MIT Technology Review

The computing issue

Volume 124
Number 6
Nov/Dec
2021
USD $9.99
CAD $10.99


"Well, the basic idea is I break into a company's network, encrypt their files and hold the keys for ransom."
Meet Ransomware
www.extrahop.com/advancedthreats
ExtraHop

"Hello, world!"

I'm Mat Honan, the new editor in chief of MIT Technology Review. This is the first issue of the magazine I've had the pleasure of working on. Maybe you have been reading Technology Review for years, like me. Or maybe this is your first issue. Either way, I'm excited by the opportunity to make this magazine something you look forward to reading every time it appears. I hope you enjoy reading this issue as much as we enjoyed putting it together.

I want to start by making you a promise about Technology Review: We're going to make it worth your time to read and worth your money to subscribe. We're going to bring you incredible stories about things at the edge of impossibility. We'll expose hidden truths, and hold the industries and people we cover to account. We're going to help you understand the ways in which science and technology are reshaping the world we all share. We're going to make you dream and wonder about the coming years. We're going to make you miss your stop because you just can't quit reading.

As you likely know, each issue has its own theme. This issue is on computing—a topic so utterly central to what we cover it seemed important to tackle it head on.

When I was young, personal computers were something entirely new. They were vaguely mysterious—you had to know the language—and utterly fascinating. I spent countless hours tinkering on the one in my mother's home office, writing simple programs, mapping out dungeons in Zork, and trying to understand the universe inside that box.

Today computers are, obviously, everywhere—in every pocket and automobile, even on the walls of our homes. And although computers, and computing, have become far more ubiquitous and accessible, their roles are often even more mysterious now than they were when I was a child in the 1980s. Virtually all aspects of modern life are now modulated by systems beyond our control. This is not merely because the network or the service or the algorithm is maintained by some unseen entity. As Will Douglas Heaven notes on page 78, the very nature of how computing works has changed with the rise of artificial intelligence. We want to help demystify things a bit.

This issue explores how we arrived where we are, and where we are going next. Margaret O'Mara's sweeping introductory essay (page 8) grounds the trajectory of computing in its greater historical context. Siobhan Roberts's exploration of the beguiling P vs. NP question (page 56) traces the long road Sisyphean researchers have traveled in trying to find a definitive answer. Chris Turner's review of *A Biography of the Pixel* (page 67) starts by exploring the complex history of "Digital Light" and builds to an unexpected, utterly delightful treatise on the triumph of Steamed Hams. (You're just going to have to read it.)

But history is meant to serve the present. Morgan Ames delves into the hype around One Laptop per Child (page 74) to help us find a better way toward ensuring that the most vulnerable in our society receive true equity of access. Fay Cobb Payton, Lynette Yarger, and Victor Mbarika explain how we can think about building true pathways into the industry for underrepresented groups (page 24). Lakshmi Chandrasekaran's examination of the triumph of silicon over other seemingly fallow technologies (remember spintronics?) shows how those alternatives may ultimately prove their worth (page 26). Meanwhile, Clive Thompson brings us the story of ASML, the Dutch company whose revolutionary process is keeping Moore's Law alive, at least for now (page 44).

Mat Honan is editor in chief of MIT Technology Review

But it's the future where things get weird and exciting. On page 38, you'll meet Alán Aspuru-Guzik, who is combining artificial intelligence and robotics in an attempt to accelerate materials discovery—with the ultimate aim of solving really thorny problems like climate change. And then there's Antonio Regalado's story on brain-computer interfaces (page 28). I frankly had to just sit down for a little while and think after I finished reading it, imagining a coming era that brings not only the ability to control machines with our minds, but also shared agency with an artificial neural network. It's wild stuff.

There is, of course, much more to explore within these pages. I hope you also find something that grabs you by the collar and makes you stop and think. And I hope to see you again soon. Let me know! I always want to hear your feedback. You can reach me on email at mat.honan@technologyreview.com, or yell at me on Twitter, where I am @mat.

Until next time,
Mat

ROBYN KESSLER


OWN
YOUR

Contents

Cover illustration by Saiman Chow

Masthead

Editorial

Editor in chief
Mat Honan

Executive editor
Michael Reilly

Editor at large
David Rotman

News editor
Niall Firth

Managing editor
Timothy Maher

Commissioning editors
Bobbie Johnson
Konstantin Kakaes
Amy Nordrum
Rachel Courtland

Senior editor, MIT News
Alice Dragoon

Senior editor, biomedicine
Antonio Regalado

Senior editor, climate and energy
James Temple

Senior editor, digital culture
Abby Ohlheiser

Senior editor, cybersecurity
Patrick Howell O'Neill

Senior editors, AI
Karen Hao
Will Douglas Heaven

Senior editor, computing
Siobhan Roberts

Senior editor, podcasts and live journalism
Jennifer Strong

Podcast producer
Anthony Green

Editor, Pandemic Technology Project
Lindsay Muscato

Senior reporters
Tanya Basu (humans and technology)
Eileen Guo (technology policy and ethics)

Reporters
Charlotte Jee (news)
Tate Ryan-Mosley (data and audio)

Copy chief
Linda Lowenthal

Social media editor
Benji Rosen

Engagement editor
Abby Ivory-Ganja

Proofreader
Barbara Wallraff

Editorial fellows
Casey Crownhart
Tatyana Woodall

Design

Chief creative officer
Eric Mongeon

Art director
Emily Luong

Marketing and events designer
Kyle Thomas Hemingway

Photo editor
Stephanie Arnett

Corporate

Chief executive officer and publisher
Elizabeth Bramson-Boudreau

Assistant to the CEO
Katie McLean

Manager of information technology
Colby Wheeler

Office manager
Linda Cardinal

Technology

Chief technology officer
Drake Martinet

Head of product
Mariya Sitnova

Senior software engineer
Molly Frey

UI product designer
Rachel Stein

Associate product manager
Allison Chase

Software engineer
Jack Burns

Head of data and analytics
Danielle Golds

Product management intern
Yufei Zheng

Events

Senior vice president, events and strategic partnerships
Amy Lammers

Director of event content and experiences
Brian Bryson

Head of international and custom events
Marcy Rizzo

Senior event content producer
Kristen Kavanaugh

Event content producer
Erin Underwood

Associate director of events
Nicole Silva

Event operations manager
Elana Wilner

Event partnership coordinator
Madeleine Frasca

Events associate
Bo Richardson

Finance

Vice president, finance
Enejda Xheblati

General ledger manager
Olivia Male

Accountant
Anduela Tabaku

Consumer marketing

Senior vice president, marketing and consumer revenue
Doreen Adger

Director of digital marketing
Emily Baillieu

Director of event marketing
Brenda Noiseux

Email marketing manager
Tuong-Chau Cai

Growth marketing manager
Em Okrepkie

Assistant consumer marketing manager
Caroline da Cunha

Circulation and print production manager
Tim Borton

Advertising sales

Senior vice president, sales and brand partnerships
Andrew Hendler
andrew.hendler@technologyreview.com
646-520-6981

Executive director, integrated marketing
Caitlin Bergmann
caitlin.bergmann@technologyreview.com

Executive director, brand partnerships
Marii Sebahar
marii@technologyreview.com
415-416-9140

Executive director, brand partnerships
Kristin Ingram
kristin.ingram@technologyreview.com
415-509-1910

Senior director, brand partnerships
Debbie Hanley
debbie.hanley@technologyreview.com
214-282-2727

Senior director, brand partnerships
Ian Keller
ian.keller@technologyreview.com
203-858-3396

Senior director, brand partnerships
Miles Weiner
miles.weiner@technologyreview.com
617-475-8078

Digital sales strategy manager
Casey Sullivan
casey.sullivan@technologyreview.com
617-475-8066

Media kit
www.technologyreview.com/media

MIT Technology Review Insights and international

Vice president, Insights and international
Nicola Crepaldi

Director of custom content, US
Laurel Ruma

Senior project manager
Martha Leibs

Content manager
Jason Sparapani

Senior manager of licensing
Ted Hu

Director of custom content, international
Francesca Fanshawe

Director of business development, Asia
Marcus Ulvne

Board of directors

Martin A. Schmidt, Chair
Peter J. Caruso II, Esq.
Whitney Espich
Jerome I. Friedman
David Schmittlein
Glen Shor
Alan Spoon

Customer service and subscription inquiries

National
877-479-6505

International
847-559-7313

Email
customer-service@technologyreview.com

Web
www.technologyreview.com/customerservice

MIT Records (alums only)
617-253-8270

Reprints
techreview@wrightsmedia.com
877-652-5295

Licensing and permissions
licensing@technologyreview.com

MIT Technology Review

One Main Street
13th Floor
Cambridge, MA 02142
617-475-8000

The mission of MIT Technology Review is to make technology a greater force for good by bringing about better-informed, more conscious technology decisions through authoritative, influential, and trustworthy journalism.

Technology Review, Inc., is an independent nonprofit 501(c)(3) corporation wholly owned by MIT; the views expressed in our publications and at our events are not always shared by the Institute.

INTRODUCTION

If you want to know what will come next in computing, look to the past.

It matters who is inventing the breakthroughs and why. By Margaret O'Mara

The history of the future of computing

1961: IBM engineers give visitors from Ames Research Center a look at the future.

NASA AMES/INTERNET ARCHIVE

If the future of computing is anything like its past, then its trajectory will depend on things that have little to do with computing itself.

Technology does not appear from nowhere. It is rooted in time, place, and opportunity. No lab is an island; machines' capabilities and constraints are determined not only by the laws of physics and chemistry but by who supports those technologies, who builds them, and where they grow.

Popular characterizations of computing have long emphasized the quirkiness and brilliance of those in the field, portraying a rule-breaking realm operating off on its own. Silicon Valley's champions and boosters have perpetuated the mythos of an innovative land of garage startups and capitalist cowboys. The reality is different. Computing's history is modern history—and especially American history—in miniature.

The United States' extraordinary push to develop nuclear and other weapons during World War II unleashed a torrent of public spending on science and technology. The efforts thus funded trained a generation of technologists and fostered multiple computing projects, including ENIAC—the first all-digital computer, completed in 1946. Many of those funding streams eventually became permanent, financing basic and applied research at a scale unimaginable before the war.

The strategic priorities of the Cold War drove rapid development of transistorized technologies on both sides of the Iron Curtain. In a grim race for nuclear supremacy amid an optimistic age of scientific aspiration, government became computing's biggest research sponsor and largest single customer. Colleges and universities churned out engineers and scientists. Electronic data processing defined the American age of the Organization Man, a nation built and sorted on punch cards.

The space race, especially after the Soviets beat the US into space with the launch of the *Sputnik* orbiter in late 1957, jump-started a silicon semiconductor industry in a sleepy agricultural region of Northern California, eventually shifting tech's center of entrepreneurial gravity from East to West. Lanky engineers in white shirts and narrow ties turned giant machines into miniature electronic ones, sending Americans to the moon. (Of course, there were also women playing key, though often unrecognized, roles.)

In 1965, semiconductor pioneer Gordon Moore, who with colleagues had broken ranks with his boss William Shockley of Shockley Semiconductor to launch a new company, predicted that the number of transistors on an integrated circuit would double every year while costs would stay about the same. Moore's Law was proved right. As computing power became greater and cheaper, digital innards replaced mechanical ones in nearly everything from cars to coffeemakers.

A new generation of computing innovators arrived in the Valley, beneficiaries of America's great postwar prosperity but now protesting its wars and chafing against its culture. Their hair grew long; their shirts stayed untucked. Mainframes were seen as tools of the Establishment, and achievement on earth overshadowed shooting for the stars. Small was beautiful. Smiling young men crouched before home-brewed desktop terminals and built motherboards in garages. A beatific newly minted millionaire named Steve Jobs explained how a personal computer was like a bicycle for the mind. Despite their counterculture vibe, they were also ruthlessly competitive businesspeople. Government investment ebbed and private wealth grew.

The ARPANET became the commercial internet. What had been a walled garden accessible only to government-funded researchers became an extraordinary new platform for communication and business, as the screech of dial-up modems connected millions of home computers to the World Wide Web. Making this strange and exciting world accessible were very young companies with odd names: Netscape, eBay, Amazon.com, Yahoo.

By the turn of the millennium, a president had declared that the era of big government was over and the future lay in the internet's vast expanse. Wall Street clamored for tech stocks, then didn't; fortunes were made and lost in months. After the bust, new giants emerged. Computers became smaller: a smartphone in your pocket, a voice assistant in your kitchen. They grew larger, into the vast data banks and sprawling server farms of the cloud.

Fed with oceans of data, largely unfettered by regulation, computing got smarter. Autonomous vehicles trawled city streets, humanoid robots leaped across laboratories, algorithms tailored social media feeds and matched gig workers to customers. Fueled by the explosion of data and computation power, artificial intelligence became the new new thing. Silicon Valley was no longer a place in California but shorthand for a global industry, although tech wealth and power were consolidated ever more tightly in five US-based companies with a combined market capitalization greater than the GDP of Japan.

It was a trajectory of progress and wealth creation that some believed inevitable and enviable. Then, starting two years ago, resurgent nationalism and an economy-upending pandemic scrambled supply chains, curtailed the movement of people and capital, and reshuffled the global order. Smartphones recorded death on the streets and insurrection at the US Capitol. AI-enabled drones surveyed the enemy from above and waged war on those below. Tech moguls sat grimly before congressional committees, their talking points ringing hollow to freshly skeptical lawmakers.

Our relationship with computing had suddenly changed.

The past seven decades have produced stunning breakthroughs in science and engineering. The pace and scale of change would have amazed our mid-20th-century forebears. Yet techno-optimistic assurances about the positive social power of a networked computer on every desk have proved tragically naïve. The information age of late has been more effective at fomenting discord than advancing enlightenment, exacerbating social inequities and economic inequalities rather than transcending them.

The technology industry—produced and made wealthy by these immense advances in computing—has failed to imagine alternative futures both bold and practicable enough to address humanity's gravest health and climatic challenges. Silicon Valley leaders promise space colonies while building grand corporate headquarters below sea level. They proclaim that the future lies in the metaverse, in the blockchain, in cryptocurrencies whose energy demands exceed those of entire nation-states.

The future of computing feels more tenuous, harder to map in a sea of information and disruption. That is not to say that predictions are futile, or that those who build and use technology have no control over where computing goes next. To the contrary: history abounds with examples of individual and collective action that altered social and political outcomes. But there are limits to the power of technology to overcome earthbound realities of politics, markets, and culture.

To understand computing's future, look beyond the machine.

1 The hoodie problem

First, look to who will get to build the future of computing.

The tech industry long celebrated itself as a meritocracy, where anyone could get ahead on the strength of technical know-how and innovative spark. This assertion has been belied in recent years by the persistence of sharp racial and gender imbalances, particularly in the field's topmost ranks. Men still vastly outnumber women in the C-suites and in key engineering roles at tech companies. Venture capital investors and venture-backed entrepreneurs remain mostly white and male. The number of Black and Latino technologists of any gender remains shamefully tiny.

Much of today's computing innovation was born in Silicon Valley. And looking backward, it becomes easier to understand where tech's meritocratic notions come from, as well as why its diversity problem has been difficult to solve.

INTRODUCTION

Silicon Valley was once indeed a place where people without family money or connections could make a career and possibly a fortune. Those lanky engineers of the Valley's space-age 1950s and 1960s were often heartland boys from middle-class backgrounds, riding the extraordinary escalator of upward mobility that America delivered to white men like them in the prosperous quarter-century after the end of World War II.

Many went to college on the GI Bill and won merit scholarships to places like Stanford and MIT, or paid minimal tuition at state universities like the University of California, Berkeley. They had their pick of engineering jobs as defense contracts fueled the growth of the electronics industry. Most had stay-at-home wives whose unpaid labor freed husbands to focus their energy on building new products, companies, markets. Public investments in suburban infrastructure made their cost of living reasonable, the commutes easy, the local schools excellent. Both law and market discrimination kept these suburbs nearly entirely white.

In the last half-century, political change and market restructuring slowed this escalator of upward mobility to a crawl, right at the time that women and minorities finally had opportunities to climb on. By the early 2000s, the homogeneity among those who built and financed tech products entrenched certain assumptions: that women were not suited for science, that tech talent always came dressed in a hoodie and had attended an elite school—whether or not someone graduated. It limited thinking about what problems to solve, what technologies to build, and what products to ship.

Having so much technology built by a narrow demographic—highly educated, West Coast based, and disproportionately white, male, and young—becomes especially problematic as the industry and its products grow and globalize. It has fueled considerable investment in driverless cars without enough attention to the roads and cities these cars will navigate. It has propelled an embrace of big data without enough attention to the human biases contained in that data. It has produced social media platforms that have fueled political disruption and violence at home and abroad. It has left rich areas of research and potentially vast market opportunities neglected.

Computing's lack of diversity has always been a problem, but only in the past few years has it become a topic of public conversation and a target for corporate reform. That's a positive sign. The immense wealth generated within Silicon Valley has also created a new generation of investors, including women and minorities who are deliberately putting their money in companies run by people who look like them.

But change is painfully slow. The market will not take care of imbalances on its own.

For the future of computing to include more diverse people and ideas, there needs to be a new escalator of upward mobility: inclusive investments in research, human capital, and communities that give a new generation the same assist the first generation of space-age engineers enjoyed. The builders cannot do it alone.

Brainpower monopolies

2 Then, look at who the industry's customers are and how it is regulated.

The military investment that undergirded computing's first all-digital decades still casts a long shadow. Major tech hubs of today—the Bay Area, Boston, Seattle, Los Angeles—all began as centers of Cold War research and military spending. As the industry further commercialized in the 1970s and 1980s, defense activity faded from public view, but it hardly disappeared. For academic computer science, the Pentagon became an even more significant benefactor starting with Reagan-era programs like the Strategic Defense Initiative, the computer-enabled system of missile defense memorably nicknamed "Star Wars."

In the past decade, after a brief lull in the early 2000s, the ties between the technology industry and the Pentagon have tightened once more. Some in Silicon Valley protest its engagement in the business of war, but their objections have done little to slow the growing stream of multibillion-dollar contracts for cloud computing and cyberweaponry. It is almost as if Silicon Valley is returning to its roots.

Defense work is one dimension of the increasingly visible and freshly contentious entanglement between the tech industry and the US government. Another is the growing call for new technology regulation and antitrust enforcement, with potentially significant consequences for how technological research will be funded and whose interests it will serve.

The extraordinary consolidation of wealth and power in the technology sector and the role the industry has played in spreading disinformation and sparking political ruptures have led to a dramatic change in the way lawmakers approach the industry. The US has had little appetite for reining in the tech business since the Department of Justice took on Microsoft 20 years ago. Yet after decades of bipartisan chumminess and laissez-faire tolerance, antitrust and privacy legislation is now moving through Congress. The Biden administration has appointed some of the industry's most influential tech critics to key regulatory roles and has pushed for significant increases in regulatory enforcement.

The five giants—Amazon, Apple, Facebook, Google, and Microsoft—now spend as much or more lobbying in Washington, DC, as banks, pharmaceutical companies, and oil conglomerates, aiming to influence the shape of anticipated regulation. Tech leaders warn that breaking up large companies will open a path for Chinese firms to dominate global markets, and that regulatory intervention will squelch the innovation that made Silicon Valley great in the first place.

Viewed through a longer lens, the political pushback against Big Tech's power is not surprising. Although sparked by the 2016 American presidential election, the Brexit referendum, and the role social media disinformation campaigns may have played in both, the political mood echoes one seen over a century ago.

