IEEE SPECTRUM

FOR THE TECHNOLOGY INSIDER | 06.20

Rising to the Challenge

TURBINE-POWERED, LONG-RANGE CARGO DRONES COULD UPEND THE AIR-FREIGHT INDUSTRY **P. 30**

IEEE

IEEE
SPECTRUM

CONTROLLING COVID-19 WITH CONTROL THEORY

With meticulous use of feedback, officials can keep hospitals' COVID-19 caseloads within manageable limits.

BY GREG STEWART, KLASKE VAN HEUSDEN & GUY A. DUMONT *Page 22*

The Institute

On the cover: Illustration for *IEEE Spectrum* by John MacNeill

IEEE SPECTRUM
(ISSN 0018-9235) is published monthly by The Institute of Electrical and Electronics Engineers, Inc. All rights reserved. © 2020 by The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016-5997, U.S.A. Volume No. 57, Issue No. 6. 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: $49.95 included in dues. Libraries/institutions: $399. POSTMASTER: Please send address changes to IEEE Spectrum, c/o Coding Department, IEEE Service Center, 445 Hoes Lane, Box 1331, Piscataway, NJ 08855. Periodicals postage paid at New York, NY, and additional mailing offices. Canadian GST #125634188. Printed at 120 Donnelley Dr., Glasgow, KY 42141-1060, U.S.A. IEEE Spectrum circulation is audited by BPA Worldwide. IEEE Spectrum is a member of the Association of Business Information & Media Companies, the Association of Magazine Media, and Association Media & Publishing. IEEE prohibits discrimination, harassment, and bullying. For more information, visit https://www.ieee.org/web/aboutus/whatis/policies/p9-26.html.

PHOTOGRAPH BY John Tlumacki/The Boston Globe/Getty Images

KEVIN KENNEDY

IT'S A TAKE

AS A KID, LARRY HERBST LOVED TINKERING. Naturally enough for someone like that, he trained as an electrical engineer. But early on he decided that a career in engineering might end up forcing him to focus more narrowly than he would like. So after graduating from Brown University, he changed course and studied filmmaking. Ever since, he's made his living as a filmmaker and videographer, and he has no regrets. "Film gave me an opportunity to do a lot of extraordinary things," he explains.

Herbst has made a specialty of covering technical and scientific topics, working frequently with the XPrize Foundation to document its technology competitions. And in 2012 he was part of James Cameron's Deepsea Challenge expedition to dive to the bottom of the Mariana Trench, for which Herbst designed one of the camera-control systems. He's also filmed at sea with Barry Clifford, the underwater explorer famous for his discovery of historic shipwrecks, and aboard Zero Gravity Corp.'s weightless flights.

Even in his creative role as a videographer, Herbst says, he finds plenty of opportunity to do technical work. He built a miniature 3D time-lapse camera that was used for filming an Imax movie, and he is a coauthor (with Cameron and others) of a technical publication on the construction of "hadal landers"–oceanographic instruments designed to descend to the bottom of deep-sea trenches.

Herbst has spent a lot of time on the water, too, with Kevin Kennedy, an Alaskan fisherman who engineered a novel oil-cleanup apparatus. Kennedy is the subject of Herbst's article in this issue [see "Fishing for Oil," page 36].

The photo above shows Herbst preparing to film Kennedy at Cook Inlet in Alaska. Herbst really enjoys working in remote locales. "I love to see the world," he says. And being able to make his living this way only makes him love it more. ■

IEEE
SPECTRUM

EDITOR IN CHIEF Susan Hassler, s.hassler@ieee.org
EXECUTIVE EDITOR Glenn Zorpette, g.zorpette@ieee.org
EDITORIAL DIRECTOR, DIGITAL
Harry Goldstein, h.goldstein@ieee.org
MANAGING EDITOR Elizabeth A. Bretz, e.bretz@ieee.org
SENIOR ART DIRECTOR
Mark Montgomery, m.montgomery@ieee.org
SENIOR EDITORS
Stephen Cass, cass.s@ieee.org
Erico Guizzo (Digital), e.guizzo@ieee.org
Jean Kumagai, j.kumagai@ieee.org
Samuel K. Moore, s.k.moore@ieee.org
Tekla S. Perry, t.perry@ieee.org
Philip E. Ross, p.ross@ieee.org
David Schneider, d.a.schneider@ieee.org
Eliza Strickland, e.strickland@ieee.org
DEPUTY ART DIRECTOR Brandon Palacio, b.palacio@ieee.org
PHOTOGRAPHY DIRECTOR Randi Klett, randi.klett@ieee.org
ONLINE ART DIRECTOR Erik Vrielink, e.vrielink@ieee.org
NEWS MANAGER Amy Nordrum, a.nordrum@ieee.org
ASSOCIATE EDITORS
Willie D. Jones (Digital), w.jones@ieee.org
Michael Koziol, m.koziol@ieee.org
SENIOR COPY EDITOR Joseph N. Levine, j.levine@ieee.org
COPY EDITOR Michele Kogon, m.kogon@ieee.org
EDITORIAL RESEARCHER Alan Gardner, a.gardner@ieee.org
ADMINISTRATIVE ASSISTANT
Ramona L. Foster, r.foster@ieee.org
CONTRIBUTING EDITORS Evan Ackerman, Mark Anderson, Robert N. Charette, Peter Fairley, W. Wayt Gibbs, Tam Harbert, Mark Harris, David Kushner, Robert W. Lucky, Prachi Patel, Morgen E. Peck, Richard Stevenson, Lawrence Ulrich

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

DIRECTOR, PERIODICALS PRODUCTION SERVICES Peter Tuohy
EDITORIAL & WEB PRODUCTION MANAGER Roy Carubia
SENIOR ELECTRONIC LAYOUT SPECIALIST Bonnie Nani
PRODUCT MANAGER, DIGITAL Shannan Dunlap
WEB PRODUCTION COORDINATOR Jacqueline L. Parker
MULTIMEDIA PRODUCTION SPECIALIST Michael Spector

ADVERTISING PRODUCTION +1 732 562 6334
ADVERTISING PRODUCTION MANAGER
Felicia Spagnoli, f.spagnoli@ieee.org
SENIOR ADVERTISING PRODUCTION COORDINATOR
Nicole Evans Gyimah, n.gyimah@ieee.org

EDITORIAL ADVISORY BOARD, *IEEE SPECTRUM*
Susan Hassler, *Chair*; Steve Blank, David C. Brock, Ronald F. DeMara, Shahin Farshchi, Lawrence O. Hall, Jason K. Hui, Leah Jamieson, Mary Lou Jepsen, Deepa Kundur, Gianluca Lazzi, Allison Marsh, Carmen Menoni, Sofia Olhede, Maurizio Vecchione, Edward Zyszkowski

EDITORIAL ADVISORY BOARD, *THE INSTITUTE*
Kathy Pretz, *Chair*; Qusi Alqarqaz, Philip Chen, Roberto Graglia, Shashank Gaur, Susan Hassler, Cecilia Metra, San Murugesan, Mirela Sechi Annoni Notare, Tapan K. Sarkar, Joel Trussell, Hon K. Tsang, Chonggang Wang

EDITORIAL CORRESPONDENCE
IEEE Spectrum, 3 Park Ave., 17th Floor,
New York, NY 10016-5997
TEL: +1 212 419 7555 FAX: +1 212 419 7570
BUREAU Palo Alto, Calif.; Tekla S. Perry **+1 650 752 6661**

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

ADVERTISING INQUIRIES Naylor Association Solutions, Erik Henson +1 352 333 3443, ehenson@naylor.com

REPRINT SALES +1 212 221 9595, ext. 319
REPRINT PERMISSION / LIBRARIES Articles may be photocopied for private use of patrons. A per-copy fee must be paid to the Copyright Clearance Center, 29 Congress St., Salem, MA 01970. For other copying or republication, contact Managing Editor, *IEEE Spectrum*.

CONTRIBUTORS

Ed De Reyes

As an ex-Air Force pilot, De Reyes knew that his startup, Sabrewing Aircraft Co., in Camarillo, Calif., had to keep its pilotless cargo plane from crashing into things. "Even the military does it only in a kind of rudimentary way–say, with a camera system," he says. "Our system has to provide a way for the aircraft to autonomously avoid obstacles." Read about how his company tackled the problem in "Can Cargo Drones Solve Air Freight's Logjams?" [p. 30].

Yulia Frumer

When Frumer first heard of Gakutensoku, an early humanoid robot, she became intrigued by its creator, Makoto Nishimura. "He wasn't an engineer but a biologist who studied marine moss balls," says Frumer, a historian of science and technology at Johns Hopkins University. In "The Short, Strange Life of the First Friendly Robot" [p. 42], she tells the improbable tale of how Nishimura built Gakutensoku and how it influenced generations of roboticists.

Stacey Higginbotham

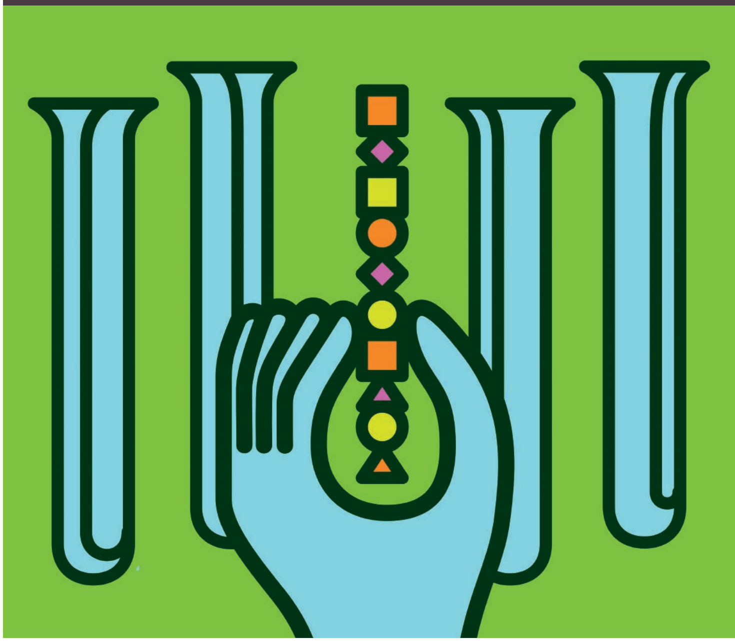
Higginbotham is the editor of Stacey on IoT, an online publication covering the Internet of Things. She also writes *IEEE Spectrum*'s Internet of Everything column. She's been covering technology for nearly two decades, and she finds that sometimes tech's best feature is also its biggest vulnerability. Case in point: In "Pandemic vs. Privacy" [p. 20], she explores how the IoT can simultaneously help coronavirus tracking efforts while also putting personal privacy at risk.

John MacNeill

MacNeill, an illustrator based in Watertown, Mass., depicted a cargo drone for this issue [p. 30]. It's his first *Spectrum* assignment since the pandemic struck. MacNeill is used to working at home, but these days his wife is telecommuting, and their two sons are studying online. "It's working out reasonably well," he says. "If somebody has a conference call, everyone cooperates–so far nobody's decided that would be a fine time to start throwing the football around."

Greg Stewart

Stewart is vice president of data science for the agriculture-technology startup Ecoation and an adjunct professor at the University of British Columbia, in Vancouver. An IEEE Fellow, he and two UBC colleagues, Klaske van Heusden and Guy A. Dumont, describe in this issue how control theory can control the spread of COVID-19 [p. 22]. Van Heusden, an IEEE Senior Member, is a research associate at UBC. Dumont, an IEEE Fellow, is a professor of electrical and computer engineering.



Researchers Are Using Algorithms to Tackle the Coronavirus Test Shortage

The scramble to develop new test kits that deliver faster results

We're hearing about the problem daily: There aren't enough test kits available to accurately track the new coronavirus. It's a global problem, although some countries—including the United States—are doing a worse job of testing than others. ¶ More testing, epidemiologists say, would give us a better understanding of how the virus is moving through the population and whether it's waxing or waning. And more testing is essential to allow us to safely return to any semblance of normal life. ¶ Part of the solution to the testing shortage, however, may come not from making more kits but from algorithms. Computer scientists and engineers who study ways to compress data, decode video, and process medical images say that biological samples from groups of people could be pooled for testing, and the results decoded by cleverly designed algorithms. This sample pooling could go a long way toward addressing the test-kit shortage, with one test kit being able to determine the health status of 5, 10, or even more people.¶ I spoke with Dror Baron, associate professor of electrical and computer engineering at North Carolina State University and an IEEE Senior Member, about his work developing algorithms for this kind of coronavirus test processing.

***IEEE Spectrum:* What exactly is a "pooled" test?**
Dror Baron: In a pooled test, you are taking biological material from multiple people and testing it once. If what you are looking for doesn't exist in the group, then you can rule out a large number of people with one measurement. This works best in situations in which the majority of people being tested don't have the disease. It works well for screening HIV in blood samples, because most people who donate blood are HIV negative.

An extended version of this article appears in the View From the Valley blog.

Post your comments at spectrum.ieee.org/testing-jun2020

***Spectrum:* Why so much interest?**
D.B.: We are going to want to test huge numbers of people for the coronavirus—like anybody who gets on a plane, maybe even anybody who gets into a major government building. If testing was available at a reasonable cost, Amazon, for example, could test its warehouse employees every day. To truly restart the economy, we need to test not tens or hundreds of thousands a day but millions a day.

***Spectrum:* How does your research relate to this?**
D.B.: Like so many of us, I've been thinking about the coronavirus. I realized I could apply one of our algorithms, one that uses a graph data structure. In that example of a group of 100, with each patient appearing in six test groups, you would have 100 nodes that are patients and 20 nodes that are tests. The algorithm passes messages containing statistical information between the patient nodes and the test nodes—revealing the probability that I, as a patient, am sick based on the information gathered so far.

For example, if most of the six groups I was tested in are positive, then the probability of my being sick is very high. For those people who tested positive in two, three, or four groups, the algorithm would adjust the probabilities of those people being sick based on whether or not I, as very likely sick, was in their groups—that is, "I tested sick five times and you only tested sick twice, both times in tests with me, so you very likely are fine." If you design the pools cleverly, every pair of patients will be in no more than two pools together.

Like many researchers around the world, my group has something that could be helpful, but we don't know who to reach out to. This is not just a problem in the United States—it is worldwide. It's a chaotic situation. —TEKLA S. PERRY

CORRECTION: In "What to Do With 177 Giant Tanks of Radioactive Sludge" [May], we mislabeled the Yakima River as the Hood River. And the top level of waste in the single-shell tank diagram is called a supernatant, not a supermutant.

ILLUSTRATION BY Greg Mably

News

HEADS UP: A crane lowers a heavy structure called a top head onto the Vogtle Unit 4 containment vessel in Georgia.

GEORGIA POWER CO.

LIMITED PROGRESS FOR U.S. NUCLEAR

Two fission projects raise questions about the long-term future of nuclear power

There are 53 nuclear reactors currently under construction around the world. Only two are in the United States, once the world's leader in nuclear energy development. And those two reactors represent expansions of a preexisting two-reactor facility, Plant Vogtle in Waynesboro, Ga.

Separately, a company in Portland, Ore., called NuScale Power is now working with the U.S. Nuclear Regulatory Commission to develop a next-generation reactor built around a smaller-scale, modular design.

These two projects together represent the leading edge of commercial U.S. nuclear-fission reactor development today. The fact that there are only two raises questions about the direction of this once-booming energy sector. Is the United States redirecting its focus onto fusion and leaving fission behind? Or could a fission renaissance be yet to come?

Congress upped the U.S. Department of Energy's nuclear *fusion* budget from US $564 million to $671 million for fiscal year 2020. And such companies as AGNI Energy in Washington state and Commonwealth Fusion Systems in Massachusetts (alongside Tokamak Energy and General Fusion in the United Kingdom and Canada, respectively) are courting venture capital for their multimillion-dollar visions of fusion's bright future.

Meanwhile, in March, construction workers at the Vogtle fission plant hoisted a 680,000-kilogram steel-and-concrete structure to cap one of the containment vessels for the new AP1000 reactors. As John Kraft, a spokesperson for Georgia Power, explained, "The shield building is a unique feature of the AP1000 reactor design for Vogtle 3 and 4, providing an additional layer of safety around the containment vessel and nuclear reactor to protect the structure from any potential impacts."

The AP1000 pressurized-water reactor, designed by Westinghouse, is a 21st-century "new" reactor. It's been deployed in just two other places, in China, with two AP1000 reactors at the Sanmen Nuclear Power Station in Zheijang province and two at the Haiyang Nuclear Power Plant in Shandong province. According to the International Atomic Energy Agency (IAEA), the AP1000 reactors at these locations operate at 1,157 megawatts and 1,126 MW, respectively.

In 2005, the Nuclear Regulatory Commission (NRC) certified the AP1000 design, clearing the way for its sale and installation at these three sites more than a decade later. Last year, Dan Brouillette, the U.S. secretary of energy, wrote in a blog post: "The U.S. Department of Energy (DOE) is all in on new nuclear energy."

NuScale's modular design–with 12 smaller reactors, each operating at a projected 60 MW–met NRC Phase 4 approval at the end of last year. According to Diane Hughes, vice president of marketing and communications for NuScale, "This means that the technical review by the NRC is essentially complete and that the final design approval is expected on schedule by September 2020."

NuScale's first customer, the Utah Associated Municipal Power Systems, plans to install a power plant with NuScale reactors at the Idaho National Laboratory site in Idaho Falls. The plant, Hughes said, is "slated for operation by the mid-2020s based on the NRC's approved design."

The idea of harnessing multiple smaller reactors in a single design is not new, dating back as far as the 1940s. At the time, the economics of the smaller, modular design could not compete with bigger, individual reactors, says M.V. Ramana, a nuclear physicist and professor at the University of British Columbia's School of Public Policy and Global Affairs.

"Nuclear power is unlike almost any other energy technology, in that it's the one tech where the costs have gone up, not down, with experience," he said. "The way to think about it is that the more experience we have with nuclear power, the more we learn about potential vulnerabilities that can lead to catastrophic accidents."

However, Hughes of NuScale counters that, unlike the 54 competing small modular reactor designs that the IAEA has records of, NuScale is "the first ever small modular reactor technology to undergo…NRC design certification review."

And in 2018, an interdisciplinary MIT report on nuclear energy found that NuScale's reactor is "quite innovative in its design. It has virtually eliminated the need for active systems to accomplish safety functions, relying instead on a combination of passive systems and the inherent features of its geometry and materials."

Of course, while the number of catastrophic nuclear accidents (such as Three Mile Island, Chernobyl, and Fukushima) is small for the amount of energy that nuclear power has generated over the past 70 years, Ramana adds, the cost of each accident is astronomical–sacrificing human lives and uprooting untold many more from disaster zones as well as requiring cleanups that cost hundreds of billions of dollars. "One every other decade is not good enough," Ramana said.
—MARK ANDERSON

POST YOUR COMMENTS AT
spectrum.ieee.org/nuclear-jun2020

NEWS

JOURNAL WATCH

Messages From a MicroLED

Wireless communications most often travel via radio waves. But some scientists see potential in using light to transmit data.

Mohamed Sufyan Islim, a researcher at the University of Edinburgh, explains why: "The radio-frequency spectrum is becoming extremely crowded," he says. Visible and infrared light waves offer 2,000 times as much bandwidth as radio waves. And radio waves are easy to intercept, whereas light can be shielded.

To show what light can do, researchers at the University of Strathclyde, in Scotland, created a custom microLED array with nine lights placed in series. Each light (or pixel) measured just 20 micrometers in diameter. Then, the group at the University of Edinburgh used advanced techniques to modulate the array at extremely high speeds.

