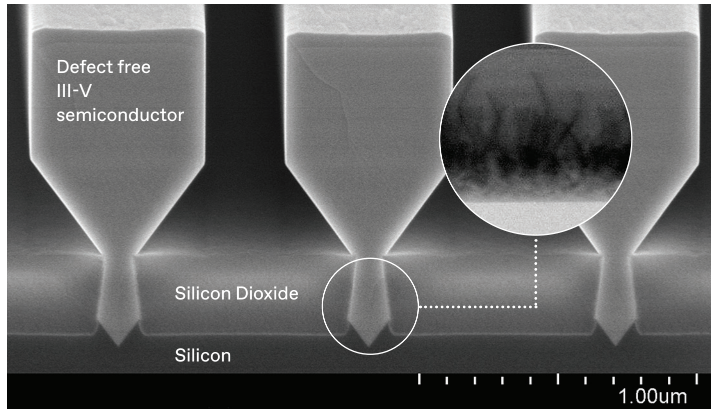
Because of this lattice mismatch, every III-V layer grown on the silicon becomes strained. After only a few nanometers of III-V film is added, defects in the crystal emerge, releasing the built-up strain. These "misfit" defects form along lines that penetrate the entire III-V layer. The defects include lines of open crystal bonds and localized crystal distortion, both of which severely degrade the performance of optoelectronic devices.

To prevent misfit defects from killing the laser, they must be confined to places far from the device. Doing that generally involves laying down a layer of III-V material that is several micrometers thick, forming a hefty buffer between the defects below and a strain-free region above, where the laser devices can be fabricated. Researchers at the University of California, Santa Barbara, have reported excellent progress using this approach, demonstrating high-efficiency gallium-arsenide-based quantum-dot lasers with promising reliability lifetimes.

Those experiments have been done only at small scales, however. Extending the technique to the 200- or 300-mm wafers used in industry will be difficult. The addition of thick buffer layers may lead to various mechanical problems, such as cracks developing inside the III-V film or the wafer bowing. In addition, with the active device on top of such a thick buffer layer, it is challenging to couple light to an underlying waveguide in the silicon substrate.

To circumvent these challenges, Imec has introduced a new approach to monolithic integration called nanoridge engineering, or NRE. The technique aims to force defects to form in such a confined space that working devices can be constructed little more than 100 nm above the interface with the underlying silicon.

NRE confines the defects to small regions using a phenomenon called aspect-ratio trapping. It starts by creating narrow and deep trenches inside a layer of silicon dioxide insulator. At the bottom of the trench, where the insulator meets the silicon, a groove cuts into the silicon, giving the void an arrowhead-shaped cross section. A thin layer of III-V crystal is then grown within the trench, and the strain-induced misfit defects are efficiently trapped at the trench sidewalls, preventing these lines of defects from penetrating farther. After the trench is filled, the growth continues to form a larger nanoridge of III-V material above the trench. The material in that nanometer-scale ridge is sufficiently free of defects so that it can be used for laser devices.



Nanoridge engineering grows laser-suitable semiconductor in specially-shaped trenches in the silicon. The shape of the trench traps defects [inset] well below the area where the laser is constructed.

IMEC

Most research on monolithic integration is done at the level of improving individual devices and identifying reasons for their failure. But Imec has already made substantial progress in demonstrating complete wafer-scale integration with this technique, producing high-quality GaAs-based photodiodes in a 300-mm silicon pilot line. The next milestone will be the demonstration of an electrically pumped laser based on a similar design to the photodiodes. Nanoridge engineering is still under development in the lab, but if it works, it will no doubt have a large impact on this industry.

THE OUTLOOK FOR LASERS ON SILICON

In the next few years, each of the approaches discussed here will surely progress further. We expect that they will eventually coexist to serve different application needs and use cases.

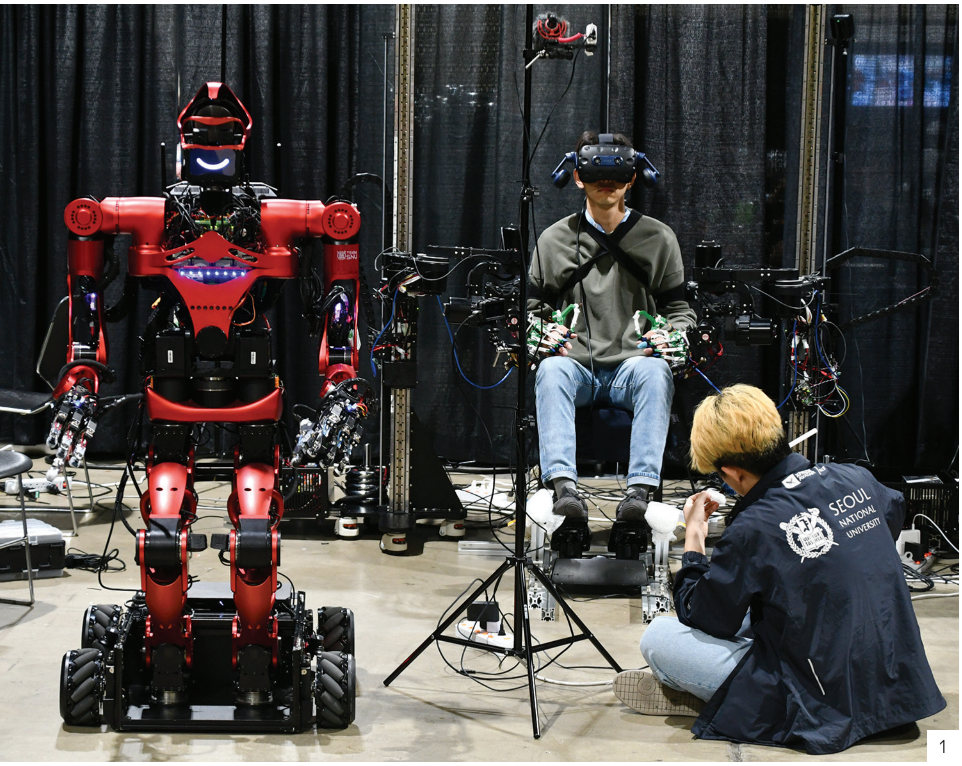
The relatively modest setup cost and readiness of flip-chip laser assembly will enable near-term products. The approach is particularly attractive for applications requiring only one or a couple of lasers per photonic IC, such as the optical transceivers used in data centers. In addition, the flexibility inherent in this approach makes it attractive for applications that require nonstandard laser wavelengths or uncommon photonics technology.

For high-volume applications that require multiple lasers or amplifiers per photonic IC, transfer printing and die-to-wafer bonding offer higher manufacturing throughput, smaller coupling losses, and the potential for deeper cost reductions. Because the setup costs here are substantially higher, the applications for which these techniques are suitable will have to have large markets.

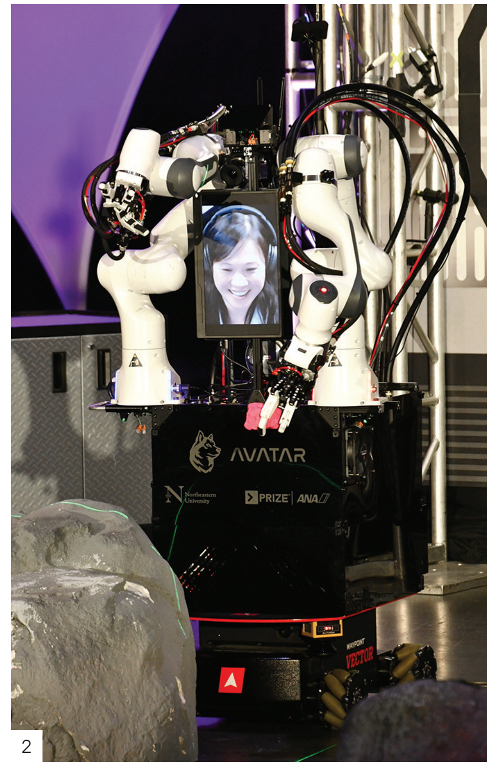
Finally, direct III-V epitaxy on silicon, such as the NRE technique, represents the ultimate level of laser integration. But we and other researchers will have to make further progress in material quality and wafer-scale integration to unlock its potential. ■

The authors wish to thank Katrien Mols.