We might be looking at a tech future where companies remain large but regulated, comparable to the technology and communications giants of the middle part of the 20th century. This model did not squelch technological innovation. Today, it could actually aid its growth and promote the sharing of new technologies.

Take the case of AT&T, a regulated monopoly for seven decades before its ultimate breakup in the early 1980s. In exchange for allowing it to provide universal telephone service, the US government required AT&T to stay out of other communication businesses, first by selling its telegraph subsidiary and later by steering clear of computing.

Like any for-profit enterprise, AT&T had a hard time sticking to the rules, especially after the computing field took off in the 1940s. One of these violations resulted in a 1956 consent decree under which the US required the telephone giant to license the inventions produced in its industrial research arm, Bell Laboratories, to other companies. One of those products was the transistor. Had AT&T not been forced to share this and related technological breakthroughs with other laboratories and firms, the trajectory of computing would have been dramatically different.

Right now, industrial research and development activities are extraordinarily concentrated once again. Regulators mostly looked the other way over the past two decades as tech firms pursued growth at all costs, and as large companies acquired smaller competitors. Top researchers left academia for high-paying jobs at the tech giants as well, consolidating a huge amount of the field's brainpower in a few companies.

More so than at any other time in Silicon Valley's ferociously entrepreneurial history, it is remarkably difficult for new entrants and their technologies to sustain meaningful market share without being subsumed or squelched by a larger, well-capitalized, market-dominant firm. More of computing's big ideas are coming from a handful of industrial research labs and, not surprisingly, reflecting the business priorities of a select few large tech companies.

Tech firms may decry government intervention as antithetical to their ability to innovate. But follow the money, and the regulation, and it is clear that the public sector has played a critical role in fueling new computing discoveries—and building new markets around them—from the start.

Location, location, location

3 Last, think about where the business of computing happens.

The question of where "the next Silicon Valley" might grow has consumed politicians and business strategists around the world for far longer than you might imagine. French president Charles de Gaulle toured the Valley in 1960 to try to unlock its secrets. Many world leaders have followed in the decades since.

Silicon Somethings have sprung up across many continents, their gleaming research parks and California-style subdivisions designed to lure a globe-trotting workforce and cultivate a new set of tech entrepreneurs. Many have fallen short of their startup dreams, and all have fallen short of the standard set by the original, which has retained an extraordinary ability to generate one blockbuster company after another, through boom and bust.

While tech startups have begun to appear in a wider variety of places, about three in 10 venture capital firms and close to 60% of available investment dollars remain concentrated in the Bay Area. After more than half a century, it remains the center of computing innovation.

It does, however, have significant competition. China has been making the kinds of investments in higher education and advanced research that the US government made in the early Cold War, and its technology and internet sectors have produced enormous companies with global reach.

The specter of Chinese competition has driven bipartisan support for renewed American tech investment, including a potentially massive infusion of public subsidies into the US semiconductor industry. American companies have been losing ground to Asian competitors in the chip market for years. The economy-choking consequences of this became painfully clear when covid-related shutdowns slowed chip imports to a trickle, throttling production of the many consumer goods that rely on semiconductors to function.

As when Japan posed a competitive threat 40 years ago, the American agitation over China runs the risk of slipping into corrosive stereotypes and lightly veiled xenophobia. But it is also true that computing technology reflects the state and society that makes it, whether it be the American military-industrial complex of the late 20th century, the hippie-influenced West Coast culture of the 1970s, or the communist-capitalist China of today.

What's next

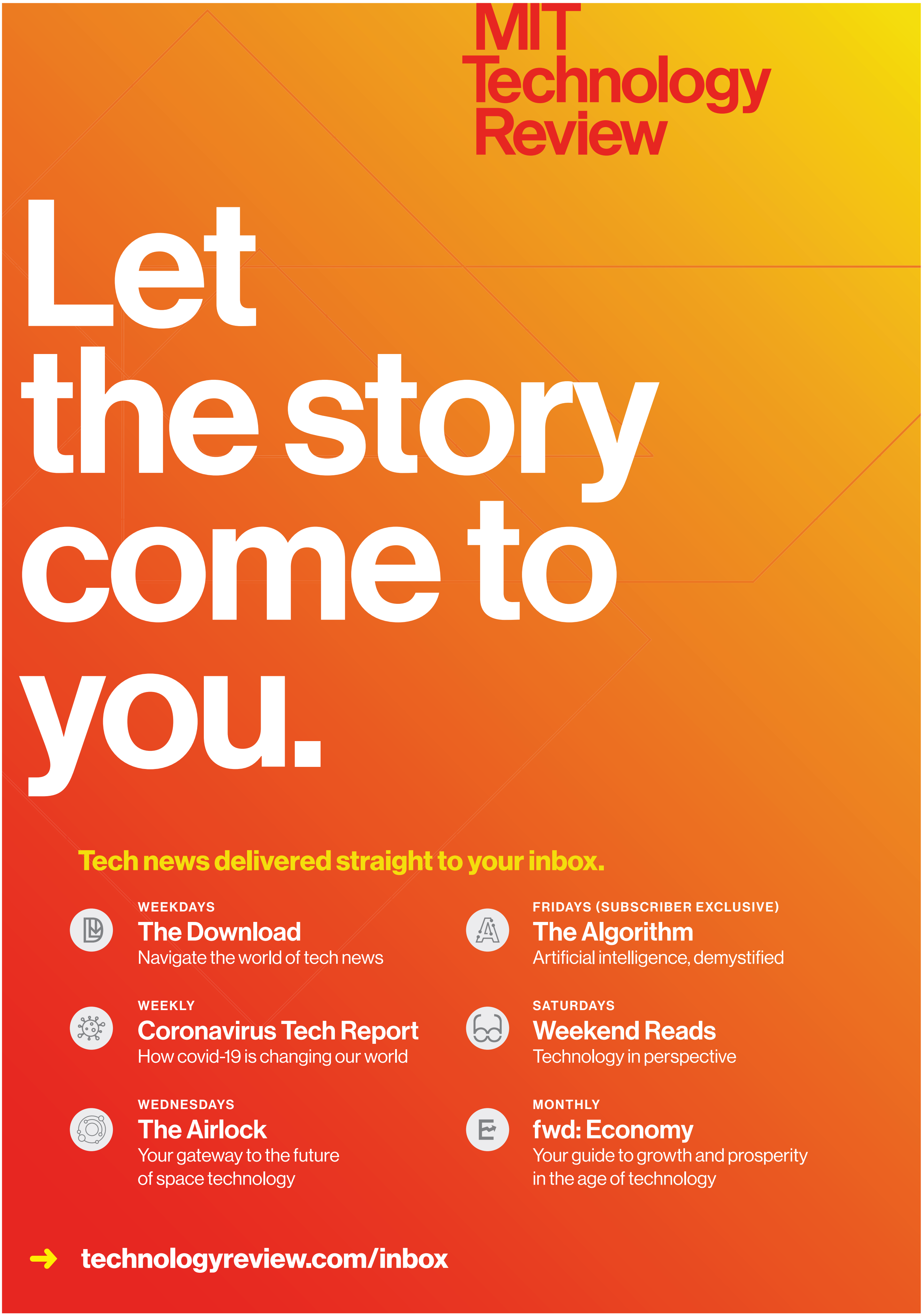
Historians like me dislike making predictions. We know how difficult it is to map the future, especially when it comes to technology, and how often past forecasters have gotten things wrong.

Intensely forward-thinking and impatient with incrementalism, many modern technologists—especially those at the helm of large for-profit enterprises—are the opposite. They disdain politics, and resist getting dragged down by the realities of past and present as they imagine what lies over the horizon. They dream of a new age of quantum computers and artificial general intelligence, where machines do most of the work and much of the thinking.

They could use a healthy dose of historical thinking.

Whatever computing innovations will appear in the future, what matters most is how our culture, businesses, and society choose to use them. And those of us who analyze the past also should take some inspiration and direction from the technologists who have imagined what is not yet possible. Together, looking forward and backward, we may yet be able to get where we need to go. ■

Margaret O'Mara is a historian at the University of Washington and author of The Code: Silicon Valley and the Remaking of America.


MIT Technology Review
Let the story come to you.
Tech news delivered straight to your inbox.
WEEKDAYS
The Download
Navigate the world of tech news
FRIDAYS (SUBSCRIBER EXCLUSIVE)
The Algorithm
Artificial intelligence, demystified
WEEKLY
Coronavirus Tech Report
How covid-19 is changing our world
SATURDAYS
Weekend Reads
Technology in perspective
WEDNESDAYS
The Airlock
Your gateway to the future of space technology
MONTHLY
fwd: Economy
Your guide to growth and prosperity in the age of technology
technologyreview.com/inbox

"Early in my career, a lot of times things would strike without warning, because the technology wasn't that good ... You don't see that as much anymore."

RE
PORT

Get up to speed on the key themes, big ideas, and major players discussed in this issue.

DANIEL ZENDER

WEATHER-WISE

Supercomputers have made it easier for meteorologists to predict deadly storms. But climate change makes it harder to get out of harm's way.

Casey Crownhart is a journalism fellow at MIT Technology Review.

When Hurricane Michael made landfall on the Gulf Coast of Florida in October 2018, it was a category 5 storm, with wind speeds over 150 miles per hour. The US National Hurricane Center had initially predicted they would reach less than half that.

Michael went through a process called rapid intensification, where a hurricane develops massively higher wind speeds in a short time. And the experts didn't see it coming.

Predicting the chaos that is the center of a hurricane, and understanding how storms strengthen, is still a challenge for forecasters. But armed with better models and more experience, they accurately predicted that Hurricane Ida, which hit New Orleans in September this year, would rapidly intensify, although the storm strengthened even more than they had expected.

Supercomputers have been part of these improvements in predicting where, when, and how storms might hit. And by the end of 2021, the US National Weather Service (NWS) will receive two brand-new supercomputers. It's an upgrade they hope will continue the steady march toward more accurate forecasts, which will become even more essential as climate change continues to fuel more intense storms.

The agency will use the new machines in operational forecasting—the system that forecasters use to make predictions like the ones on the nightly news. Once the agency has fully vetted them, probably in July 2022, the new supercomputers should help meteorologists better predict everything from the chance of rain in Denver to the odds that a hurricane will hit Miami.

Each supercomputer (one in Virginia and one in Arizona, so there's always a backup) is about the size of 10 refrigerators and has a capacity of 12.1 petaflops. "Flops" stands for "floating point operations per second," so 12.1 petaflops means the supercomputers can make just over 12 quadrillion calculations every second. It's a huge upgrade—nearly triple the size of the old system—and will cost roughly $300 million to $500 million over the next decade.

Computing capacity upgrades are a big piece of recent improvements in forecasts of hurricane path and intensity, says Michael Brennan, head of the Hurricane Specialist Unit at the National Hurricane Center.

Forecasts like the ones Brennan's team releases are made by humans who sort through different models and decide how to synthesize the information.

Projections of hurricane paths have gotten steadily more accurate over the past 30 years as large-scale weather models, and the computers running them, have improved. Average errors in hurricane path predictions dropped from about 100 miles in 2005 to about 65 miles in 2020. The difference might seem small when storms can be hundreds of miles wide, but when it comes to predicting where the worst effects from a hurricane will hit, "every little wiggle matters," Brennan says.

Understanding and predicting hurricanes' intensity has been more challenging than predicting their paths, since the strength of a hurricane is driven by more local factors, like wind speed and temperature at the center of a storm. Still, intensity predictions have also started to improve in the past decade. Errors in the intensity forecast within 48 hours decreased by 20% to 30% between 2005 and 2020.

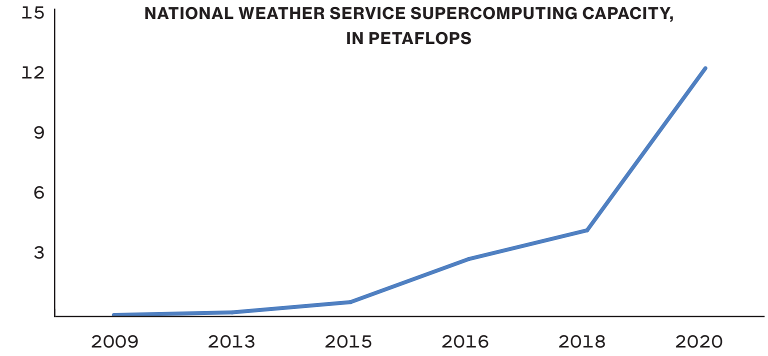
Powering predictions

When building models to predict something as complicated as the weather, "it's easy to suck up additional computer resources," says Brian Gross, director of the NWS Environmental Modeling Center.

Models can benefit from computing power in multiple ways, and each model can quickly slurp up huge amounts of capacity. A model can get more complex by digesting more information or by using more complicated physics to better represent the world. In a weather model, this might mean more details about processes in the ocean when considering the frequency of hurricanes.

More computing power might also allow a model to get more

Each of the NWS supercomputers, at 12.1 PETA-FLOPS, is about 150,000 TIMES more powerful than an ordinary laptop.



geographically precise. Weather models work by splitting the globe up into a bunch of pieces and trying to calculate what will happen in each of them. A higher-resolution model will break up the globe into smaller fragments, which means there are more of them to consider.

Finally, researchers can put together what's called an ensemble model, which they run as many as 20 or 30 times. Each of those runs is performed under slightly different conditions to see how the predictions differ. The results are then tallied up and considered together.

You've probably seen ensemble models in hurricane predictions. Consider a storm that starts in the Atlantic Ocean. If an ensemble forecast contains a wide variety of results, with the storm heading to Texas in some and skirting up the East Coast in others, it has many plausible paths. But if the ensemble members all show the storm hitting Florida's Gulf Coast, forecasters can be more certain about where it will land.

Previous supercomputer upgrades have led to improvements in multiple areas for some models, like the all-purpose Global Ensemble Forecast System (GEFS). In 2018, the last time the system was upgraded, NWS increased its resolution from 34 kilometers to 25 kilometers and the number of models in the ensemble from 21 to 31.

The changes, and the resulting forecast accuracy, have made life easier for officials like Jim Stefkovich, a meteorologist for the Alabama Emergency Management Agency who helps the state government prepare for weather risks. "Early in my career, a lot of times things would strike without warning, because the technology wasn't that good, and the warnings weren't always that accurate," Stefkovich says. "You don't see that as much anymore."

But even today, signals from weather forecasters aren't always clear to the public, which may not be tuned in for every storm, or may not understand the difference between a storm watch (hazardous conditions are expected) and a storm warning (they've already been observed). If people don't know how to react, a better forecast won't really help them stay safe, Stefkovich says.

This happened during Ida, he says. The forecasts for the storm's path and its strength were accurate a few days before. But partially because of poor communication, dozens of people lost their lives, and millions lost power or saw their homes or cars damaged.

Climate change means the risk of extreme weather will likely get worse. Forecasters hope that better predictions, communicated in the right way, can help people make better decisions when the storms come. ■

NATIONAL WEATHER SERVICE

EN ROUTE

A food delivery service in China is using Bluetooth to track orders more accurately.

MR Fu, a driver in Beijing for the food delivery service Eleme, makes about a dozen deliveries per shift. But he could make more—and spill less—if he didn't have to constantly get his phone out to update his status. "I have to log in every few minutes on the app to avoid being penalized if the delivery is late because of the restaurant," he says.

In China, fierce competition and the promise of instant delivery drive delivery apps to seek a technological edge. Now Eleme, one of the main players, has rolled out a vast indoor detection system to track drivers and ensure that customers receive their food on time. Wireless advances and the explosion of connected devices—including smartphones—have made this system possible.

Eleme, which has 83 million monthly active users, is owned by the tech giant Alibaba, which also owns Taobao, one of the world's biggest e-commerce platforms. Since launching the new system in hundreds of Chinese cities starting in 2018, Eleme says, it has saved merchants $8 million in refunds to customers for problems with their deliveries, including lateness.

To build it, Eleme had to find a cost-effective system that works indoors. GPS is accurate to five meters outside, but walls, furniture, and even people disrupt its signals. "It's also really bad at elevation," says Pat Pannuto, a computer science professor at the University of California, San Diego. That's a problem because most urban retailers in China are in multistory buildings.

Indoor localization systems based on Wi-Fi and radio-frequency identification do work, but Bluetooth is by far the cheapest, most reliable option. Its accuracy is roughly 10 meters, good enough to detect people walking into a shop or restaurant.

In early 2018, Alibaba placed more than 12,000 Bluetooth beacons in shops across Shanghai. Beacons emit signals that are picked up by drivers' phones in the form of "ID tuples." The app uploads each tuple to the platform's servers, where it's matched with merchant IDs, and the system logs where and when the signal was sent.

Similar networks are widely used for tracking goods, people, and services. One of the largest is in London's Gatwick Airport, where around 2,000 Bluetooth beacons are installed. But Eleme's is one of the first to be built out on a city scale.

To take its system to more cities in China, Alibaba exploited the fact that mobile phones can also act as Bluetooth beacons. Apple introduced this function for iOS devices in 2013, and similar features are now widely available on other smartphones.

Using this technology, more than 3 million merchants and a million drivers signed on to a pilot program to use their phones as beacons or receivers, delivering 3.9 billion orders to 186 million customers in 364 Chinese cities.

In early 2018, Alibaba placed more than
12,000
Bluetooth beacons in shops across Shanghai.

Yuan Ren is a science and technology journalist based in London.

For now, the system simply acts as a check on the driver's own logs, and it still requires drivers and merchants to have the app open on their phones to guarantee a connection. If drivers try to report their arrival before their phone receives a signal from a merchant's phone, the app sends a "too early" prompt.

It's not perfect—merchants can game the system by disabling Bluetooth, so couriers' apps don't log the time they spent waiting to pick up an order. And virtual beacons are less reliable than physical ones.

In theory, having more accurate data on drivers' locations means the system can better assign upcoming jobs and ensure that drivers can complete their deliveries in time. Alibaba says automation makes delivery drivers' work easier, but it may also intensify pressures on the job.

Alibaba hopes drivers' phones could one day become both beacons and receivers so their handsets could locate each other without relying on merchants' virtual beacons. The security and privacy implications of operating so many beacons are unclear.

Pannuto says the scale with which Alibaba has expanded the system is impressive. He isn't convinced it will be replicated elsewhere, but in China, where delivery remains cheap and demand high, companies are eager to find any way to outshine the competition. ■

DANIEL ZENDER

COMPUTING THE STARS

Astronomers are using AI, supercomputing, and the cloud to organize a universe of data.

Tatyana Woodall is a journalism fellow at MIT Technology Review.

AS space scientists collect more and more data, observatories around the world are finding new ways to apply supercomputing, cloud computing, and deep learning to make sense of it all. Here are some examples of how these technologies are changing the way astronomers study space.

What happens when black holes collide?

As a postdoctoral student in the US, astrophysicist Eliu Huerta started to think about how technology might help more breakthroughs happen in his field. Then researchers detected gravitational waves for the first time in 2015 with LIGO (the Laser Interferometer Gravitational-Wave Observatory).

Scientists have since charted these observations and scrambled to learn all they can about these elusive forces. They've detected dozens more gravitational-wave signals, and advances in computing are helping them to keep up.

As a postdoc, Huerta searched for gravitational waves by tediously trying to match data collected by detectors to a catalogue of potential waveforms. He wanted to find a better way.

Earlier this year Huerta, who is now a computational scientist at Argonne National Laboratory near Chicago, created an AI ensemble that's capable of processing a month's worth of LIGO data in just seven minutes.