Together, the universities built a system that achieved record rates of data transmission using LED lights, with 11.74 gigabits per second at 0.3 meter and 1.61 Gb/s at 20 meters. They recently described their work in *IEEE Photonics Technology Letters*.

In previous research, Islim says, his group showed that "the small dimensions of the microLEDs mean that they can be modulated at much higher frequencies than standard LEDs—several hundreds of megahertz compared with a few tens of megahertz." Being able to modulate at higher frequencies allowed them to transmit more data.

However, most microLEDs manufactured today are made for visual displays, and not designed to transmit data. That's one thing that would have to change for this technique to be used more widely. —**MICHELLE HAMPSON**

An extended version of this article appears on our website in the Journal Watch section.

THE RISE OF CONFIDENTIAL COMPUTING

Big tech companies are adopting a new security model to protect data while it's in use

A handful of major technology companies are going all in on a new security model they're calling confidential computing in an effort to better protect data in all its forms.

The three pillars of data security involve protecting data at rest, in transit, and in use. Protecting data at rest means using methods such as encryption or tokenization so that even if data is copied from a server or database, a thief can't access the information. Protecting data in transit means making sure unauthorized parties can't see information as it moves between servers and applications. There are well-established ways to provide both kinds of protection.

Protecting data while in use, though, is especially tough because applications need to have data in the clear—not encrypted or otherwise protected—in order to compute. But that means malware can dump the contents of memory to steal information. It doesn't really matter if the data was encrypted on a server's hard drive if it's stolen while exposed in memory.

Proponents of confidential computing hope to change that. "We're trying to evangelize there are actually practical solutions" to protect data while it's in use, said Dave Thaler, a software architect from Microsoft and chair of the Confidential Computing Consortium's Technical Advisory Council.

The consortium, launched last August under the Linux Foundation, aims to define standards for confidential computing and support the development and adoption of open-source tools. Members include technology heavyweights such as Alibaba, AMD, Arm, Facebook, Fortanix, Google, Huawei, IBM (through its subsidiary Red Hat), Intel, Microsoft, Oracle, Swisscom, Tencent, and Vmware. Several already have confidential computing products and services for sale.

Confidential computing uses hardware-based techniques to isolate data, specific functions, or an entire application from the operating system, hypervisor or virtual machine manager, and other privileged processes. Data is stored in the trusted execution environment (TEE), where it's impossible to view the data or operations performed on it from outside, even with a debugger. The TEE ensures that only authorized code can access the data. If the code is altered or tampered with, the TEE denies the operation.

Many organizations have declined to migrate some of their most sensitive applications to the cloud because of concerns about potential data exposure. Confidential computing makes it possible for different organizations to combine data sets for analysis without accessing each other's data, said Seth Knox, vice president of marketing at Fortanix and the outreach chair for the Confidential Computing Consortium. For example, a retailer and credit card company could cross-check customer and transaction data for potential fraud without giving the other party access to the original data.

Confidential computing may have other benefits unrelated to security. An image-processing application, for example, could store files in the TEE instead of sending a video stream to the cloud, saving bandwidth and reducing latency. The application may even divide up such tasks on the processor level, with the main CPU handling most of the processing, but relying on a TEE on the network interface card for sensitive computations.

Such techniques can also protect algorithms. A machine-learning algorithm, or an analytics application such as a stock trading platform, can live inside

NEWS

ILLUSTRATION BY Dan Page

the TEE. “You don’t want me to know what stocks you’re trading, and I don’t want you to know the algorithm,” said Martin Reynolds, a technology analyst at Gartner. “In this case, you wouldn’t get my code, and I wouldn’t get your data.”

Confidential computing requires extensive collaboration between hardware and software vendors so that applications and data can work with TEEs. Most confidential computing performed today runs on Intel servers (like the Xeon line) with Intel Software Guard Extension (SGX), which isolates specific application code and data to run in private regions of memory. However, recent security research has shown that Intel SGX can be vulnerable to side-channel and timing attacks.

Fortunately, TEEs aren’t available only in Intel hardware. OP-TEE is a TEE for nonsecure Linux Kernels running on Arm Cortex-A cores. Microsoft’s Virtual Secure Mode is a software-based TEE implemented by Hyper-V (the hypervisor for Windows systems) in Windows 10 and Windows Server 2016.

The Confidential Computing Consortium currently supports a handful of open-source projects, including the Intel SGX SDK for Linux, Microsoft’s Open Enclave SDK, and Red Hat’s Enarx. Projects don’t have to be accepted by the consortium to be considered confidential computing: For example, Google’s Asylo is similar to Enarx, and Microsoft Azure’s confidential computing services support both Intel SGX and Microsoft’s Virtual Secure Mode.

Hardware-based TEEs can supplement other security techniques, Thaler said, including homomorphic encryption and secure element chips such as the Trusted Platform Module. “You can combine these technologies because they are not necessarily competing,” he said. “Are you looking at the cloud or looking at the edge? You can pick which techniques to use.” —FAHMIDA Y. RASHID

POST YOUR COMMENTS AT
spectrum.ieee.org/confidential-jun2020

AMANDA JOHNSON

COVID-19 DISRUPTS RESEARCH PLANS

The coronavirus pandemic has interrupted fieldwork and lab experiments

The COVID-19 pandemic has affected virtually every facet of life, including scientific research carried out at companies and universities around the world.

As biomedical researchers scramble to find a treatment or vaccine, other scientists and engineers try to continue their own work in the midst of a pandemic. In some cases, this means writing a paper or grant from home rather than at the office. But in many others, the disruption is more pronounced.

Most academic, government, and corporate labs in the United States have scaled back operations or closed temporarily to comply with stay-at-home orders. The impacts of these changes can vary greatly from one field to the next, depending on the nature of the work.

John Verboncoeur, a director of IEEE and associate dean for research and graduate studies at Michigan State University, says, “Our surveys indicate that theoretical research teams—my own included—are operating at around 80 to 90 percent efficiency, with the main challenge being the ability to explain complicated concepts without our traditional ‘waving hands about’ and interactive work at the white- or blackboard.”

For experimentalists, the pandemic is more disruptive, although some experiments may be completed from home. “The early focus [for experimentalists] was on catching up on the literature, completing manuscripts, analyzing existing data, and so on, which led to a productivity of 50 percent or so,” says Verboncoeur. “However, much of that is coming to completion, and we are seeing productivity drop as the activities narrow down to designing upcoming experiments and protocols.”

Engineers in many fields are looking for new ways to remain inno-

MAKING IT WORK: Brian Johnson of the University of Washington continues his research from his home office.

vative and productive. Take, for example, those in the green energy sector. While some climate and energy research may continue from home, other projects are more difficult or impossible to complete remotely.

Sally Benson's lab at Stanford University does a mix of theory, modeling, and experiments to support the transition to a low-carbon future, including studies related to carbon capture and storage. While the theory and modeling aspects of this research are easy enough to continue, the experimental work involves analyzing rock samples at the extreme temperatures and pressures found in underground reservoirs—tests that aren't feasible to carry out at home.

Despite this limitation, Benson's group is still finding ways to continue with some aspects of their experimental work. "The good news is that as experimentalists, we tend to collect way, way more data than we can assimilate," she says. "We generate these immensely rich data sets, where there's plenty more we can mine out of those data sets."

The group is now returning to its old data sets and reanalyzing the data to answer new, unexplored questions, in part by applying machine learning. By doing so, the researchers have uncovered previously unknown ways that carbon dioxide interacts with rock. Benson acknowledges, however, that this reuse of old experimental data can't go on forever.

Further up the coast, at the University of Washington, Brian Johnson is leading two projects funded by the U.S. Department of Energy. Both are designed to facilitate a major shift from electromechanical power grids to grids based on power-electronics systems that will better support renewable energy.

One project involves the design of controllers for these new power grids. The effort launched in April just as the pandemic was taking hold in the United States, but the team was able to get the research started by focusing on pen-and-paper designs and software simulations.

However, the pandemic may prove more problematic for Johnson's second endeavor. It involves the design of a new breed of high-efficiency power electronics that converts DC power from solar cells into grid-compatible AC power. "For that project, we have a heavy set of milestones coming up in the summer months to actually demonstrate the hardware," says Johnson. "If we can't do [tests] in the summer, we're going to have to start coming up with some contingency plans. Since these experiments necessitate a power lab with specialized equipment, they cannot be done in our homes."

While the pandemic affects each research project to varying degrees, its overall impact on the broader shift toward green tech—and on the state of engineering research more generally—is still unclear.

Benson says she's slightly concerned that the pandemic may cause some researchers to shift their focus from climate change to medicine. "To me, the COVID-19 pandemic is sort of a multiyear challenge and a short-term nightmare," she says. "If we're not careful, climate change will be a decadal-scale nightmare. So this work needs all of the attention it can get."

Johnson is less concerned that the pandemic will interfere with the advancement of green tech, saying: "I think that energy is such an integral part of modern life itself and infrastructure that I don't perceive [the COVID-19 pandemic] fundamentally altering the fact that we all need energy, and cheap energy."
—MICHELLE HAMPSON

POST YOUR COMMENTS AT
spectrum.ieee.org/research-jun2020

STORING DATA IN GLASS

Optical storage media is advancing from CDs and DVDs to small glass squares

Magnetic tape and hard disk drives hold much of the world's archival data. Compared with other memory and storage technologies, tape and disk drives cost less and are more reliable. They're also nonvolatile, meaning they don't require a constant power supply to preserve data. Cultural institutions, financial firms, government agencies, and film companies have relied on these technologies for decades, and will continue to do so far into the future.

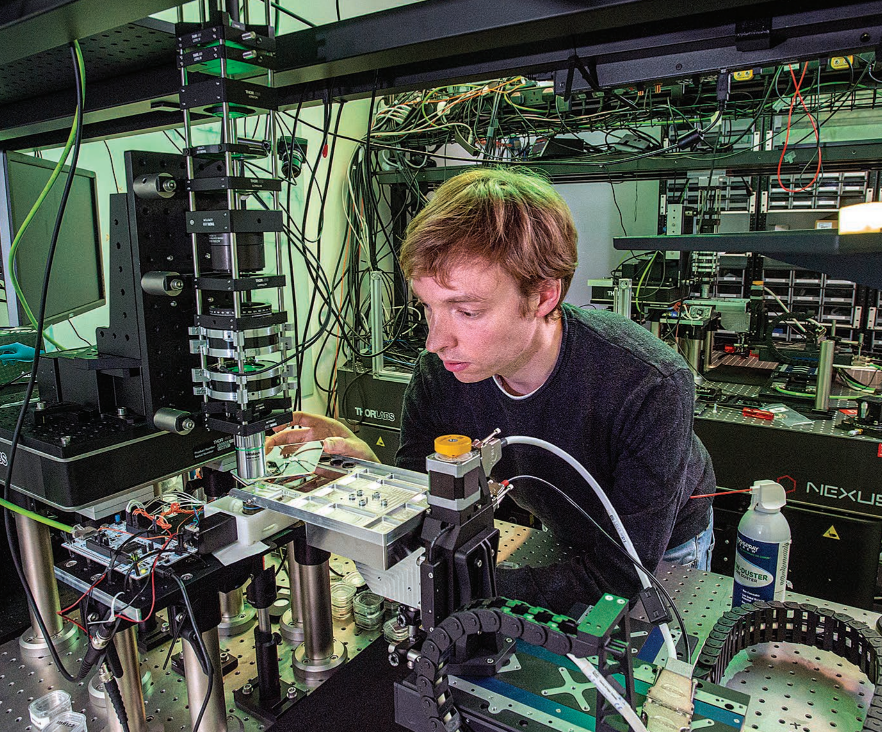
But archivists may soon have another option—using an extremely fast laser to write data into a 2-millimeter-thick piece of glass, roughly the size of a Post-it note, where that information can remain essentially forever.

This experimental form of optical data storage was demonstrated in 2013 by researchers at the University of Southampton in England. Soon after, that group began working with engineers at Microsoft Research in an effort called Project Silica. Last November, Microsoft completed its first proof of concept by writing the 1978 film *Superman* on a single small piece of glass and retrieving it.

With this method, researchers could theoretically store up to 360 terabytes of data on a disc the size of a DVD. For comparison, Panasonic aims to someday fit 1 TB on conventional optical discs, while Seagate and Western Digital are shooting for 50- to 60-TB hard disk drives by 2026.

International Data Corp. expects the world to produce 175 zettabytes of data by 2025—up from 33 ZB in 2018. Though only a fraction of that data will be stored, today's methods may no longer suffice.

NEWS



LEARNING TO READ: James Clegg, a senior optical scientist at Microsoft, loads a piece of glass into a system used to retrieve information stored in the glass.

MICROSOFT

"We believe people's appetite for storage will force scientists to look into other kinds of materials," says Waguih Ishak, chief technologist at Corning Research and Development Corp.

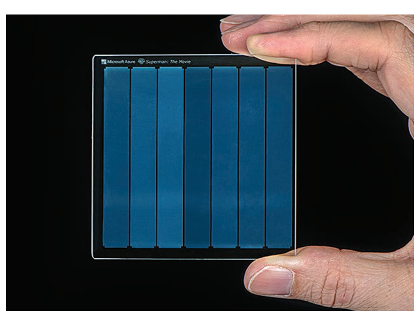
Microsoft's work is part of a broader company initiative to improve cloud storage through optics. "I think they see it as potentially a distinguishing technology from something like [Amazon Web Services] and other cloud providers," says James Byron, a Ph.D. candidate in computer science at the University of California, Santa Cruz, who studies storage methods.

Microsoft isn't alone–John Morris, chief technology officer at Seagate, says researchers there are also focused on understanding the potential of optical data storage in glass. "The challenge is to develop systems that can read and write with reasonable throughput," he says.

Writing data to glass involves focusing a femtosecond laser, which pulses very quickly, on a point within the glass. The glass itself is a sort known as fused silica. It's the same type of extremely pure glass used for the Hubble Space Telescope's mirror as well as the windows on the International Space Station.

The laser's pulse deforms the glass at its focal point, forming a tiny 3D structure called a voxel. Two properties that measure how the voxel interacts with polarized light–retardance and change in the light's polarization angle–can together represent several bits of data per voxel.

Microsoft can currently write hundreds of layers of voxels into each piece of glass. The glass can be written to once and read back many times. "This is data in glass, not on glass," says Ant Rowstron, a principal researcher and deputy lab director at Microsoft Research Lab in Cambridge, England.



MADE TO LAST: Microsoft and Warner Bros. stored the 1978 film *Superman* on this piece of glass in a proof of concept for Project Silica. The glass contains 75.6 gigabytes of data plus error-correcting codes.

Reading data from the glass requires an entirely different setup, which is one potential drawback of this method. Researchers shine different kinds of polarized light–in which light waves all oscillate in the same direction, rather than every which way–onto specific voxels. They capture the results with a camera. Then, machine-learning algorithms analyze those images and translate their measurements into data.

Ishak, who is also an adjunct professor of electrical engineering at Stanford University, is optimistic about the approach. "I'm sure that in the matter of a decade, we'll see a whole new kind of storage that eclipses and dwarfs everything that we have today," he says. "And I firmly believe that those pure materials like fused silica will definitely play a major role there."

But many scientific and engineering challenges remain. "The writing process is hard to make reliable and repeatable, and [it's hard] to minimize the time it takes to create a voxel," says Rowstron. "The read process has been a challenge in figuring out how to read the data from the glass using the minimum signal possible from the glass."

The Microsoft group has added error-correcting codes to improve the system's accuracy and continues to refine its machine-learning algorithms to automate the read-back process. Already, the team has improved writing speeds by several orders of magnitude from when they began, though Rowstron declined to share absolute speeds.

The team is also considering what it means to store data for such a long time. "We are working on thinking what a Rosetta Stone for glass could look like to help people decode it in the future," Rowstron says. –AMY NORDRUM

POST YOUR COMMENTS AT
spectrum.ieee.org/glass-jun2020

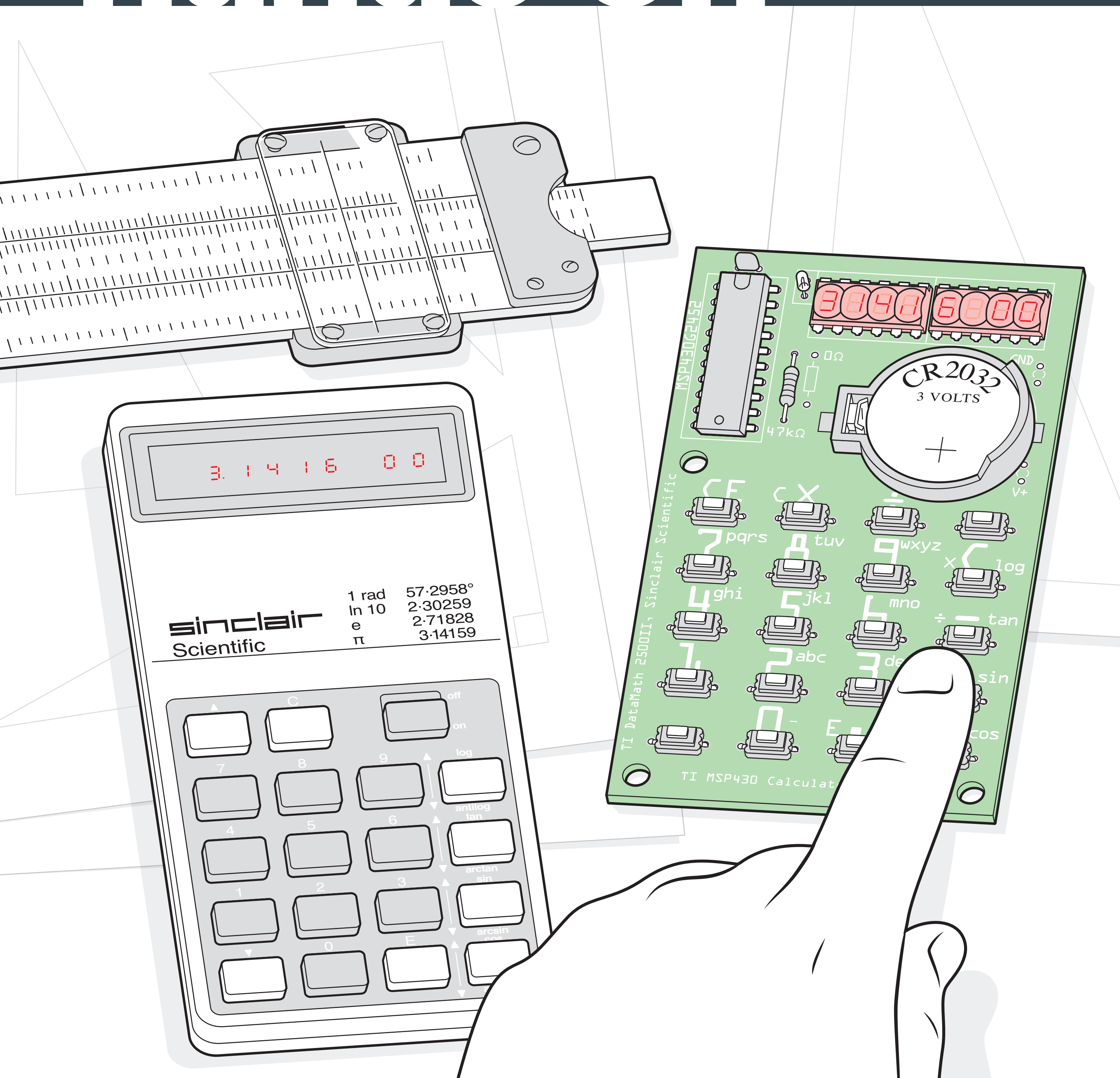
PHOTOGRAPH BY Marcel Van Hoorn/AFP/Getty Images

EMERGENCY BROADCAST

DRONES ARE USED FOR WORK, play, and even war. But now they've found a new application. Desperate authorities–responding to public gatherings where people make themselves possible vectors for the spread of the coronavirus–have turned to flying drones equipped with loudspeakers that play messages telling people to disperse. This drone was photographed in early April as it hovered over shoppers at a commercial boulevard in Heerlen, Netherlands, reminding them to stay at least 2 meters apart. Though the presence of remotely operated aircraft has led to suspicion that they are being used to identify people and record their movements, local governments insist that the drones are not recording or using facial recognition.