HUMAN in the LOOP

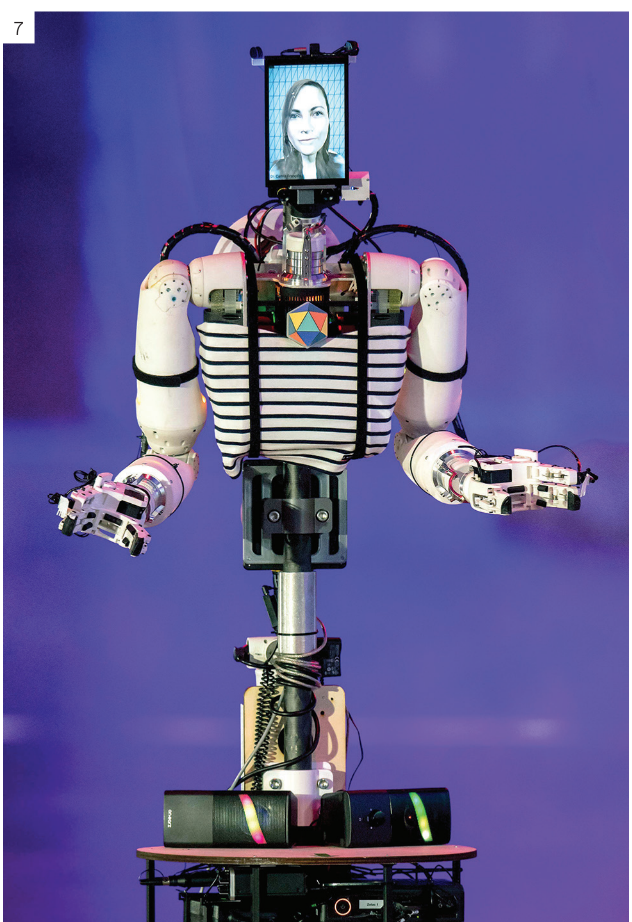


1

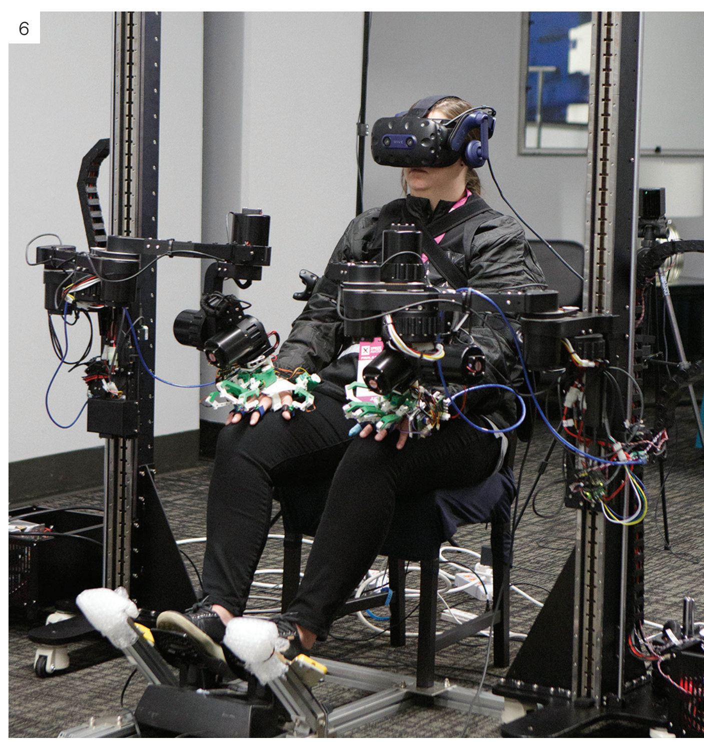


2

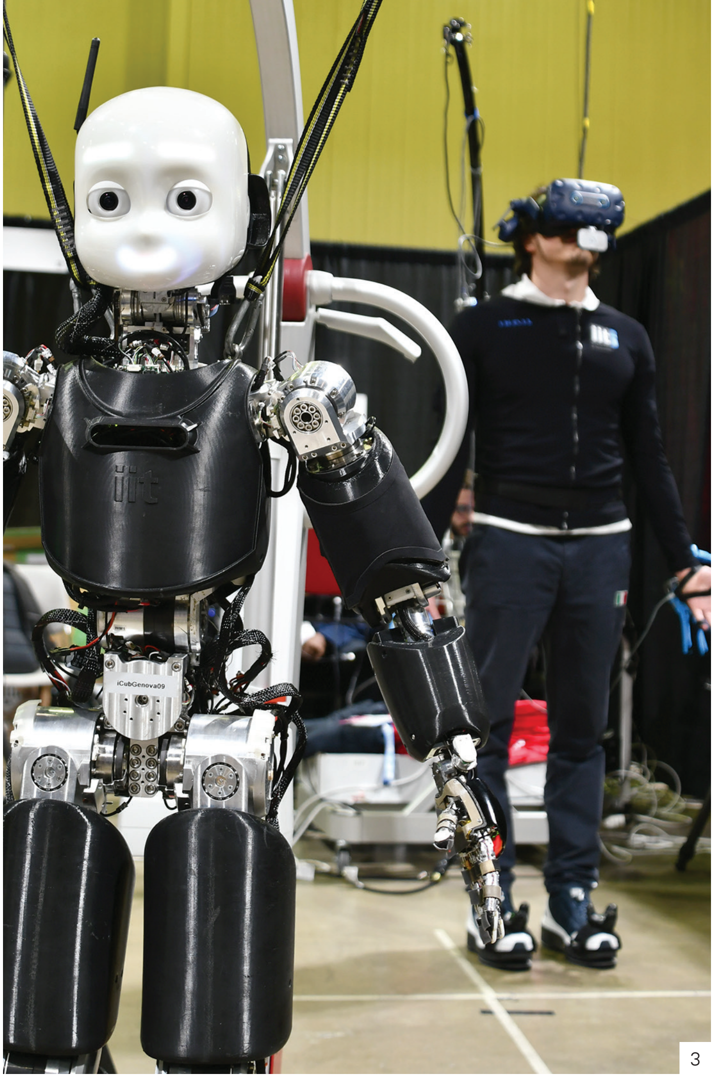
7



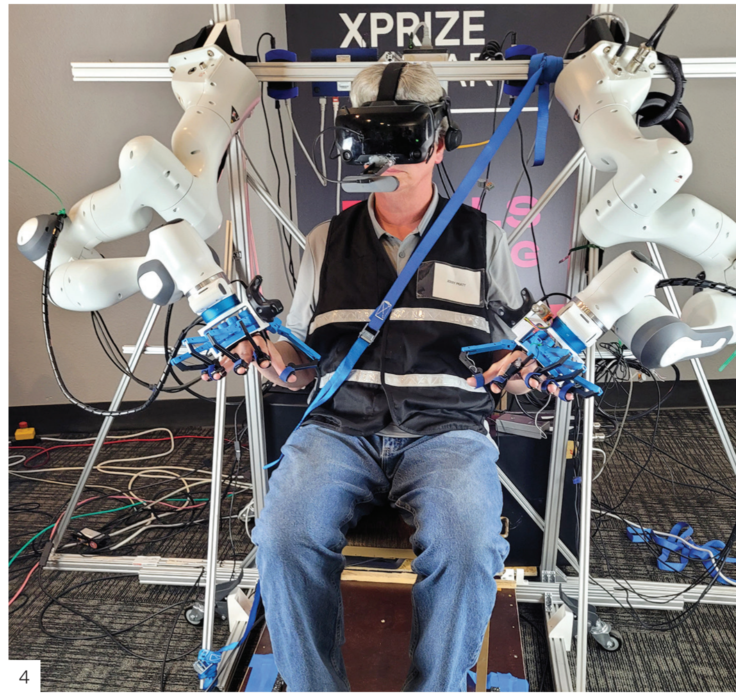
6



By Evan Ackerman



3



4

What the **Avatar XPrize** revealed about the future of telepresence robots

5

[1] Legs were not an asset on the Avatar XPrize course, so Seoul National University adapted by having its legged, human-size robot sit on top of a wheeled base. [2] Northeastern's unique microfluidic actuators gave its operator a detailed sense of touch. [3] Team iCub, from the Istituto Italiano di Tecnologia, believed its bipedal avatar was the most intuitive way to transfer natural human motion to a robot. [4] XPrize judge Jerry Pratt operated NimbRo's robot on the course. [5] Many teams, including i-Botics, relied on commercial virtual-reality headsets as part of their interfaces. [6] Avatar interfaces were made as immersive as possible to help operators control their robots effectively. [7] French startup Pollen Robotics used a modified version of its own commercial platform.