The 10-year project will deliver a
500-PETABYTE
set of data and images to the cloud.

His algorithms—which run on special processors called GPUs—combine advances in artificial intelligence and distributed computing. Using either separate computers or networks that act as a single system, Huerta can identify gravitationally dense places like black holes, which produce waves when they merge.

Huerta's collection of AI models is open source, which means anyone can use them. "Not everybody has access to a supercomputer," he says. "This is going to lower the barriers for researchers to adopt and to use AI."

How has the night sky changed?

As much as astronomy has expanded, the field has been slow to integrate cloud computing. The Vera C. Rubin Observatory, currently under construction in Chile, will become the first astronomical institution of its size to adopt a cloud-based data facility.

When the observatory starts up in 2024, the data its telescope captures will become available as part of the Legacy Survey of Space and Time (LSST) project, which will create a catalogue thousands of times larger than any previous survey of the night sky. Past surveys were almost always downloaded and stored locally, which made it hard for astronomers to access each other's work.

"We are making a map of the full sky," says Hsin-Fang Chiang, a member of the Rubin's data management team. And in the process, they are building "a huge data set that's going to be useful for many different kinds of science in astronomy."

Although Chiang's PhD is in astronomy, her initial research had nothing to do with the survey. Years later, she got a chance to be involved thanks to the sheer size of the project. She's proud that her work could improve the way scientists collaborate.

The 10-year project will deliver a 500-petabyte set of data and images to the cloud, to help astronomers answer questions about the structure and evolution of the universe.

"For each position in the sky, we'll have more than 800 images there," says Chiang. "You could even see what happened in the past. So especially for supernovas or things that change a lot, then that's very interesting."

The Rubin Observatory will process and store 20 terabytes of data every night as it maps the Milky Way and places beyond. Astronomers affiliated with the project will be able to access and analyze that data from anywhere via a web browser. Eventually, the images the telescope takes every night will be converted into an online database of stars, galaxies, and other celestial bodies.

> *By comparing 4,000 simulations, scientists could rewind time and ask why some places in the universe are rife with cosmic activity while others are barren.*

DANIEL ZENDER

What did the early universe look like?

Advances in computing could help astronomers turn back the cosmic clock. Earlier this year, Japanese astronomers used ATERUI II, a supercomputer that specializes in astronomy simulations, to reconstruct what the universe may have looked like as early as the Big Bang.

ATERUI II is helping the researchers investigate cosmic inflation—the theory that the early universe expanded exponentially from one moment to the next. Astronomers agree that this expansion would have left extreme variations in the density of matter that would have affected both the distribution of galaxies and the way they developed.

By comparing 4,000 simulations of the early universe—all with different density fluctuations—against the real thing, scientists could rewind time and ask why some places in the universe are rife with cosmic activity while others are barren.

Masato Shirasaki, an assistant professor at the National Astronomical Observatory of Japan, says that question would be almost impossible to answer without these simulations. The project requires a huge amount of data storage (about 10 terabytes, equivalent to 22,000 episodes of *Game of Thrones*).

Shirasaki's team developed a model of how the universe is thought to have evolved and applied it to each of the simulations to see which result may be closest to how it looks today. This method made it easier to explore the physics of cosmic inflation.

In the next few years, Shirasaki's methods could help shorten the observation time needed for future efforts like SPHEREx, a two-year mission slated for 2024 involving a spacecraft that will orbit Earth and gaze at nearly 300 million galaxies across the sky. With these leaps in computing, our understanding of the universe is expanding, bit by bit. ■

The project requires a huge amount of data storage (about 10 TERABYTES, equivalent to 22,000 episodes of *Game of Thrones*)

IMPOSSIBLE INSTRUMENTS

Researchers wanted to build hyper-realistic digital instruments. Musicians had other plans.

When Gadi Sassoon met Michele Ducceschi backstage at a rock concert in Milan in 2016, the idea of making music with mile-long trumpets blown by dragon fire, or guitars strummed by needle-thin alien fingers, wasn't yet on his mind. At the time, Sassoon was simply blown away by the everyday sounds of the classical instruments that Ducceschi and his colleagues were re-creating.

"When I first heard it, I couldn't believe the realism. I could not believe that these sounds were made by a computer," says Sassoon, a musician and composer based in Italy. "This was completely groundbreaking, next-level stuff."

What Sassoon had heard were the early results of a curious project at the University of Edinburgh in Scotland, where Ducceschi was a researcher at the time. The Next Generation Sound Synthesis, or NESS, team had pulled together mathematicians, physicists, and computer scientists to produce the most lifelike digital music ever created, by running hyper-realistic simulations of trumpets, guitars, violins, and more on a supercomputer.

Sassoon, who works with both orchestral and digital music, "trying to smash the two together," was hooked. He became a resident composer with NESS, traveling back and forth between Milan and Edinburgh for the next few years.

It was a steep learning curve. "I would say the first year was spent just learning. They were very patient with me," says Sassoon. But it paid off. At the end of 2020, Sassoon released *Multiverse*, an album created using sounds he came up with during many long nights hacking away in the university lab.

Computers have been making music for as long as there have been computers. "It predates graphics," says Stefan Bilbao, lead researcher on the NESS project. "So it was really the first type of artistic activity to happen with a computer."

But to well-tuned ears like Sassoon's, there has always been a gulf between sounds generated by a computer and those made by acoustic instruments in physical space. One way to bridge that gap is to re-create the physics, simulating the vibrations produced by real materials.

The NESS team didn't sample any actual instruments. Instead they developed software that simulated the precise physical properties of virtual instruments, tracking things like the changing air pressure in a trumpet as the air moves through tubes of different diameters and lengths, the precise movement of plucked guitar strings, or the friction of a bow on a violin. They even simulated the air pressure inside the virtual room in which the virtual instruments were played, down to the square centimeter.

Tackling the problem this way let them capture nuances that other approaches miss. For example, they could re-create the sound of brass instruments played with their valves held down only part of the way, which is a technique jazz musicians use to get a particular sound. "You get a huge variety of weird stuff coming out that would be pretty much impossible to nail otherwise," says Bilbao.

Sassoon was one of 10 musicians who were invited to try out what the NESS team was building. It didn't take long for them to start tinkering with the code to stretch

Will Douglas Heaven is a senior editor for AI at MIT Technology Review.

> *One downside is that fewer people will learn to play physical instruments. On the other hand, computers could start to sound more like real musicians—or something different altogether.*

the boundaries of what was possible: trumpets that required multiple hands to play, drum kits with 300 interconnected parts.

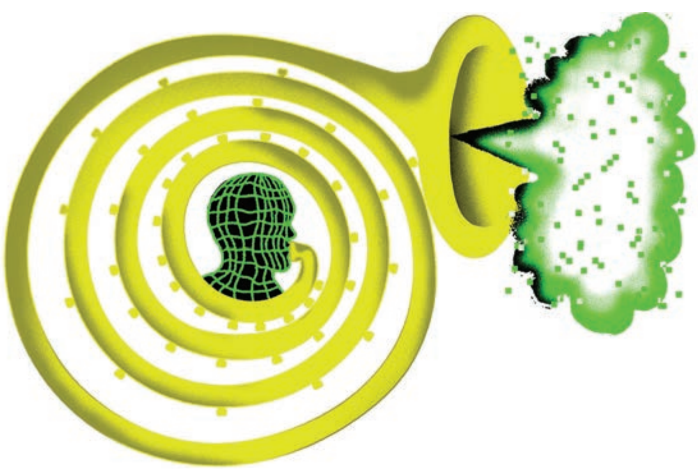
At first the NESS team was taken aback, says Sassoon. They had spent years making the most realistic virtual instruments ever, and these musicians weren't even using them properly. The results often sounded terrible, says Bilbao.

Sassoon had as much fun as anyone, coding up a mile-long trumpet into which he forced massive volumes of air heated to 1,000 °K—a.k.a. "dragon fire." He used this instrument on *Multiverse*, but Sassoon soon became more interested in more subtle impossibilities.

By tweaking variables in the simulation, he was able to change the physical rules governing energy loss, creating conditions that don't exist in our universe. Playing a guitar in this alien world, barely touching the fretboard with needle-tip fingers, he could make the strings vibrate without losing energy. "You get these harmonics that just fizzle forever," he says.

The software developed by NESS continues to improve. Their algorithms have sped up with the help of the university's parallel computing center, which operates the UK's supercomputer Archer. And Ducceschi, Bilbao, and others have spun off a startup called Physical Audio, which sells plug-ins that can run on laptops.

Sassoon thinks this new generation of digital sound will change the future of music. One downside is that fewer people will learn to play physical instruments, he says. On the other hand, computers could start to sound more like real musicians—or something different altogether. "And that's empowering," he says. "It opens up new kinds of creativity." ■



DANIEL ZENDER

VIRTUAL WORLDS

IMMERSION TRIP

The metaverse might sound like an internet catchphrase, but the concept has potential to change how we think about pain management, grieving, and systemic prejudice.

By **John David N. Dionisio**

The first person to write about the "metaverse" was Neal Stephenson in his 1992 novel *Snow Crash*, but the concept of alternative electronic realms, including the "cyberspace" of William Gibson's 1984 novel *Neuromancer*, was already well established.

In contrast to what we typically think of as the internet, a metaverse is a 3D immersive environment shared by multiple users, in which you can interact with others via avatars. A metaverse can, with the support of the right technology, feel like real life, with all the usual elements of work, play, trade, friendship, love—a world of its own.

Perhaps the best-known prototype metaverse is the online virtual world Second Life, whose very name implies an alternate existence. Other games might also be said to be metaverses in their own right: World of Warcraft, Everquest, Fortnite, Animal Crossing. Each of these offers its own version of an immersive world, although they don't quite have the ability to completely take over your senses. Most users experience these games from the outside looking in: screens front and center, with speakers on the sides. Actions are mediated by a keyboard, mouse, trackpad, or game controller instead of players' hands and feet.

John David N. Dionisio is a professor of computer science in the Seaver College of Science and Engineering at Loyola Marymount University.

Technology is starting to change that. High-density screens, virtual-reality goggles and glasses, surround sound, and spatial audio are putting more genuinely immersive experiences within reach. Cameras are gaining 3D capabilities, and single microphones are giving way to microphone arrays that capture sound with better depth and position. Augmented reality, which overlays virtual objects onto a video feed of the real world, provides a bridge between purely virtual and analog or real experiences. There is progress toward adding a sense of touch, too, in the form of multitouch screens, haptic technologies, control gloves, and other wearables. Wraparound environments like Industrial Light and Magic's Stagecraft are within reach only to certain industries for now but may see wider use as technology follows the typical curve of adoption and commoditization.

The tech giants weigh in

The core ideas of a metaverse can be found most readily in games. But that's likely to change, as evidenced by the way certain tech CEOs are now talking openly about how a metaverse might work for them. Facebook's Mark Zuckerberg and Microsoft's Satya Nadella have already publicly mused about the possibilities.

Zuckerberg uses the term "embodied internet" for his version of the metaverse: he imagines a system that is already much like Facebook's now-familiar communities, photos, videos, and merchandise, but instead of looking at that content, in Zuckerberg's vision you'd feel as if you were inside and surrounded by the content—an experience he presumably aims to deliver with technologies from Facebook-owned Oculus VR.

Nadella, meanwhile, has called Microsoft's Azure cloud services and other offerings a "metaverse stack"—he's used the phrase "digital twin" in reference to a system in which users can engage with data, processes, and each other as richly in virtual form as in reality, only with greater speed and flexibility. Microsoft's Surface and HoloLens technologies would then play the role that Oculus would for Facebook.

JULIA DUFOSSÉ

Let's say these visions come to fruition. Would that be a good thing? Given all the misinformation and loss of privacy produced by the old-fashioned internet, it's easy to be skeptical about what massive technology companies might do with a metaverse. Just like the internet and social media, the metaverse can and will be misused. Deepfake technology can already produce images that are indistinguishable from photographs. People can be misled by much less. How much more powerful might an immersive environment prove to be?

What's it really good for?

Those important caveats aside, though, there is reason to believe a metaverse could actually empower us to do quite a bit.

Metaverses have already been employed to let potential customers experience real estate and merchandise (Ikea furniture, Apple phones and computers)—but those functions are nice-to-haves, not essentials. A truly immersive metaverse could go much further.

For one thing, there's therapeutic potential for the likes of PTSD, anxiety, and pain. Programs for burn victims at the University of Washington, pediatric patients at Children's Hospital Los Angeles, and women in labor at Cedars-Sinai indicate that virtual reality helps mitigate pain in a very real way. These initiatives involve synthetic environments where patients connect alone; a fully realized metaverse, with family and caregivers also "dialing in," could have additional benefits.

Immersive environments can also help people experience things that would otherwise be out of reach. Projects at Rensselaer Polytechnic Institute and Penn State, as an example, have sought to change attitudes toward climate change by letting people viscerally experience the results of irreversible global warming.

Immersion might also help us understand each other. The National Center for Civil and Human Rights in Atlanta has an exhibit where participants experience being the target of racist taunts and threats. With audio alone, this is revelatory; if similar experiences were made available to more people, in a manner that included visuals and haptics, metaverse technologies could be used to advance the cause of diversity, equity, and inclusion by helping people empathize with marginalized groups and understand the effects of systemic prejudice.

The metaverse can and should become newsworthy for reasons other than being some privileged executive's dream. A fully realized metaverse can stand not only as a feat of technological innovation and engineering but also, with the right applications, as a vehicle for good in the real world we all inhabit. ■

DIVERSITY

Fay Cobb Payton is a University Faculty Scholar professor at North Carolina State University.

Lynette Yarger is an associate professor and assistant dean at Pennsylvania State University.

Victor Mbarika is Stallings Distinguished Scholar at East Carolina University.

WHO GETS TO SHAPE THE FUTURE OF COMPUTING?

Tech’s diversity and inclusion problems are preventing people and organizations from reaching their full potential.

By **Fay Cobb Payton, Lynette Yarger,** and **Victor Mbarika**

Last year, in response to Black Lives Matter, many US organizations published diversity statements and made bold claims about fostering social change. As Black scholars in computing, we saw these statements and pledges as reactionary and largely ineffective.

Corporate America pledged $50 billion to address racial justice but allocated only a fraction of a percent of those funds to direct grants, the best way to bring about systemic change. Meanwhile, at least 230 higher-education institutions issued statements within two weeks of George Floyd’s murder. Many mentioned solidarity, equality, and greater inclusion, but only one in 10 included concrete action items to address racial issues.

The track record of these institutions does not engender confidence that they will follow through on whatever promises they did make. There is little accountability, and no way to assess whether these commitments have actually improved the lives and livelihoods of Black people.

Diversity and inclusion (especially of Black people) can improve product development, spur innovation, and spark creativity and entrepreneurship, all of which drive the nation’s economy. Research shows that more diverse teams are more innovative and generate more revenue.

We often hear the path to a technology career described as a pipeline. Most diversity efforts in our field have focused on getting more people from diverse backgrounds into this pipeline. And yet representation remains stubbornly low. Between 2014 and 2020, the proportion of Black and Hispanic tech professionals at Facebook increased by less than two percentage points.

Why? The pipeline metaphor ignores the realities of racism, classism, and sexism faced by those historically excluded from technology careers. Individuals who leak out are often deemed deficient. This kind of thinking screams: “‘Fix’ the people and not the system.”

Enter the “pathway” model, an alternative to the pipeline metaphor. Pathway advocates try to create multiple entry points that can lead someone to a technology career. The idea is that people will flow in from other fields, such as engineering, the arts, mathematics, and even the humanities. One way to promote this flow is for two-year and four-year schools to make it easy for people to start in one program and finish in a different one.

Even when pathways provide more entry points, getting through remains challenging, particularly for minorities in America. One still has to be familiar with the opportunities for academic success and career readiness—and aware of the barriers that can stand in the way. Those vary between schools, and even between departments within the same school. And students must also be able to apply that knowledge to navigate antiquated processes and complex power structures.

The question is, what would be better? We advocate for an ecosystem approach in which many organizations work together to address the lack of representation in tech. The tech ecosystem should involve K–12 schools, higher-education institutions, companies, nonprofits, government agencies, and venture capitalists. Public-private partnerships could help design environments that would be inclusive from

COURTESY PHOTOS

the time people start their education to the day they finish their careers.

This might require us to rebuild systems like gateway mathematics courses (classes such as pre-calculus that students must pass in order to continue their program of study) and registration holds (which prevent a student from registering for classes until tuition and fees are fully paid). These systems slow student progress and perpetuate disparate outcomes.

Universities and technology companies could provide professional development opportunities for students from underrepresented groups. But these organizations would have to first change their own cultures to be more inclusive. That means reimagining recruitment practices, which typically rely on professional networks and result in a homogeneous pool of applicants, and addressing sources of algorithmic bias, such as automated résumé screeners that select candidates from particular schools and avoid those with ethnic-sounding names.

Organizations and fields of study that adopt this approach will foster excellence, innovation, and creativity. Georgia State University is a good model. The university has eliminated achievement gaps by introducing meta-majors that students select when they enroll. A biology major who chooses a meta-major like STEM takes classes together with students who are pursuing careers in other STEM fields, like medicine or math. Today, African-American and Hispanic students at Georgia State graduate at the same rate as white students.

Ecosystems depend on both universities and companies to go beyond diversity statements. What we need is sustainable, intentional change. Donating money to a cause can help, but it must be paired with policies that can make technology more equitable.

Most important, we must hold today's leaders accountable by implementing policies and procedures that emphasize transparency, compliance, and enforcement. The best way to fix systems that benefit some and exclude others is to address the underlying structures, not just the people. ■

Setting an example

A computer science educator on why she s talking about identity in the classroom.

By **Melba Newsome**

COURTESY PHOTO

Nicki Washington could be mistaken for a social scientist. Washington has long argued that computer scientists like her should better understand how their own identities affect their work. She joined Duke University in June 2020 and launched a groundbreaking course that analyzes how race, gender, and class influence the way technologies get developed.

We spoke about computing's challenges around diversity, equity, and inclusion (DEI) and how she found her way in the field. This interview has been condensed and lightly edited for clarity.

Q: The students in Duke s computer science department are overwhelmingly white or Asian men. What made you decide to discuss race, gender, and class in your course?

A: Being well aware of the lack of diversity in tech, I always worked to get more Black and brown students into computing early. After a while, I realized that it no longer made sense, because they are not going to stay if they re experiencing racism. We have to change the mindset of a workforce that is overwhelmingly white, Asian, and male and get them to recognize that new perspectives lead to more innovation.

Q: You say computer science needs a stronger dose of social sciences. Why?

A: The problems in technology don t begin with technology; they begin with the environment where people are learning and working. Some disciplines, like health care, teach cultural competence because they work with clients and patients from a range of backgrounds and identities. Why are we not doing this in computing when the technology we re developing impacts the same people in ways that are equally as harmful?

Q: Why haven t diversity efforts been more successful?

A: Because they re focused on increasing numbers and representation. But once you have more minoritized people, then what? If you haven t changed the environment, you ll lose them as quickly as you get them.

From college to industry, every effort has been focused on the deficit of people with the most marginalized or minority identities. They are not the issue; the issue is everyone who marginalizes them.

Q: What can those in power do?

A: People in positions of power have to recognize that they re not the most knowledgeable about these things. So, first and foremost, listen to Black women! Black women have been telling us on so many levels what the issues are. Second, allow yourself to be as uncomfortable as possible and sit with that discomfort, which means unlearning and yielding space.