THE BIG PICTURE | NEWS

Hands On



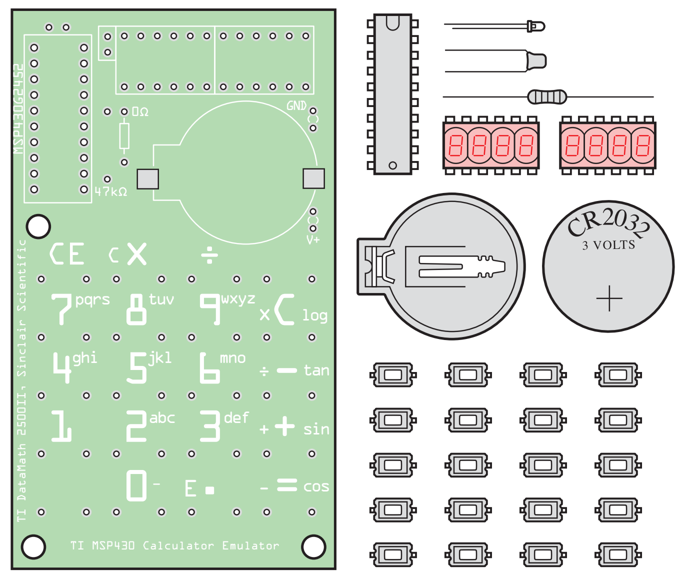
HANDS ON BY STEPHEN CASS

ILLUSTRATIONS BY James Provost

THE SINCLAIR SCIENTIFIC, REMADE

A CLASSIC CALCULATOR WAS REVERSE ENGINEERED FROM THE BARE METAL

Was the Sinclair Scientific calculator elegant? It certainly was a hit, gracing the cover of publications like *Popular Mechanics* after its release in 1974. Cleverly written firmware dragooned its limited processor, intended only for basic arithmetic, into performing way beyond specifications. This allowed Sinclair to sell a scientific calculator to countless folks who otherwise could not have afforded one. But it was also slow and sometimes inaccurate, provided barely enough mathematical functions to qualify as a scientific calculator, and was difficult for the uninitiated to use.

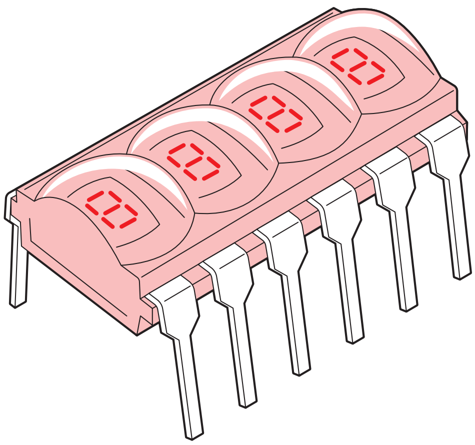


REFACTORED REPLICA: The original Sinclair Scientific was also sold in kit form. Chris Chung's version is smaller, uses fewer components, and can actually replicate *two* calculators based on the TMS080*x* family of CPUs: the TI Datamath, a simple arithmetical calculator, and the Sinclair Scientific. Hence, the printed silkscreen on the circuit board has layouts for both calculators.

I'd vaguely known of this calculator because of its place as a milestone toward the birth of the British microcomputer industry and such beloved machines as the Sinclair ZX Spectrum. So I clicked when I came across Chris Chung's replica kit of the calculator on the Tindie marketplace. Then I read the description of how the original calculator worked—scientific notion only? no "equals" button?—and how the replica reproduced its behavior, by using an emulator running firmware that had been reverse engineered by *visually examining the metal of an original processor.* This I had to try.

Let's first get to the hardware. The kit is one of a number of Sinclair calculator replicas, but it wins points for simplicity: One chip and a credit-card-size printed circuit board are combined with a small handful of discrete components. Chung actually offers two kits: his original, developed in 2014 but put on Tindie in late 2019, displays numbers using two small QDSP-6064 bubble LED modules, which have the classic look of 1970s calculators but are long discontinued and hard to get. His 2020 update of the kit uses modern seven-segment LED displays. The difference in rarity is reflected in the price: The 2020 version costs US $39, while the original version costs $79. However, in a nice touch, the original version lets you have your cake and eat it with regard to the bubble LEDs: The through holes in the PCB are sized to create a friction fit. This means you don't have to solder in the modules, preserving them for some future project.

Both versions are functionally identical, based around a MSP430 mi-



THAT 70'S LOOK: Bubble LED modules were common in early calculators. The bubbles were lenses that magnified the small segments.

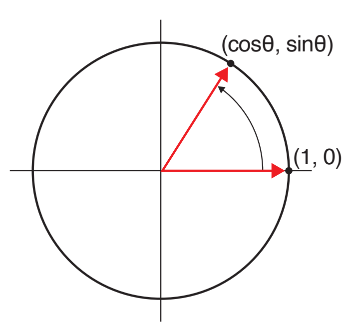
crocontroller. The MSP430 eliminates the need for most of the other components found in the Sinclair Scientific, and runs an emulator of the TMS080*x* family of chips. The TMS080*x* family was built by Texas Instruments (TI), and specific versions, like the TMS0805 used in the Sinclair Scientific, were differentiated by the contents of the chip's 3,520-bit ROM.

For many years, how Sinclair coaxed this chip into performing magic remained locked away in TMS0805 chip ROMs. That began to change in 2013, when Ken Shirriff got wind of the Sinclair Scientific after getting involved with the Visual 6502 team. This group likes to reverse engineer classic chips, for which the original design drawings are often lost. Sometimes this involves etching chip packages away with acid and carefully photographing the exposed silicon dies with a microscope to see the individual transistors. Shirriff was able to create a generic simulator of a TMS080*x* chip in JavaScript just by studying Texas Instruments' patent filings, but the specific code used in the Sinclair's TMS0805's ROM still eluded him, until Visual 6502 team member John McMaster took photographs of an exposed die in 2014.

Shirriff is no stranger to *IEEE Spectrum* due to his extensive work in the history of computing, so I emailed him to ask how he'd gone from a micrograph to working code. By looking at how the metal oxide gates were laid down in the ROM section of the chip, "I was able to extract the raw bits the same day," Shirriff wrote back. "Phil Mainwaring, Ed Spittles, and I spent another day figuring out how the raw bits corresponded to the code….The code was 320 words of 11 bits, but the ROM was physically 55 rows and 64 columns….Through a combination of examining the circuitry, analyzing patterns in the bits, and brute-force trying a bunch of combinations, we figured out the arrangement and could extract the code."

Once he had the code loaded into his simulator, Shirriff could tease out its workings. The algorithms used throughout "were essentially the simplest brute-force algorithms that could get an answer. But there were some interesting mathematical tricks they used to get more accuracy, not to mention programming tricks to get the code to fit," he explained. (If you want detailed explanations, Shirriff maintains an online version of his simulator that lets you step through the code, line by line.)

The Sinclair Scientific was able to reduce complexity by using reverse Polish notation, in which mathematical operators come after the numbers they are operating on—for instance, "5 + 4 =" becomes "5 4 +." The trigonometric functions used an iterated approximation technique that could take several seconds to produce results and were often accurate only to the first three significant figures. The calculator also used fixed scientific notation for everything—there was no way to enter a decimal point. So rather than entering "521.4," you'd enter "5214," which is displayed as "5.214"; then you'd press "E" and enter "2," making the number "5.214×10^{2}." Only one number could be entered at a time.



DO THE TWIST: The Sinclair Scientific calculated trigonometric functions by repeatedly rotating an initial vector until the target angle was reached. It took several seconds to calculate most angles.

On paper this looks terrible, something you'd have used only if you couldn't afford, say, the HP-35, whose designers prided themselves on accuracy and functionality (although the HP-35 also used reverse Polish notation, it did so in a more sophisticated way).

But Sinclair wasn't trying to compete against other calculators; he was competing against *slide rules.* I'd read that point in histories before, but I didn't really understand its meaning until I held this kit in my hand. By chance, I'd also recently purchased a vintage Pickett slide rule and learned how to do basic operations with it, thanks to the lessons available on the International Slide Rule Museum website. As I used the Sinclair Scientific, I was struck by how conceptually similar it was to using my slide rule. There, two- to three-figure accuracy is the norm, the rule's sliding "cursor" means you're passing just one number between scales, and you typically don't bother about magnitudes—you work with the most significant figures and make a mental estimate to insert the decimal point at the end, so that 52 × 2 and 5,200 × 20 are calculated in exactly the same way.

This realization is why I feel replica kits are so important—they are an easy way to provide the understanding that only comes from operating a device with your own hands. It was a great reminder that, as folks like Henry Petroski have pointed out, good design is not really something that exists as an abstract ideal but as something that exists only within a specific context. So, was the Sinclair Scientific elegant? For me, the answer is a resounding yes. **—STEPHEN CASS**

POST YOUR COMMENTS AT
spectrum.ieee.org/sinclair-jun2020

Careers

WAGE DISCRIMINATION INCREASES IN TECH

WOMEN LOSE GAINS MADE IN 2018

After falling slightly in 2018, the number of women in tech being offered less than men for the same job at the same company jumped back up in 2019—to 63 percent. Meanwhile, the average salary differential between U.S. tech workers identifying as male and those identifying as female held steady at 3 percent, according to a 2020 study by the job search firm Hired that was released at the end of March.

Last month, a study by Dice also reported that the gender pay gap continues across most specialties, but may be narrowing in cloud engineering, systems architecture, and network engineering. The Hired study instead looked at changes in the overall U.S. tech industry to find that the gender gap is growing, not narrowing.

That's the data. In terms of the perception, only 63 percent of men who responded to the Hired survey indicated that they believe a wage gap between genders exists, with 16 percent of men convinced there is no gap and 21 percent unsure. For women, 84 percent are convinced of the gap, with only 6 percent indicating that it doesn't exist and 10 percent unsure.

According to Hired's data, the gap in major metropolitan areas is largest in New York (10 percent) and smallest in Los Angeles (5 percent). The San Francisco Bay Area fell in the middle (7 percent).

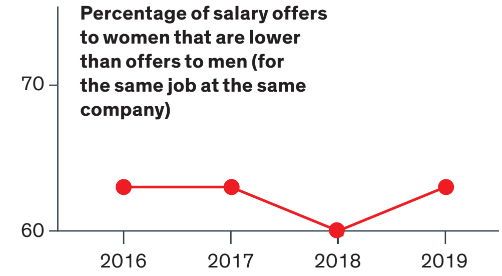
With such a pervasive salary gap, it's not surprising that women in tech have lower expectations for their own salaries—which is helping to cement the wage gap, Hired's analysts suggest. And expectations this year were worse than last year's: Sixty-five percent of women asked for lower salaries than men in 2020, compared with 61 percent in 2019.

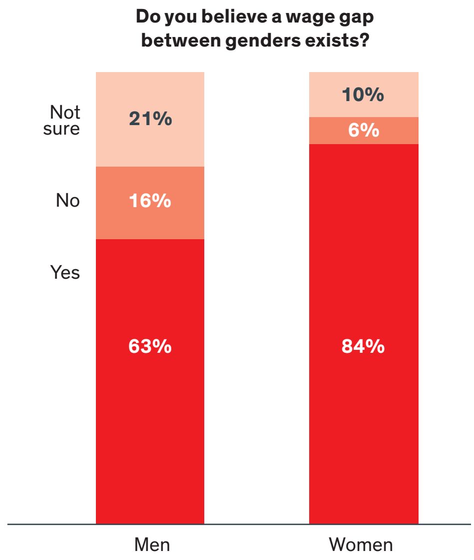
More pay transparency and open conversations about salaries among peers are needed to break through this barrier, the Hired report indicated. And most tech employees would welcome salary transparency: Sixty-eight percent of women and 63 percent of men think transparency would increase their interest in working for a company, according to the report.

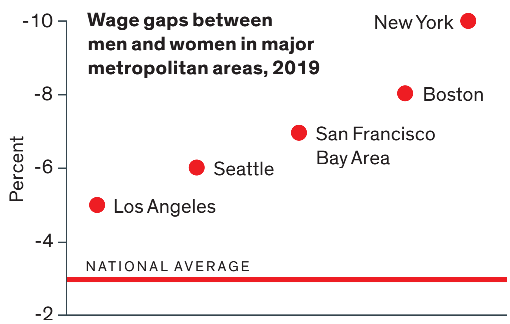
—TEKLA S. PERRY

An extended version of this article appears on the View From the Valley blog.

POST YOUR COMMENTS AT spectrum.ieee.org/wagegap-jun2020







SOURCE: 2020 STATE OF WAGE INEQUALITY IN THE WORKPLACE, HIRED

DEPARTMENTS

BUILDING THE GREAT PYRAMID

GIVEN THAT SOME 4,600 years have elapsed since the completion of the Great Pyramid of Giza, the structure stands remarkably intact. It is a polyhedron with a regular polygon base, its volume is about 2.6 million cubic meters, and its original height was 146.6 meters, including the lost pyramidion, or capstone. We may never know exactly how the pyramid was built, but even so, we can say with some confidence how many people were required to build it.

We must start with the time constraint of roughly 20 years, the length of the reign of Khufu, the pharaoh who commissioned the construction (he died around 2530 B.C.E.). Herodotus, writing more than 21 centuries after the pyramid's completion, was told that labor gangs totaling 100,000 men worked in three-month spells a year to finish the structure in 20 years. In 1974, Kurt Mendelssohn, a German-born British physicist, put the labor force at 70,000 seasonal workers and up to 10,000 permanent masons.

These are large overestimates; we can do better by appealing to simple physics. The potential energy of the pyramid—the energy needed to lift the mass above ground level—is simply the product of acceleration due to gravity, mass, and the center of mass, which in a pyramid is one-quarter of its height. The mass cannot be pinpointed because it depends on the specific densities of the Tura limestone and mortar that were used to build the structure; I am assuming a mean of

PATRA KONGSIRIMONGKOLCHAI/EYEEM/GETTY IMAGES

NUMBERS DON'T LIE BY VACLAV SMIL

2.6 metric tons per cubic meter, hence a total mass of about 6.75 million metric tons. That means the pyramid's potential energy is about 2.4 trillion joules.

To maintain his basal metabolic rate, a 70-kilogram (154-pound) man requires some 7.5 megajoules a day; steady exertion will raise that figure by at least 30 percent. About 20 percent of that increase will be converted into useful work, which amounts to about 450 kilojoules a day (different assumptions are possible, but they would make no fundamental difference). Dividing the potential energy of the pyramid by 450 kJ implies that it took 5.3 million man-days to raise the pyramid. If a work year consists of 300 days, that would mean almost 18,000 man-years, which, spread over 20 years, implies a workforce of about 900 men.

A similar number of workers might be needed to place the stones in the rising structure and then to smooth the cladding blocks (many interior blocks were just rough-cut). And in order to cut 2.6 million cubic meters of stone in 20 years, the project would have required about 1,500 quarrymen working 300 days a year and producing 0.25 cubic meter of stone per capita. The grand total of the construction labor would then be some 3,300 workers. Even if we were to double that number to account for designers, organizers, and overseers and for labor needed for transport, tool repair, the building and maintenance of on-site housing, and cooking and laundry work, the total would be still less than 7,000 workers.

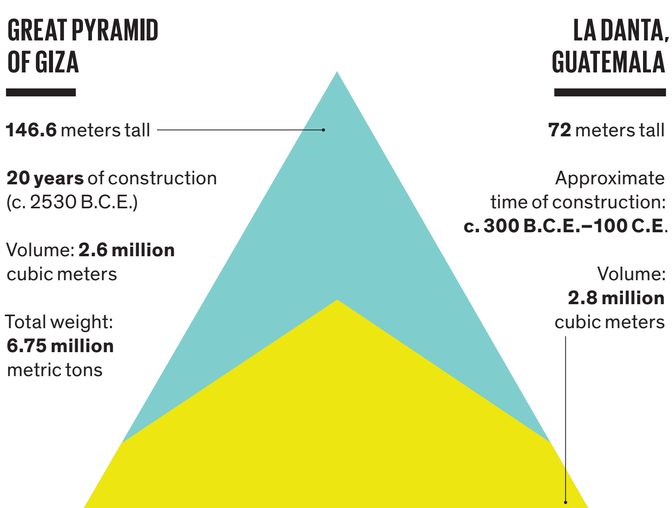
During the time of the pyramid's construction, the total population of the late Old Kingdom was 1.5 million to 1.6 million people, and hence such a labor force would not have been an extraordinary imposition on the country's economy. The challenge was to organize the labor, plan an uninterrupted supply of building stones, and provide housing, clothing, and food for labor gangs on the Giza site.

In the 1990s, archaeologists uncovered a cemetery for workers and the foundations of a settlement used to house the builders of the two later pyramids at the site, indicating that no more than 20,000 people lived there. That an additional two pyramids were built in rapid succession at the Giza site (for Khafre, Khufu's son, starting at 2520 B.C.E., and for Menkaure, starting 2490 B.C.E.) shows how quickly early Egyptians mastered the building of pyramids: The erection of those massive structures became just another series of construction projects for the Old Kingdom's designers, managers, and workers. If you build things, it becomes easier to build things—a useful lesson for those who worry about the sorry state of our infrastructure. ■

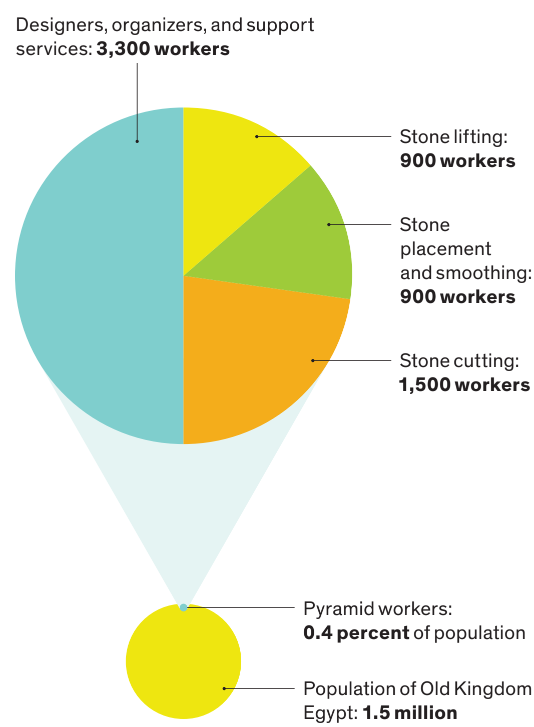
POST YOUR COMMENTS AT
spectrum.ieee.org/pyramids-jun2020

SWEAT EQUITY

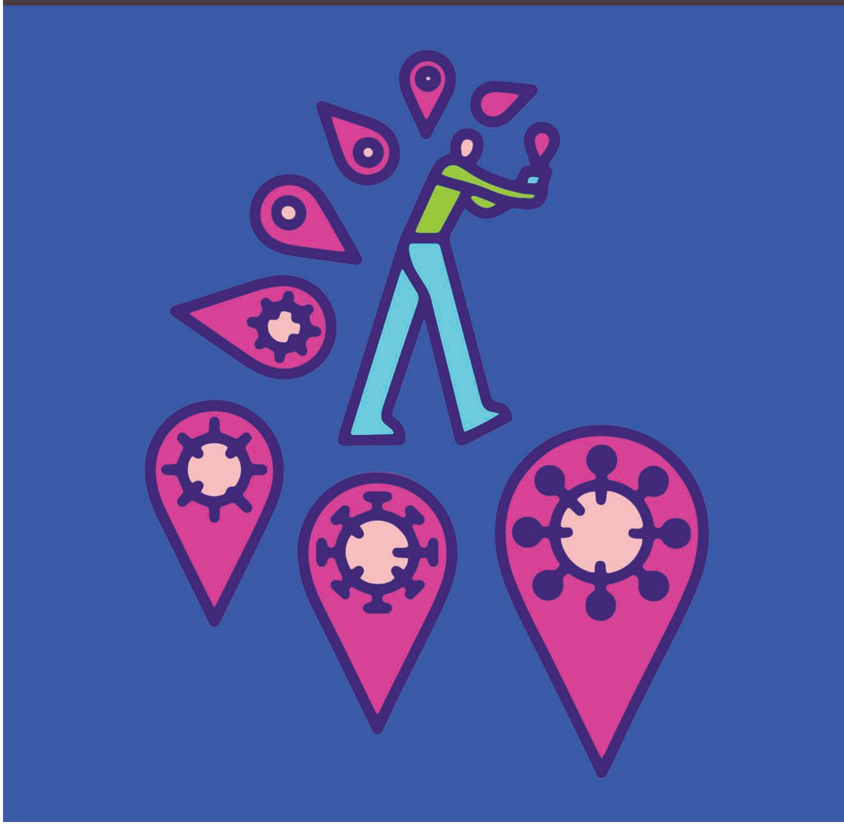
Sheer human muscle can raise a great pyramid because most of the mass is near the bottom and the work can be divided among thousands of people and distributed over decades. It was thus no great drain on national resources—and think of the spin-off effects!