LEFT PAGE: CLOCKWISE FROM BOTTOM LEFT: ZUMA/ALAMY; EVAN ACKERMAN (2); XPRIZE FOUNDATION. RIGHT PAGE: CLOCKWISE FROM BOTTOM: EVAN ACKERMAN (2); TEAM NIMBRO

Robots are NOT READY for the real world.

It's still an achievement for autonomous robots to merely *survive* in the real world, which is a long way from any kind of useful generalized autonomy. Under some fairly specific constraints, autonomous robots are starting to find a few valuable niches in semistructured environments, like offices and hospitals and warehouses. But when it comes to the unstructured nature of disaster areas or human interaction, or really any situation that requires innovation and creativity, autonomous robots are often at a loss.

For the foreseeable future, this means that humans are still necessary. It doesn't mean that humans must be *physically present*, however—just that a human is in the loop somewhere. And this creates an opportunity.

In 2018, the XPrize Foundation announced a competition (sponsored by the Japanese airline ANA) to create "an avatar system that can transport human presence to a remote location in real time," with the goal of developing robotic systems that could be used by humans to interact with the world anywhere with a decent Internet connection. The final event took place last November in Long Beach, Calif., where 17 teams from around the world competed for US $8 million in prize money.

The competition showcased the power of humans paired with robotic systems, transporting our experience and adaptability to a remote location. While the robots and interfaces were very much research projects rather than systems ready for real-world use, the Avatar XPrize provided the inspiration (as well as the structure and funding) to help some of the world's best roboticists push the limits of what's possible through telepresence.

» A ROBOTIC AVATAR

A robotic avatar system is similar to virtual reality, in that both allow a person located in one place to experience and interact with a different place using technology as an interface. Like VR, an effective robotic avatar enables the user to see, hear, touch, move, and communicate in such a way that they feel like they're actually somewhere else. But where VR puts a human into a virtual environment, a robotic avatar brings a human into a physical environment, which could be in the next room or thousands of kilometers away.

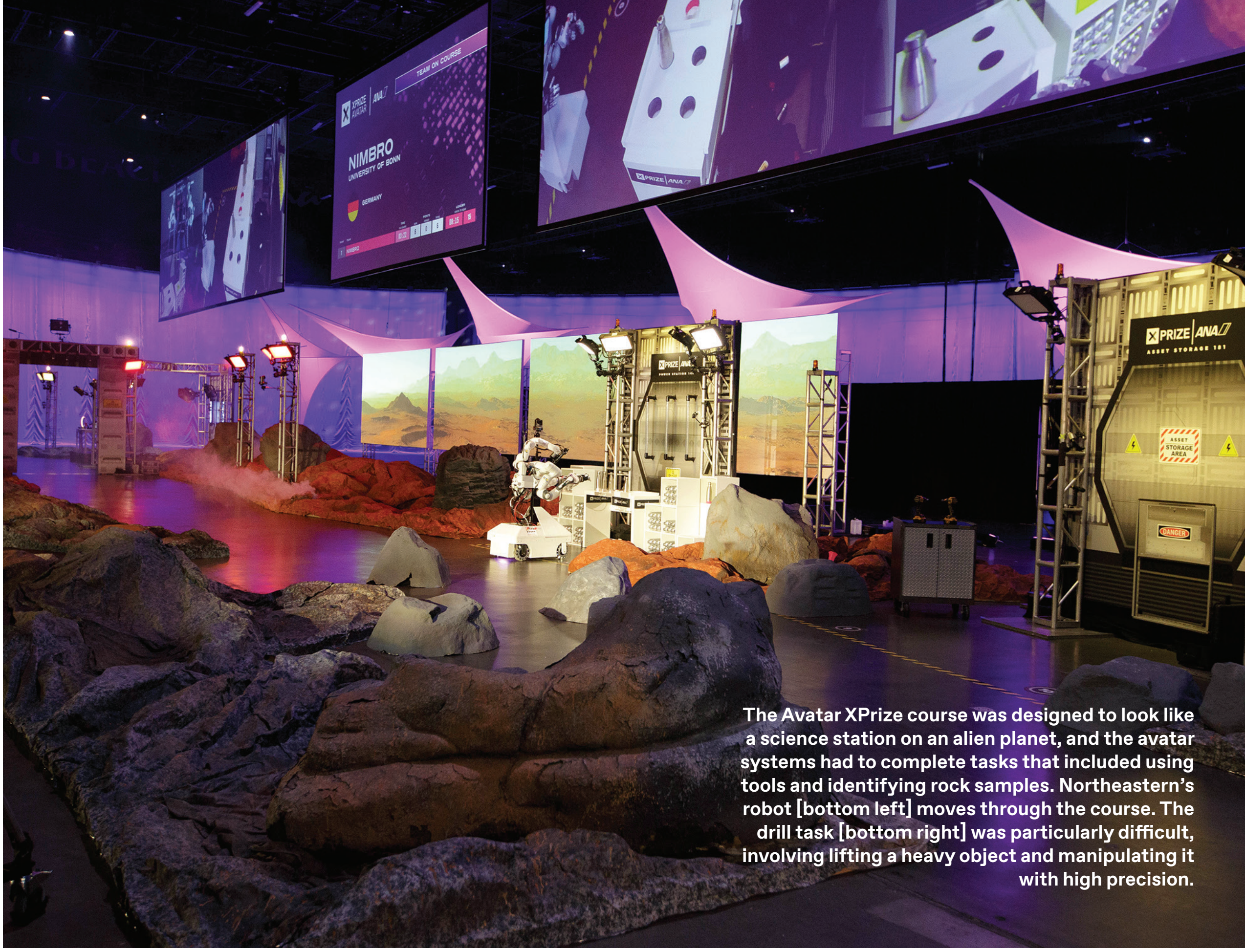
The XPrize Foundation hopes that avatar robots could one day be used for more practical purposes: **providing care** to anyone instantly, regardless of distance; **disaster relief** in areas where it is too dangerous for human rescuers to go; and **performing critical repairs**, as well as maintenance and other hard-to-come-by services.

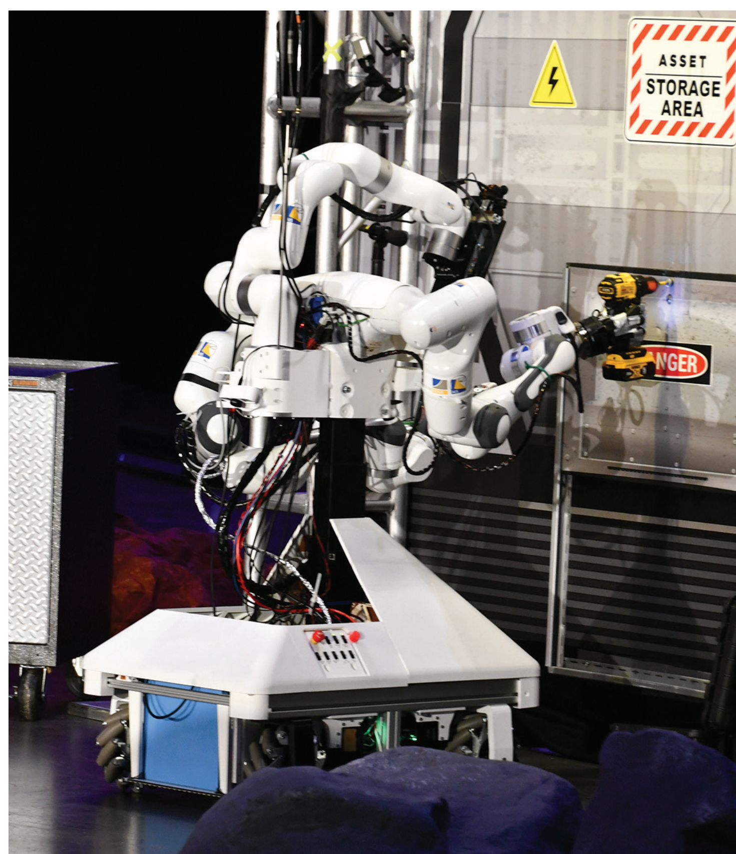
"The available methods by which we can physically transport ourselves from one place to another are not scaling rapidly enough," says David Locke, the executive director of Avatar XPrize. "A disruption in this space is long overdue. Our aim is to bypass the barriers of distance and time by introducing a new means of physical connection, allowing anyone in the world to physically experience another location and provide on-the-ground assistance where and when it is needed."