Q: How did having a mom who was a computer scientist affect your career decisions?

A: When Mom graduated college in 1973, she was marginalized, ignored, and had to deal with problematic managers. She also had a small group of friends who graduated from Black colleges and started at IBM at the same time. So when people talk about how representation matters, it really did for me. It was normal for me to see programmers, engineers, and managers who looked like me. ■

SILICON AND ON AND ON ...

WHAT EVER HAPPENED TO

DNA COMPUTING?

Silicon has bested every potential replacement. But let's not count them all out just yet.

By **Lakshmi Chandrasekaran**

When the first transistor was created, in 1947, few could have imagined the eventual impact of this device—the switch that lies at the heart of logic chips.

We have silicon to thank for computing's great takeover. Add a minute pinch of impurities to the element, and silicon forms a material almost ideal for transistors in computer chips.

For more than five decades, engineers have shrunk silicon-based transistors over and over again, creating progressively smaller, faster, and more energy-efficient computers in the process. But the long technological winning streak—and the miniaturization that has enabled it —can't last forever. "There is a need for technology to beat silicon, because we are reaching tremendous limitations on it," says Nicholas Malaya, a computational scientist at AMD in California.

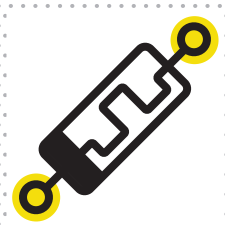
What could this successor technology be? There has been no shortage of alternative computing approaches proposed over the last 50 years. Here are five of the more memorable ones. All had plenty of hype, only to be trounced by silicon. But perhaps there is hope for them yet.

Spintronics

Computer chips are built around strategies to control the flow of electrons—more specifically, their charge. In addition to charge, however, electrons also have angular momentum, or spin, which can be manipulated with magnetic fields. Spintronics emerged in the 1980s, with the idea that spin can be used to represent bits: one direction could represent *1* and the other *0*.

In theory, spintronic transistors can be made small, allowing for densely packed chips. But in practice it has been tough to find the right substances to construct them. Researchers say that a lot of basic materials science still needs to be worked out.

Nevertheless, spintronic technologies have been commercialized in a few very specific areas, says Gregory Fuchs, an applied physicist at Cornell University in Ithaca, New York. So far, the biggest success for spintronics has been nonvolatile memory, the sort that prevents data loss in the case of power failure. STT-RAM (for "spin transfer torque random access memory") has been in production since 2012 and can be found in cloud storage facilities.



Memristors

Classic electronics is based on three components: capacitor, resistor, and inductor. In 1971, the electrical engineer Leon Chua theorized a fourth component he called the memristor, for "memory resistor." In 2008, researchers at Hewlett-Packard developed the first practical memristor, using titanium dioxide.

It was exciting because memristors can in theory be used for both memory and logic. The devices "remember" the last applied voltage, so they hold onto information even if powered down. They also differ from ordinary resistors in that their resistance can change depending on the amount of voltage applied. Such modulation can be used to perform logic operations. If done within a computer's memory, those operations can cut down on how much data needs to be shuttled between memory and processor.

Memristors made their commercial debut as nonvolatile storage, called RRAM or ReRAM, for "resistive random access memory." But the field is still moving forward. In 2019, researchers developed a 5,832-memristor chip that can be used for artificial intelligence.

Carbon nanotubes

Carbon isn't an ideal semiconductor. But under the right conditions it can be made to form nanotubes that are excellent ones. Carbon nanotubes were first crafted into transistors in the early 2000s, and studies showed they could be 10 times more energy efficient than silicon.

In fact, of the five alternative transistors discussed here, carbon nanotubes may be the farthest along. In 2013, Stanford researchers built the world's first functional computer powered entirely by carbon nanotube transistors, albeit a simple one.

But carbon nanotubes tend to roll into little balls and clump together like spaghetti. What's more, most conventional synthesis methods make semiconducting and metallic nanotubes in a messy mix. Material scientists and engineers have been researching ways to correct and work around these imperfections. In 2019, MIT researchers used improved techniques to make a 16-bit microprocessor with more than 14,000 carbon nanotube transistors. That's still far from a silicon chip with millions or billions of transistors, but it's progress nonetheless.

DNA computing

In 1994, Leonard Adleman, a computer scientist at the University of Southern California in Los Angeles, made a computer out of a soup of DNA. He showed that DNA could self-assemble in a test tube to explore all possible paths in the famous "traveling salesman" problem. Experts predicted DNA computing would beat silicon-based technology, particularly with massively parallel computing. Decades later, researchers concluded that DNA computing isn't fast enough to do that.

But DNA holds some advantages. Researchers have shown that it's possible to encode poetry, GIFs, and digital movies into the molecules. The potential density is staggering. All of the world's digital data could be stored in a coffee mug full of DNA, biological engineers at MIT estimated in a paper earlier this year. The catch is cost: one coauthor later said that DNA synthesis would need to be six orders of magnitude cheaper to compete with magnetic tape.

Unless researchers can cut the cost of DNA storage, the stuff of life will stay stuck in cells.

Molecular electronics

It's a compelling vision: transistors keep getting smaller and smaller, so why not jump ahead and make them out of individual molecules? Nanometer-scale switches would make for a supremely cost-effective, densely packed chip. The chips might even be able to assemble themselves thanks to interactions between molecules.

Groups at Hewlett-Packard and elsewhere in the early 2000s raced to make the chemistry and electronics work together.

But after decades of work, the dream of molecular electronics is still just that. Researchers have found that single molecules can be finicky, working as transistors under only very narrow conditions. "No one has shown how single-molecule devices can be reliably integrated into massively parallel microelectronics," says Richard McCreery, a chemist at the University of Alberta.

The dream of molecular electronics has not completely died, but these days it is largely relegated to the chemistry and physics labs, where researchers continue struggling to make endlessly fickle molecules behave.

What comes next?

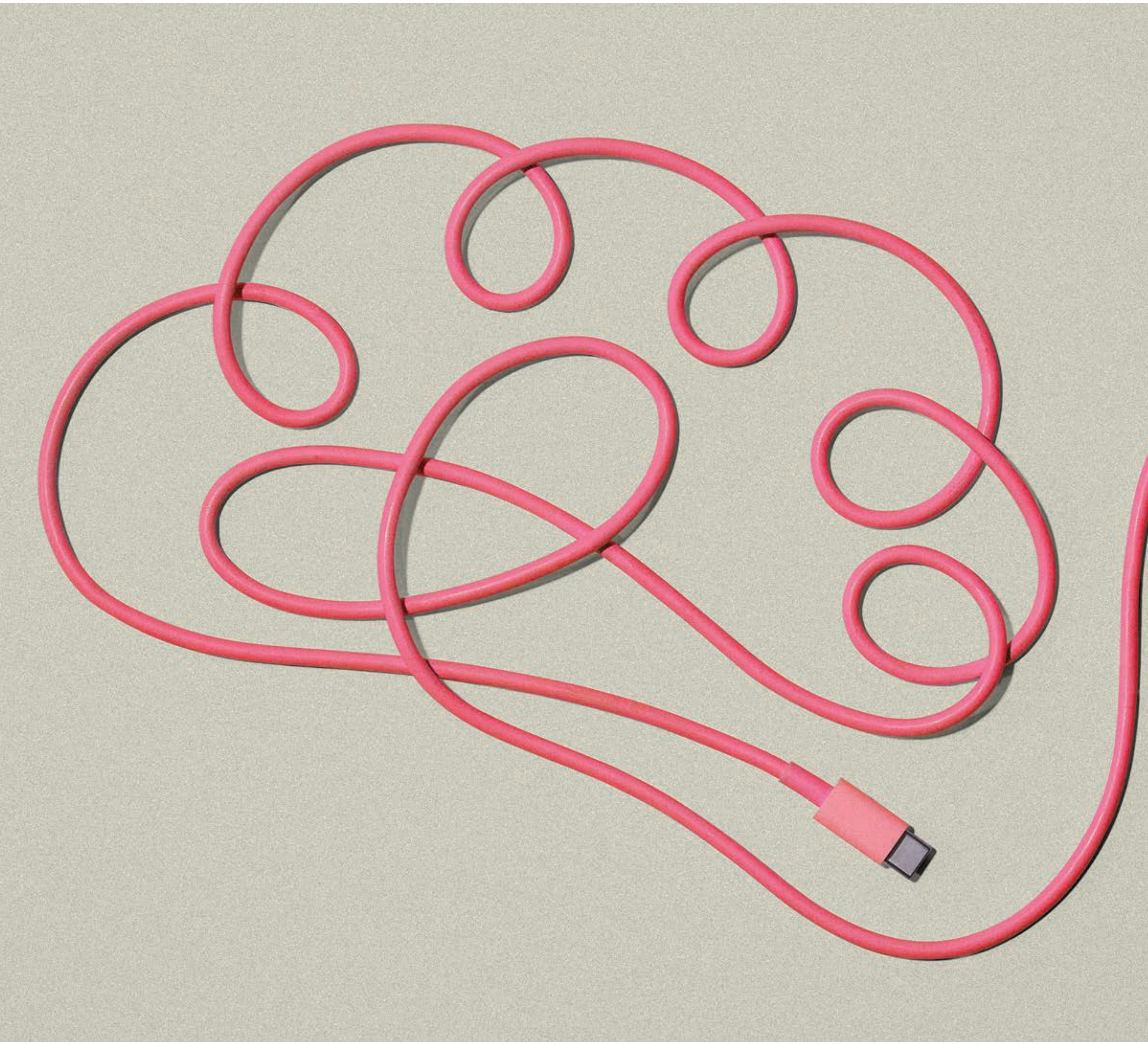
Silicon still reigns supreme, but time is running out for everyone's favorite semiconductor. The latest International Roadmap for Devices and Systems (IRDS) suggests that transistors are expected to stop shrinking after 2028 and that integrated circuits will need to be stacked in three dimensions to keep making faster and more efficient chips possible.

This might be the time when other computing devices find an opening, but only in conjunction with silicon technology. Researchers are exploring hybrid approaches to making chips. In 2017, researchers who had made progress with carbon nanotube transistors integrated them with layers of nonvolatile memristors and silicon devices—a prototype for an approach to improving speed and energy consumption in computing by moving away from traditional architecture.

Classic silicon-based chips will still make some progress, says AMD's Malaya. But, he adds, "I think the future will be heterogeneous, in which all the technologies are used probably in a complementary way to traditional computing."

In other words, the future will still be silicon. But it will be other things as well. ■

Lakshmi Chandrasekaran is a freelance science writer based in Chicago.



A computer mouse inside your head

PARALYZED PEOPLE ARE USING BRAIN IMPLANTS TO TYPE AND MOVE ROBOTIC ARMS. IS THIS THE NEXT GREAT COMPUTER INTERFACE FOR ALL OF US?

In a 12-by-20-foot room at a skilled-nursing facility in Menlo Park, California, researchers are testing the next evolution of the computer interface inside the soft matter of Dennis Degray’s motor cortex. Degray is paralyzed from the neck down. He was hurt in a freak fall in his yard while taking out the trash and is, he says, “as laid up as a person can be.” He steers his wheelchair by puffing into a tube.

But Degray is a virtuoso at using his brain to control a computer mouse. For the last five years, he has been a participant in BrainGate, a series of clinical trials in which surgeons have inserted silicon probes the size of a baby aspirin into the brains of more than 20 paralyzed people. Using these brain-computer interfaces, researchers can measure the firing of dozens of neurons as people think of moving their arms and hands. And by sending these signals to a computer, the scientists have enabled those with the implants to grasp objects with robot arms and steer planes around in flight simulators.

Degray is the world’s fastest brain typist. He first established the mark four years ago, using his brain signals to roam over a virtual keyboard with a point-and-click cursor. Selecting letters on a screen, he reached a rate of eight correct words

BY ANTONIO REGALADO

ILLUSTRATIONS BY SELMAN DESIGN

in a minute. Then, right before the covid-19 pandemic began, he demolished his own record, using a new technique where he imagined he was handwriting letters on lined paper. With that approach, he managed 18 words per minute.

One of the people responsible for the studies with Degray is Krishna Shenoy, a Stanford University neuroscientist and electrical engineer who is among the leaders of the BrainGate project. While other brain-interface researchers grabbed the limelight with more spectacular demonstrations, Shenoy's group has stayed focused on creating a practical interface that paralyzed patients can use for everyday computer interactions. "We had to persevere in the early days, when people said *Ah, it's cooler to do a robotic arm—it makes a better movie*," says Shenoy. But "if you can click, then you can use Gmail, surf the Web, and play music."

Shenoy says he is developing the technology for people with "the worst afflictions and the most need." Those include patients who are utterly locked in and unable to speak, like those in the end stage of ALS.

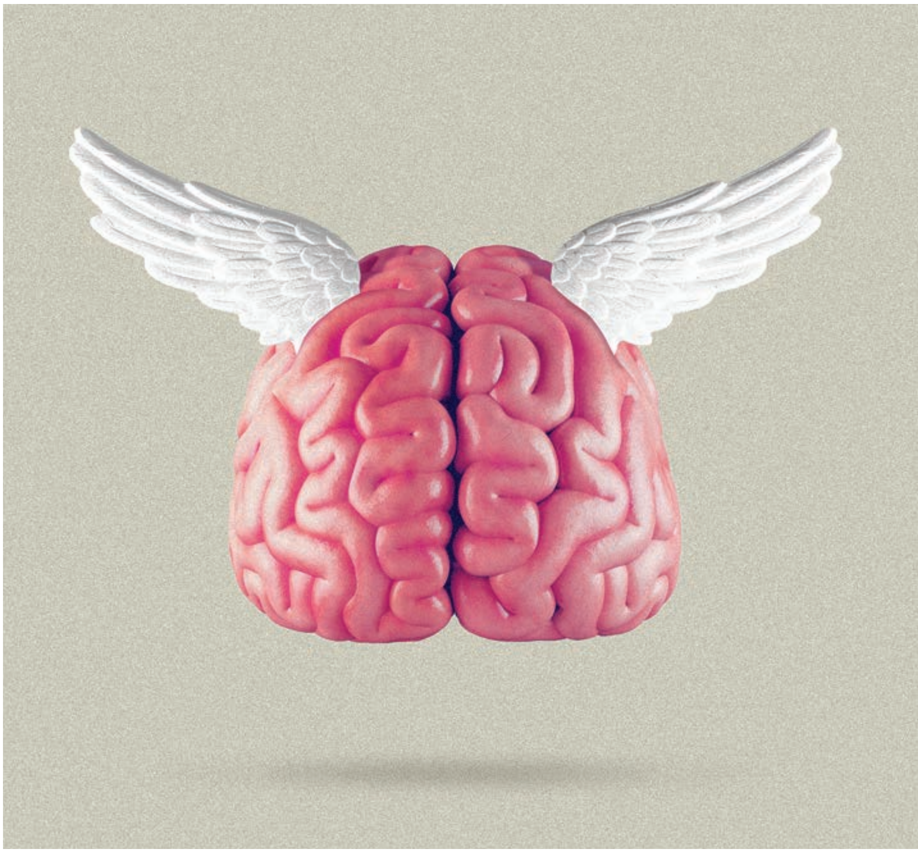
But if the technology allows people like Degray to link their brain directly to a computer, why not extend it to others? In 2016, Elon Musk started a company called Neuralink that began developing a neural "sewing machine" to implant a new type of threaded electrode. Musk said his goal was to establish a high-throughput connection to human brains so that society could keep pace with artificial intelligence.

The same month Neuralink went public with its plans, Facebook announced it would develop a "noninvasive" brain-reading helmet to translate thoughts into social media posts. What's followed has been a huge influx of investment in brain interfaces of all kinds, including EEG readers, magnetic headbands, and new types of high-density implanted probes capable of measuring signals from tens of thousands of neurons at a time.

More than $300 million has been raised by such companies in the last 12 months, even though Facebook this year dropped its quest (it determined a brain-reading helmet won't be a feasible way to send texts for years). "The field was un-investable until Elon entered. That is what sent shock waves through the venture capital world," says Shenoy. "Now there are nearly infinite resources."

The money comes with a catch, though. Medical researchers like Shenoy want to help desperate cases. But entrepreneurs want the next interface for everybody. Musk has said he is aiming for brain implants that would be available to any consumer who wants one—Neuralink even designed a sleek white surgical chair where he imagines people will sit for a routine 30-minute implant procedure.

Shenoy, who is a paid consultant to Neuralink, told me he's living a scientific paradox. He is opposed to consumer brain implants; he worries about everything from their impact on inequality (what if only some people can afford one?) to the consequences of directly linking people's brains to social media. But he has made a Faustian bargain in working with Neuralink, which is bringing much-needed resources to commercializing an interface that—at first, at least—promises huge benefits for paralyzed people.



"It's not comfortable, but welcome to science," says Shenoy. "Anything that is therapeutic and restorative, I am into. Anything that is elective, enhancement—I don't want to work on that. But when the technology is so early, you can't pursue the restorative stuff without being generally aligned with the people who want to take it beyond. We are on the early part of the same path."

Monkey Pong

Neuralink is a secretive company that communicates with the public mostly via theatrical presentations.

PBS NEWSHOUR STILL / NEWSHOUR PRODUCTIONS LLC (OPPOSITE)

The latest, released in April 2021, showed a rhesus monkey named Pager playing the video game Pong with its mind. The demo led to an excited response on social media—as well as a lawsuit by animal rights activists—but mind Pong was not new. A BrainGate subject named Matt Nagle had played the game against a Wired editor in 2005.

Shenoy says he is developing the technology to restore a digital existence to people with "the worst afflictions and the most need."

The real advance made by Neuralink was something not visible in the video—the implant itself. Chip designers at the company have built a soda-cap-size disc, containing processors and a wireless radio, that connects to electrodes stitched into the monkey's cortex. The disc lies flush with the monkey's skull and is covered with skin—giving the implant a more practical footprint than the cables that protrude from Degray's head.

In a blog post, Neuralink said that Pong was just a demonstration—and also articulated for the first time what its implant would be used for, at least in the near term. It read, "Our first goal is to give people with paralysis their digital freedom back: to communicate more easily via text, to follow their curiosity on the web, to express their creativity through photography and art, and, yes, to play video games." A Neuralink engineer later told IEEE Spectrum that the company had the specific aim of beating Degray's brain-communication record.

But Musk's long-term plans are equally clear: he thinks human brains need to be directly connected to phones, computers, and applications. You could run Google searches directly from your brain. Or you can even imagine connecting to someone else's mind, seeing and hearing what the other person is doing.

Musk says all this is part of a strategy to offset existential risks he thinks future artificial intelligence will pose to humankind—like a scenario in which an AI decides to wipe out humanity, Terminator style. His view is that to prevent such an outcome, humans should become cyborgs and merge with AI. "If you can't beat em, join em," Musk typed into Twitter in July 2020, describing the phrase as the "Neuralink mission statement."

Neuralink says its eventual goal is "creating a whole brain interface capable of more closely connecting biological and artificial intelligence." Technologically, achieving that goal means developing a high-bandwidth brain-computer connection that can tap into thousands or millions of neurons at once.

The technology isn't there yet. The system used on Degray measures from around 100 electrodes at once. In general, brain implants use each electrode to listen to one neuron. Neuralink's N1 implant measures from 1,024 electrodes that lie along thin metal threads; that means it's listening to about a thousand neurons. And it has only been tested in monkeys and pigs so far.