6,600 EGYPTIANS WORKING 300 DAYS PER YEAR FOR 20 YEARS



CROSSTALK



PANDEMIC VS. PRIVACY

THE INTERNET of Things makes the invisible visible. That's the IoT's greatest feature, but also its biggest potential drawback. More sensors on more people means the IoT becomes a visible web of human connections that we can use to, say, track down a virus.

Track-and-trace programs are already being used to monitor outbreaks of COVID-19 and its spread. But because they would do so through easily enabled mass surveillance, we need to put rules in place about how to undertake any attempts to track the movements of people.

In April, Google and Apple said they would work together to build an opt-in program for Android or iOS users. The program would use their phones' Bluetooth connection to deliver exposure notifications—meaning that transmissions are tracked by who comes into contact with whom, rather than where people spend their time. Other proposals use location data provided by phone applications to determine where people are traveling.

All of these ideas have slightly different approaches, but at their core they're still tracking programs. Any such program that we implement to track the spread of COVID-19 should follow some basic guidelines to ensure that the data is used only for public health research. This data should not be used for marketing, commercial gain, or law enforcement. It shouldn't even be used for research *outside* of public health.

Let's talk about the limits we should place around this data. A tracking program for COVID-19 should be implemented only for a prespecified duration that's associated with a public health goal (like reducing the spread of the virus). So, if we're going to collect device data and do so without requiring a user to opt in, governments need to enact legislation that explains what the tracking methodology is, requires an audit for accuracy and efficacy by a third party, and sets a predefined end.

Ethical data collection is also critical. Apple and Google's Bluetooth method uses encrypted tokens to track people as they pass other people. The Bluetooth data is people-centric, not location-centric. Once a person uploads a confirmation that they've been infected, their device can issue notifications to other devices that were recently nearby, alerting users—anonymously—that they may have come in contact with someone who's infected.

This is good. And while it might be possible to match a person to a device, it would be difficult. Ultimately, linking cases anonymously to devices is safer than simply collecting location data on infected individuals. The latter makes it easy to identify people based on where they sleep at night and work during the day, for example.

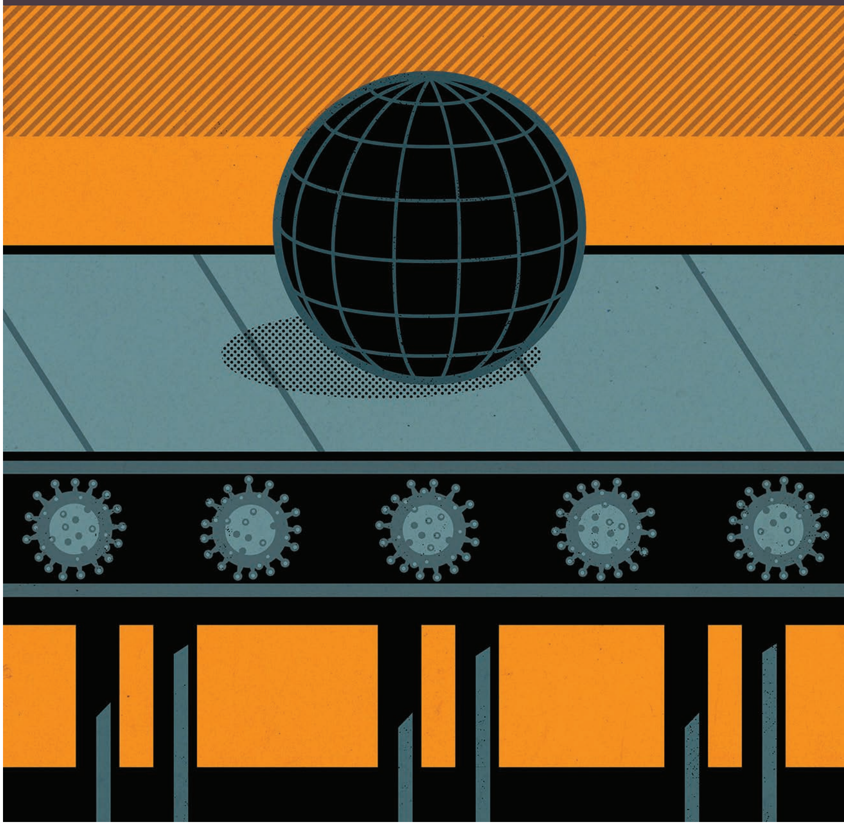
Going further, this data must be encrypted on the device, during transit and when stored on a cloud or government server, so that random hackers can't access it. Only the agency in charge of track-and-trace efforts should have access to the data from the device. This means that police departments, immigration agencies, or private companies can't access that data. Ever.

However, researchers should have access to some form of the data after a few years have passed. I don't know what that time limit should be, but when that time comes, institutional review boards, like those that academic institutions use to protect human research subjects, should be in place to evaluate each request for what could be highly personal data.

If we can get this right, we can use the lessons learned during COVID-19 not only to protect public health but also to promote a more privacy-centric approach to the Internet of Things. ■

POST YOUR COMMENTS AT
spectrum.ieee.org/pandemic-jun2020

ILLUSTRATION BY Greg Mably



NOT BUSINESS AS USUAL

BACK IN 2012, Netflix released Chaos Monkey, an open-source tool for triggering random failures in critical computing infrastructure. Similar stress testing, at least in a simulated environment, has been applied in other contexts such as banking, but we've never stress tested industrial production during a viral pandemic. Until now.

COVID-19 has demonstrated beyond doubt the fragility of the global system of lean inventories and just-in-time delivery. Many nations have immediate need for critical medical supplies and equipment, even as we grope for the switch that will allow us to turn the global economy back on. That means people have to be able to manufacture stuff, where it's needed, when it's needed, and from components that can be locally sourced. That's a big ask, because most of our technology comes to us from far away, often in seamlessly opaque packaging–like a smartphone, all surface with no visible interior.

The manufacture of even basic products can be so encumbered by secrecy or obscurity that it quickly becomes difficult to learn how to make them or to re-create their functionality in some other way. While we normally tolerate such impediments as part of normal business practice, they have thrown up unexpected roadblocks to keeping the world operating through the present crisis.

We must do whatever we can to lower the barriers to getting things built, and that begins by embracing a newfound flexibility in our approaches to both manufacturing and intellectual property. Companies are already rising to this challenge.

For example, Medtronic shared the designs and code for its portable ventilator at no charge, enabling other capable manufacturers to take up the challenge of building and distributing enough units to meet peak demand during the pandemic. Countless other pieces of electronic equipment–everything from routers to thermostats–operate in critical environments and need immediate replacement should they fail. Where they cannot be replaced, we will fall deeper into the ditch we now find ourselves in.

It would be ideal if any sort of equipment could be "printed" on demand. We already have the capacity for such rapid manufacturing in some realms, but to address the breadth of the present crisis would require a comprehensive database of product designs, testing, firmware, and much else. Little of that infrastructure exists at present, highlighting a real danger: If we don't construct a distributed, global build-it-here-now capacity, we might burn through our existing inventories without any way to replenish them. Then we will be truly stuck.

Many firms will no doubt have reservations about handing over their intellectual property, even to satisfy critical needs. This tension between normal business practice and public good echoes the contours of the dilemmas facing personal privacy and public health. We nevertheless need urgently to find a way to share trade secrets–temporarily–to preserve the kind of world within which business can one day operate normally.

Some governments have already signaled their greater flexibility in enforcing both patent protection and intellectual property protections during this crisis. Yet more is needed. Like Medtronic, businesses should take the plunge, open up, share their trade secrets, provide guidance to others (even former competitors) to help us speed our way into a post-pandemic economy. Sharing today will make that return much faster and far less painful. To paraphrase a wise old technologist, we either hang together, or we will no doubt hang separately. ■

POST YOUR COMMENTS AT
spectrum.ieee.org/sharing-jun2020

ILLUSTRATION BY Dan Page

HOW CONTROL THEORY CAN HELP US CONTROL COVID-19

Using feedback, a standard tool in control engineering, we can manage our response to the novel coronavirus pandemic for maximum survival while containing the damage to our economies

BY GREG STEWART, KLASKE VAN HEUSDEN & GUY A. DUMONT

CLOSED FOR BUSINESS: The streets of Manhattan are quiet on 10 April as most people comply with the city's self-quarantine rules.

As we write these words, countless people, perhaps the majority of the world's population, are subject to physical-distancing policies or confined to their homes in an attempt to contain one of the worst pandemics of modern times. Economic activity has plummeted, hundreds of millions of people are out of work, and entire industries have ground to a halt.

Quite understandably, a couple of questions are on everyone's mind: What is the exit strategy? How will we know when it's safe to implement it?

Around the globe, epidemiologists, statisticians, biologists, and health officials are grappling with these questions. Though engineering perspectives are uncommon in epidemiological modeling, we believe that in this case public officials could greatly benefit from one. Of course, the COVID-19 pandemic isn't an obvious or typical engineering problem. But in its basic behavior it is an unstable, open-loop system. Left alone, it grows exponentially. However, there's good news, too: Like many such systems, it can be stabilized effectively and efficiently by applying the principles of control theory, most notably the use of feedback.

Inspired by the important work of epidemiologists and others on the front lines of this global crisis, we have explored how feedback can help stabilize and diminish the rate of propagation of this deadly virus that now literally plagues us. We relied on feedback-based mechanisms to devise a system that would bring the outbreak under control and then adeptly manage the longer-term caseload.

It is during this longer-term phase, the inevitable relaxing of physical distancing that is required for a functioning society, that the strengths of a response grounded in control theory are most crucial. Using one of the widely available computer models of the disease, we tested our proposal and found that it could help officials manage the enormous complexity of trade-offs and unknowns that they will face, while saving perhaps hundreds of thousands of lives.

Our goal here is to share some of our key findings and to engage a community of control experts in this vital and fascinating problem. Together, we can contribute vitally to the international efforts to manage this outbreak.

THE COVID-19 PANDEMIC is unlike any other recent disease outbreak for several reasons. One is that its basic reproduction number, or R0 ("R naught"), is relatively high. R0 is an indication of how many people, on average, an infected person will infect during the course of her

PREVIOUS PAGES: SPENCER PLATT/GETTY IMAGES

illness. R0 is not a fixed number, depending as it does on such factors as the density of a community, the general health of its populace, its medical infrastructure and resources, and countless details of the community's response. But a commonly cited R0 figure for ordinary seasonal influenza is 1.3, whereas a figure calculated for the experience in Wuhan, China, where COVID-19 is understood to have originated, is 2.6. Figures for some outbreaks in Italy range from 2.76 to 3.25.

The goal of infectious-disease intervention is reducing the R0 to below 1, because such a value means that new infections are in decline and will eventually reach zero. But with the COVID-19 outbreak, the level of urgency is quite high due to the disease's relatively high fatality rate. Fatality rates, too, are quite variable and depend on such factors as age, physical fitness, present pathologies, region, and access to health care. But in general they are much higher for COVID-19 than for ordinary influenza. A surprisingly large percentage of people who contract the disease develop a form of viral pneumonia that sometimes proves fatal. Many of those patients require artificial ventilation, and if their number exceeds the capacity of intensive care units to accommodate them, some number of them, perhaps a majority, will die.

For that reason, enormous worldwide efforts have focused on "flattening the curve" of infections against time. A high, sharp curve indicating a surge of infections in a short time period, as occurred in China, Italy, Spain, and elsewhere, means that the number of serious cases will swamp the ability of hospitals to treat them and result in mass fatalities. So to reduce the peak demand on health care, the first priority must be to bring the caseload under control. Once that's done, the emphasis shifts to managing a long-term return to normalcy while minimizing both death rates and economic impact.

The two basic approaches to controlling the spread of disease are mitigation, which focuses on slowing but not necessarily stopping the spread, and suppression, which aims to reverse epidemic growth. For mitigation, R0 is reduced but remains greater than 1, while for suppression, R0 is smaller than 1. Both obviously require changing R0. Officials accomplish that by introducing social measures such as restricted travel, home confinement, social distancing, and so on. These restrictions are referred to as nonpharmaceutical interventions, or NPIs. What we are proposing is a systematically designed strategy, based on feedback, to change R0 through modulation of NPIs. In effect, the strategy alternates between suppression and mitigation in order to maintain the spread at a desired level.

It may sound straightforward, but there are many challenges. Some of them arise from the fact that COVID-19 is a very peculiar disease. Despite enormous efforts to characterize the virus, biologists still do not understand why some people experience fairly mild symptoms while others spiral into a massive, uncontrolled immune response and death. And no one can explain why, among fatalities, men predominate. Other mysteries include the disease's long incubation period—up to 14 days between infection and symptoms—and even the question of whether a person can get re-infected.

These perplexities have helped bog down efforts to deal with the pandemic. As a recent Imperial College London research paper notes: "There are very large uncertainties around the transmission of this virus, the likely effectiveness of different policies, and the extent to which the population spontaneously adopts risk reducing behaviours." Consider the long incubation time and apparent spreading of the virus before symptoms are experienced. These undoubtedly contributed to the relatively high R0 values, because people who were infectious continued to interact with others and transmitted the virus without being aware that they were doing so.

This lag before the onset of symptoms corresponds to time delay in control-system theory. It is notorious for introducing oscillations into closed-loop systems, particularly when combined with substantial uncertainty in the model itself.

In addition to delays, there are very significant uncertainties. Testing, for example, has been spotty in some countries, and that inconsistency has obscured the number of actual cases. Even NPIs are not immutable. The extent to which the public is complying with policies is never 100 percent.

The point is, a pandemic is a dynamic, fast-moving situation, and inadequate local attempts to monitor and control it can be disastrous. In the Spanish flu pandemic of 1918, cities took widely varying approaches to the lockdown and release of their citizens, with wildly varying results. Some recovered straightforwardly, others had rebound spikes larger than the initial outbreak, and still others had multiple outbreaks after the initial lockdown.

IN THE ABSENCE of widespread immunity or vaccination, the only way to suppress the disease is total confinement—obviously not a viable long-term solution. A reasonable middle ground is to implement a policy that makes extensive use of feedback to keep R0 close to 1, with perhaps small oscillations on either side. In so doing we could maintain the critical caseload within the capacity of health care institutions while slowly and safely building immunity in our communities, and returning to normal social and economic conditions as quickly as is safely possible.

What exactly do we mean by "feedback" here? Consider what's already happening now: Public officials are now taking hospital caseloads into account before imposing or lifting restrictions. This is an example of the use of feedback, where the feedback variable is the number of cases in local hospitals. However, the use of feedback now is typically relatively coarse, and its results less satisfactory, in comparison with what could be achieved if control principles were more finely and systematically applied.

Here's an obvious way that coarseness would cause problems. If the tuning mechanism is too aggressive—for example, switching between full and zero social distancing—it would lead to severe oscillations and overwhelmed hospitals. On the other hand, tuning that is too timid also courts fiasco. An example of such tuning might be a policy requiring a full month in which no new cases are recorded before officials relax restrictions. Such a hypercautious approach risks needlessly prolonging the pandemic's economic devastation, creating a catastrophe of a different sort.

But a properly designed feedback-based policy that takes into account both dynamics and uncertainty can deliver a stable result while keeping the hospitalization rate within a desired approximate range. Furthermore, keeping the rate within such a range for a prolonged period allows a society to slowly and safely increase the percentage of people who have some sort of antibodies to the disease because they have either suffered it or they have been vaccinated—preferably the latter.

Clearly, tried-and-true principles of control theory, particularly feedback, can help officials plot more robust and optimal strategies as they attempt to safely ease the social distancing that has helped mitigate the COVID-19 pandemic. But how to make officials aware of these powerful tools?

Imagine an online interactive tool offering detailed, specific guidance in plain language and aimed at public officials and others charged with mounting a response to the pandemic in their communities. The guidance would be based on strategies developed by a small group of control theorists, epidemiologists, and people with policy experience. The site could review the now-familiar initial response, in which nonessential workers are confined to their homes except for essential needs. Then the site could go on to give some guidance on how and when the tightest restrictions could be lifted.

The biggest challenge to the designers of such a Web-based tool will be enabling nonspecialists to visualize how the various components of the epidemiological model interact with the various feedback policy options and model uncertainty. How exactly should the main feedback measure—likely some aspect of hospital or intensive care occupancy—be implemented? Which restrictions should be lifted in the first round of easing? How should they be eased in the first round? In the second round? While monitoring the feedback measure, how frequently should officials consider whether to implement another round of easing? Feedback will help officials determine when to time various phases of interventions.

Such an interactive tool that could assess different policy approaches, vividly illustrating what conditions must be in place to lessen the effects of uncertainty and shrink the projected caseload, would be of incalculable value. It could save untold lives.

Material in this article originally appeared as a post on Medium, "Coronavirus: Policy Design for Stable Population Recovery," and in several other outlets.

THE WOMAN IN THE WINDOW: Having tested positive for the novel coronavirus, Marietta Diaz self-quarantines in her Florida home on 23 March.

ZAK BENNET/AFP/GETTY IMAGES

Five Scenarios

To explore how feedback can save lives, we devised a series of scenarios, each indicative of a recovery strategy with a different level of feedback, and simulated the resulting policies against a commonly used infectious-disease computer model. We plotted the results in a series of graphs showing COVID-19 hospital cases as a function of time. Hospital occupancy is probably a more reliable and tangible measure than total case count, which depends on extensive testing that many countries (such as the United States) do not have at the moment. Furthermore, hospital ICU bed occupancy or ventilator availability is arguably an important measure of the ability of the local health care system to treat those who are suffering from respiratory distress acute enough to require intensive care and perhaps assisted breathing.