» GLOBAL COMPETITION

In the Long Beach convention center, XPrize did its best to create an atmosphere that was part rock concert, part sporting event, and part robotics research conference and expo. The course was set up in an arena with stadium seating (open to the public) and extensively decorated and dramatically lit. Live commentary accompanied each competitor's run. Between runs, teams worked on their avatar systems in a convention hall, where they could interact with each other as well as with curious onlookers. The 17 teams hailed from France, Germany, Italy, Japan, Mexico, Singapore, South Korea, the Netherlands, the United Kingdom, and the United States. With each team preparing for several runs over three days, the atmosphere was by turns frantic and focused as team members moved around the venue and worked to repair or improve their robots. Major academic research labs set up next to small robotics startups, with each team hoping their unique approach would triumph.

The competition course included a series of tasks that each robot had to perform, based around a science mission on the surface of an alien planet. Completing the course involved communicating with a human mission commander, flipping an electrical switch, moving through an obstacle course, identifying a container by weight and manipulating it, using a power drill, and finally, using touch





The Avatar XPrize course was designed to look like a science station on an alien planet, and the avatar systems had to complete tasks that included using tools and identifying rock samples. Northeastern's robot [bottom left] moves through the course. The drill task [bottom right] was particularly difficult, involving lifting a heavy object and manipulating it with high precision.

TOP: XPRIZE FOUNDATION; BOTTOM: EVAN ACKERMAN (2) to categorize a rock sample. Teams were ranked by the amount of time their avatar system took to successfully finish all tasks.

There are two fundamental aspects to an avatar system. The first is the robotic mobile manipulator that the human operator controls. The second is the interface that allows the operator to provide that control, and this is arguably the more difficult part of the system. In previous robotics competitions, like the DARPA Robotics Challenge and the DARPA Subterranean Challenge, the interface was generally based around a traditional computer (or multiple computers) with a keyboard and mouse, and the highly specialized job of operator required an immense amount of training and experience. This approach is not accessible or scalable, however. The competition in Long Beach thus featured avatar systems that were essentially operator-agnostic, so that anyone could effectively use them. "Ultimately, the general public will be the end user," explains Locke. "This competition forced teams to invest time into researching and improving the operator-experience component of the technology. They had to open their technology and labs to general users who could operate and provide feedback on the experience, and the teams who scored highest also had the most intuitive and user-friendly operating interfaces."

During the competition, team members weren't allowed to operate their own robots. Instead, a judge was assigned to each team, and the team had 45 minutes to train the judge on the robot and interface. The judges included experts in robotics, virtual reality, human-computer interaction, and neuroscience, but none of them had previous experience as an avatar operator.

Once the training was complete, the judge used the team's interface to operate the robot through the course, while the team could do nothing but sit and watch. Two team members were allowed to remain with the judge in case of technical problems, and a live stream of the operator room captured the stress and helplessness that teams were under: After years of work and with millions of dollars at stake, it was up to a stranger they'd met an hour before to pilot their system to victory. It didn't always go well, and occasionally it went very badly, as when a bipedal robot collided with the edge of a doorway on the course during a competition run and crashed to the ground, suffering damage that was ultimately unfixable.

» HARDWARE AND HUMANS

The diversity of the teams was reflected in the diversity of their avatar systems. The competition imposed some basic design requirements for the robot, including mobility, manipulation, and a communication interface, but otherwise it was up to each team to design and implement their own hardware and software. Most teams favored a wheeled base with two robotic arms and a head consisting of a screen for displaying the operator's face. A few daring teams brought bipedal

While avatar systems are all able to move and interact with their environment, the Avatar XPrize competition showcased a variety of different hardware and software approaches to creating the most effective system.

humanoid robots. Stereo cameras were commonly used to provide visual and depth information to the operator, and some teams included additional sensors to convey other types of information about the remote environment.

For example, in the final competition task, the operator needed the equivalent of a sense of touch in order to differentiate a rough rock from a smooth one. While touch sensors for robots are common, translating the data that they collect into something readable by humans is not straightforward. Some teams opted for highly complex (and expensive) microfluidic gloves that transmitted touch sensations from the fingertips of the robot to the fingertips of the operator. Other teams used small, finger-mounted vibrating motors to translate roughness into haptic feedback that the operator could feel. Another approach was to mount microphones on the robot's fingers. As its fingers moved over different surfaces, rough surfaces sounded louder to the operator while smooth surfaces sounded softer.

In addition to perceiving the remote environment, the operator had to efficiently and effectively control the robot. A basic control interface might be a mouse and keyboard, or a game controller. But with many degrees of freedom to control, limited operator training time, and a competition judged on speed, teams had to get creative. Some teams used motion-detecting virtual-reality systems to transfer the motion of the operator to the avatar robot. Other teams favored a physical interface, strapping the operator into hardware (almost like a robotic exoskeleton) that could read their motions and then actuate the limbs of the avatar robot to match, while simultaneously providing force feedback. With the operator's arms and hands busy with manipulation, the robot's movement across the floor was typically controlled with foot pedals.

Another challenge of the XPrize competition was how to use the avatar robot to communicate with a remote human. Teams were judged on how natural such communication was, which precluded using text-only or voice-only interfaces; instead, teams had to give their robot some kind of expressive face. This was easy enough for operator interfaces that used screens; a webcam that was pointed at the operator and streamed to display on the robot worked well.

But for interfaces that used VR headsets, where the operator's face was partially obscured, teams had to find other solutions. Some teams used in-headset eye tracking and speech recognition to map the operator's voice and facial movements onto an animated face. Other teams dynamically warped a real image of the user's face to reflect their eye and mouth movements. The interaction wasn't seamless, but it was surprisingly effective.

> **"Teleportation might not be an option, but telepresence and telexistence is, where you can actually feel physically present in a location, using the robot as your conduit."**
>
> **DAVID LOCKE**, EXECUTIVE DIRECTOR OF THE ANA AVATAR XPRIZE

XPRIZE FOUNDATION

» HUMAN FORM OR HUMAN FUNCTION?

With robotics competitions like the Avatar XPrize, there is an inherent conflict between the broader goal of generalized solutions for real-world problems and the focused objective of the competing teams, which is simply to win. Winning doesn't necessarily lead to a solution to the problem that the competition is trying to solve. XPrize may have wanted to foster the creation of "avatar system[s] that can transport human presence to a remote location in real time," but the winning team was the one that most efficiently completed the very specific set of competition tasks.

For example, Team iCub, from the Istituto Italiano di Tecnologia (IIT) in Genoa, Italy, believed that the best way to transport human presence to a remote location was to embody that human as closely as possible. To that end, IIT's avatar system consisted of a small bipedal humanoid robot—the 100-centimeter-tall iCub. Getting a bipedal robot to walk reliably is a challenge, especially when that robot is under the direct control of an inexperienced human. But even under ideal conditions, there was simply no way that iCub could move as quickly as its wheeled competitors.

XPrize decided against a course that would have rewarded more humanlike robots—there were no stairs on the course, for example—which prompts the question of what "human presence" actually means. If it means being able to go wherever able-bodied humans can go, then legs might be necessary. If it means accepting that robots (and some humans) have mobility limitations and consequently focusing on other aspects of the avatar experience, then perhaps legs are optional. Whatever the intent of XPrize may have been, the course itself ultimately dictated what made for a successful avatar for the purposes of the competition.

» AVATAR OPTIMIZATION

Unsurprisingly, the teams that focused on the competition and optimized their avatar systems accordingly tended to perform well. Team Northeastern won third place and $1 million using a hydrostatic force-feedback interface for the operator. The interface was based on a system of fluidic actuators first conceptualized a decade ago at Disney Research.

Second place went to Team Pollen Robotics, a French startup. Their robot, Reachy, is based on Pollen Robotics' commercially available mobile manipulator, and it was likely one of the most affordable systems in the competition, costing a mere €20,000 (US $22,000). It used primarily 3D-printed components and an open-source design. Reachy was an exception to the strategy of optimization, because it's intended to be a generalizable platform for real-world manipulation. But the team's relatively simple approach helped them win the $2 million second-place prize.