When it comes to consumer implants installed via elective brain surgery, regulators, public opinion, and even the medical profession may also stand in the way. In 2016, a Pew Research poll found that 69% of Americans were either very or somewhat worried about the prospect of brain chips that offered an improved ability to concentrate or process information. According to Pew, this opposition was strongly related to a fear of "loss of human control."

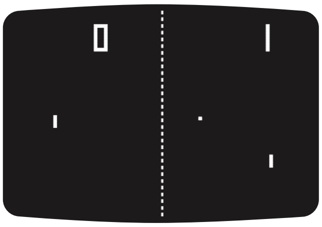
And brain surgeons will still need some convincing before they drill holes in the heads of healthy people. Jaimie Henderson, the Stanford neurosurgeon who put in Degray's implants and co-leads the project with Shenoy, says he thinks small implants done with minimal trauma are "fairly low risk," with the main hazard being a 3% to 5% chance of infection—a risk that may be worth it to improve a severely disabled person's life. The question will be whether healthy people gain enough from an implanted computer mouse to offset the dangers, even if they're small.

"It's unclear to me what benefits able-bodied people would be able to get from any current brain-computer interface system," Henderson says. "Our goal has been to try to restore function for people who have lost it, as best we can—not to provide some sort of 'superhuman' capability."

<--- Dennis Degray uses brain signals to type on a virtual keyboard. He manages 18 words per minute, about half as fast as the average able-bodied person texting from a smartphone.

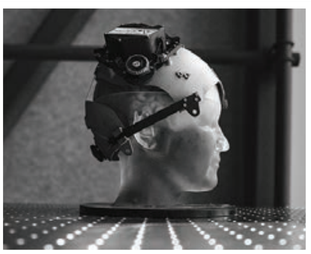
5 milestones in the history of brain-machine interfaces

2005 – Matt Nagle, who is paralyzed, uses a brain-computer interface to play Pong against a Wired editor.



2008 – Early brain-interface company Cyberkinetics ceases operations. Academic experiments continue under the name "BrainGate."

2016 – Elon Musk founds Neuralink with the aim of developing a high-bandwidth brain-machine interface connecting humans to artificial intelligence.



2017 – Facebook says it has assigned 60 engineers to developing a thought-reading helmet. It abandons the project four years later as technically infeasible.

2021 – Implant recipient Dennis Degray's brain-typing record of 18 words a minute is reported in Nature.

Still, Shenoy was one of several academic scientists who told me that, like it or not, they do think consumer brain implants are going to be possible. Enough subjects like Degray have lived with implants for years, with few ill effects, and they're achieving useful mastery of the brain mouse. "Technologically, I see no barrier. I would not have said that 10 years ago and might not have said it five years ago," Shenoy says. "It's basically electrodes, chips, and a radio."

To some, such an interface is intriguing because of the sheer amount of time we now spend on phones, playing video games, listening to podcasts, or scrolling through social media. That's propelling investments in new ways of interfacing with the brain, says Nita Farahany, a law professor at Duke University who is writing a book on consumer neurotechnology.

"The question of why seemingly disparate companies are investing is that if you could use your brain as the controller, instead of a mouse or joystick, it's not so crazy to want to invest," says Farahany. "This may be the next revolution in the computer interface."

Nathan Copeland is another paralyzed person living with a brain implant—he's part of a study in Pittsburgh. Last year he became the first to plug his head into a tablet computer at home, on his own time, not as part of scientific session (it normally takes a small team of medical workers in a clinical setting). Copeland told me at first he was using the device eight hours a day, playing video games and using drawing programs. He later tired of it—his tablet is a medical device that uses an older version of Windows, and its battery doesn't last long.

Still, Copeland told me he believes paralyzed people are "test pilots" for future consumer brain interfaces. In his own case, he says, he's mostly interested in being able to play more video games—one of his favorite pastimes—at a higher level.

Game changer

Of the 35 or so people who have received a long-term brain implant to interface with a computer, 29 of them, including Degray, have electrode implants built by a company called Blackrock Neurotech, based in Salt Lake City. The implant, aptly called the Utah array, is a silicon square with 100 small needles, which is pushed into the surface of the brain. Blackrock mostly sells systems to researchers experimenting on animals, but as investors have flocked to implants, observers have sometimes called Blackrock and Neuralink the Lyft and Uber of brain interfaces.

The president of Blackrock, an electrical engineer named Florian Solzbacher, thinks the timing is right to take implants forward for paralyzed people. "People would say *Oh my God, it's brain surgery*, but actually we haven't seen any problems," he says. Every time there is a video of someone controlling a robot or eating a Twinkie with a robotic hand, Solzbacher says, he gets calls from paralyzed people wondering when a commercial product could be available. It's only recently that he's been willing to say it could happen soon: "It's always been 15 years away, and now what I can say for the first time is soon you will be able to take it home."

That's due to several factors, including the development of a wireless version of the BrainGate hardware. Instead of cables, subjects have a hockey-puck-size wireless transmitter screwed onto their brain ports. It's nothing as compact and sleek as Neuralink's electronics, but it works. Solzbacher says his company plans to seek approval to sell its own improved wireless system to people with ALS or severe paralysis.

Solzbacher says Degray's typing points to the potential of the technology—he can tap out words much faster than anyone using an EEG headband, for example. "That means you are 10 times faster than anything that is out there," he says. "Now you can start being productive, and you have performance close to an able-bodied person."

However, Solzbacher is being financed by people who aren't only interested in helping paralyzed people. This year his company raised $10 million from investors including the German billionaire Christian Angermayer, who invests widely in psychedelics, longevity treatments, and mental health. In a tweet, Angermayer

WIKIMEDIA COMMONS; FACEBOOK; STANFORD HEALTH CARE/SCHOOL OF MEDICINE (KEYBOARD) left no doubt he thinks a general-purpose brain mouse is the ultimate goal: "It's fundamentally an input-output device for the brain, and it can benefit ALL. We can unlock truly astonishing use cases & I believe Blackrock will be the one to take us there. Ppl will communicate with each other, get work done + even create artwork, directly with their minds."

The "mixed reality" experiments done in virtual space hint at how able-bodied people might experience computer worlds through brain interfaces.

Solzbacher says for now, none of Blackrock's internal plans or projections involve consumer brain implants. Still, he acknowledges that could happen: "I expect there is part of society that may really want it, even though there is nothing wrong with them."

I asked Solzbacher whether any able-bodied person had ever requested such a device. He says he's hasn't gotten such a request, yet.

Mixed reality

Robert "Buz" Chmielewski had his head down in concentration, and because of a screen, he couldn't see which of two toy-size soccer balls had been placed in the robotic hand he was controlling. Using his thoughts, Chmielewski closed the plastic and metal hand and squeezed the ball. "Pink ball," he called back. When the researcher swapped it for another, stiffer ball, Chmielewski could sense the change. "Black ball," he said.

Chmielewski, 50, got his Utah arrays implanted in 2019, 30 years after a surfing accident in Ocean City, Maryland, left him in a wheelchair. During the two years the experiment lasted (it ended in September), he had more implants put in than any other patient—a total of six, in both hemispheres of his brain. Because of this, he was able to control two robot arms simultaneously. What's more, three of the probes placed into his sensorimotor cortex sent signals back into his brain, allowing him to receive tactile information from the robots.

Chmielewski was part of a project at Johns Hopkins University's Applied Physics Laboratory that's testing new forms of perception. He also tried out the Microsoft HoloLens headset and used his sense of virtual touch to arrange blocks in virtual space. "If you would have told me three years ago I would be controlling things with my thoughts, I would have said you're crazy," Chmielewski said during a recent online presentation. "Some of the applications we are working on have blown my mind."

The researchers at APL include Michael Wolmetz, manager of the Human and Machine Intelligence program. Wolmetz says the demonstrations are a glimpse of "fundamental" changes ahead in human-computer interaction, especially the concept of "mixed reality." The experiments in virtual space hint at how able-bodied people might experience computer worlds through brain interfaces—making the APL project one of the most explicit explorations of how such technology might lead to human enhancement.

"For all of biological history, the only way we've interacted with the environment is with senses and motor function," Wolmetz says. "We have, for the first time, the ability to go outside that paradigm. It's the first time any biological organism has done that."

Wolmetz doesn't know whether surgically implanted brain interfaces will ever be widely used, but he says these devices are a "sneak peek" at how consumers might use future non-invasive systems like brain-reading helmets or headbands, should accurate ones be developed.

When I asked Wolmetz what he thought people might use such interfaces for in the future, he said it is hard to predict. "It's like asking what the computer is going to be for," he says. "I think that in our lifetimes it will be for everything. But in the next five years, it's hard to answer."

Some want not only the computer mouse but the entire interface—including the screen, or whatever replaces a screen—in the brain. One of them is Max Hodak, the former president of Neuralink. He was fired by Musk in March—it's not clear why—but quickly formed a new company, called Science Corp., with financial backing from the cryptocurrency billionaire Jed McCaleb. Hodak says he plans to develop a new type of implant that rests on the retina and can send information to the visual cortex at the back of the brain.

Key players

Blackrock Neurotech
Markets the "Utah array," the implant most often used in brain-computer experiments.

BrainGate
An academic consortium has put implants in the brains of over 20 people.

DARPA
The US defense agency has spent $120 million on brain interfaces in the last five years.

Kernel
Has developed a type of wearable brain scanner that uses infrared beams.

Neuralink
Elon Musk's company raised a further $200 million in 2021.

Paradromics
The startup is developing high-density brain electrodes.

Synchron
The Australian company started a trial of a brain sensor inserted through a blood vessel. It allows simple computer control.

Shenoy says his concern is that putting computer interfaces into people's minds will lead to inequality and the same sorts of information abuses seen on the internet.

Initially Hodak's new company will be looking to help people, like his grandfather, who went blind from retinal diseases. But a medical product is a stalking horse for a bigger ambition, which is to create a device that can produce images in the eyes of healthy people as well.

"It could just be a computer screen that looks as solid as any ever has, and it's just floating in front of you," he says. "When your eyes are open, you would see the world of atoms. When you close your eyes, you see the world of bits." Hodak thinks that in a generation, children will be "baffled when we tell them that there used to be just nothing there when we closed our eyes."

Ethics questions

Before Musk and venture capitalists arrived on the scene, DARPA, an R&D agency of the US Department of Defense, was the world's largest funder of brain-interface research.

Andy Schwartz, a researcher at the University of Pittsburgh, told me he is convinced the military's fascination with the technology springs from a 1982 Clint Eastwood film, *Firefox*, whose plot involves an effort to steal a thought-controlled Soviet MiG jet. After the military had one of his research subjects fly a simulated war jet, Schwartz says, he stopped collaborating with the agency.

John Donoghue, a professor at Brown University and one of the founding scientists of BrainGate, is also concerned about a "circus-like atmosphere" around brain implants. He spent time in a wheelchair as a child, which is one reason he has pursued the goal of restoring movement to paralyzed people. But when he gave a talk at Google a few years ago, an engineer approached him and said he was an avid gamer. The engineer wanted to know if it would be possible to have a third thumb.

"That's taking things to an extreme. I don't want to implant electrodes into people so they can be better gamers," says Donoghue. "I always challenge all of these ideas because I don't see what it gets you. But I don't dismiss it, either ... that is what is driving people. It's the cool factor, that you could have this new interface."

Donoghue doubts that implants will provide superpowers, or that you'll be able to download French for Beginners directly into your head anytime soon. The brain has evolved to receive and send information at the speed that it does—not at the rate of an Ethernet cable. "Have you listened to a podcast at 4x speed? It doesn't work very well," he says. "Our brains are made to make and intake speech at a level that we can use it."

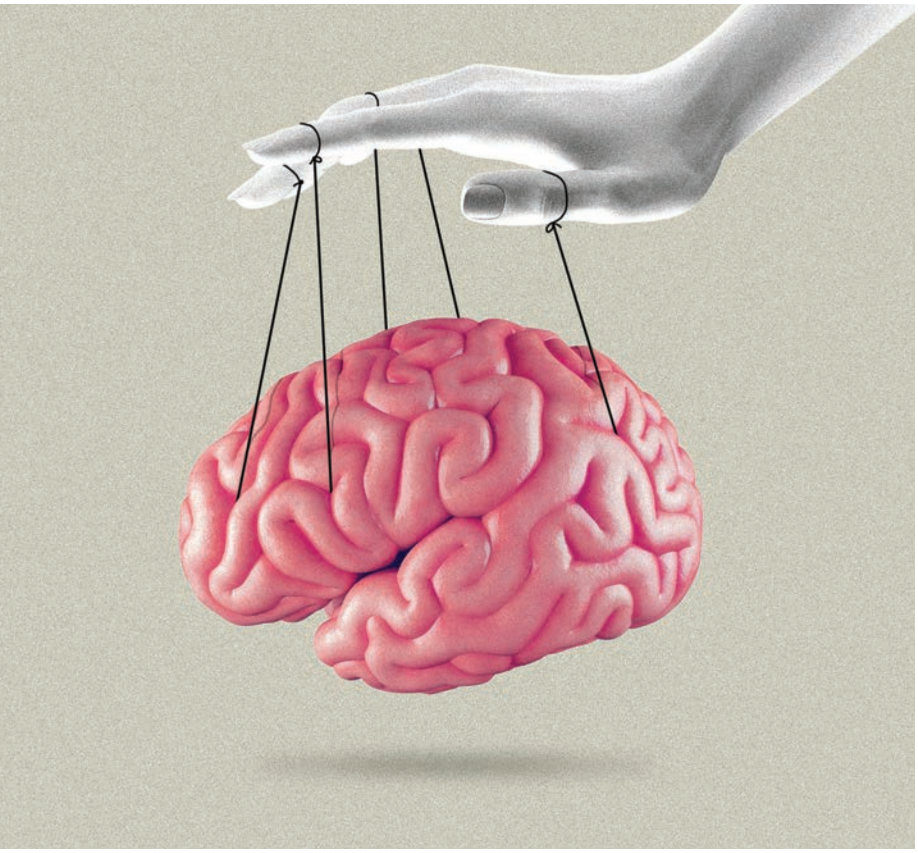
But others believe that mind reading and mind control are rising dangers. In 2017, the same year that both Neuralink and Facebook's brain-interface plans were unveiled, a group of researchers calling itself the Morningside Group published a manifesto in the journal Nature. It sounded alarm bells about a "convergence" between brain technology and AI advances.

The group formed at the urging of Rafael Yuste, a neuroscientist at Columbia University, who became alarmed over experiments in his own lab, in which he could not only read from the visual center of a mouse brain but also use a laser to make the animal perceive things that were not there. "We had control over the visual perceptions of the mice, and we could run them like puppets," says Yuste.

Yuste keeps a list of experiments he thinks point to how neurotechnology could compromise human autonomy. For instance, there's the work of Jack Gallant, in California, who has used MRI scanners to deduce what images people are seeing. Then there's the scientist who wired one monkey's brain to control the arm of different monkey, calling one the "master" and the other its "avatar."

The fundamental fear is that everything bad about the internet—disinformation, malicious hackers, government control, corporate manipulation, endless harassment—could get much worse if technology were to breach what the Morningside Group calls "the last frontier of privacy" and know our thoughts. "There is a huge problem, and it's the problem of mental privacy," says Yuste.

<--- Robert "Buz" Chmielewski, 50, had implants in both hemispheres of his brain. When they were in place, he was able to control two robot arms simultaneously.



In May, Yuste hosted a day-long online gathering of ethicists and neurotech entrepreneurs to discuss responsible neural-interface design.

Several participants said they believed there was a need to establish rules before it becomes possible to collect brain information easily. “We don’t want to go through this cycle of big corporations harvesting data to profit from and then, in the end, facing regulations and asking for forgiveness,” Ryan Field, CTO of Kernel, which is developing a noninvasive headset to read brain activity, said during the event.

Yuste wants far stricter privacy rules than those governing internet data or what’s on your iPhone. He would like to see brain data treated like transplant organs—carefully tracked and with a ban on any profit-making. At a minimum, he says, brain data should be protected like medical information. He also says the military should be forbidden from employing brain implants.

“I have to change me”

In certain ways, the field of brain-computer interfaces is already beginning to realize its loftiest goal and some people’s greatest fear: the merger of humans and AI.

That’s certainly the case with research volunteers like Degray. The buzzing of his neurons is interpreted by AI software called a recurrent neural network. Each day that Degray uses his implant, he starts by imagining some simple movements, like drawing a circle. The neural network listening to his neurons then calibrates the statistical map that relates each neuron’s activity to the movement. And most brain-computer interfaces won’t only use software to interpret brain signals, but also to improve on them—for example, programs might predict what word someone is trying to spell on the basis of the first few letters.

This results in what Blackrock’s Solzbacher calls “shared agency,” or outputs that are picked partly by a person and partly by a machine. “That is scientifically interesting but is also an ethics question,” he says. “Because who is actually making the decisions when the systems adapt?”

Currently, the closest thing there is to brain-interface experience design is the experiments being carried out with Degray in California. Most recently, the team has been trying to get Degray to try mental touch-typing. If software can track what movements he’s thinking of making with his fingers, that could increase his communication speed even more. The problem is that before his accident, Degray was never more than a hunt-and-peck typist. He now has paper keyboards pasted on the ceiling above his bed so he can practice thinking about typing.

One thing I wanted to know from Degray is what it feels like to operate a computer with his brain. He described what he calls a “meeting of the minds” with the cartful of machines and software reading his thoughts. This was particularly true when he was performing the imagined-handwriting task.

“It’s a very personal interaction. You have to feel for where the movements are in your own body,” he says. “You are trying to write the letters, and it is trying to understand you. I wouldn’t call it a relationship, but it’s close. It’s almost a conversation between the device and myself. Some days it’s a little bit surly at first—it’s hard to wake it up. Of course, the machine is perfectly constant. So I have to change me to get it to work.”

One day Degray imagined writing 5,000 words. He worked so hard at it the researchers had to remind him to breathe. “I just pounded it out,” he says. “Over the course of doing so many words, you can get consistent. You quickly lapse into a pattern that the computer can recognize.” ■

Antonio Regalado is the senior editor for biomedicine at MIT Technology Review.

JOHNS HOPKINS APL (OPPOSITE)

During her time at Microsoft and in academia, Jennifer Chayes has been fighting to use data science and computing to make artificial intelligence more fair and less biased.

From dropping out of school at the age of 15 to becoming the doyen of data science at the University of California, Berkeley, Chayes has had quite the career path. She joined UCLA in 1987 as a tenured professor of mathematics. Ten years later, Microsoft lured her to cofound its interdisciplinary Research Theory Group.

It was in her Microsoft lab in New York City that researchers discovered bias in the company's facial recognition software, showing that the system classified white faces more accurately than it did brown and Black faces. This finding caused the company to turn down a lucrative contract with a police department and start working to remove the bias from such algorithms. The FATE (Fairness, Accountability, Transparency and Ethics in AI) group was created at the lab.

Anil Ananthaswamy asked Chayes, now associate provost of the Division of Computing, Data Science, and Society and dean of the School of Information at Berkeley, how data science is transforming computing and other fields.

Jennifer Chayes

Q: What was it like to transition from academia to industry?

A: That was quite a shock. The VP of research at Microsoft, Dan Ling, called me to try to convince me to go for an interview. I talked to him for about 40 minutes. And I finally said, "Do you really want to know what's bothering me? Microsoft is a bunch of adolescent boys, and I don't want to spend my life with a bunch of adolescent boys."