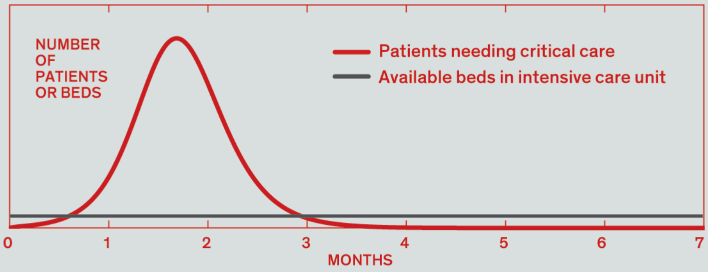
The model we used was created by Jeffrey Kantor, professor of chemical engineering at the University of Notre Dame (Kantor's model is available on GitHub). The model assumes we can suppress disease transmission to a very low level by choosing appropriate policy levers. Although worldwide experience with COVID-19 is still limited, at the time of this writing this assumption appears to be a realistic one.

To make the model more reflective of our current understanding of COVID-19, we added two types of uncertainty. We assumed different values of R0 to see how they affected outcomes. To consider how noncompliance with nonpharmaceutical interventions would affect results, we programmed for a range of effectiveness of these NPIs.

Our first, simplest simulation confirms what we all know by now, which is that not doing anything was not an option [Figure 1, below].

1. CURVE OF DEATH

Failing to respond to the outbreak of a deadly and highly contagious illness results in a sharp spike of serious cases that overwhelms the capacity of local hospitals.

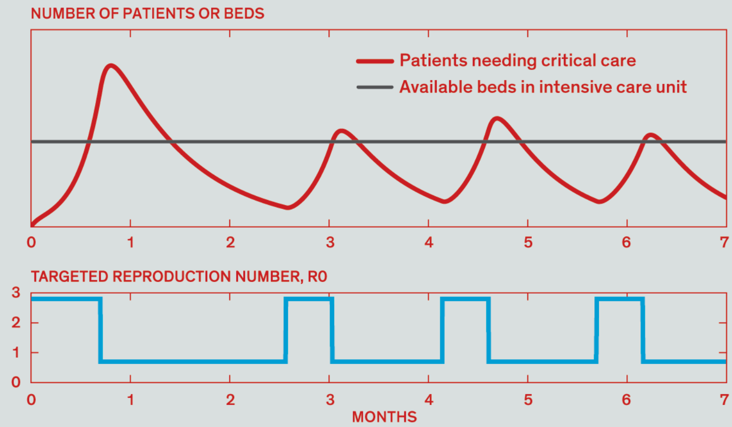


The large and lengthy peak well above the available bed capacity in the intensive care unit indicates a huge number of cases that will likely result in death. This is why, of course, most countries have put aggressive measures in place to flatten the curve.

So what do we do when the number of infections comes down? We considered one possibility: relaxing all restrictions when the number of infections has come down. We simulated such a tactic and confirmed what epidemiologists have long known: It would only lead to a second surge in infections. Not only could this second surge overwhelm our hospitals, it could also lead to an even higher mortality rate than the first surge, as occurred repeatedly in several U.S. cities during the Spanish flu epidemic of 1918.

2.
AN ON-OFF APPROACH

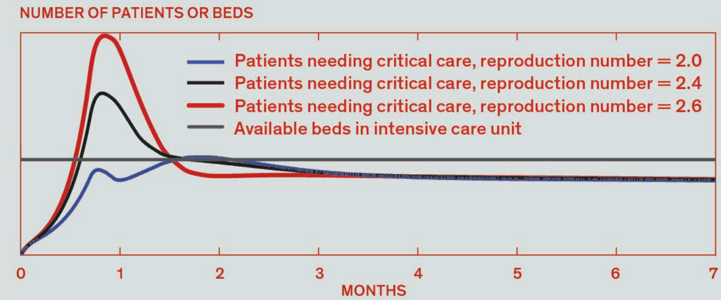
Imposing and lifting strong social restrictions causes abrupt swings in the reproduction number, R0 [in blue]. Those swings in turn lead to oscillations in the caseload [red, upper graph], which can exceed the capacity of local hospitals [black].



In the simple on-off approach to confinement, most of the usual restrictions on gatherings, travel, and social interaction are lifted entirely when the number of new ICU cases drops below a lower threshold, and then are put back into place when this number exceeds a higher threshold. In this case, the R0 swings sharply between two levels, a high above 2 and a low below 1, as shown in blue in the graph. This approach leads to oscillations, and if it is applied too aggressively, the high points of these oscillations will exceed the health care system's capacity to treat patients. Another likely problem with this approach has been labeled "social distancing fatigue." People become weary of the repeated changes to their routine—going back to work for a couple of weeks, then being told to stay at home for a few weeks, then being given the all-clear to go back to work, and so on.

3.
THE BEAUTY OF FEEDBACK

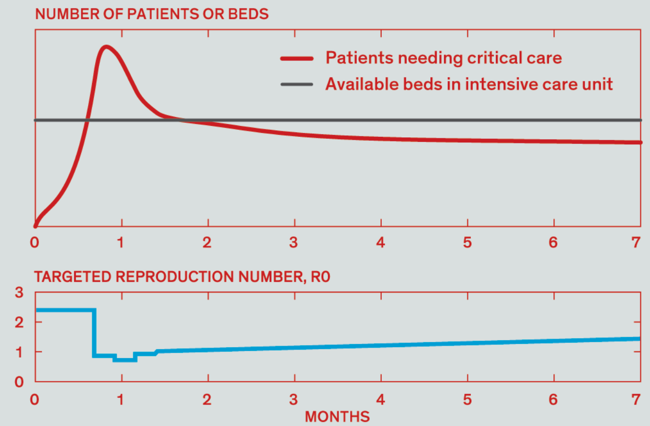
In real-world scenarios, officials will typically have only an approximate value for the basic reproduction number, R0. However, with appropriately applied feedback, this uncertainty won't matter. For R0s between 2.0 [blue] and 2.6 [red], the caseload stabilizes at an acceptable level within a couple of months as a result of feedback from actual hospital conditions.



Using feedback to finely and systematically modulate the restrictions imposed on a population to modify R0 leads to a policy that is robust. For example, early on in the outbreak, there will be a great deal of uncertainty about R0 because testing will still be spotty, and because an unknown number of people will likely have the disease without realizing it. That uncertainty will inevitably fuel a surge in initial cases. However, once the case count is stabilized by the initial restrictive regime, a policy based on feedback will prove very tolerant of variations in the base-level R0, which is the R0 that prevailed before the restrictions were put in place. As the graph shows, after a few months it doesn't matter whether the base-level R0 is 2.0 or 2.6 because the total case count stays well below the number of available hospital beds due to the use of feedback.

4. SYSTEMATIC APPROACH

A response based on control-theory principles strives to maintain the reproduction number, R0, at close to 1 [shown in blue]. As the caseload comes down, officials use feedback painstakingly to ease restrictions and control R0 so that it very gradually approaches 1, perhaps slightly and briefly exceeding it from time to time. As shown in the upper graph, this strategy keeps the caseload within the capacity of the local health care system to accommodate it indefinitely.



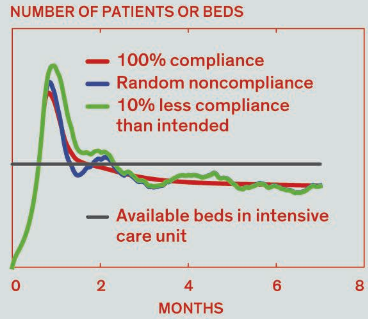
For our third experiment, we developed a scenario in which we targeted 90 percent occupancy of hospital intensive care units. To achieve this, we designed a simple feedback-based policy using the principles of control systems theory.

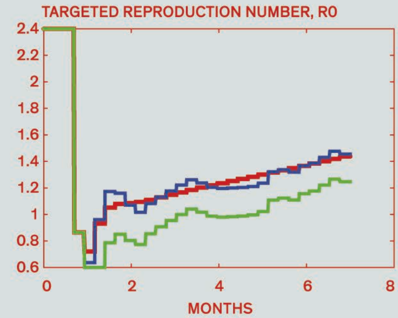
When R0 is high, many restrictions are put into place. People are largely confined to their homes, and services are limited to the bare minimum needed for society to function—utilities, police, sanitation, and food distribution, for example. Then, as conditions begin to improve, as revealed by our feedback measure of hospital-bed occupancy, other services are gradually phased in. Recovered people are allowed to move freely as they can no longer contract, or transmit, the virus. Perhaps people are allowed to visit restaurants within walking distance, some small businesses are allowed to reopen under certain conditions, or certain age groups are subject to less-stringent restrictions. Then geographical mobility might be loosened in other ways. The point is that restrictions are eased gradually, with each new gradation based carefully on feedback.

This strategy results in a stable response that maximizes the rate of recovery. Furthermore, the demand for hospital ICU beds never exceeds a threshold, thanks to a "set point" target below that threshold. The health care capacity limit is never breached. In addition, note the general upward trend for the release of restrictions, as the number of recovered and immune people grows and nonpharmaceutical interventions are gradually phased out.

5. COMPLIANCE CONUNDRUM

Officials can never know the exact extent to which people are disregarding their restrictions. But if their policies are based closely on feedback, realistic levels of noncompliance won't cause anything more than minor deviations from the expected levels of illness.





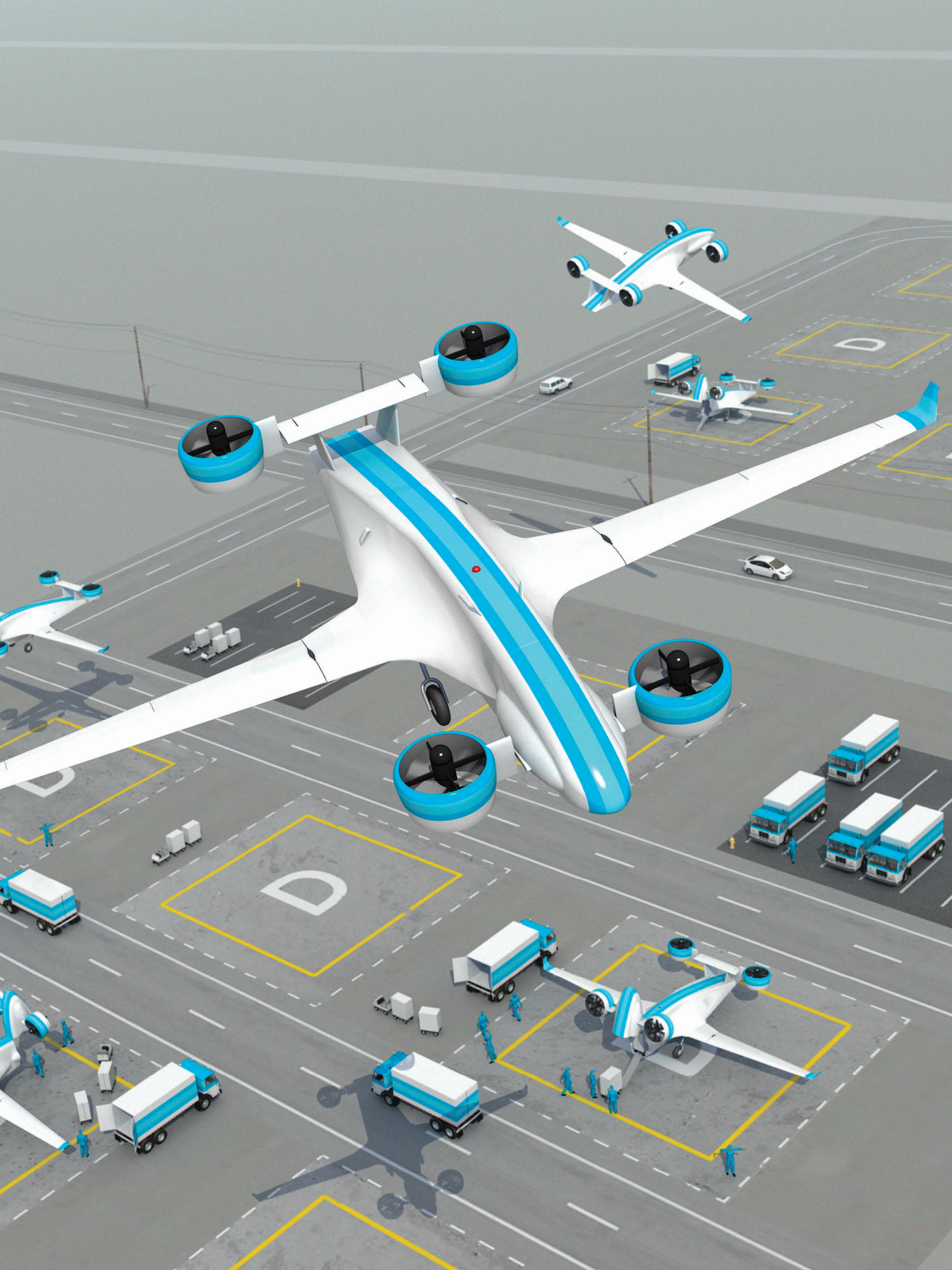
Using feedback creates a policy that restricts social interaction effectively even in the face of likely degrees of noncompliance.

In practice, what noncompliance means is that a given level of restrictions will result in an R0 that is slightly higher than expected, which in turn causes fluctuations in the number of people who are infected. Noncompliance might, for instance, result in the restrictions being 10 percent less effective than intended. However, through feedback, the policy will automatically tighten to compensate.

CAN CARGO DRONES SOLVE AIR FREIGHT'S LOGJAMS?

A DRONE STARTUP SAYS ITS BIG VERTICAL-TAKEOFF FLIER WOULD BE QUICK TO LAND, LOAD, AND TAKE OFF AGAIN

BY ED DE REYES
ILLUSTRATIONS BY JOHN MACNEILL



It's 4 a.m. on 23 December, and an MD-11 freighter has just landed at the logistics base near San Bernardino, Calif. It landed late because of heavy fog; many of the aircraft in San Bernardino won't be able to depart because the other airports in the Los Angeles Basin are fogbound as well.

Ground crews remove cargo from the MD-11 and get the many packages into delivery trucks and small planes as quickly as possible before the start of the morning rush of cargo from Amazon, Walmart, and other online retailers scrambling to deliver Christmas gifts. Meanwhile, construction is snarling traffic on the I-10 freeway into Los Angeles, and because of the fog, the I-15 is no better. Similar messes are often seen at choke points throughout the world, where the package-delivery business faces a raft of problems.

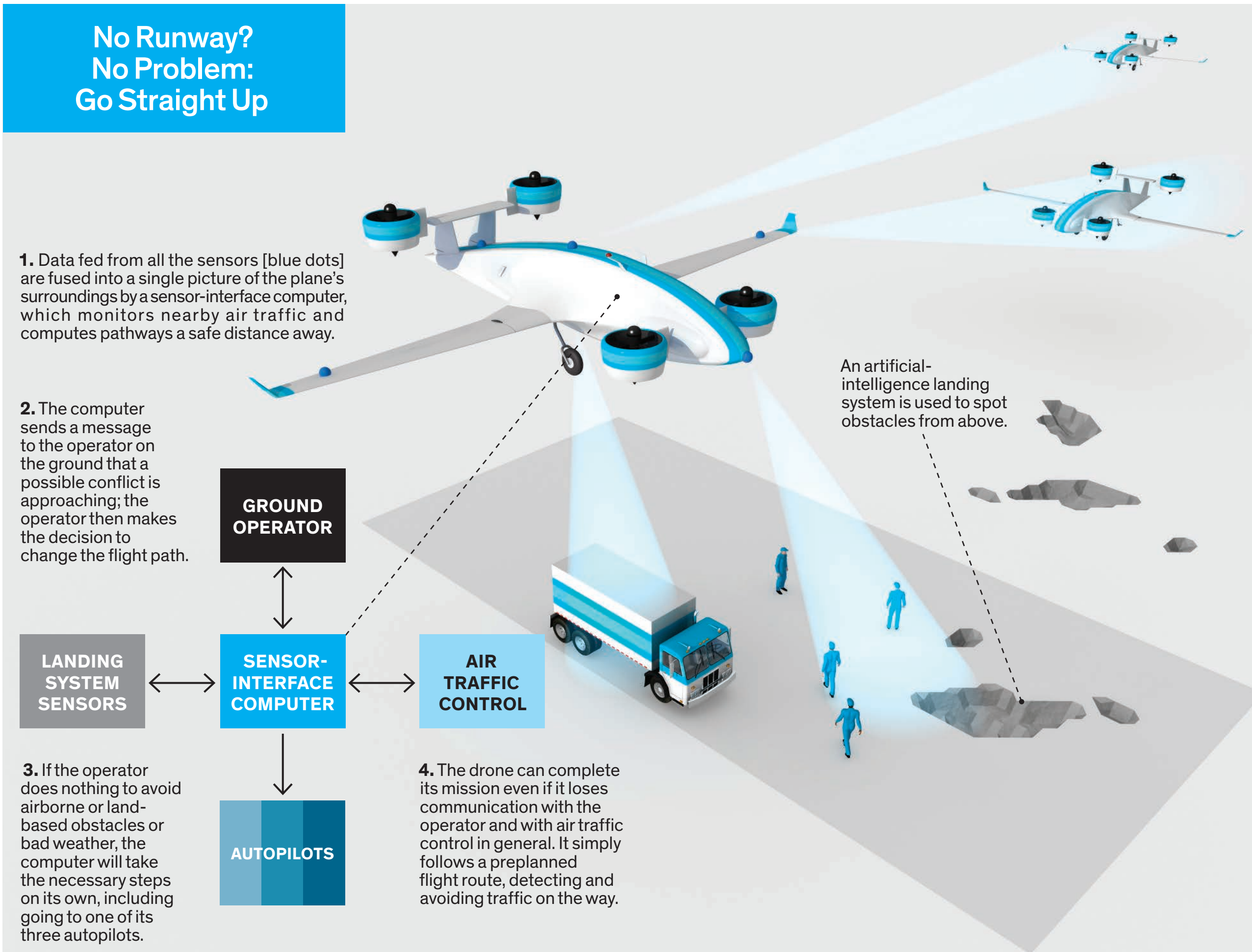
Now imagine that the MD-11 landing in San Bernardino meets up with a number of small aircraft, each fully fueled for a day of flying, with no pilots on board and no weather restrictions. The ground crew slides four fully loaded LD-2 containers with 2,400 kilograms of cargo through the nose of each craft, closes them up, and *poof*, they're cleared for takeoff, despite the thickening fog. The aircraft all take off, arriving 30 minutes later at various nearby airports, like the one in Oxnard, Calif., where they taxi to the ramps of the relevant package-delivery companies. There the ground crew removes the cargo, sorts it, and loads the small aircraft with new cargo, bound for the Beverly Hills package-sorting facility.

After being cleared for takeoff, the first of these aircraft rises like a helicopter before heading to the Beverly Hills site, which is just a parking lot with two ground handlers. As this odd-looking uncrewed aircraft comes in, the wings fold to help the plane clear any branches or power lines. Just then, a truck pulls right into the intended landing spot, but the aircraft perceives the obstruction and rises to a holding position until the ground crew can clear the area. Only then does it land, give up its cargo, accept new cargo, and get cleared again for takeoff.

And so it goes, right down the supply chain. Weather conditions that today would paralyze operations are shrugged off—fog, freezing rain, even failures of ground vehicles at improvised landing areas. Say a forklift can't cross a muddy field to unload the cargo container. No problem; one of the ground crew simply tilts the aircraft's nose up and unlatches the hooks and the cargo container slides right out with the help of a cargo winch. Throughout the day, the process of sorting, loading, and takeoff is repeated 15 more times at Oxnard and other airports in Southern California before a final 4,500 kg of payload is loaded into the aircraft and it returns to San Bernardino. Refueling takes place just once, at the end of the day.