In first place, completing the entire course in under 6 minutes with a perfect score, was Team NimbRo, from the University of Bonn, in Germany. NimbRo has a long history of robotics competitions; they participated in the DARPA Robotics Challenge in 2015 and have been involved in the international RoboCup competition since 2005. But the Avatar XPrize allowed them to focus on new ways of combining human intelligence with robot-control systems. "When I watch human intelligence operating a machine, I find that fascinating," team lead Sven Behnke told *IEEE Spectrum*. "A human can see deviations from how they are expecting the machine to behave, and then can resolve those deviations with creativity."

Team NimbRo's system relied heavily on the human operator's own senses and knowledge. "We try to take advantage of human cognitive capabilities as much as possible," explains Behnke. "For example, our system doesn't use sensors to estimate depth. It simply relies on the visual cortex of the operator, since humans have evolved to do this in tremendously efficient ways." To that end, NimbRo's robot had an unusually long and flexible neck that followed the motions of the operator's head. During the competition, the robot's head could be seen shifting from side to side as the operator used parallax to understand how far away objects were. It worked quite well, although NimbRo had to implement a special rendering technique to minimize latency between the operator's head motions and the video feed from the robot, so that the operator wouldn't get motion sickness.

The team also put a lot of effort into making sure that using the robot to manipulate objects was as intuitive as possible. The operator's arms were directly attached to robotic arms, which were duplicates of the arms on the avatar robot. This meant that any arm motions made by the operator would be mirrored by the robot, yielding a very consistent experience for the operator.

» THE FUTURE OF HYBRID AUTONOMY

The operator judge for Team NimbRo's winning run was Jerry Pratt, who spent decades as a robotics professor at the Florida Institute for Human and Machine Cognition (IHMC) before joining humanoid robotics startup Figure last year. Pratt had led Team IHMC (and a Boston Dynamics Atlas robot) to a second-place finish at the DARPA Robotics Challenge Finals in 2015. "I found it incredible that you can learn how to use these systems in 45 minutes," Pratt said of his XPrize run. "And operating them is super fun!" Pratt's winning time of 5:50 to complete the Avatar XPrize course was not much slower than human speed.

At the 2015 DARPA Robotics Challenge finals, by contrast, the Atlas robot had to be painstakingly piloted through the course by a team of experts. It took that robot 50 minutes to complete the course, which a human could have finished in about 5 minutes. "Trying to pick up things with a joystick and mouse [during the DARPA competition] is just really slow," Pratt says. "Nothing beats being able to just go, 'Oh, that's an object, let me grab it' with full telepresence. You just do it."

Both Pratt and NimbRo's Behnke see humans as a critical component of robots operating in the unstructured environments of the real world, at least in the short term. "You need humans for high-level decision making," says Pratt. "As soon as there's something novel, or something goes wrong, you need human cognition in the world. And that's why you need telepresence."

Behnke agrees. He hopes that what his group has learned from the Avatar XPrize competition will lead to hybrid autonomy through telepresence, in which robots are autonomous most of the time but humans can use telepresence to help the robots when they get stuck. This approach is already being implemented in simpler contexts, like sidewalk delivery robots, but not yet in the kind of complex human-in-the-loop manipulation that Behnke's system is capable of.

"Step by step, my objective is to take the human out of that loop so that one operator can control maybe 10 robots, which would be autonomous most of the time," Behnke says. "And as these 10 systems operate, we get more data from which we can learn, and then maybe one operator will be responsible for 100 robots, and then 1,000 robots. We're using telepresence to learn how to do autonomy better."

While the Avatar XPrize final competition was based around a space-exploration scenario, Behnke is more interested in applications in which a telepresence-mediated human touch might be even more valuable, such as personal assistance. Behnke's group has already demonstrated how their avatar system could be used to help someone with an injured arm measure their blood pressure and put on a coat. These sound like simple tasks, but they involve exactly the kind of human interaction and creative manipulation that is exceptionally difficult for a robot on its own. Immersive telepresence makes these tasks almost trivial, and accessible to just about any human with a little training—which is what the Avatar XPrize was trying to achieve.

Still, it's hard to know how scalable these technologies are. At the moment, avatar systems are fragile and expensive. Historically, there has been a gap of about five to 10 years between high-profile robotics competitions and the arrival of the resulting technology—such as autonomous cars and humanoid robots—at a useful place outside the lab. It's possible that autonomy will advance quickly enough that the impact of avatar robots will be somewhat reduced for common tasks in structured environments, but it's hard to imagine that autonomous systems will ever achieve human levels of intuition or creativity. That is, there will continue to be a need for avatars for the foreseeable future. And if these teams can leverage the lessons they've learned over the four years of the Avatar XPrize competition to pull this technology out of the research phase, their systems could bypass the constraints of autonomy through human cleverness, bringing us robots that are much more helpful in our daily lives. ■

>>> THE EARLY INTERNET'S DO-OR-DIE MOMENTS

One of the key industry conferences that helped shape the Internet was the 1988 TCP/IP Interoperability Exhibition and Solutions Showcase in Santa Clara, Calif., which was given the shorter, catchier name "Interop."

MARGOT SIMMONS

>>> A handful of demos shaped today's networked world

BY JAMES L. PELKEY, LORING G. ROBBINS & ANDREW L. RUSSELL

>>> **NEW TECHNOLOGIES ARE OFTEN** introduced through spectacle: Think of the historic demonstrations carried out by Faraday, Edison, Morse, and Bell, or more recently, by Steve Jobs onstage in his black turtleneck at Macworld 2007, holding the first iPhone. Indeed, hyped-up product announcements at industry events like the Consumer Electronics Show (now CES) and the Game Developers Conference have become regular features of the digital world.

There's also a parallel tradition—less flashy but no less important—of industry events that focus attention on digital infrastructure. Several of these events, such as the first public demo of the ARPANET in 1972, or the mid-1980s conferences now known as Interop, alerted experts to new technologies and altered the balance between competing approaches.

Although many of these gatherings have escaped the attention of historians, our view is that these events should be recognized more fully as moments when experts could glimpse possible futures and judge for themselves what was most likely to happen. Here we describe a few of these do-or-die moments. You may not have heard of any of these events—but if you were there, you will never forget them.

THE ARPANET WAS one of the first networks to apply packet switching, an approach to communications that breaks messages into discrete chunks, or packets, of data. It was a major departure from circuit-switched networks, such as telephone networks, for which communication partners were linked through a dedicated circuit.

>>>

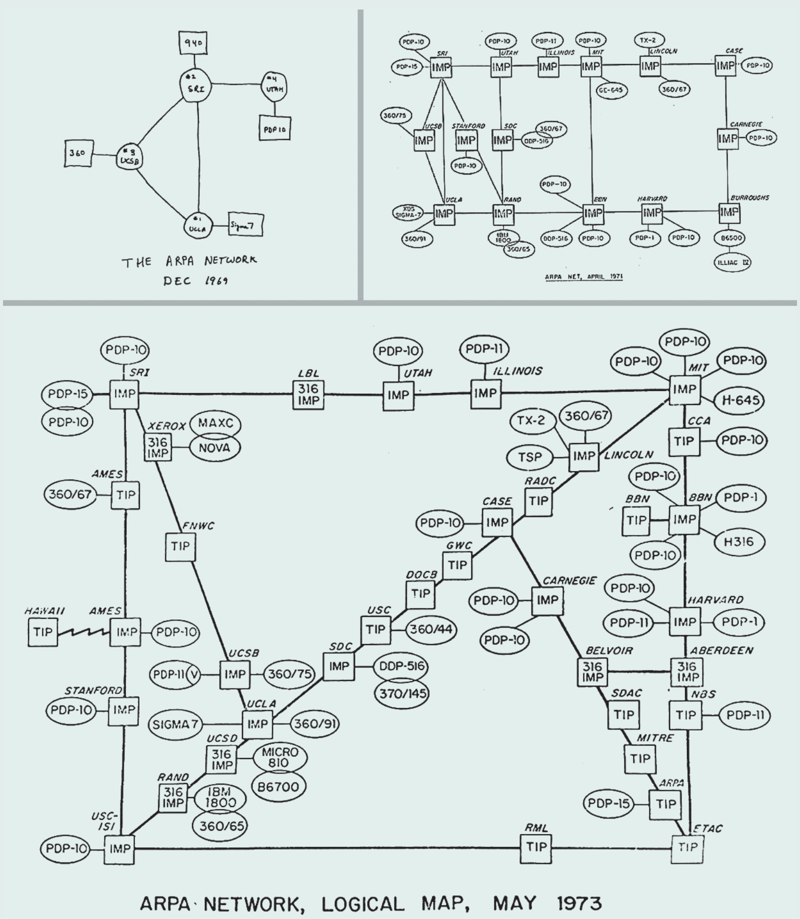
The first node of the ARPANET was installed at the University of California, Los Angeles, in 1969. But the ARPANET didn't take off immediately. And by mid-1971, program director Lawrence Roberts of the Advanced Research Projects Agency (ARPA) was becoming impatient with the slow pace at which ARPA-funded researchers were getting connected. One of these researchers, Bob Kahn, suggested that Roberts organize a public demonstration of the ARPANET, both to educate other researchers about the network's capabilities and to encourage new partners to support the initiative. Once Kahn found a venue for the demo—at the International Conference on Computer Communications (ICCC), to be held in Washington, D.C., in late October of 1972—he worked feverishly to get it organized.