Q: How did he react to that?

A: He said, "Oh, no, we are not. Come and meet us." I met some incredible women there when I visited, and I met phenomenally open-minded people who wanted to try things to change the world.

Q: How has data science changed computing?

A: As we've gotten more data, computer science has begun to look outward. I think of data science as a marriage of computing, statistics, ethics, and a domain emphasis or a disciplinary emphasis, be it biomedicine and health, climate and sustainability, or human welfare and social justice, and so on. It is transforming computing.

Q: Is there a difference in how data scientists solve problems?

A: With the advent of all of this data, we have the opportunity to learn from the data without having a theory of why something is happening. Especially in this age of machine learning and deep learning, it enables us to draw conclusions and make predictions without an underlying theory.

Q: Can that cause problems?

A: Some consider it a problem in cases in which you have, [for example], biomedical data. The data very accurately predicts what's going to work and what's not going to work, without an underlying biological mechanism.

Q: Any advantages?

A: What the data has allowed us to do now, in many cases, is to run what an economist would call a counterfactual, where you actually see random variation in the data that allows you to draw conclusions without doing the experiments. That's incredibly useful.

Do I really want to try out different educations on different populations? Or do I want to see [that] there was random variation at some point that will allow me to draw a really good causal inference, and therefore I can base policy on it?

Q: Do you see a problem in how data is being used, especially by big companies?

A: There are myriad problems. It's not only being used by tech corporations. It's being used by insurance companies. It's being used by government platforms, public health platforms, and educational platforms. If you do not explicitly understand what biases can be creeping in, both in the data sets themselves and in the algorithms, you will likely exacerbate bias.

These biases sneak in [when] there isn't much data. And it can also get correlated with other factors. I personally worked on interpreting bios and CVs automatically. We are not allowed to use gender or race. Even if I don't look at [these] protected attributes, there are many things [in the data] that are proxies for gender or race. If you have gone to certain schools, if you grew up in certain neighborhoods, if you played certain sports and you had certain activities, they are correlated [with gender or race].

Q: Do algorithms pick up on these proxies?

A: They exacerbate it. You must explicitly understand this, and you must explicitly prevent it in writing the algorithm.

Q: How can we address such issues?

A: There is this whole area of FATE: fairness, accountability, transparency, and ethics in AI, which is the design of these algorithms and understanding what they are. But there is so much more that we need to do.

Q: And data science helps?

A: This is absolutely data science. There is part of the web called the "manosphere," where a lot of hate is originating. It's kind of hard to trace. But if you use natural-language processing and other tools, you can see where it's coming from. You can also try to build interfaces that allow advocacy groups and others to find this and to help root it out. This goes beyond just being fair. This is turning the tables on the way in which these platforms have been usurped to increase bias and hate and saying, "We are going to use the power of computing and data science to identify and mitigate hate." ■

CHRISTIE HEMM KLOK

>RE-IMAGINING MATERIALS DISCOVERY

Faced with the world's need for better materials to address climate change, Alán Aspuru-Guzik has an audacious vision of digitizing the discovery process.

By Simon Lewsen

Photographs Derek Shapton

Opposite:
Aspuru-Guzik is a leading evangelist for using computer science to transform chemistry.

When Alán Aspuru-Guzik, a Mexico City–born, Toronto-based chemist, looks at climate-change models, his eyes gravitate to the error bars, which show the range of uncertainty surrounding any given prediction. "As scientists," he says, "we have a duty to contemplate worst-case scenarios." If climate change proceeds as expected, humanity might have a couple of decades or so to come up with materials that don't yet exist: molecules that enable us to quickly and cheaply capture carbon, and batteries—made of something other than lithium, a metal that is costly and difficult to mine—to store the global supply of renewable energy.

And what if the situation gets worse than we expected it to? The need for new materials will go from urgent to extremely urgent to dire. Could we quickly come up with the things we need?

Aspuru-Guzik (one of MIT Technology Review's 35 Innovators Under 35 in 2010) has devoted much of his life to versions of this question. Materials discovery—the science of creating and developing useful new substances—often moves at a frustratingly slow pace. The typical trial-and-error approach, whereby scientists produce new molecules and then test each one sequentially for the desired properties, takes an average of two decades, making it too expensive and risky for most companies to pursue.

Aspuru-Guzik's objective—which he shares with a growing number of computer-savvy chemists—is to shrink that interval to a matter of months or years, enabling humanity to quickly amass an arsenal of resources for fighting climate change, like batteries and carbon-capture filters. The goal is to revive the moribund materials industry by incorporating digital simulations, robotics, data science, artificial intelligence, and even quantum computing into the discovery process.

Imagine computer programs that use precise knowledge of molecules' electronic structure to create new designs; imagine robots that make and test these molecules. And imagine the software and robots working together—testing molecules, tweaking designs, and testing again—until they produce a material with the properties we're looking for.

That's the idea, at least. Actually executing it is another matter. The structures of molecules are mind-bogglingly complex, and chemical synthesis is often more art than science, defying efforts to automate the process. But advances in AI, robotics, and computing are bringing new life to the vision.

Aspuru-Guzik cochaired a 2017 workshop in Mexico City where 133 participants—including Nobel Prize–winning scientists and representatives from 17 national governments—came together to focus the global research community on this goal. The conference was a pivotal moment, helping take the field of accelerated materials discovery from a niche area of inquiry to a worldwide priority for many of those attendees. After the event, Canada, India, and the EU, among others, began investing in initiatives to speed up material research.

The work itself is ambitious and technically difficult because it spans so many disciplines. But as a chemist, software engineer, AI pioneer, quantum computer programmer, robotics enthusiast, and serial entrepreneur, Aspuru-Guzik just may have the right mix of computational expertise and imagination to connect the multiple tools essential to making it happen. He has emerged as one of the more convincing evangelists for the new way of doing chemistry.

"Alán can see beyond what people think is possible," says Joshua Schrier, a Fordham University chemist and frequent collaborator. He is the kind of innovator, says Schrier, who changes the way everybody around him practices science.

For Ryan Babbush, head of the quantum algorithms team at Google, Aspuru-Guzik's most prominent character trait is his creative restlessness. "Alán spends his time and energy on the newest thing, the most uncharted territory," he says. "He doesn't stick around and focus on incremental developments."

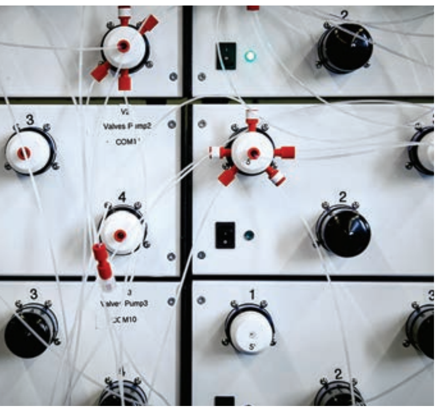
That can be a problem given the time and hard work it takes to bring a new material to market—an undertaking that requires dogged, narrowly focused research and endless business patience. But ultimately, Babbush says, Aspuru-Guzik is interested in reimagining the process of materials discovery, equipping scientists in the community with the computational and automation tools they need to speed up their job.

Today, Aspuru-Guzik is building a lab in Toronto where AI algorithms design novel molecules, and robots quickly make and test them. The lab is a kind of prototype, meant to demonstrate how materials discovery might work in the future. "I want to enable a whole new era, the age of materials on demand, where every lab can easily create new compounds," he says. In the future, he hopes, we'll be better positioned to address the next global crisis. "The problems of the world require molecules," he adds. "And right now, we suck at making them."

Opposite:
Aspuru-Guzik's passions (from top left) range from stickers for street art to lab robotics to Mexican lucha libre masks to automated fluid handling.

Battle scars

Aspuru-Guzik speaks exuberantly, digressively, and very quickly. When I first visited his office at the University of Toronto, he showed me a collection of lucha libre (Mexican wrestling) masks—bright blue, green, and pink balaclavas adorned with Aztec patterns. “The masks are a humanization tool,” he says. “You bring a Nobel Prize winner or an executive from Hitachi into your office, and after talking for a while, it’s good to stop and say, ‘Pick a mask. Take a selfie.’” It’s hard not to view the masks as a metaphor for his multifaceted life.



Aspuru-Guzik grew up in a half Catholic, half Jewish family of writers, musicians, and architects. As a 19-year-old chemistry student at the National Autonomous University of Mexico, he was returning from an overnight rave in the city of Cuernavaca when the car he was riding in veered off the road and crashed. Surgeons had to open his belly to repair his intestines and cauterize the ruptured blood vessels, leaving him with a scar that runs, like a median line, down the center of his abdomen.

After this early brush with mortality, he became committed to a life of intellectual adventurousness. If a field of inquiry intrigued him, he’d pursue it, even if it was esoteric or beyond his expertise.

At the time, there was great excitement over the possibility of using computer-based modeling to design molecules with desired properties, forgoing slow and tedious experiments. Scientists talked about a new era of virtual chemistry, only it didn’t work very well. Computers were too slow and molecules too complex.

While browsing journals in the university library, Aspuru-Guzik came across a paper about the challenges of doing molecular chemistry inside a computer. In 1926, the physicist Erwin Schrödinger had published an equation to predict the behavior of subatomic particles, like electrons and protons. If you can mathematically model a molecule at the subatomic level, you can begin to make inferences about the resulting material: how it combines with other materials, how hard or soft it is, or how quickly it decomposes. At least that’s the idea. But for most materials the Schrödinger equation becomes too complicated for even today’s largest supercomputer.

To make the math doable, Aspuru-Guzik set about creating versions of the equation that require fewer approximations, making them more accurate—a project that became the focus of his doctoral studies at the University of California, Berkeley. The goal was to streamline the calculations to the point where a computer could handle them but not so far that the model became scientifically useless. Using Aspuru-Guzik’s algorithms, a researcher could model—that is, simulate—a random molecule and immediately make predictions about the properties of the resulting substance.

Other scientists had designed similar algorithms, but the ones Aspuru-Guzik came up with as a grad student were impressive enough to get him a job at Harvard when he finished as a postdoc at Berkeley. As an assistant professor at Harvard—and as director of the Aspuru-Guzik research group, a 40-person team of computer scientists, biologists, engineers, physicists, and chemists—he threw himself into an initiative called the Harvard Clean Energy Project. Most solar panels use silicon to transform sunlight into electricity. But were there cheap, easy-to-make organic substances that could do the job?

Over six years, Aspuru-Guzik and his team ran simulations of 2.3 million different organic molecules to see which might have photovoltaic properties. He was hardly the first researcher to practice virtual chemistry, but he was doing it at an unprecedented scale. The increased computing capacity of the era meant that a single molecule could be simulated in a matter of minutes; in the 1990s, such simulations had taken days. Most important, he had access to seemingly limitless server space, much of it borrowed from other people’s devices. In a system akin to the old SETI@Home program, people who wanted to support the project could download a screen saver that would temporarily lend their hard drive to Aspuru-Guzik and his team. “We had one of the biggest supercomputers in the world,” he says, “but it was distributed all over the planet.”

In the end, Aspuru-Guzik discovered many organic materials that could, theoretically, be used for photovoltaic cells. The problem was that these winning molecules were too complicated to be manufactured cheaply. “My mistake,” he says, “was not consulting with organic chemists at the beginning to find out which molecules were easily makeable.”

With the Clean Energy Project, Aspuru-Guzik had basically been doing combinatorial chemistry—the old trial-and-error approach—inside computers instead of inside a lab. Then, beginning in 2012, researchers in Toronto and elsewhere made a series of breakthroughs in deep learning and other methods of machine learning. Like many chemists looking for new materials, Aspuru-Guzik transitioned to AI, which enabled him to discover molecules in a faster, more deliberate way. “The computer simulations are like a machine gun shooting randomly in the air in the hopes of getting a hit,” he says. “AI is a sniper. It chooses a target and takes aim.”

First, he had to train a neural network by feeding it a data set describing the molecular composition and chemical properties of 100,000 organic substances. The AI program could start recognizing patterns—that is, correlations between a given molecule and the substance it forms. It could then use this knowledge to invent candidate molecules to be synthesized and tested in the lab. With the help of AI, Aspuru-Guzik discovered new organic light-emitting diodes, or OLEDs, that were brighter than typical LEDs. He also identified new chemicals to be used in future organic flow batteries, massive industrial batteries that won’t require metals like lithium.

Meanwhile, he threw himself into the nascent field of quantum computing. The Schrödinger equation is hard to run on classical computers precisely because electrons and protons don’t obey the laws of classical physics. They operate, instead, according to quantum mechanics: they can be entangled (behaving in concert with one another, even if they aren’t connected), and they can exist in so-called superposition (occupying multiple opposing states at the same time). The math required to model these complex phenomena is dizzyingly complex, too. But because the qubits in quantum computers also obey the laws of quantum mechanics, the devices are better suited, at least in theory, to simulating molecules.

In practice, though, somebody had to figure out how to make the simulations work. In 2014, Aspuru-Guzik and a team of researchers released the Variational Quantum Eigensolver (VQE), a program to model molecules, albeit on small, error-prone quantum devices that, unlike all-purpose quantum computers, actually exist today. While the Schrödinger equation is a kind of abstraction—a mathematical formula meant to describe subatomic particles—the VQE uses quantum bits to mimic the behavior of the particles in a molecule, much as players in a reenactment might perform the Battle of Gettysburg.

In time, as companies develop more powerful quantum computers, the VQE could enable chemists to run strikingly accurate simulations. These models might be so precise that scientists won’t need to synthesize and test the materials at all. “If we ever reach this point,” Aspuru-Guzik says, “my work in materials science will be done.”

When Donald Trump was elected president of the United States in 2016, Aspuru-Guzik’s career was flourishing, but suddenly the prospect of remaining in the country no longer appealed to him. One week after the election, he began emailing colleagues in Australia and Canada, looking for a new job.

The University of Toronto offered him a prestigious government-funded position meant to lure top-tier researchers to the country and a cross-appointment at the Vector Institute for Artificial Intelligence, a nonprofit corporation cofounded by machine-learning pioneer Geoffrey Hinton that is quickly making Toronto a global hub for AI. The biggest inducement, however, was a promise to build a radical new materials lab called the Matter Lab, a project Aspuru-Guzik had dreamed of for years.

“Fuck it”

“In the Matter Lab, we only attack a problem after asking three questions,” says Aspuru-Guzik. “Does it matter for the world? If not, then fuck it. Has somebody else already done it? If the answer is yes, there’s no point. And is it remotely possible?" Here, the word "remotely" is key. Aspuru-Guzik wants to tackle challenges that are within the range of feasibility, but barely so. "If a material is too easy," he says, "let other people find it."

Located in a postwar brick building in downtown Toronto, the lab is unlike any other at the university. The ceiling is adorned with maroon and burgundy acoustic panels, an homage to the beloved Mexican architect Luis Barragán. Tucked away in an inconspicuous corner is a typical lab bench—a table with flasks, scales, and beakers beneath a fume hood—where graduate students can practice chemistry in much the same way their grandparents' generation did. One gets the sense that this workstation isn't often used.

In the center is a $1.5 million robot—a nitrogen-filled glass-and-metal enclosure housing a mechanical arm that moves back and forth along a track. The arm can select powders and liquids from an array of canisters near the sides of the enclosure and deposit the contents, with exacting accuracy, in one of a number of reactors. "The robot is like a tireless lab assistant who mixes chemicals 24/7," says Aspuru-Guzik. It can make 40 compounds in a mere 12 hours.

Opposite:
The new materials lab in Toronto combines conventional chemistry equipment and the latest in automation and AI.

Two additional features make the Matter Lab's experimental setup unique. The first is software that Aspuru-Guzik and his collaborators designed, called ChemOS. It includes an AI system that generates candidate molecules and a program that interfaces with the robot, directing it to synthesize candidates on demand.

The second distinct feature is the "closed loop" nature of the production process. To explain how this works, Aspuru-Guzik points to a pair of narrow hoses at the back of the robot. "That's where the pee-pee comes out," he says. Once a reaction is finished, the resulting liquid runs through the plastic hoses to an analytical machine the size and shape of a mini-fridge, which separates out unwanted by-products. The refined material will flow into another robot that will test it to learn about its properties. Then the robot will feed the results of the experiment back into the ChemOS program, enabling the AI to update its data and instantly generate a new, better slate of candidate molecules, until—after rounds of predictions, syntheses, and tests—a winner emerges.

The idea of an automated, closed-loop discovery system has, partly because of Aspuru-Guzik's tireless advocacy, become increasingly popular among the new practitioners of chemistry. Peers in Vancouver, New York, Champaign-Urbana, and Glasgow are building similar facilities. These labs are intended as all-purpose, automated spaces of molecular creation. That's why Aspuru-Guzik doesn't speculate too much about what, specifically, the Matter Lab will produce next. Such decisions will be dictated by curiosity, perhaps, or by the imperatives of a global crisis.

Making a mark

In 2020, Aspuru-Guzik experienced a period of early-pandemic weight gain, which caused his surgical wound to reopen. At the same time, he felt trapped and bored by the 2D world of Zoom calls and frustrated at not being able to roam freely about his lab. His harried work life had left little space for the kind of aimless—or seemingly aimless—pursuits that, in the past, had fostered creative breakthroughs. He needed a change.

A few months later, he began doodling on his computer, drawing a lucha libre mask resembling Screamin' Jay Hawkins, the rock 'n' roll pioneer known for his operatic vocals and macabre stage antics. He named the character Bruho (a variation of "brujo," Spanish for sorcerer) and decided to impose his artwork on the urban landscape. He bought a sticker printer and began plastering the Bruho avatar on mailboxes and streetlights. Soon he was part of the city's bustling street-art scene.

Today, Aspuru-Guzik has two goals for the near future. The first is to design a modular, affordable version of his closed-loop system that can serve as a model to scientists around the world. He wants to build an all-in-one lab box, containing the ChemOS package along with synthesis and characterization robots. With this device, a user will punch in a description of a given material, and the system will immediately simulate and test candidate molecules. If we are to usher in a new era of materials on demand, Aspuru-Guzik reasons, the technology has to proliferate—and it has to be easy to use.

His second medium-term goal is to make his mark, artistically, on the city of Toronto.

A few days after my visit to the lab, I joined him and his crew for a night of stickering and postering. Like his materials work, this too was collaborative. Our eight-person group included Soap Ghost, an aloof young woman with full-sleeve tattoos; Urban Ninja, a wiry middle-aged man who arrived pulling a cart with a bucket of wheat paste, a homemade liquid adhesive; and Life, a flinty insomniac, his hair split down the middle, one half dyed blond like Cruella de Vil's. "I'll keep going until sunrise," he told me, gamely. Everyone had bundles of stickers or rolls of posters they'd designed themselves.

In Toronto, this kind of street art—which doesn't require spray paint—is punishable by fines (even though the police often look the other way), so we moved quickly and furtively. Ninja took us down an alleyway to a bare plywood wall of a boarded-up building, and we descended on it with our brushes, covering the surface with the paste and papering it with images—a bearded Buddha, a ukulele-playing rat, a Bruho figure, robed like a Jedi. The assemblage didn't make a whole lot of visual sense, but it had a kind of anarchic beauty to it. Within an impossibly short time frame, emptiness had given way to multiplicity, and Aspuru-Guzik was thrilled. "This wall was blank a minute ago," he exclaimed. "Look at it now." ■

Simon Lewsen is a Toronto-based magazine writer.

The upper module of ASML's next-generation EUV machine was built from a 17-ton piece of milled aluminum.