THIS VISION OF URBAN air mobility, built on the promise of electric propulsion and on autonomous flight, is no sci-fi dream but a practical project, one that a number of companies are pursuing. Airbus has finished testing its Vahana, a concept electric vertical takeoff and landing (VTOL) aircraft that is meant to fly passengers at low altitudes within and between cities and towns. There is also the Cora, a creation of Google impresarios; it, too, is meant only for short distances and low altitudes. Neither of those two aircraft can carry much cargo, though, particularly in bad weather.

Consider how much easier it would be to use such methods to move cargo



instead of people. If there are no passengers on board, you can lose the heavy, bulky gear that assures passenger safety. Replace pilots and you can also dispense with the instruments that help them see where they're going, as well as the equipment that soundproofs the cabin and supports the windows, floor beams, bulkheads, and so forth. In some cases, an aircraft can weigh 25 percent more with human-factor equipment than without it.

My company, Sabrewing Aircraft, in Camarillo, Calif., was founded to exploit these advantages. By starting with a clean-sheet concept that was never meant to fly people, only cargo, and thus with no one on board to be put at risk, the aircraft can go to and from places no crewed rival can safely reach.

We call it the Rhaegal. If need be, it can lift almost 2,500 kg (5,500 pounds) of cargo straight up from the ground, like a helicopter; if a short runway is available, it can take off in the standard way, then fly straight ahead carrying as much as 4,500 kg (10,000 pounds). That's more than the new Cessna 408 SkyCourier can manage, and the Rhaegal flies much faster and higher. Also, it is designed to load and unload without the help of forklifts, pallet jacks, or other specialized equipment.

The Rhaegal sits low to the ground, whether on tarmac or even a sand dune, then tilts its nose upward so that either containerized or bulk cargo can be quickly loaded and secured. The aircraft's high-flotation "tundra tires" and four-post landing-gear arrangement allow it to land in mud, snow, sand, marsh, or deep puddles, and an integral loading ramp with rollers can be used to ease loading of pallets or containers.

Because the Rhaegal has a maximum gross weight above 600 kg (1,320 pounds) it falls under U.S. Federal Aviation Administration Regulation Part 23, which requires that it be remotely monitored and controlled and that it remain in contact with air traffic control at all times. Its operator, who can be hundreds or even thousands of miles away, controls the aircraft via a satellite link. In this way, the local air traffic control authority speaks to the operator through the aircraft, just as if the operator were sitting in the cockpit itself.

Prior to takeoff, the operator loads into the computer an exact flight plan, provided by the air traffic control authorities, that includes procedures for departing in any weather and also establishes the frequencies, routes, and a clearance to the aircraft's final destination. That way it can find its way home even if it loses communication with the operator or air traffic control.

The U.S. Federal Aviation Administration (FAA) requires that a human pilot of a conventional aircraft must see and avoid any air traffic that may be following an intersecting flight path. The same rule applies to the Rhaegal: It must do this job by itself, without operator input. This system, known as the Detect and Avoid (DAA) system, uses a mix of sensors, among them an anticollision radar (made by Garmin), a camera-based system that can spot conflicting air traffic and provide autopilot commands to avoid it (made by Iris Automation), and a lidar, or laser-ranging system, to detect power lines and other small obstructions at close range (made by Attollo Engineering). The DAA system also uses what is called automatic dependent surveillance-broadcast (ADS-B), a satellite-navigation system now mandated by the FAA for virtually all aircraft of any size operating in controlled airspace. This system tracks all flights no matter what paths they take, allowing for far more flexible routing than the older, ground-based radars could manage.

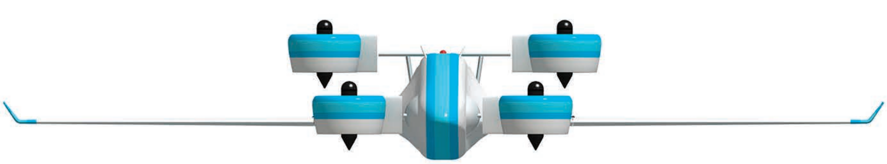
NOT ALL TRAFFIC PROBLEMS are in the air; some are on the ground. Cars or trucks may be moving around or even parked, for instance, when a parking lot itself serves as the landing zone. The Rhaegal uses an artificial-intelligence landing system to spot obstacles from above, including vehicles, people, rocks, and uneven surfaces. This landing system can recognize many types of obstacles and clearings, including landing pads aboard ships at sea.

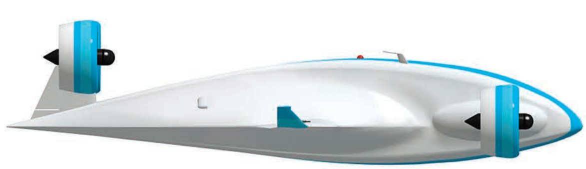
Rhaegal, by the Numbers

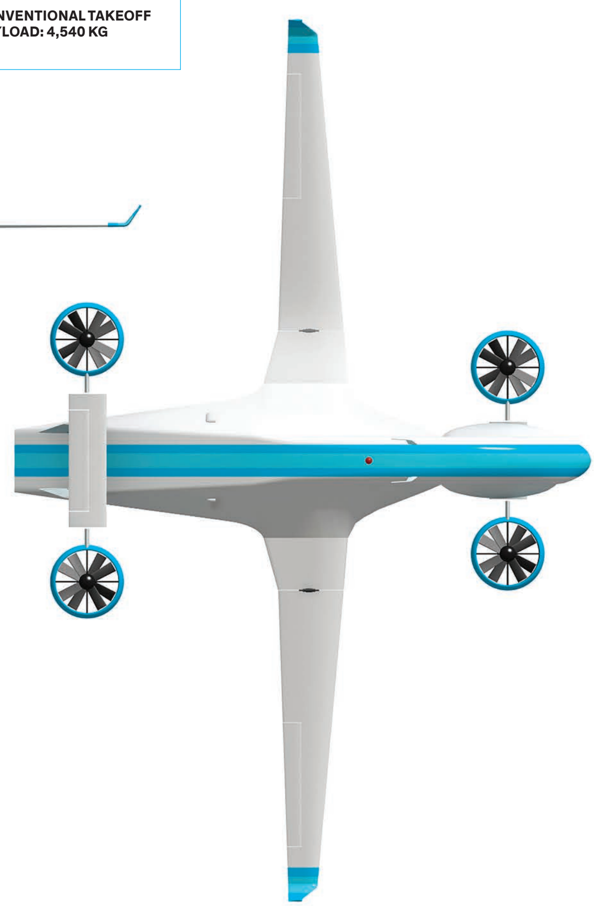
WINGSPAN: 18 METERS
LENGTH: 18 METERS
HEIGHT: 3.7 METERS

CRUISE SPEED: 180 KNOTS
RANGE: 1,000 NAUTICAL MILES (1,850 KILOMETERS)
CEILING: 6,700 METERS

VERTICAL TAKEOFF PAYLOAD: 2,450 KILOGRAMS
CONVENTIONAL TAKEOFF PAYLOAD: 4,540 KG







RHAEGAL'S ROLLOUT CAME in early May, and flight tests will begin at Edwards Air Force Base, in California. Oliver Garrow [left], Sabrewing's chief technology officer, is standing next to Lukas Flenner, the design leader.

SABREWING AIRCRAFT COMPANY

Data fed from all the sensors are fused into a single picture of the plane's surroundings by a sensor-interface computer, which monitors nearby air traffic and computes how to keep a safe distance away. When that happens, the computer sends a message to the operator on the ground that a possible conflict is approaching; the operator then makes the decision to change the flight path. If the operator does nothing, the computer will take the necessary steps on its own. Wherever the aircraft goes, the computer can detect bad weather up ahead and provide the data to the operator, who together with air traffic controllers can make changes to avoid storms, in some cases by flying well above them.

What's more, the Rhaegal is semiautonomous, meaning that it can complete its mission even if it loses communication with the operator and with air traffic control in general. It simply follows a preplanned flight route, detecting and avoiding traffic on the way and then landing at a remote location.

The Rhaegal's all-composite airframe is built in sections that can be quickly and easily repaired or even replaced in the field, with a minimum of hand tools. This modular design means that inspections that used to ground aircraft for weeks or even months can now be accomplished in hours. The Rhaegal is well suited for military applications: It can fly high and fast enough to avoid ground fire or fly low to avoid radar, enabling it to bring vital supplies to isolated units. It's even versatile enough to whisk four casualties and two medics to a mobile hospital within the "golden hour" after an injury occurs, greatly increasing the patient's chances of survival. In addition, the Rhaegal has a proprietary system that allows it to land safely if its propulsion system is damaged: It can either glide to a safe landing spot or, if the craft is hovering, it can land even if it loses the thrust of an entire duct unit.

RHAEGAL GETS ITS POWER from a turboshaft engine, which is basically a gas turbine designed specifically to turn a rotor rather than generate thrust, as a jet engine would do. This engine drives a generator that sends power to electric motors, which turn rotor blades. These are like propellers, but they are shrouded to provide more thrust than an open rotor would and to protect both people on the ground and the blades themselves when landing near bushes or trees. The point of this turboelectric drivetrain is to provide high efficiency in cruise flight and also high power during takeoff and landing. That heightened efficiency allows it to emit an estimated 70 percent less carbon than the Cessna 408 SkyCourier while carrying twice the load four times farther. And because of its turbo design, it can be made "greener" still by using biofuel.

The airframe of the first Rhaegal was completed in March 2020, and we expect flight testing will have begun by the time this article appears. Sabrewing has been in discussions with the FAA since 2017, and permission to start type certification–which assures the safety of a new type of aircraft–could come shortly.

Certification is no small administrative matter: Type certification for even a small private airplane can easily cost US $50 million to $100 million, but a cargo UAV should cost just a fraction as much to certify. And the Rhaegal stands first in line for such certification, ahead of all other electric cargo carriers using vertical takeoff and landing.

So don't be surprised if you look up sometime soon and see a Rhaegal cruising above your head. And some December in the not-too-distant future it might be playing Santa Claus to kids in both remote villages and large industrial centers throughout the world. ■

POST YOUR COMMENTS AT spectrum.ieee.org/cargodrones-jun2020

THE SEAS HAD TURNED ROUGH as a sudden squall whipped up the winds enough to howl through the rigging. And with those winds came a powerful smell of oil. Soon I could see the characteristic rainbow sheen from my position on the rail of this fishing trawler. It was May of 2016, and we were in the Gulf of Mexico, about 16 kilometers off the southeast coast of Louisiana. ● "Skimmer in the water," bellowed Kevin Kennedy, an Alaskan fisherman turned oil-spill remediation entrepreneur. Ropes groaned as the boat's winches lowered his prototype oil-recovery system into the heaving seas. As the trawler bobbed up and down, Kennedy's contraption rode the waves, its open mouth facing into the slick, gulping down a mix of seawater and crude oil. ● The stomach of Kennedy's device, to continue the analogy, was a novel separator, which digested the mixture of seawater and oil. By virtue of its clever engineering, it excreted essentially nothing but water. At the top of the separator's twin tanks, the collected oil began to swell. When enough had accumulated, the oil was sucked safely into a storage tank. Then the cycle would begin again. >

Fishing for Oil

With the right technology, oil spills on the water could be cleaned up much more effectively

BY LARRY HERBST

How much oil there was on the water here was a subject of great dispute. But its source was clear enough. In 2004, Hurricane Ivan blasted through the Gulf of Mexico triggering submarine landslides that toppled a drilling platform erected by New Orleans–based Taylor Energy. The mangled tops of Taylor's subsea oil wells were then buried under some 20 meters of mud. But all that loose mud didn't do much to stem the flow of oil and gas from many of these wells.

Efforts to contain the flow from the shattered wells, conducted between 2009 and 2011, saw only partial success. Oil continued to flow from some of these wells and rise to the surface for years to come.

While this oil spill posed a nasty threat to the marine environment, it served as a valuable testing ground for Kennedy's invention. This former fisherman has spent a small fortune to prove he has created an effective system for cleaning up oil spilled on the water, one that works well in real-world conditions. But for all he can tell so far, it's a system that nobody wants.

"I thought if I built a better mousetrap, everyone would want one," Kennedy says. "Instead, the world has decided they're okay with mice."

There are countless oil tankers, barges, rigs, and pipelines that operate in, around, and through U.S. coastal waters. Every year, some of them leak some of their contents. In a typical year the leakage amounts to no more than a million gallons or so. But every now and then a monster mishap spills considerably more: In 1989, the *Exxon Valdez* tanker ran aground on a reef and gushed some 11 million U.S. gallons (42,000 cubic meters) of oil into the pristine waters of Prince William Sound, Alaska. In 2005, Hurricane Katrina unleashed more than 8 million gallons (30,000 cubic meters) from Louisiana storage facilities. And even those incidents pale in comparison with the *Deepwater Horizon* disaster in 2010, in which a drilling rig leased by BP exploded in the Gulf of Mexico, killing 11 people and ultimately releasing some 210 million gallons (almost 800,000 cubic meters) of oil.

Such disasters not only ravage huge, complex, and delicate marine ecosystems, but they are also economically devastating: The damage to tourism and commercial fisheries is often measured in the hundreds of millions of dollars.

To deal with such fiascoes, engineers, chemists, and inventors have devised, sometimes on the fly, a grab bag of equipment, systems, chemicals, and procedures for collecting the oil and removing it, or for breaking it up or burning it in place to lessen its environmental impacts.

Today, the oil-spill-management market is a roughly US $100-billion-a-year industry with hundreds of companies. But multiple studies of the biggest episodes, including the *Deepwater Horizon* disaster, have questioned the industry's motives, methods, track record, and even its utility.

After decades in the industry, Kennedy, a small player in a big business, has unique perspectives on what ails it. His involvement with oil spills stretches back to 1989, when he bought his first fishing boat and paid for a license to trawl for shrimp near Prince William

PREVIOUS PAGES AND THESE PAGES: LARRY HERBST (4)

THICK OF THINGS: One of Kevin Kennedy's oil skimmers was deployed at a Superfund site on the shores of Lake Superior, where it was able to separate a highly viscous creosote oil from lake waters.

Sound. He couldn't have chosen a worse time to begin a fishing career. On the first day of the shrimping season, the *Exxon Valdez* ran aground on Bligh Reef, and Kennedy found himself drafted as a first responder. He spent more than four months corralling oil with his nets and using his fish pumps to transfer the unending gobs of sticky mess into his boat's hold.

Meanwhile, millions of dollars of oil-skimming equipment was airlifted to the nearby port of Valdez, most of it ultimately proving useless. Kennedy has witnessed something similar every time there's a spill nearby: There may be lots of cleanup activity, but often, he insists, it's just to put on a good show for the cameras. In the end, most of the oil winds up on the beach—"Nature's mop," he calls it.

In 2004, Kennedy participated in the cleanup that followed the grounding of the cargo ship *Selendang Ayu*—a tragic accident that cost six sailors their lives and released 336,000 gallons (1,272 cubic meters) of fuel oil into Alaskan waters. After that incident, he became convinced he could design gear himself that could effectively recover the oil spilled on the water before it hit the beach. His design employed fishing nets and fish pumps, normally used to transfer fish from the nets into the holds of fishing vessels. (Fish pumps use vacuum instead of whirling impellers, meaning no chopped-up fish.)

Fast forward to 2010 and the *Deepwater Horizon* disaster. The amount of oil released into the water by that well blowout seemed limitless—as did the money available to try to clean things up. Desperate for solutions, BP was exploring every avenue, including leasing oil-water separators built by actor Kevin Costner's oil-cleanup company, Ocean Therapy Solutions, now defunct. BP ultimately spent some $69 billion on its response to the disaster, including legal fees.

In the midst of those frenzied cleanup efforts, Kennedy packed up a hastily assembled oil-recovery system and drove from Anchorage to Louisiana. He presented himself to BP, which worked out a contract with Kennedy. But before he could sign it, the oil well was capped.

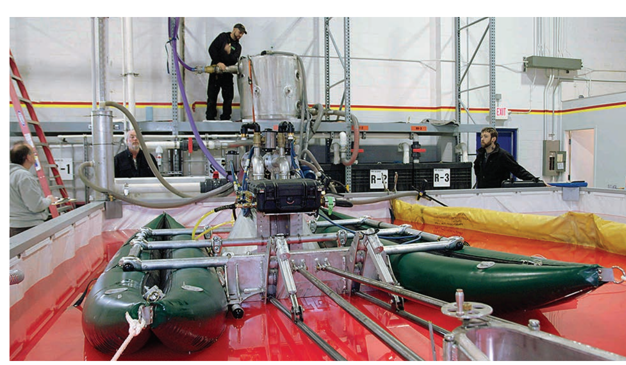
Although enormous oil slicks still covered the sea, Kennedy was no longer allowed to participate in the cleanup. Only an emergency, the relevant regulators felt, gave them the flexibility to try out new technology to address a spill. With the emergency now officially over, cleanup would be done strictly by the book, following specific U.S. Coast Guard guidelines.

First and foremost, only equipment from a very short list of certified systems would be used. So Kennedy watched from the shore as others went to work. Less than 3 percent of the oil from the BP spill was ever recovered, despite billions spent on efforts that mostly involved burning the oil in place or applying chemical dispersants—measures for addressing the problem that pose environmental hazards of their own.

In 2011, in the wake of the *Deepwater Horizon* spill, the XPrize Foundation mounted the Wendy Schmidt Oil Cleanup XChallenge, named for the philanthropist and wife of Eric Schmidt, the former executive chairman of Google. The purpose of the contest was to foster technical advances in the oil-spill cleanup industry. Kennedy submitted his system on something of a lark and was startled to learn he was chosen as one of 10 finalists from among hundreds of entrants, including some of the biggest names in the oil-spill field.

"All the global players were there: Lamor, Elastec, Nofi. Some of these are hundred-million-dollar companies," says Kennedy. "When I finished packing up the shipping container to go down to the competition, I think I had $123 left in my checking account." His life savings depleted, Kennedy was forced to ask friends for donations to afford the plane ticket to New Jersey, where the competition was being held.

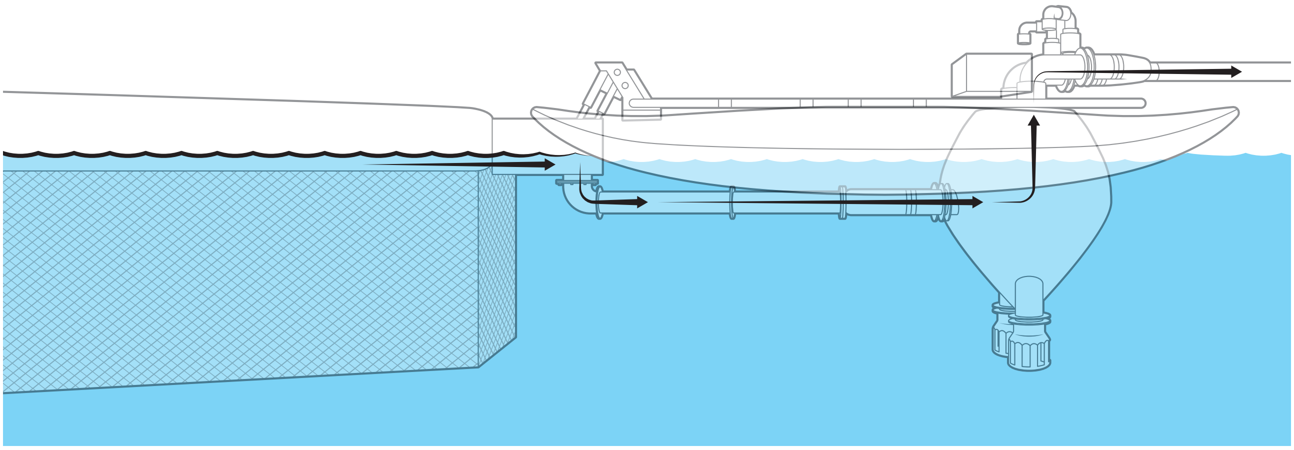
Located in Leonardo, N.J., the Department of the Interior's Ohmsett facility houses a giant pool more than 200 meters long and almost 20 meters wide. Researchers there use computerized wave generators to simulate a variety of open-water environments. Whereas



TRIAL WITHOUT FIRE: Kevin Kennedy's apparatus has been subjected to various controlled tests. In 2011, such tests were conducted as part of the Wendy Schmidt Oil Cleanup XChallenge [left]. In 2017, a smaller version of Kennedy's apparatus underwent testing to obtain an ASTM rating [right]. In both cases, the tests were performed with a thick layer of oil [red] on the water's surface, a condition that Kennedy points out rarely occurs in a real spill.