Kahn recruited about 50 people to act as facilitators, including the ARPA-funded researchers Vint Cerf, Robert Metcalfe, and Jon Postel, all of whom were destined for networking fame. [For more on Cerf, see "Mr. Internet," in this issue.] Kahn's plan called for a TIP—short for Terminal Interface Processor—to be installed at the Hilton Hotel, the site of the ICCC. From there, attendees could log on to one of the ARPANET hosts and run an application remotely.

For this, Kahn arranged for various applications (called "scenarios") to be created and tested. He also had to convince manufacturers to loan, install, and configure terminals. And he had to work with the hotel to prepare the room for the demo and arrange with AT&T to run leased lines to the Hilton's ballroom.

The ICCC would prove to be for packet switching what the 1876 Centennial Exposition in Philadelphia was for the telephone: the public unveiling of what would eventually lead to a technological discontinuity.

For the hundreds of computer-communications professionals, government employees, and academic researchers attending the ICCC, the demo permanently changed their perceptions of a computer as a single machine locked in an air-conditioned room. The TIP was on a raised floor in the middle of the ballroom, with dozens of connected computer terminals circled around it and dozens of ARPA scientists milling about, eager to show off their pride and joy.

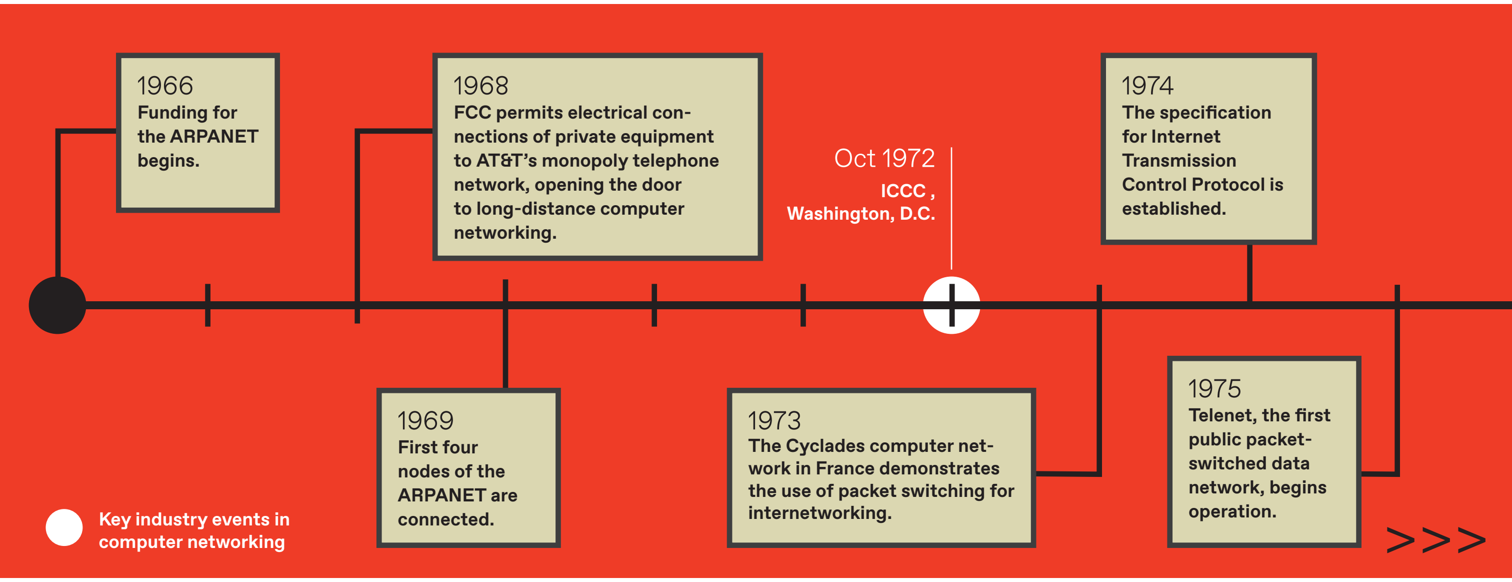


As this first hand drawing, from December 1969, shows, the ARPANET had just four nodes at the time [top left]. That number grew to 15 by April 1971 [top right]. In an effort to speed the expansion further, network advocates organized a demonstration at the International Conference on Computer Communications in Washington, D.C., in October 1972. That meeting helped to grow the ARPANET, which by May of 1973 included some three dozen nodes [bottom].

COMPUTER HISTORY MUSEUM

The Internet's 20th-Century Roots

Here are some of the key events that took place before the turn of the millennium.



To sit at a terminal and with a few keystrokes be connected through the TIP, to the ARPANET, and then to applications running on computers at dozens of universities and research facilities must have felt like a visit to an alien world. And for the ARPA scientists involved, the bonds formed from staging the demonstration left them heady and optimistic about the future they were creating.

RESEARCHERS IN government, academia, and industry struggled over the next several years to realize the potential of what they had seen. How could they scale up and simplify the capabilities that Kahn and company spent a year bringing to the Hilton ballroom? One major problem was the cost and fragility of stringing a dedicated cable from every computer to every terminal. Several parties converged on a similar solution: a local area network, where one "local" cable could traverse an entire facility with all terminals and computers connected to it.

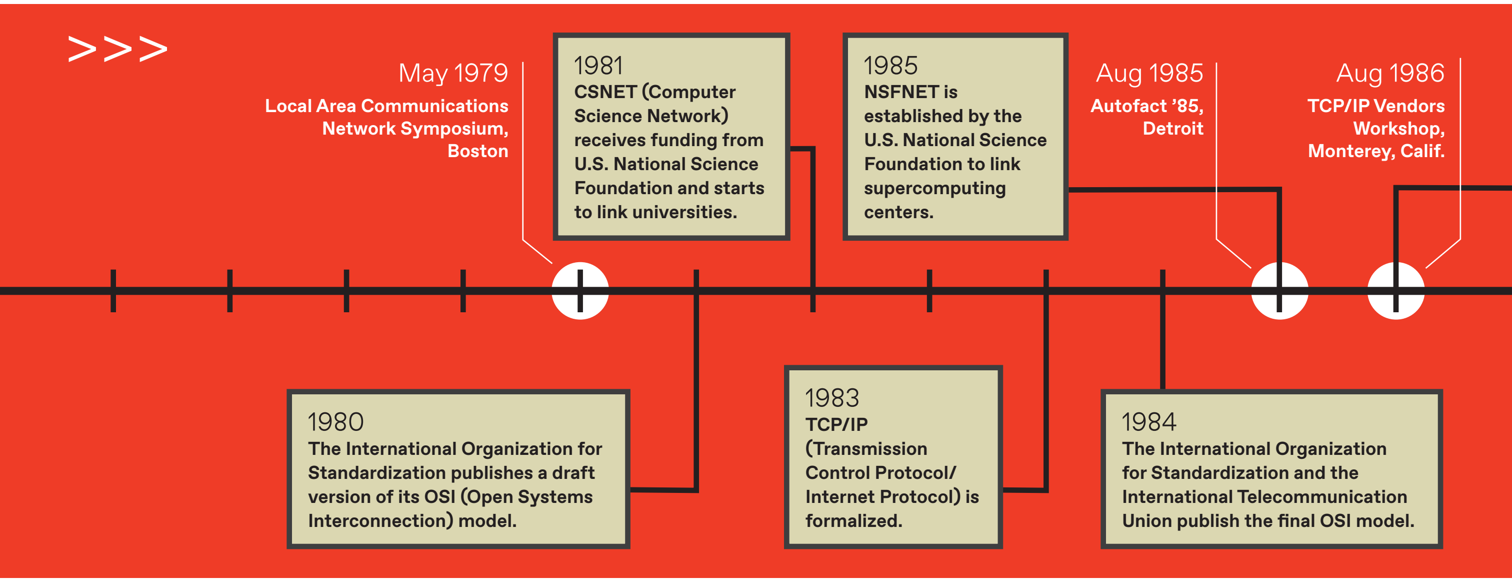
Users in large organizations—including the U.S. Air Force, which had decades of experience and investments in computer communications—had the most to gain from solutions to these problems. To promote cooperation, Robert Rosenthal at the U.S. National Bureau of Standards and Norman Meisner at Mitre, a federally funded R&D organization, arranged a series of workshops in early 1979 to explore "Local Area Network Protocols." Their goal was to provide a mechanism for sharing and obtaining results from the latest research—especially knowledge that was not available in the published literature. When Rosenthal and Meisner contacted potential participants, it became clear that while virtually everyone working on local area networking sensed its importance, they all expressed confusion over what to do about it.