INSIDE THE MOST COMPLICATED MACHINE ON THE PLANET

Moore's Law was grinding to a halt—until a small Dutch firm pushed the limits of physics.

by **Clive Thompson**
Photographs by **Christopher Payne**

Patrick Whelan peers through the faceplate of his clean-room bunny suit to see how things are going.

Before him is a gleaming chunk of glass, roughly the size of a toaster oven, that is carved with so many scooped-out sections to reduce its weight that it looks like an alien totem. Whelan's team is gluing it to a large, coffee-table-size piece of aluminum. Both metal and glass are eerily smooth, having been polished for weeks to remove minute imperfections. Over the next 24 hours, as the glue solidifies, workers will neurotically monitor the position of the glass and metal to make sure they fuse together just so.

"These will be placed together to microns of precision," Whelan tells me, gesturing at the apparatus.

A nearby technician worries he's too close, and yelps: *Back up!*

"I'm not touching! I'm not touching!" Whelan says, laughing.

Precision is serious business here. I'm in Wilton, Connecticut, in a clean room of the Dutch company ASML, which makes the world's most sophisticated machine for lithography—a crucial process used to create the transistors, wires, and other essential components of microchips. It's a coveted device, with models costing as much as $180 million, that is used in making microchip features as tiny as 13 nanometers at a rapid clip. That level of precision is crucial if you're Intel or TSMC and want to manufacture the world's fastest cutting-edge computer processors. The final machine, assembled at ASML's headquarters in the Netherlands, is the size of a small bus and filled with 100,000 tiny, coordinated mechanisms, including a system that generates a specific wavelength of high-energy ultraviolet light by blasting molten drops of tin with a laser 50,000 times a second. It takes four 747s to ship one to a customer.

"It's a very difficult technology—in terms of complexity it's probably in the Manhattan Project category," says Sam Sivakumar, Intel's director of lithography.

Here in Wilton, the glass-and-metal module that Whelan and his team are building is particularly critical. It will carry the patterns needed to make a microchip, and it'll whiz back and forth while the machine blasts it with extreme ultraviolet (EUV) light, illuminating different parts of the chip pattern. The light will then bounce down to a dinner-plate-size wafer of silicon, burning the pattern in place.

Whelan walks over to a video monitor that shows one of these glass-metal contraptions zipping back and forth while being tested. It weighs 30 kilograms, but it moves in a blur.

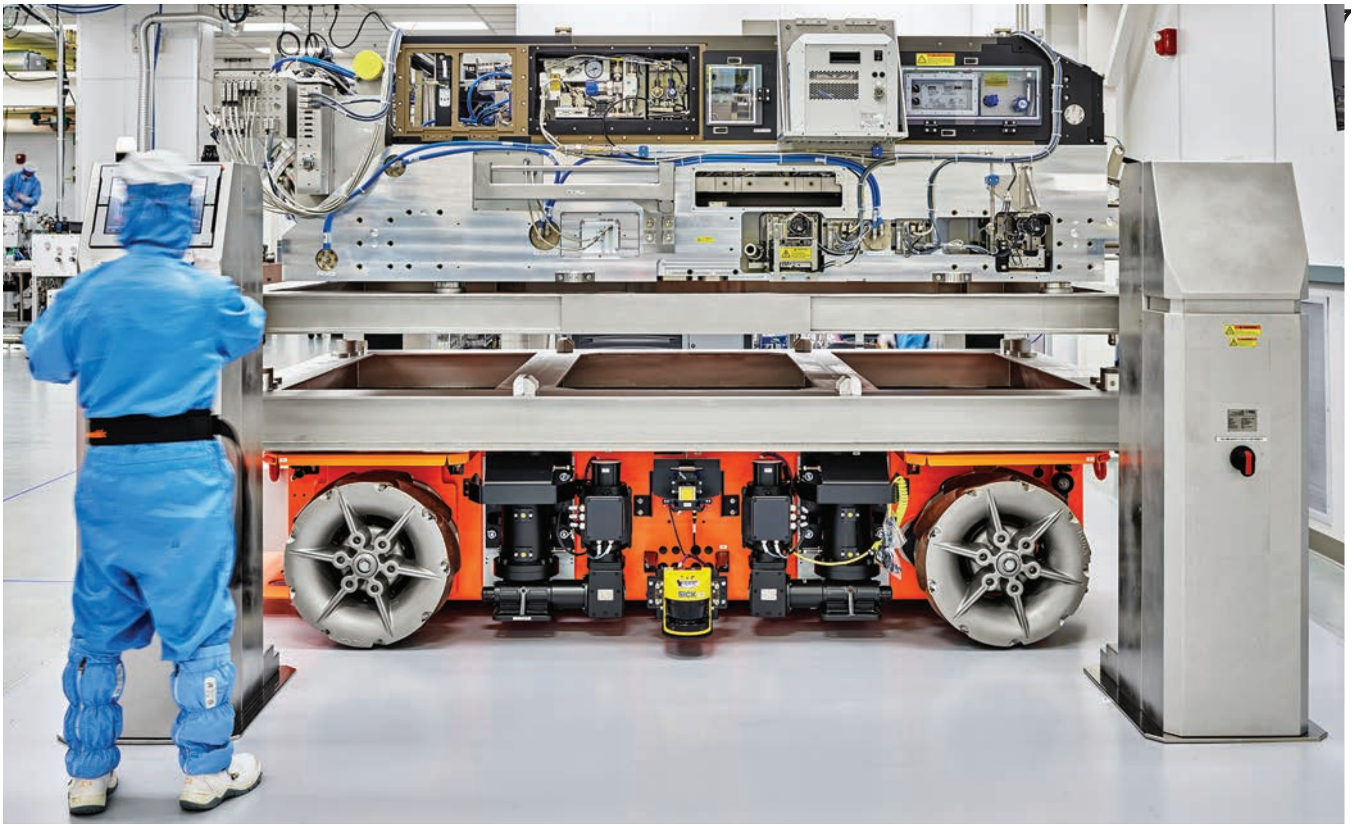
"This is accelerating faster than a fighter jet," Whelan says, his close-cropped beard and glasses obscured by his gear. "If there's anything that's loose, it'll fly apart." What's more, he says, the apparatus has to stop on a spot the size of a nanometer—"so you have one of the fastest things on earth settling at pretty much the smallest spot of anything."

This combination of speed and accuracy is key to keeping up with Moore's Law—the observation that the number of transistors crammed into a microchip doubles roughly every two years as components become ever smaller, making the chips cheaper and more powerful. The more tightly you pack transistors, the faster electrical signals can zip around the chip. Since the '60s, chipmakers have shrunk the components by switching, every decade or so, to a new form of light with a smaller wavelength. But by the late '90s, manufacturers were stuck at 193-nanometer light—and they were hotly debating what to do next. The situation grew more and more dire. Chipmakers had to use increasingly complex designs and techniques to keep Moore's Law going, but they managed to eke out another two decades of increasing performance.

Then, in 2017, ASML unveiled its production-ready EUV machine, which uses light with a wavelength of just 13.5 nanometers. With a wavelength that short, chipmakers could pack transistors more densely than ever before. CPUs can crunch numbers faster, use less power, or just get smaller. The first generations of chips with tiny EUV features are already at work for huge firms like Google and Amazon, improving language translation, search-engine results, photo recognition, and even AI that, like GPT-3, talks and writes with an eerily human quality. The EUV revolution is also reaching everyday consumers, since ASML's machines are being used to make chips for products

THIS PAGE: This glass clamp (black rectangle, upper center) is used to hold masks, which contain chip patterns to be transferred to a wafer.

FACING PAGE: ASML uses this orange robot, built by KUKA Robotics, to move heavy pieces of EUV machines around the clean-room floor.



including some Apple smartphones and Macs, AMD processors, and Samsung's Note10+ phone. As EUV machines become more common, it'll boost the performance and reduce the power demands on ever more everyday devices. EUV technology also enables simpler designs, which lets chipmakers move faster and produce more chips per wafer, resulting in cost savings that can be passed on to consumers.

The success of EUV lithography was far from guaranteed. The light is so devilishly hard to manipulate that for years experts predicted ASML would never figure it out. In fact, ASML's rivals, Canon and Nikon, both gave up trying years ago. So ASML now has a corner on the market: if you want to create the most cutting-edge processors, you need one of its machines. ASML makes only 55 of them a year, and they sell briskly to the industry's chip giants; currently over 1,000 are installed.

"Moore's Law is basically falling apart, and without this machine, it's gone," says Wayne Lam, a director of research at CCS Insight. "You can't really make any leading-edge processors without EUV."

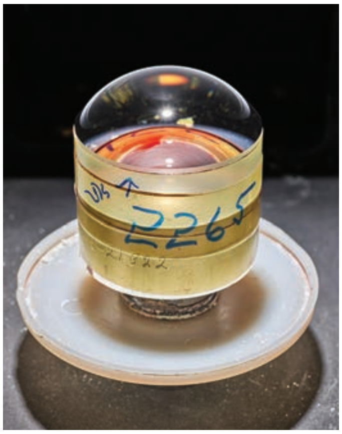
It's extremely rare for a single firm to possess a monopoly on such a key part of microchip production. Even more astonishing is the sheer grind of work: it took ASML $9 billion of R&D and 17 years of research, a nonstop spree of experimentation, tweaking, and "aha" breakthroughs. EUV is now here—it's working. But the effort and time it took to make it happen—and its late entry on the scene—raises some inevitable questions. How long will EUV be able to keep Moore's Law going? And what will happen next?

When Jos Benschop joined ASML in 1997, he'd come off a long stint with Phillips and landed smack dab in a chip industry worried about its future. Over decades, engineers in chip fabrication had mastered the art of lithography. The concept is simple. You design the components of a chip—its wires and semiconductors—and then etch them into a series of "masks," much as you make a stencil to put a pattern on a T-shirt. Then you put each mask over a silicon wafer and shine light through it (roughly equivalent to spraying paint over the stencil). The light hardens the "resist," a chemical layer on the surface of the wafer; then other chemicals etch that pattern into the silicon. In the '60s, chipmakers used visible light for this process, with a wavelength as small as 400 nanometers. Then they shifted to ultraviolet light, at 248 nm, and gradually reduced it to 193 nm—what's often called deep UV. Each switch bought them several years' extension of Moore's Law.

But by the late '90s, they'd focused deep UV as narrowly as they could manage, and they weren't sure how to go smaller. It seemed that a new light source was needed. ASML at the time was a small firm of 300 people that had been successfully selling its deep-UV lithography tools. But to stay relevant, they realized, they'd need to do some serious R&D.

Benschop—a tall, angular executive with an exuberant but wry manner—was hired as the first research employee. He started going to big conferences, held twice a year, where deep thinkers from major chip firms and government agencies would stroke their chins and argue about what form of light to use next.

"What would be the next kid on the block?" was how Benschop put it when we spoke on Zoom this past summer. The experts pondered several options, all of which had huge problems. One idea was to use a spray of ions to draw patterns onto chips; that would work, but nobody could figure out how to do it rapidly at scale. The same went for shooting beams of electrons. Some advocated for using x-rays, which have a tiny wavelength, but they had challenges of their own. The final idea was extreme ultraviolet, with a wavelength that



ABOVE: This polished optic is part of an energy sensor that helps control the intensity of light inside lithography machines.

BELOW: These polishing units are used to smooth down components that go into ASML's EUV machine. A component can spend many weeks in a multi-stage polishing process, with technicians checking smoothness down to nanometer precision.

RIGHT: A closer look at a polishing unit. The pieces of glass shown here are set at angles to achieve the correct bevel.

OPPOSITE: A few optics like the one shown above are mechanically polished.


2268
4975
1-21322

can go as low as 13.5 nanometers—pretty close to x-rays. It looked good.

The problem was that EUV would require an entirely new form of lithography machine. The existing ones used traditional glass lenses to focus light onto the wafer. But EUV light is absorbed by glass; it stops dead. If you wanted to focus it, you'd have to develop curved mirrors like the ones used in space telescopes. Worse, EUV is even absorbed by air, so you'd need to make the inside of your machine a perfectly sealed vacuum. And you'd need to generate EUV light reliably; nobody was sure how to do that.

Intel had tinkered with the idea, as had the US Department of Energy. But these were mostly lab experiments. To create a viable chipmaking lithography machine, you'd need to develop reliable techniques that could work quickly and produce chips in bulk.

After three years of pondering, in 2000 ASML decided to gamble the company and try to master EUV. They were a tiny firm, but if they could pull it off, they'd become a giant.

There were so many engineering problems to solve that, as Benschop recalls, "we didn't have the momentum to do it ourselves." So ASML's executives began calling up the firms that had made components for their existing machines. One call went to Zeiss, the German optics firm that had for years made glass lenses for ASML.

Zeiss's engineers had experience with EUV—including making extremely precise lenses and mirrors for x-ray telescopes. The trick was to coat the surface of the EUV mirrors with alternating layers of silicon and molybdenum, each only a few nanometers thick. Together they produce a pattern that reflects back as much as 70% of the EUV light that hits it.

The problem was in how to polish them. The machine would wind up needing 11 mirrors to bounce the EUV light around and focus it on the chip, rather like 11 Ping-Pong players bouncing a ball from one to another toward a target. Since the goal was to etch chip components measured in nanometers, each mirror had to be mind-bendingly smooth. The tiniest flaw would send EUV photons astray.

To give a sense of scale, if you took the mirror in your bathroom and blew it up to the size of Germany, it would have bumps about five meters high. Blown up to the same size, the smoothest EUV mirror Zeiss's engineers had yet made—for space telescopes—would have bumps only two centimeters high. These mirrors for ASML would have to be orders of magnitude smoother: if they were the size of Germany, their biggest imperfections could be less than a millimeter high. "These are really the most precise mirrors in the world," says Peter Kürz, who is responsible for the development of the next generation of EUV optics at Zeiss.

A big part of Zeiss's work would be inspecting the mirrors to look for imperfections and then using an ion beam to knock individual molecules off, gradually smoothing the surface over months and months of work.

While Zeiss was developing the mirrors, Benschop and other ASML suppliers were working on their other big challenge: how to create a light source that would produce a steady flow of EUV.

It would haunt them for years.

To generate EUV, you need to create a plasma, a finicky phase of matter that exists only at extremely high temperatures. After early experiments zapping lithium with laser pulses to produce EUV light, they switched to tin, which produced bigger bursts.

By the early 2000s, working with the San Diego firm Cymer and the German laser firm Trumpf, ASML had built something of a Rube Goldberg contraption. There's a heated vessel that keeps tin in a liquid state. It feeds into a nozzle that shoots a droplet of molten tin—"a third of the diameter of a human hair," says Danny Brown, the company's Australian-born vice president of technical development—out into the bottom part of the machine, camera systems tracking its progress. When it reaches the center of the light-producing chamber, a laser pulse strikes the tin droplet. Immolated in a burst that reaches a temperature of about 500,000 K, the tin produces a plasma that glows with EUV light. The mechanism repeats this process, shooting and destroying tin droplets, 50,000 times a second.

"It's non-straightforward, let's put it that way," Brown says drily.

Though they could now generate EUV light, Brown and his team quickly discovered new problems. Ions from the tin explosions would clog up the optics. To clean things up, they realized, they could pump hydrogen into the light chamber, where it would react with the tin ions and help scoop them away.

But they were rapidly falling behind schedule. Benschop had initially predicted that they'd have EUV machines "in volume" by 2006. In reality, by that year they had produced only two prototypes. The prototypes worked, etching patterns more finely than any lithography machine in history. But they were achingly slow. The light source was still too meager. In lithography, every photon matters; the more thickly you can generate them, the faster you can place a pattern down onto silicon.

This is a closer-in view of the glass clamp used to hold masks, shown on p. 46.

Meanwhile, the machine was growing to unbelievably complex dimensions. It contained robot arms moving wafers, motors that accelerated the reticle—that big piece of glass that holds the pattern—to 32 times Earth's gravity, and fully 100,000 parts, 3,000 cables, 40,000 bolts, and two kilometers of hosing. Worse, everything was interlinked: get one part working, and it'd create a problem somewhere else. It turned out, for example, that heat from the EUV light microscopically altered the dimensions of the mirrors. That forced Zeiss and ASML to develop sensors that would detect any change, triggering software that would shift the mirrors' positions using precision actuators.

"As we corrected one problem, we moved on to the next," Benschop says. "Every mountain you climbed, you saw the next mountain that was even higher."

Many observers in the microchip industry, watching ASML fall behind schedule again and again, figured they'd fail.

"Ninety-five percent of the smart money thought that there was no way EUV would ever work," says C.J. Muse, a semiconductor industry analyst with Evercore.

While ASML beavered away at EUV, they and the rest of the industry were performing ever more elaborate tricks to extend the performance of deep UV light as much as possible, to pack more transistors onto chips. One technique, called "immersion," involved putting a layer of water over the chip, which refracted incoming light and allowed it to be focused in a tighter pattern.

Lithography engineers also developed a technique for patterning and carving away at a chip layer multiple times—what's known as "multiple patterning"—to produce finer details. Together, these approaches pushed chip components down to 20 nanometers.

But these oddball innovations also made the act of chipmaking much more complex. Immersion required managing the presence of water in the delicate lithography process, no easy task. And chip designers found it onerous to change their designs to work with multiple patterning. Deep UV was running out of steam—and everyone knew it.

By the mid-2010s, though, it began to seem as if EUV might finally come to the rescue. Brown and his team had dived into the scientific literature, looking for ways to get more out of each tin droplet. As a former university researcher who studied plasma physics, he was known inside ASML for raising pointy-headed scientific issues; the CTO jokingly gave him a plaque emblazoned with the words "Scientifically Accurate But Practically Useless."

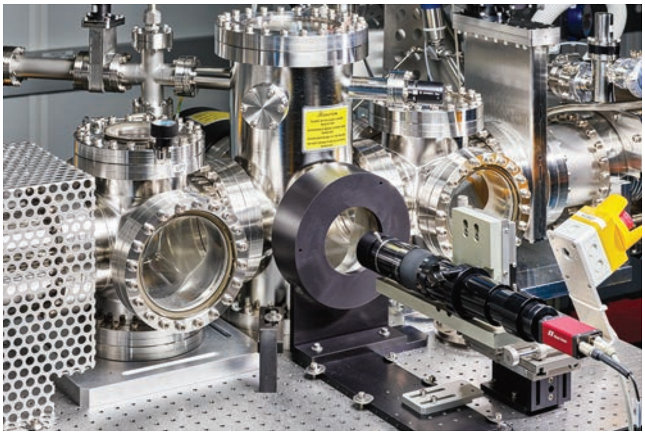
This time, though, soaking in the scientific literature paid off. It suggested the concept of hitting each tin droplet with the laser twice. A first blast would flatten the droplet into a pancake shape, which enabled a second blast—millionths of a second later—to produce far more EUV. Brown's team devised a way to do this at scale.

Other discoveries came by happy accident. As their ability to immolate tin improved, the process produced more debris than the hydrogen could clean up. Mirror performance was degrading. Then one day they noticed something funny: the mirrors didn't degrade as quickly after the machine had been opened for maintenance. As it turned out, oxygen in the air that came in helped reverse the contamination. ASML built the occasional addition of small amounts of oxygen into the design.

By the middle of 2017, the company finally had a working demo that etched chips at an industry-friendly pace—125 wafers per hour. From his office in San Diego, Brown watched the demo in the Netherlands. He was elated; he'd changed into a Hawaiian shirt, proclaiming that he'd finally be able to go on vacation.