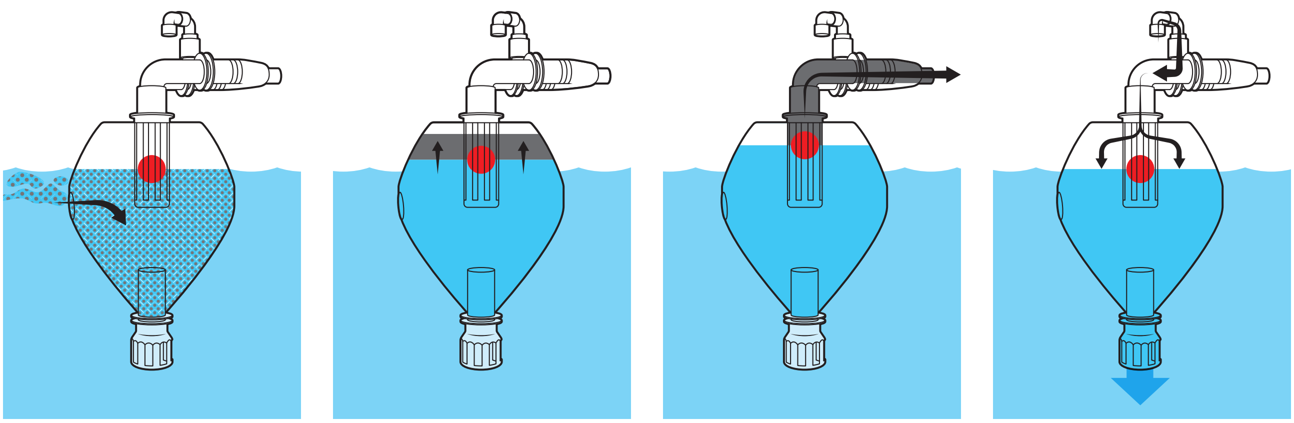
WHEN OIL AND WATER DO MIX

Kevin Kennedy's open-water oil-spill cleanup system is being marketed under the name Sea Otter by his company, PPR Alaska. Here's how it works.



1 Two floating arms, arranged in the shape of a V, channel oil floating on the surface as the system is towed forward into a spill.

2 Oil and water flow through a pipe to the separator chambers. A one-way valve in the pipe prevents backflow.



3 Oily water flowing horizontally from the intake pipe fills one of the two chambers, which are situated side by side.

4 Oil and water separate, while both are being raised above sea level by virtue of a vacuum applied by a pump through the pipe above.

5 An open float valve (whose ball floats above water but below oil) allows oil to be extracted, closing before water can be sucked out.

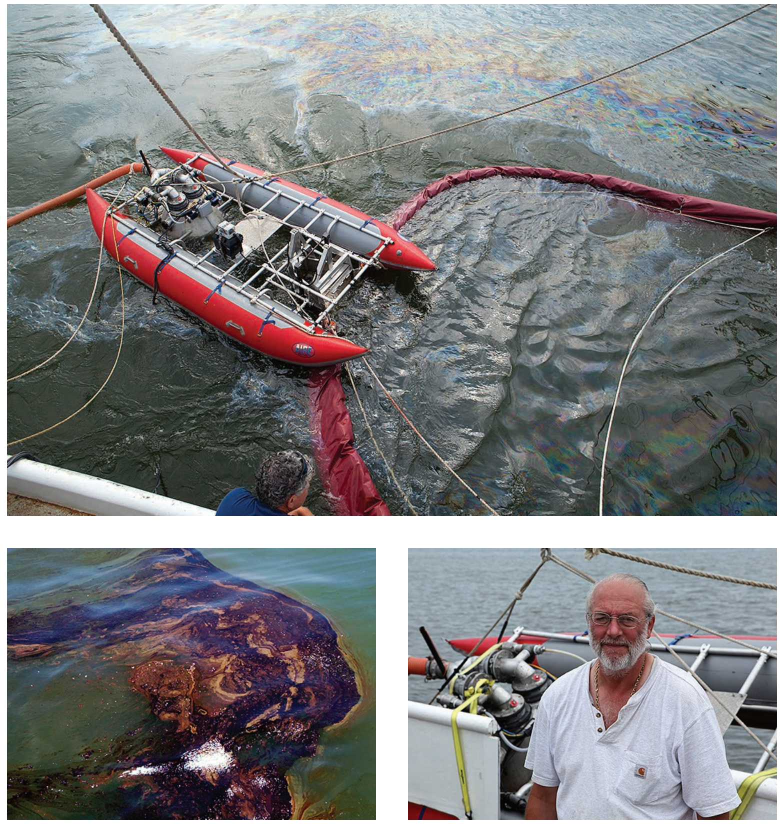
6 Releasing the vacuum allows water to flow out the bottom one-way valve until the water level in the chamber reaches sea level.

the industry standard for an oil skimmer was 1,100 gallons of oil per minute, the organizers of this XPrize competition sought a machine that could recover upwards of 2,500 gallons per minute with no more than 30 percent water in the recovered fluid.

Kennedy had cobbled together his skimmer from used fishing gear, including a powerful 5,000-gallon-per-minute fish pump. In addition, Kennedy's system used lined fishing nets to capture the oil at the surface. This equipment would be familiar to just about anyone who has worked on fishing boats, which are often the first on the scene of an oil spill. So there would be a minimal learning curve for such first responders.

When the XPrize competition began, Kennedy's team was the second of the 10 finalists to be tested. Perhaps due to inexperience, perhaps to carelessness, Ohmsett staff left the valves to the collection tank closed. Kennedy's equipment roared to full power and promptly exploded. The massive fish pump had been trying to force 5,000 gallons a minute through a sealed valve. The pressure ruptured pipes, bent heavy steel drive shafts, and warped various pressure seals.

ILLUSTRATION BY Chris Philpot



OVERLOOKED GULF TRAGEDY: A Taylor Energy platform located 16 kilometers from the Louisiana coast was toppled during Hurricane Ivan in 2004 and leaked oil for the next 14 years. The total volume of oil spewed into the Gulf of Mexico rivaled that of the *Deepwater Horizon* spill of 2010. Kevin Kennedy [bottom right] visited the Taylor spill in 2016 to test his Sea Otter oil skimmer [top] under real-world conditions, where the floating oil varied from a thin sheen to thick patches [bottom left].

LARRY HERBST (3)

Replacement parts arrived with just an hour to spare, narrowly allowing Kennedy to finish his test runs. Although his damaged pump could no longer run at full capacity, his skimmer delivered impressive efficiency numbers. "On some of the runs, we got 99 percent oil-to-water ratio," he says.

Kennedy didn't win the contest–or the $1 million dollar prize his fledgling company sorely needed. The team that took first place in this XPrize competition, Elastec, fielded a device that could pump much more fluid per minute, but what it collected was only 90 percent oil. The second-prize winner's equipment, while also pumping prodigious volumes of fluid, collected only 83 percent oil.

Although Kennedy's system demonstrated the best efficiency at the XPrize competition, buyers were not forthcoming. It wasn't surprising. "The real problem is you don't get paid by the gallon for recovered oil," says Kennedy, who soon discovered that the motivations of the people carrying out oil-spill remediation often aren't focused on the environment. "It's a compliance industry: You're required to show you have enough equipment to handle a spill on paper, regardless of whether the stuff actually works or not."

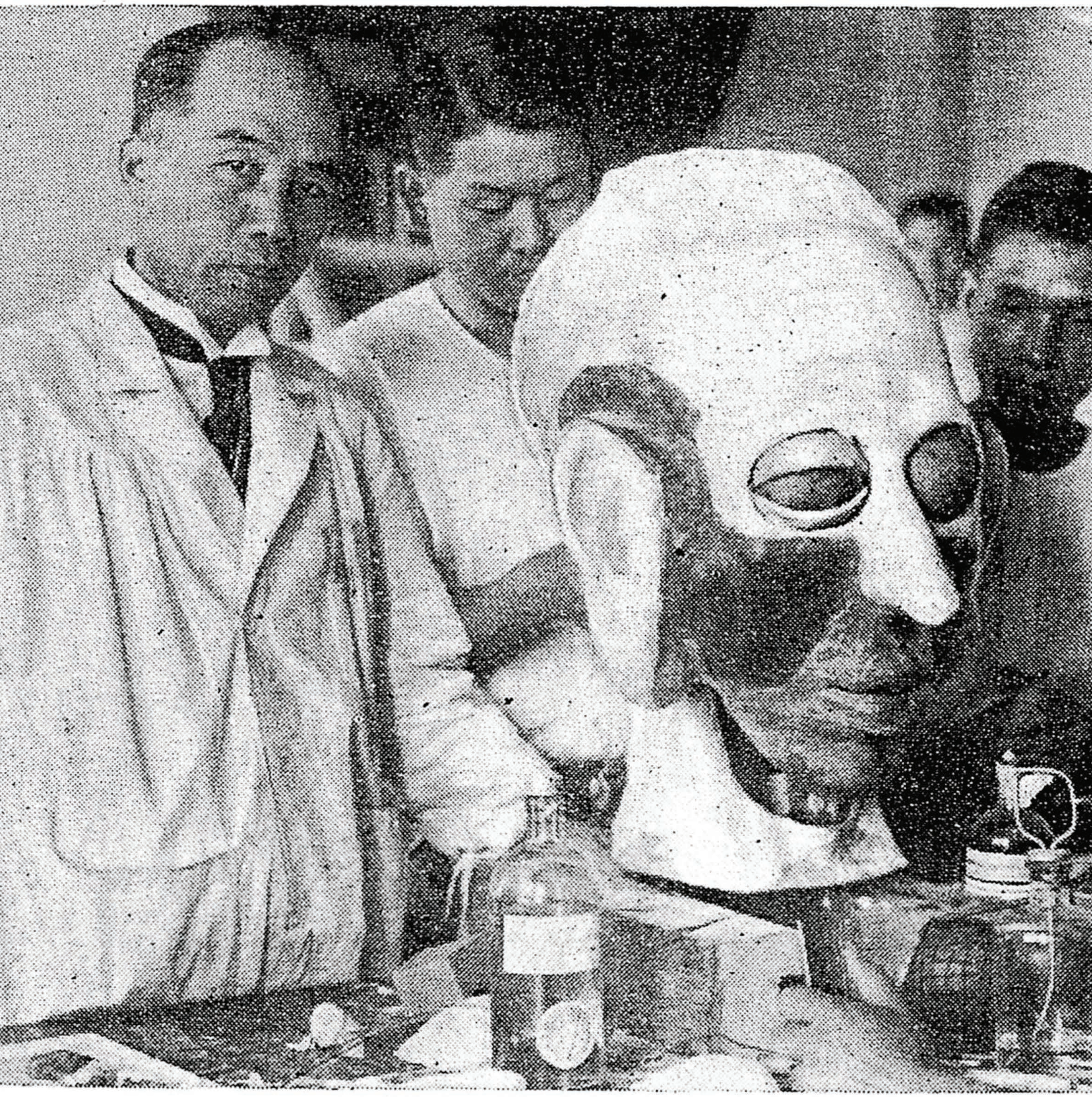
The problem, in a nutshell, is this: When there's an oil spill, responders are typically hired by the day instead of being paid for the amount of oil they collect. So there's little incentive for them to do a better job or upgrade their equipment to a design that can more efficiently separate oil from water. If anything, there's a reverse incentive, argues Kennedy: Clean up a spill twice as quickly, and you'll make only half as much money.

The key term used by regulators in this industry is EDRC, which stands for Effective Daily Recovery Capacity. This is the official estimate of what a skimmer can collect when deployed on an oil spill. According to the regulations of the Bureau of Safety and Environmental Enforcement, EDRC is computed by "multiplying the manufacturer's rated throughput capacity over a 24-hour period by 20 percent...to determine if you have sufficient recovery capacity to respond to your worst-case discharge scenario."

Reliance on the equipment's rated throughput, as determined by tank testing, and *assumed* effectiveness is at the heart of the agreement hammered out between government and oil companies in the wake of the *Exxon Valdez* disaster. It's a calculation of what theoretically would work to clean up a spill. Unfortunately, as Kennedy has seen time and again on actual spills, performance in the field rarely matches those paper estimates, which are based on tests that don't reflect real-world conditions.

Even though he thought the rules made no sense, Kennedy needed to get his equipment certified according to procedures established by ASTM International (an organization formerly known as American Society for Testing and Materials). So in 2017 he paid to have his equipment tested to establish its official ratings.

Those recovery ratings are determined by placing skimmers in a test tank with a 3-inch-thick (almost 8-centimeter) layer of floating oil. They are powered up for a minimum of 30 seconds, and the amount of oil they transfer is measured. It's an unrealistic test: Oil spills almost never result in a 3-inch layer of oil. Oil slicks floating on the ocean are usually measured in | CONTINUED ON PAGE 49



FACE TIME: Makoto Nishimura [left] and his team designed Gakutensoku's head so that it could express human affect.

THE SHORT, STRANGE LIFE OF THE FIRST FRIENDLY ROBOT

Japan's Gakutensoku was a giant pneumatic automaton that toured through Asia—until it mysteriously disappeared

BY YULIA FRUMER

IN 1923, A PLAY FEATURING ARTIFICIAL HUMANS OPENED IN TOKYO.

Rossum's Universal Robots—or *R.U.R.*, as it had become known—had premiered two years earlier in Prague and had already become a worldwide sensation. The play, written by Karel Čapek, describes the creation of enslaved synthetic humans, or robots—a term derived from *robota*, the Czech word for "forced labor." Čapek's robots, originally made to serve their human masters, gained consciousness and rebelled, soon killing all humans on Earth. In the play's final scene, the robots reveal that they possess emotions just like we do, and the audience is left wondering whether they would also achieve the ability to reproduce—the only thing still separating robots from humans.

The play was deeply disturbing for Makoto Nishimura, a 40-year-old professor of marine biology at the Hokkaido Imperial University, in the northern Japanese city of Sapporo. As Nishimura would later explain in a newspaper article, he was troubled because the play convincingly portrayed "the emergence of a perverse world in which humans become subordinate to artificial humans." A machine modeled on a human being but designed to work as a slave implies that the model itself (that is, we humans) are slaves, too, he argued. More concerning for Nishimura was that a struggle between humans and artificial humans was an aberration, something that went against nature.

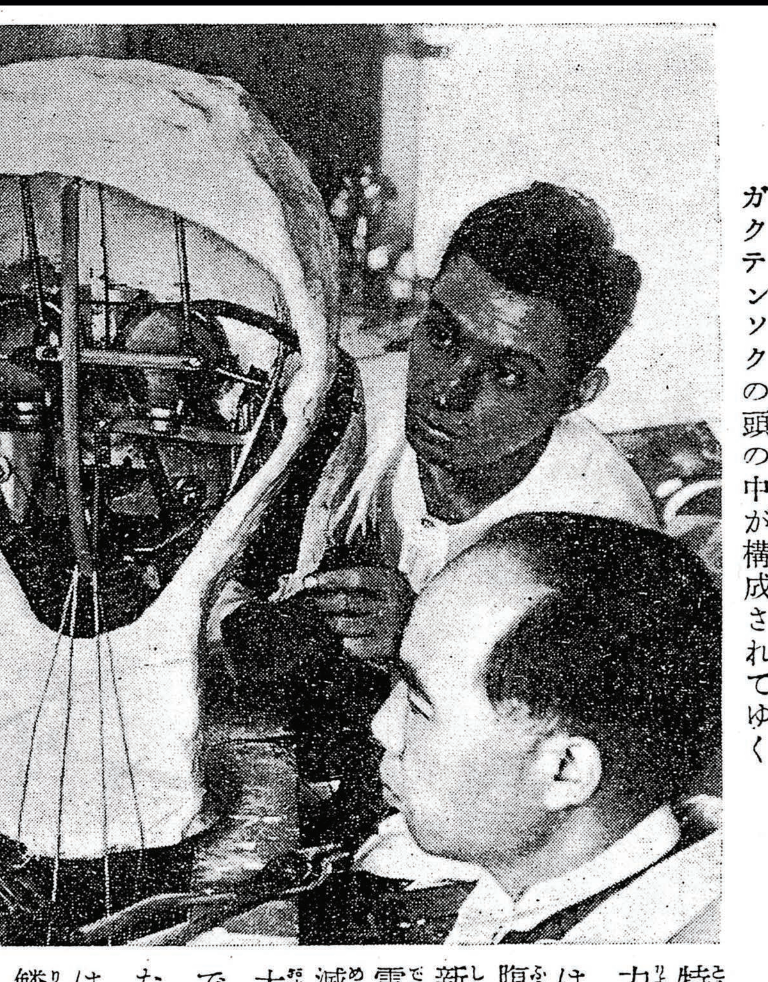
Perhaps he wouldn't have been so upset by a work of fiction if he hadn't witnessed how this fantasy was already becoming a reality in his world. He was familiar with early European and Japanese automata—mechanical figures devised to exhibit autonomous behaviors. While some of these mechanical wonders displayed noble or creative abilities, like playing musical instruments, drawing, writing calligraphy, or shooting arrows, others carried out mindless tasks. The latter were the ones that troubled Nishimura.

By the end of the 19th century, the world had also seen a variety of "steam men"—walking humanoids powered by internal steam engines (later models were powered by electric motors or gasoline engines) that appeared to pull carriages or floats. Many of these were developed in the United States. In the early 20th century, mechanical nurses, beauty queens, and policemen were introduced, and in his writings Nishimura also mentions mechanical receptionists, boat operators, and traffic cops. He may have seen some of these machines while he was living in New York City from 1916 to 1919 and pursuing a doctorate at Columbia University.

From our perspective, those machines seem more like curiosities than working devices. But for Nishimura, their potential to perform useful labor was as convincing as today's new AI and robots are to us. As a scientist he had witnessed attempts to create artificial cells in the laboratory, which only reinforced his belief that artificial human beings would one day populate Earth. The question was, what *kind* of artificial humans would those be, and what kind of relationship would they have with biological humans?

In Nishimura's opinion, the character of artificial humans—or any technology, for that matter—was determined by the intentions of their creators. And the intent behind carriage-pulling, steam-powered humanoids was to create workers enslaved by their human masters. *That*, in Nishimura's mind, would

PREVIOUS PAGES AND ABOVE: KAGAKU CHISHIKI



GENTLE GIANT: Makoto Nishimura [above, to the left of Gakutensoku] and an assistant, Bōji Nagao, pose with the robot, which was more than 3 meters tall, including its pedestal. The mechanism inside its head [left] allowed it to move its eyes, mouth, and neck.

HIROSHI MATSUO/OSAKA SCIENCE MUSEUM

inevitably lead to the formation of an *R.U.R.*-style exploited underclass that was bound to rebel.

Fearing a scenario that he described as humanity "destroyed by the pinnacle of its creation," Nishimura decided to intervene, hoping to change the course of history. His solution was to create a different kind of artificial human, one that would celebrate nature and manifest humanity's loftiest ideals—a robot that was not a slave but a friend, and even an inspirational model, to people. In 1926, he resigned his professorship, moved to Osaka, and started building his ideal artificial human.