When it came to sorting out the solution, a meeting Rosenthal and Meisner organized in May 1979 proved to have enduring significance. The Local Area Communications Network Symposium, held at the Copley Plaza Hotel in Boston, featured five formal sessions, panel discussions, and twelve workshops. Rosenthal was astonished when about 400 people showed up. For most, it was a formative event, comparable in importance to the ARPANET demonstration in 1972. "There was electricity in the air," Rosenthal recalled in a 1988 interview with one of us (Pelkey). "You had leaders [like] Bob Metcalfe saying: 'The world's going to be a better place.' "

Bruce Hunt of Zilog remembers "being amazed at how many people were really interested in local area networks," and feeling satisfied that the instinct of the researchers involved—that they were onto something really important—was validated. And it wasn't just hype by academics: Within a couple of months, three new companies were formed—Sytek, 3Com, and Ungermann-Bass. Emboldened by the clear demand for commercial networking equipment, these startups raised millions from investors and immediately began selling products for local area networking.

More and more professionals came to realize that networking technology would generate important benefits. But the engineers involved had not settled many technical details about how these networks would work. And a growing number of alternatives soon would be considered for standardization by the IEEE, including a now well-known technology called Ethernet, the concept for which was put forward at Xerox PARC 50 years ago this month.

IN THE MEANTIME, work was underway on a broader approach to the challenge of creating standards for computer communications—one that could serve to link up different computer networks, a concept that began at this time to be called "internetworking." In 1978, a few dozen experts from around the world held the first meeting for an ambitious



project to create a comprehensive suite of standards and protocols for disparate networking technologies. This effort, known as Open Systems Interconnection (OSI), was hosted first by the International Organization for Standardization (ISO) and later, jointly, by ISO and the International Telecommunication Union. OSI's founding premise was that a layered architecture would provide a way to pull together the standards, applications, and services that diverse groups around the world were developing.

The lower layers of OSI concerned the formatting, encoding, and transmission of data over networks. The upper layers included advanced capabilities and applications, such as electronic mail and directory services.

Several initiatives examined proposals for standards and applications within OSI's seven-layer framework. One arose at General Motors, which had a strategic goal of using computer-based automation to combat growing competition from abroad. In 1981, GM held exploratory conversations with Digital Equipment Corp., Hewlett-Packard, and IBM. These discussions culminated in the release of GM's Manufacturing Automation Protocol (MAP) version 1.0 in 1982.

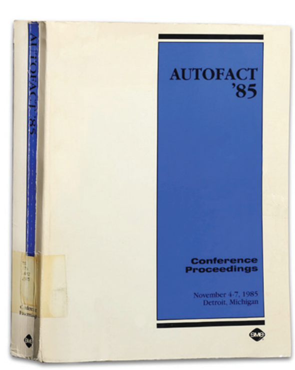
Boeing, with similar goals, announced that it would work with the National Bureau of Standards to lead the creation of an OSI protocol stack for technical and office environments, later to be named Technical and Office Protocols (TOP).

Once again, potential users and customers sought out live demonstrations so that they could judge for themselves what was hype and what was reality. One highly anticipated demo took place at Autofact '85, a conference whose name reflects the era's deep preoccupation with factory automation.

Autofact '85 drew about 30,000 people to Detroit, with some 200 vendors exhibiting MAP-compatible and other kinds of automation products. In addition to data-processing equipment such as computers and terminals, a variety of factory-automation systems, including robots, vision systems, and engineering workstations, were on display. With them, attendees explored a custom-designed version of the Towers of Hanoi game, and an application for interactive file transfer, access, and management.

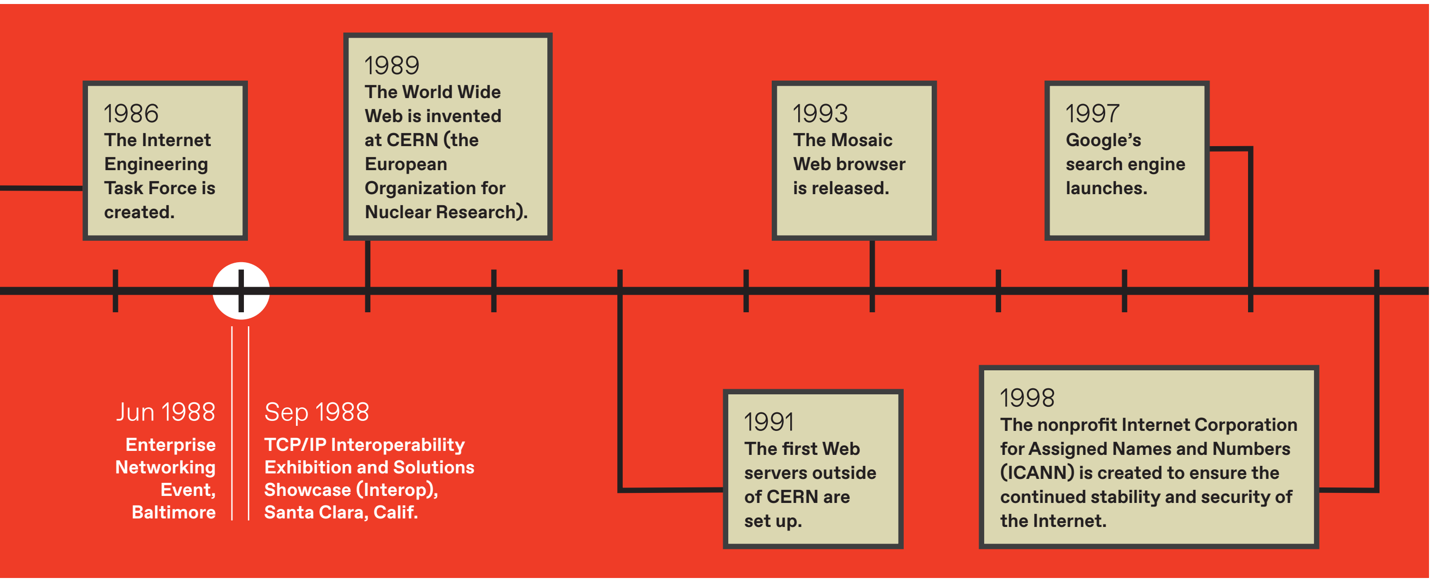
Although Autofact '85 was well attended and generally hailed in the trade press as a success, some were put off by its focus on things to come. As one press account put it, "On the show floor, there are plenty of demonstrations but few available products." The lingering questions around actual commercial applications gave promoters reason to continue organizing public demonstrations, such as the Enterprise Networking Event (ENE) in Baltimore, in June 1988.

The hope for ENE was to provide demos as well as showcase products that were actually available for purchase. All the U.S. computing giants—including IBM, HP, AT&T, Xerox, Data General, Wang Laboratories, and Honeywell—would be there, as well as leading European manufacturers and some smaller and younger companies with OSI-compatible products, such as Apple, Micom, Retix, Sun Microsystems, 3Com,



Autofact '85, a conference whose name reflects people's keen interest in factory automation during that time, brought about 30,000 people and 200 vendors to Detroit in November of 1985.