Nishimura's creation was a direct response to one particular machine: Westinghouse's Televox, which debuted in 1927. Televox was a clunky, vaguely primate-shaped being designed to connect telephone calls. It exemplified all that Nishimura had come to fear. The creation of such a slavelike artificial human was not just shortsighted but also contrary to what Nishimura, the scientist, conceived of as the laws of nature. It was an abomination.

A curious thing about Nishimura's decision to build a robot was that he wasn't an engineer and possessed no expertise in mechanical or electrical systems. He was a marine biologist with a Ph.D. in botany. When he first encountered Čapek's *R.U.R.,* he was finishing an article on the cytology of *marimo*—aquatic moss balls endemic to the cold Lake Akan in northeast Hokkaido.

Yet it was precisely his background in biology that motivated Nishimura. An avid supporter of evolutionary theory, he was nevertheless skeptical of the idea of "survival of the fittest" and loathed the rhetoric of social Darwinism, which pitted humans against one another. Instead he favored "mutual aid" as the key driver of evolutionary change. He claimed that the core natural dynamic was collaboration, and that success by one individual (or one species) could be broadly beneficial.

"Today, human advances are framed in terms of a 'conquest of nature,' " Nishimura wrote in *Earth's Belly* (*Daichi no harawata*), a 1931 book detailing his philosophy of nature. "Rather than reflecting awe of the natural world, such *triumph* is more about kindling the struggle between humans." But look at human society, he urged: "We cannot ignore the fact that humans achieved civilization through collaborative effort."

Nishimura's take on evolution and natural hierarchies had a profound effect on his views of artificial humans. This set him apart from European writers like Samuel Butler, H.G. Wells, and Čapek, who favored the "survival of the fittest" model and postulated that the success of humanoid machines would spell doom for humankind. Nishimura insisted that flesh-and-blood humans could benefit from the evolution of artificial humans—*if* the artificial humans were designed to be inspirational models rather than slaves.

The artificial human Nishimura created was indeed meant to awe. Picture a giant seated on a gilded pedestal with its eyes closed, seemingly lost in thought. In its left hand, it holds an electric light with a crystal-shaped bulb, which it slowly lifts into the air. At the exact moment the lightbulb illuminates, the giant opens its eyes, as if having a realization. Seemingly pleased with its eureka moment, it smiles. Soon, it shifts its attention to a blank piece of paper lying in front of it and begins to write, eagerly recording its newfound insights.

Nishimura refused to call his creation a "robot." Instead, he bestowed upon it the name Gakutensoku, which means "learning from the rules of nature." He considered his creation the first member of a new species, whose raison d'être was to inspire biological humans and facilitate human evolution by expanding our intellectual horizons. Nishimura rendered the word *gakutensoku* in katakana, a Japanese script also used for the scientific names of biological organisms. Nishimura envisioned future *gakutensoku* continuously evolving and becoming more complex.

CROWD PLEASER: Gakutensoku attracted and awed visitors at exhibitions in Japan, Korea, and China. To support the robot's body, Makoto Nishimura [above, in hat and coat] and his team built a pedestal that displayed the robot's name on the front panel.

HIROSHI MATSUO/OSAKA SCIENCE MUSEUM (2)

Exactly how Gakutensoku operated has long intrigued historians and roboticists. Just a few years after its completion, the machine was lost under somewhat mysterious circumstances (more on that later). A few photos of the original design exist. But our only glimpse into its inner workings comes from an article Nishimura wrote in 1931. A talented writer, he often sacrificed technical detail for the sake of captivating prose and poetic expression.

Gakutensoku's chief mechanism was set in motion by an air compressor. Presumably the compressor was powered by electricity. The airflow was controlled by a rotating drum

REBUILDING THE GIANT ROBOT THAT SMILED

There were no blueprints or detailed specifications, but Japanese researchers were able to create a faithful copy of Gakutensoku

In 1992, Gakutensoku was featured in a promotional film titled *Osaka—the Dynamic City*. The movie portrayed the city as a futuristic technological hub, with Gakutensoku symbolizing Osaka's long history of innovation. For the purpose of the movie, producers commissioned a small-scale model. But even though the model looked like Gakutensoku, the inner mechanism was entirely redesigned with more modern technology, such as electric motors and digital controllers. After the movie was shot, the model was moved to the Osaka Science Museum.

In 2007, one of the museum's researchers, Nōzō Hasegawa, decided that having a "fake" model didn't do justice to the technological legacy of Gakutensoku, and set about creating a more faithful, historically accurate reconstruction. The reconstruction process began in April 2007 and took a little over a year at a cost of 20 million yen (about US $200,000).

Hasegawa didn't have much to go on—just a few photos and articles by Nishimura and others. The photos were black-and-white, old, and somewhat blurry, and most were taken from the same angle. Moreover, looking at the different photos, Hasegawa realized that Nishimura must have made changes to the design over time, since the photos were not all consistent.

Nishimura's own articles offered only superficial details on Gakutensoku's construction and operation. For example, he described the height of Gakutensoku as "8 *shaku*"—a Japanese unit of length, equal to 30.3 centimeters (11.9 inches)—"above the chest," which isn't very specific if you don't know the dimensions of the rest of the body. Hasegawa had to analyze a photo in which Nishimura appears next to Gakutensoku in order to arrive at his estimation of 3.2 meters (from the floor, including the pedestal). Then there was the question of Gakutensoku's original colors. Nishimura said that it was painted gold, but with only black-and-white photos available, Hasegawa didn't know, for example, whether the wreath of leaves that adorned its head was gold or green.

Hasegawa was able to re-create Gakutensoku's serene smile [above, left], produced by the flow of air into the corners of the eyes and eyelids. He also enabled Gakutensoku to move its head and neck, as described in various articles and documents, as well as lift its crystal "inspiration light" and simulate a writing motion.

The reconstruction of Gakutensoku's inner mechanism [above, right], however, was tricky. Particularly challenging was the rotating drum designed to gradually increase and decrease the flow of air, to prevent Gakutensoku's facial expression from collapsing "like a mad man," as Nishimura described it. Replicating this mechanism proved beyond the means of the Osaka Science Museum team. And so, although the motions of the 21st-century Gakutensoku are produced by compressed air, as in the original, the flow of air is controlled by a computer.

The reconstructed Gakutensoku still greets visitors to the Osaka Science Museum, reminding us of the genesis of the wondrous humanoid robots produced in Japan today, and echoing the motto Nishimura inscribed on the wall of his workshop: "According to the origins and the evolution of living beings—from simple to complex." *—Y.F.*

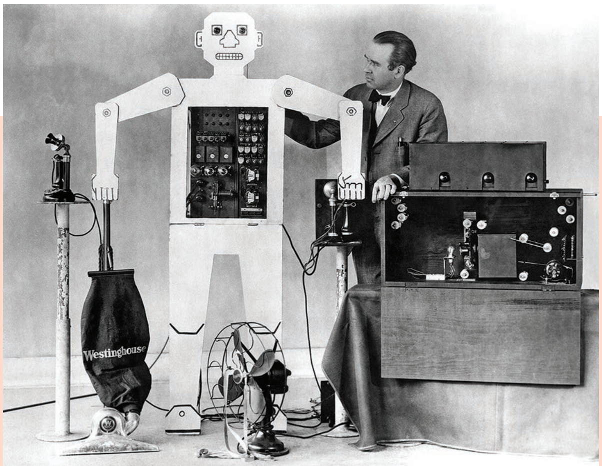
OSAKA SCIENCE MUSEUM (2)

affixed with pegs. When the mechanism was activated, the pegs opened and closed the valves on numerous rubber pipes that sent air to particular parts of Gakutensoku's body and caused them to move. Similar to the mechanisms of classic automata, the arrangement of the pegs on the drum allowed for a rudimentary programming of the sequence of motions. Unlike in traditional automata, only the pegs and the drum were mechanical, while the rest of the mechanism was pneumatic.

Nishimura strove to create "naturalness" in his machine. As he described in his article, he sought to "transcend the mechanical look" and the noisiness and clumsiness of "tin-can" bodies by eliminating as much metal as possible from Gakutensoku's frame. Only the robot's "skeleton" was made of metal. For his creature's soft tissue, he used rubber, the elasticity of which "made the movement much more natural, smooth, and without any sense of forced action." Additionally, "unlike the American robots" that relied on steam, Gakutensoku was set in motion by compressed air, which Nishimura considered to be a more "natural" power. He claimed that he arrived at the idea of using pneumatics after playing the *shakuhachi* (traditional Japanese bamboo flute) and experimenting with differences in the airflow. By varying the airflow and using different kinds of rubber with different elasticities, Nishimura was able to achieve complex, layered movement, which he described "as if within a big wave there were a medium-size wave and inside it another tiny wave as well."

For Nishimura, the most important feature of Gakutensoku was its ability to express human affect. Again, this was the result of carefully modulating the compressed air flowing through the rubber tubes. Prolonged pressure put on the outer bottom corners of the eyes and to the side of the mouth resulted in a smile. Slight airflow applied to one side of the neck produced a contemplative head tilt. Pleased with the results, Nishimura wrote that "compared to the American artificial humans, only ours has the ability to be expressive."

But when the flow of air to Gakutensoku's face was turned off, its facial expression suddenly and disturbingly collapsed. Nishimura and his team had to invent a device to allow the



CLUNKY CONTRAPTION: In the late 1920s, Westinghouse unveiled the Televox, portraying it as a futuristic robot servant. The Televox exemplified the idea of robots as the slavelike machines that Makoto Nishimura had come to fear, motivating him to build Gakutensoku.

gradual release of air pressure, which they described as "multiple wart-shaped convex parts aligned on one rotating axis.... Only when this modification was put in place did Gakutensoku finally stop looking like a mad man."

Nishimura emphasized the similarities between the structure of his artificial human and the anatomy of real human bodies. He claimed, for instance, that the compressed air that circulated inside Gakutensoku had a function similar to that of blood. Humans acquire energy by consuming food, and they distribute this energy through their circulatory system. Artificial humans, according to Nishimura, did something similar: They acquired electrical energy, then distributed this energy to different parts of their bodies by means of compressed air flowing in rubber tubes.

Nishimura's deep interest in biologically inspired design sometimes pushed him into philosophical semantics. "Humans naturally get energy from their mother's womb, while artificial humans get it from humans," he wrote in *Earth's Belly*. "So if humans are Nature's children, artificial humans are born by the power of the human hand and thus might be referred to as Nature's grandchildren."

Gakutensoku debuted in September 1928 at an exhibition in Kyoto celebrating the recently crowned Emperor Shōwa (known in the West as Hirohito). Reflecting on the exhibition several years later in *Earth's Belly*, Nishimura reported that Gakutensoku awed the crowds. Despite its being over 3 meters tall, observers said that it "looked more human than many expressionless humans." The following year, Gakutensoku was taken on tour and exhibited in Tokyo, Osaka, and Hiroshima, as well as in Korea and China, where the artificial human "work[ed] from 6 a.m. to 8 p.m." greeting spectators. Newspapers in Japan, China, and Korea reported on the exhibition and published photographs of the gentle giant, so that even those who could not see it in person had an idea of what it looked like.

Then Gakutensoku disappeared.

Nishimura himself never explained what happened. In an interview published in 1991, the scientist's son Kō Nishimura said the automaton vanished en route to Germany during the 1930s. But Kō was only a child at the time of the disappearance. While I corroborated that Gakutensoku indeed traveled to Korea and China, I have found no record of its being sent to Germany. Even if the claim is true, we don't know where exactly it vanished or who took it.

Despite its mysterious disappearance, Gakutensoku left a legacy that continues to reverberate in Japanese pop culture and robotics. During World War II, Japanese animators produced propaganda-heavy cartoons in which robots were depicted as inspirational heroes that used their supreme powers to assist humans. In the 1950s, Osamu Tezuka's *Astro Boy* (Tetsuwan Atomu, or "Mighty Atom") solidified the image of robots as emotionally sophisticated saviors driven by their empathy for other beings. I haven't seen any concrete evidence that Tezuka knew of Gakutensoku, but he grew up in the same suburb of Osaka where Nishimura lived and worked as a schoolteacher during the war.

Gakutensoku itself was featured as a character in a 1988 science-fiction movie titled *Teito Monogatari* (sometimes translated as *Tokyo: The Last Megalopolis*), in which the robot helps defeat the magical powers of a demonic villain. Nishimura, who died in 1956 at age 72, makes an appearance, played by his son Kō, who by then was one of Japan's best-known actors. The movie in turn inspired several books and TV shows exploring the origins of Japanese robots. In 1995 an asteroid spotted by Japanese astronomers was named 9786 Gakutensoku.

Then there's Gakutensoku's impact on Japanese robotics. Above all, many robot makers are guided by the underlying philosophy that machines are not in opposition to nature but rather part of it. Robots built in Japan from the 1970s on have a number of design characteristics reminiscent of those outlined by Nishimura—a preference for silent motion and the use of pneumatics; an emphasis on the textures of the robot's "skin" and "face"; and most important, an attentiveness to how people perceive and respond to a robot's humanoid features.

Since the '90s Japanese roboticists have channeled increasing efforts into cognitive robotics, exploring how humans think and behave in attempts to create likeable robot designs. Their aim is to build robots that not only perform tasks but also elicit positive emotional reactions in their users—robots that are friendly, in other words. It would be a stretch to say that Nishimura single-handedly shaped Japan's view of robotics, but the fact remains that few Japanese today fear *R.U.R.*-like scenarios of humanity being annihilated by their robotic overlords. As if fulfilling Nishimura's vision for a desirable relationship between humans and artificial humans, the common adage in Japan nowadays is that "robots are friends." ■

POST YOUR COMMENTS AT spectrum.ieee.org/japaneserobot-jun2020

UNDERWOOD ARCHIVES/GETTY IMAGES

FISHING FOR OIL

CONTINUED FROM PAGE 41

millimeters. And the thickness of an oil sheen, like that seen at the Taylor spill, is measured in *micrometers*.

"So many tests are really just a pumping exercise," says Robert Watkins, a consultant with Anchorage-based ASRC Energy Services who specializes in spill response. "But that isn't a true demonstration of response." The value of ASTM ratings, he explains, is allowing a reproducible "apples-to-apples" comparison of oil-spill equipment. He doesn't argue, however, that apples are the right choice in the first place.

Kennedy knows that it's not difficult to game the system to get high numbers. According to ASTM's testing rules, even a monstrous pump that doesn't separate oil from water at all can still get credit for doing the job. If a company stockpiles enough of those pumps in a warehouse somewhere–or maintains enough barges loaded with oil-absorbent pads–it will be certified as having a compliant spill-response plan, he contends. In the event of an actual spill, Kennedy says, most of that gear is useless: "Good luck cleaning anything up with pumps and diapers!"

In recent years, the Bureau of Safety and Environmental Enforcement (BSEE) and a consultancy called Genwest have worked to develop a better guide, hoping to replace Effective Daily Recovery Capacity with a different metric: Estimated Recovery System Potential (ERSP). This new measure looks at the entire system functionality and delivers far more realistic numbers.

A highly efficient system like Kennedy's would stack up favorably according to ERSP calculations. But according to Elise DeCola, an oil-spill contingency planning expert with Alaska-based Nuka Research and Planning Group, there has been limited adoption of the ERSP calculator by the industry.

"While BSEE recommends ERSP as an industry best practice, they do not require its use," says DeCola. "Operators that have deep inventories of low-efficiency skimmers–equipment that is still compliant with current guidelines–could suddenly be forced to invest in new skimmers." For many, moving the goal posts would simply cost too much.

The current rules, with their lack of emphasis on efficiency, accept pumping a large amount of oily water into your tanks–a mixture that must then be disposed of as hazardous waste. The better goal is to remove only the oil, and Kennedy's equipment is about as good as you can get in this regard, with its most recent ASTM-certified oil-to-water rating being 99 percent.

What's more, that "test tank" rating matches Kennedy's experiences with his equipment under real-world conditions. Whether on the Taylor slick with its micrometer-thick sheen, a Lake Superior Superfund site with spilled creosote as viscous as peanut butter, or a toxic spill in California's Long Beach Harbor, his efficiency numbers have always been very high, he claims.

Kennedy attributes this performance to his unique separation system. It uses a pair of collection vessels, in which the oil floats to the top of the mixture taken in. A specially designed float valve closes once the oil is drawn off the top. That extraction is done by a vacuum pump, which has the virtue of creating a partial vacuum, causing any water that gets caught up in the oil to boil off. The resultant water vapor is exhausted to the air before it can condense and dilute the recovered oil. The fuel oil his system collects often has even lower moisture content than it did when it came fresh out of the refinery.

Yet even with a skimmer that has remarkable performance, Kennedy has faced an uphill climb to find buyers. In 2016, he offered his equipment to Taylor Energy, only to be turned down. For the next two years, he repeatedly approached the Coast Guard, offering evidence that the Taylor Spill was larger than reported and insisting he had a potential solution. Even on a no-cure-no-pay basis, the Coast Guard wasn't interested.

"The Coast Guard shines when it tackles major crises, like Hurricane Katrina or the devastation in Puerto Rico," says retired Coast Guard Capt. Craig Barkley Lloyd, now president and general manager of Alaska Clean Seas. "But this was a slowly boiling frog."

It wasn't until 2018 that the Coast Guard was finally goaded to act. The Justice Department had hired Oscar Garcia-Pineda, a consultant who had studied the Taylor Spill, to do an independent assessment, which found the spill to be far more expansive than previously reported. According to the new calculations, the total volume of oil released over time rivaled that of the epic *Deepwater Horizon* spill. Picking up on that analysis, in October 2018 a *Washington Post* story labeled it "one of the worst offshore disasters in U.S. history."

In response to that newspaper article, the Coast Guard began to look for solutions in earnest. It quickly hired the Louisiana-based Couvillion Group to build a giant collection apparatus that could be lowered onto the seafloor to capture the leaking oil before it reached the surface. In the spring of 2019, Couvillion installed this system, which has since been collecting more than 1,000 gallons of oil a day.

For 14 years after the Taylor spill commenced, oil covered large swaths of the sea, and not a single gallon of that oil was recovered until Kennedy demonstrated that it could be done. The incentives just weren't there. Indeed, there were plenty of disincentives. That's the situation that some regulators, environmentalists, and spill-cleanup entrepreneurs, including this former fisherman, are trying to change. With the proper equipment, oil recovered from a spill at sea might even be sold at a profit.

During Kennedy's trial runs at the site of the Taylor spill in 2016, the crew of the shrimp boat he hired began to realize spilled oil could be worth more than shrimp. With the right technology and a market to support them–those same men might someday be fishing for oil. ■

POST YOUR COMMENTS AT spectrum.ieee.org/oilcleanup-jun2020

MATT FLYNN/COOPER HEWITT/SMITHSONIAN DESIGN MUSEUM

THE FIRST LAPTOP IN ORBIT

It was about the size of a three-ring binder and weighed 4.5 kilograms. Introduced in 1982, the Graphical Retrieval Information Display (GRiD) Compass was the first portable computer to feature a clamshell design. The bright amber of the laptop's 21.6-centimeter plasma screen could be "viewed from any angle and under any lighting conditions," according to its maker, GRiD Systems Corp. The computer was also exceedingly rugged. Sales reps claimed they would drop one in front of prospective buyers to show off its durability. With a price tag of US $8,150 (about $23,000 today), the Compass wasn't intended for the consumer market. Instead, it found a home with NASA, becoming standard equipment on space shuttle missions through the early 1990s. The space agency's trust in the computer was not misplaced: Reportedly, the GRiD flying aboard *Challenger* survived the January 1986 crash. ■

↗ **For more on the GRiD Compass, see spectrum.ieee.org/pastforward-jun2020**

Semantic segmentation for wildlife conservation.