SUNY POLYTECHNIC INSTITUTE



and Touch Communications. Keynote speakers from the upper levels of the U.S. Department of Defense, Arthur Andersen, and the Commission of European Communities reinforced the message that all major stakeholders were behind the global adoption of OSI.

ENE confirmed both the hopes of OSI's supporters and the fears of its critics. Vendors were able to demonstrate OSI standards for network management and electronic mail, but instead of products for sale, the 10,000 or so attendees saw mostly demonstrations of prototypes—a marginal improvement on Autofact '85.

THERE WAS A PAINFUL reality to the computer networks of the mid-1980s: On the one hand, they held a vast potential to improve business practices and enhance productivity; on the other, actual products that could integrate the diversity of installed equipment and networks—and thus provide a robust means of internetworking—were very limited. The slow progress of MAP and TOP products left an opening for alternative approaches.

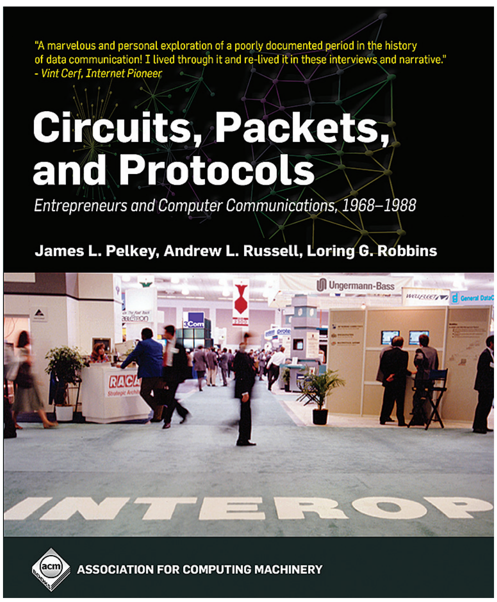
And the most promising of those approaches was to rely on the core protocols then in use for the ARPANET: Transmission Control Protocol and Internet Protocol (known to insiders as TCP/IP). A broad market for suitable equipment hadn't yet developed, but the community of experts that had grown around the ARPANET was increasingly active in promoting the commercial adoption of such products.

One of the chief promoters was Dan Lynch, a consultant who was instrumental in managing the ARPANET's transition to TCP in 1983. Lynch led the planning of a workshop in Monterey, Calif., in August 1986, where equipment vendors could learn about TCP/IP. Lynch wanted to get the apostles of TCP "to come out of their ivory towers" and provide some guidance for vendors implementing their protocols. And they did, as Lynch recalled in a 1988 interview, where he called the workshop "outrageously successful."

This meeting, the first TCP/IP Vendors Workshop, featured a mix of leaders from the TCP/IP research community and representatives from 65 vendors, such as Ungermann-Bass and Excelan. Lynch continued this trade-show-like approach with the TCP/IP Interoperability Conference, in Monterey, in March 1987 and the 2nd TCP/IP Interoperability Conference, in Arlington, Va., in December of the same year.

Lynch's strategy for TCP/IP seemed to be gaining momentum, as evidenced by an article in *Data Communications* in November 1987, which neatly summarized the state of affairs: "By the end of 1986, there were more than 100 vendor offerings of TCP/IP and its associated DARPA protocols. Moreover, major vendors, including IBM and Digital Equipment Corp. have recently begun to offer

Vendors could participate in TCP/IP "bake-offs," where they could check to see whether their equipment interoperated with other vendors' products.



This article is based on parts of the authors' recent book *Circuits, Packets, and Protocols: Entrepreneurs and Computer Communications, 1968-1988* (ACM Books, 2022). Readers interested in purchasing the book will receive a 25 percent discount with the publisher at bit.ly/3Lhj2NK using the promo code SPECTRUM25.

ACM

TCP/IP as part of their product lines.... While the long-term strategic direction taken by most companies is in the implementation of the OSI model and its protocols, TCP/IP appears to be solving the short-term problems of connections between networks."

The market-research firm Infonetics published a report in May 1988 that documented a "dramatic increase in the commercialization of TCP/IP" and noted that increasing numbers of users were seeking solutions to integrate diverse computer equipment and networks. "Every sector of the market is planning to purchase TCP/IP products in the next year," the report stated. "There is no indication that OSI is affecting purchase intent."

AT THE TIME, Lynch was planning a new venue to promote the adoption of the protocols used for the ARPANET: the TCP/IP Interoperability Exhibition and Solutions Showcase, to be held in Santa Clara, Calif., in September 1988. And he gave the event a slick new title: Interop.

Interop featured lots of products: "every medium, every bridge box, every router you can imagine," according to Peter de Vries of the Wollongong Group, which was responsible for putting together the network at Interop. That network provided connections among all vendors on display, including equipment available for purchase from Cisco Systems, Proteon, and Wellfleet Communications.

Using TCP/IP, attendees could traverse links to NSFNET, the regional BARRNET in San Francisco, and a variety of other networks. Vendors could participate in TCP/IP "bake-offs," where they could check to see whether their equipment interoperated with other vendors' products. Self-appointed "net police" went so far as to hand out "tickets" to implementations that did not comply with the TCP/IP specifications.

In many respects, Interop '88 was far more successful than ENE. It featured working products from more vendors than did ENE. And whereas ENE carried the burden of people's expectations that it would provide comprehensive solutions for large-scale manufacturing, office, and government procurement, Interop took on the immediate and narrower problems of network interconnection. In the "age of standards," as an article in *Data Communications* referred to that time, this focus on product compatibility, interoperability, and connectivity energized the estimated 5,000 attendees as well as the market for TCP/IP products.

The stage was now set for innovations that would change global society: the invention of the World Wide Web the following year and the privatization of the NSFNET/Internet backbone in the mid-1990s. The advances in global computer networking that have come since then all rest on that initial foundation.

Accounts of the beginnings of modern computing often include dramatic descriptions of a conference that has since become known as "the Mother of all Demos"—a 1968 joint meeting of the Association for Computing Machinery and the IEEE Computer Society where ARPA-funded researcher Douglas Engelebart gave a 90-minute presentation that included the use of windows, hypertext, videoconferencing, and the computer mouse, among other innovations. His demo is rightly recognized as a turning point for expanding the realm of the possible in personal computing.

But mind-expanding possibilities were also on display—and sometimes even for sale—at the five meetings we've described here. In our view, the contribution of these industry events to the development of today's world of computing shouldn't be forgotten, because the *connection* of different kinds of computers is the advance that has transformed our lives. ■

Loring Robbins and Andrew Russell dedicate this article to their coauthor and longtime friend James Pelkey, who died shortly before it was published.

IEEE WIE is one of the world's leaders in changing the face of engineering. Our global network connects over 20,000 members in over 100 countries to advance women in technology at all points in their life and career.

wie.ieee.org

#IEEEWIElead

IEEE

BY ALLISON MARSH

Past Forward

A cattle rancher found this chunk of Skylab more than a decade after the space station reentered Earth's atmosphere on 12 July 1979.

Skylab's Great Fall

On 14 May 1973, a modified Saturn V rocket launched from the Kennedy Space Center carrying Skylab, the United States' first space station. Skylab's fate was sealed moments after lift-off when the sun shield and main solar panel were significantly damaged, making it questionable whether the space station could fulfill its planned missions. Eleven days later, after engineers hurriedly designed and tested some possible fixes, the first Skylab crew launched, rendezvoused, and started emergency repairs. Skylab went on to host two more missions over the next year.

Eventually, though, Skylab's orbit decayed. The craft reentered Earth's atmosphere in a fiery blaze in the early hours of 12 July 1979, spreading debris across the Indian Ocean and Western Australia.

More than a decade later, a rancher found this titanium and fiberglass end cap from one of Skylab's oxygen tanks in the dirt, after he spotted his cattle drinking rainwater from its shallow bowl. The piece, one of the surviving remnants of the US $2.2 billion NASA program, now resides at the Museum of Applied Arts and Sciences in Sydney. ■

FOR MORE ON THE HISTORY OF SKYLAB, SEE spectrum.ieee.org/pastforward-may2023

PHOTO: RYAN HERNANDEZ; OBJECT: POWERHOUSE COLLECTION