"This thing was like *zzzt zzzt zzzt zzzt*," he recalls, mimicking the speed of the reticle zipping around, and the robotic arm sliding in a new wafer about every 30 seconds. "It was the last domino to basically say, 'Yeah, EUV lithography will happen.'"

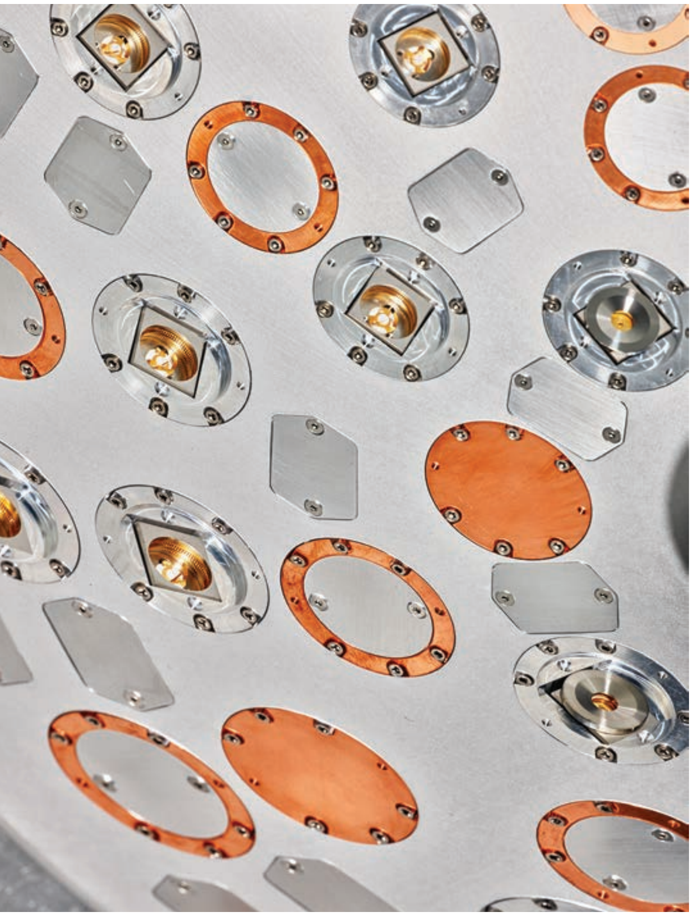
That year, ASML began finally shipping out machines that would revolutionize chipmaking. Once the market realized that ASML had a monopoly on the cutting-edge tools, its stock began to soar, reaching $549 and making the company's market cap almost the size of Intel's.



ABOVE: This tabletop experimental setup at ASML's San Diego factory is used to test droplet generator assemblies—part of the EUV machine's light source.

BELOW: These turbomolecular pumps remove air and other gases to produce a vacuum inside the EUV machine—crucial because EUV light is absorbed by air. The pumps spin at 30,000 RPM and knock out individual molecules of gas, one by one.

OPPOSITE: The mirrors inside the lithography machine can accumulate tin debris from the EUV light source. After the mirrors have been cleaned and polished, this machine is used to examine them.



If you're a gearhead like me, the machine is truly gorgeous to behold—a marvel of engineering. When I visited Wilton, they walked me over to view a massive block of milled aluminum that forms the top part of the device. It is eight feet long, six feet wide, and two feet thick. Gleaming like the chassis of a spaceship, it holds the glass reticle and also has mounted on it huge, barrel-shaped molecular pumps. Each pump contains tiny blades that spin at 30,000 RPM, sucking all gases out of the machine to produce a vacuum within. "They actually smack the molecules of the gas out of the way, one at a time," Whelan told me.

One could argue that ASML's chief success has not been so much in making machinery as in measuring it. When I pulled off my bunny suit, I visited the machine shop, where huge chunks of glass were being carved for the reticle. After each piece of glass is milled, it's placed on machines that gradually smooth it for hundreds of hours over several weeks. As machine-shop manager Guido Capolino told me, they measure the glass all along to see how many imperfections are being removed, starting with coarse microns. He gestured at a polishing machine behind us, where glass pieces slowly revolved on top a slurry of wet polishing mix.

"We're down at angstroms and nanometers for the variability here," he said. Using glass in the reticle is crucial; it doesn't deform under heat as much as metal. But it's devilishly hard to carve—yet another problem the engineers had to slowly solve.

ASML's success with EUV has won the company deep respect across the microchip industry. Chris Mack, a four-decade veteran of chip lithography, is currently the chief technology officer for Fractilia, a firm that makes software for chipmaking. He says the reason ASML and its partners succeeded—where others never even dared to try—is sheer, dogged persistence.

"They peeled the onion," he told me. "They go, *Oh, now I got the next layer.* And then they pull that layer. And then nobody really knows whether it's rotten in the core or it's going to be good. They just keep peeling it. And to their credit, they just never gave up."

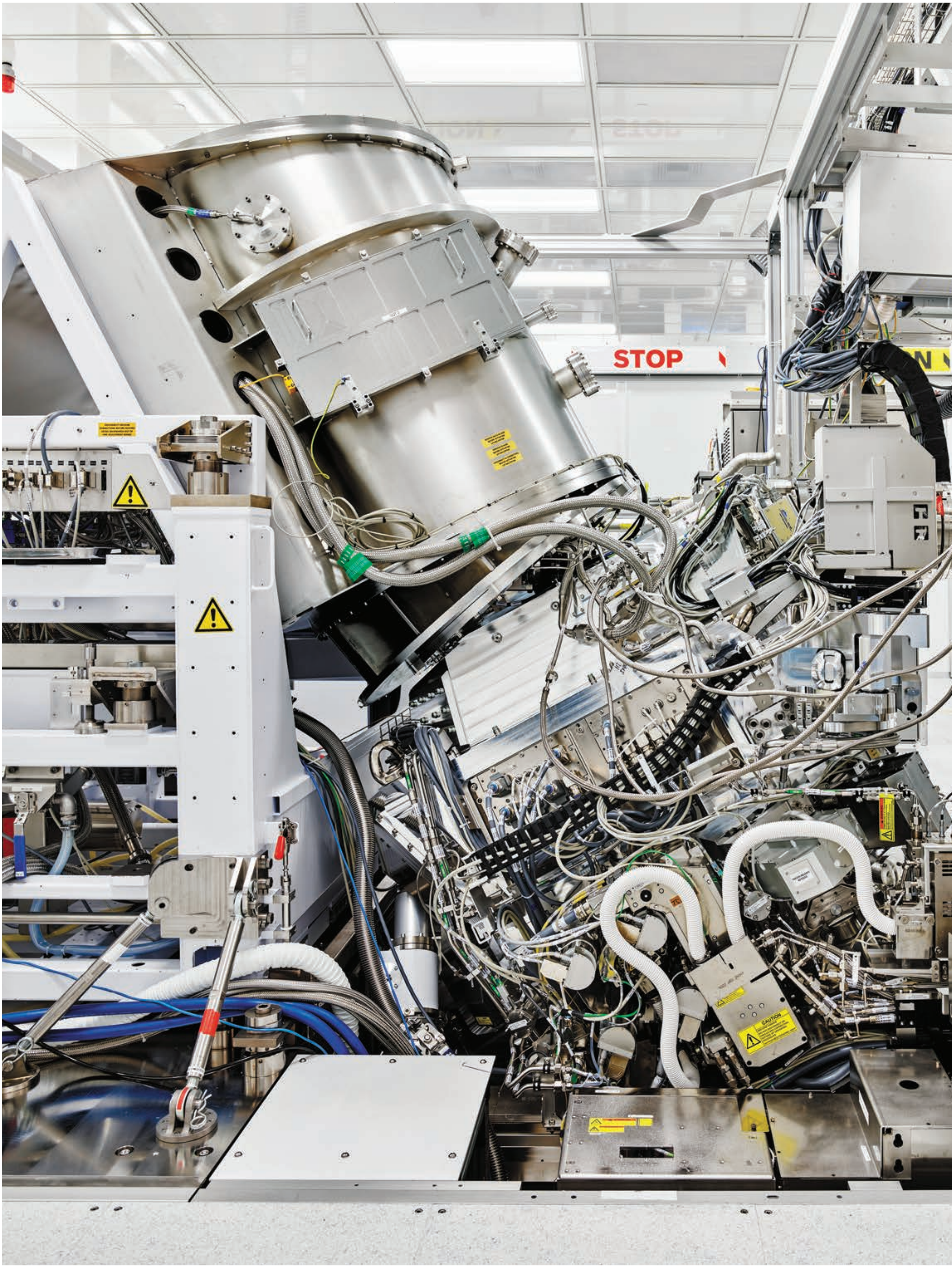
Now that they have the ability to keep crafting smaller and smaller components, major firms like Intel and TSMC and Samsung can build ever faster and more power-conserving chips.

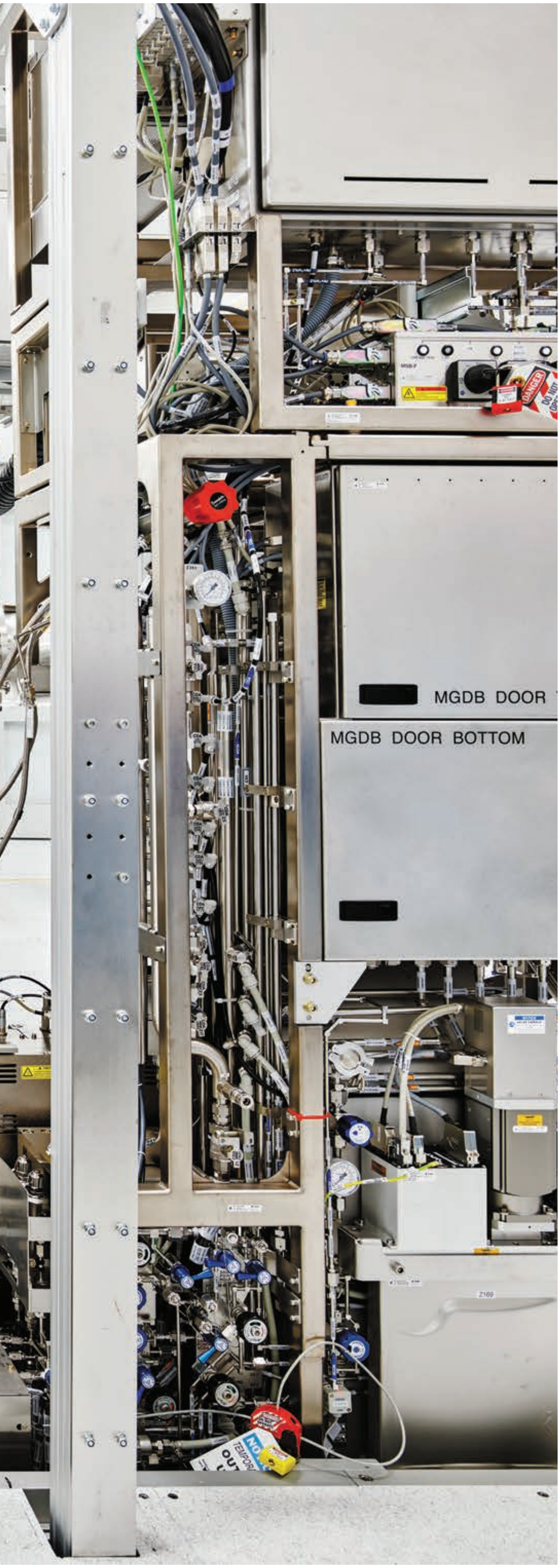
"Our designers can breathe a sigh of relief," says Intel's Sam Sivakumar. "Moore's Law is alive."

As more EUV machines come online and their cost amortizes, the technology will trickle down to an increasing number of everyday devices. The one place that won't benefit from the EUV revolution—at least in the short term—is China.

Worried that China poses a technological threat, both the Trump and Biden administrations successfully pressured the Netherlands to prevent ASML from selling EUV machines to customers there.

Can China simply make its own EUV devices? Some industry observers suspect it can't. ASML's success with EUV required enormous collaboration with firms based everywhere from Germany and the US to Japan (which makes chemicals critical to the lithographic masks). China, being relatively isolated, stands little chance on its own, according to Will Hunt, an


STOP



analyst with Georgetown University's Center for Security and Emerging Technology. "It can't really close that gap," he says.

What's possible, other observers suggest, is that there'll simply be a delay in China's ability to buy EUV machines. Typically, China's chipmakers work with last-generation tools that are a step behind what's used by TSMC in Taiwan, Samsung in Korea, or Intel in the US, C.J. Muse says. So when ASML's first generation of EUV machines become a bit older—a few years from now—and the industry moves on to newer models, China might be allowed to buy them.

And in fact, ASML is already working on an improved version of the device. It will be able to focus EUV light to an even sharper degree thanks to what's known as a higher numerical aperture, allowing it to etch components that could be under 10 nanometers wide. This "high-NA" EUV machine will have larger mirrors, requiring the entire machine to get larger too. Intel is currently the first customer for one of these next-generation machines, and it expects to sell its first chips built with them by 2025.

ASML and most observers figure EUV will help chips progress until at least 2030, and possibly longer. After all, some of the tricks that chip designers developed to keep deep UV going for so long should be repeatable with EUV.

But at some point in the next decade or so, the chip industry's desire to shrink features will start bumping up against some physical limitations that are even harder than the ones they've currently bested. For one thing, quantum problems begin to emerge. Indeed, they already have: chipmakers using ASML's EUV machines have to wrestle with "stochastic errors"—rays of EUV light naturally go astray, producing incorrect patterns on chips. These aren't show-stopping problems yet, but they'll furrow brows more and more the smaller chipmakers go.

An EUV light source sits in a test bay in an ASML clean room.

Assuming "high NA" keeps Moore's Law going to 2030, what will take over then? Industry experts figure ASML will continue to explore even higher-numerical-aperture devices, allowing them to focus EUV on smaller and smaller points. At the same time, chip designers are looking into strategies for improving chips that aren't so dependent on further miniaturization, such as extending architectures upward and building into the third dimension by stacking chip layers. As to what lithography technology might come after EUV, no one yet knows. Intel's Sivakumar wouldn't speculate; Mack said that outside of high-NA EUV, "nothing else" is under intensive development.

Inside the Wilton clean room, Whelan gave me a peek at their high-NA EUV machine. He rolled up a huge garage-style door and ushered me into a massive new clean room the size of a football field. In the corner was a shiny aluminum reticle bed. It was just like the one I'd seen for the original EUV machine, but it could no longer fit comfortably in a living room; it was almost as big as a subway car and weighed fully 17 tons. They had to install cranes in the roof to move it.

"So this," Whelan said, "is going to be the machine that helps us continue pushing Moore's Law into the future." ■

Clive Thompson is a science and technology journalist based in New York City and author of Coders: The Making of a New Tribe and the Remaking of the World.

THE PR

NP-COMPLETE PROBLEM

The traveling salesman problem

Find the shortest possible route that visits each city once, ultimately returning to the city of origin.

TO END ALL

OBLEM

P VS NP WAS ONCE DESCRIBED AS COMPUTER SCIENCE'S 'FAVORITE PARADIGM, FAD, PUNCHING BAG, BUZZWORD, ALIBI, AND INTELLECTUAL EXPORT.'

NO MATTER THE MONIKER, IT IS A VERY HARD PROBLEM.

By SIOBHAN ROBERTS

Photographs by DEREK BRAHNEY

1.

On Monday, July 19, 2021, in the middle of another strange pandemic summer, a leading computer scientist in the field of complexity theory tweeted out a public service message about an administrative snafu at a journal. He signed off with a very loaded "Happy Monday."

PROBLEMS

In a parallel universe, it might have been a very happy Monday indeed. A proof had appeared online at the esteemed journal ACM Transactions on Computational Theory, which trades in "outstanding original research exploring the limits of feasible computation." The result purported to solve the problem of all problems—the Holy Grail of theoretical computer science, worth a $1 million prize and fame rivaling Aristotle's forevermore.

This treasured problem—known as "P versus NP"—is considered at once the most important in theoretical computer science and mathematics and completely out of reach. It addresses questions central to the promise, limits, and ambitions of computation, asking:

Why are some problems harder than others?

Which problems can computers realistically solve?

How much time will it take?

And it's a quest with big philosophical and practical payoffs.

"Look, this P versus NP question, what can I say?" Scott Aaronson, a computer scientist at the University of Texas at Austin, wrote in his memoir of ideas, *Quantum Computing Since Democritus*. "People like to describe it as 'probably the central unsolved problem of theoretical computer science.' That's a comical understatement. P vs NP is one of the deepest questions that human beings have ever asked."

One way to think of this story's protagonists is as follows:

"P" represents problems that a computer can handily solve.

"NP" represents problems that, once solved, are easy to check—like jigsaw puzzles, or Sudoku. Many NP problems correspond to some of the most stubborn and urgent problems society faces.

The million-dollar question posed by P vs. NP is this: Are these two classes of problems one and the same? Which is to say, could the problems that seem so difficult in fact be solved with an algorithm in a reasonable amount of time, if only the right, devilishly fast algorithm could be found? If so, many hard problems are suddenly solvable. And their algorithmic solutions could bring about societal changes of utopian proportions—in medicine and engineering and economics, biology and ecology, neuroscience and social science, industry, the arts, even politics and beyond.

Sometimes the classifications evolve—hard problems are revealed to be easy when researchers find more efficient solutions. Testing whether a number is prime, for instance, has been known to be in the class NP since the mid-1970s. But in 2002, three computer scientists at the Indian Institute of Technology Kanpur devised an unconditional proof and a clever algorithm that finally confirmed the problem was also in P.

If *all* the tricky problems could be transformed with such algorithmic sleight of hand, the consequences for society—for humanity and our planet—would be enormous.

For starters, encryption systems, most of which are based on NP problems, would be cracked. We'd need to find a completely different approach to sending secure communications. Protein folding, a 50-year-old grand challenge in biology, would become more tractable, unlocking newfound abilities to design drugs that cure or treat disease and discover enzymes that break down industrial waste. It would also mean finding optimal solutions to everyday hard problems, such as mapping out a road trip to hit all destinations with minimal driving, or seating wedding guests so that only friends share the same dinner table.

Since the P vs. NP problem's inception 50 years ago—emerging from the momentous intersection of mathematical logic and electronic computing technology—researchers around the world have made Herculean attempts at a solution. Some computer scientists have suggested that the efforts might be better likened to those of Sisyphus, who labored without resolution. But while those who first explored the problem are running out of time to see a solution, the newer generations are happily taking up the quest.

For Manuel Sabin, a computer scientist just finishing a PhD at UC Berkeley, the allure is in probing the impossibility of problems where "you won't know the answer until the sun engulfs the earth." The search might be quixotic, but Sabin would regret not tilting at these windmills.

Timothy Gowers, a mathematician at the University of Cambridge, calls it "one of my personal mathematical diseases." He lost the summer of 2013 to the pursuit, after he asked students for an essay about the subject on a test. As he recounted on his blog: "After marking the essays in June, I thought I would just spend an hour or two thinking about the problem again, and that hour or two accidentally turned into about three months."

The quest has even stumped the University of Toronto computer scientist Stephen Cook, who framed the problem and launched the field of computational complexity with a seminal paper in 1971. For this work, he won the Turing Award, computer science's equivalent of the Nobel Prize. But he's had no luck finding a solution. Cook says he never had any good ideas—"It's just too difficult."

2.

Michael Sipser, an MIT computer scientist, estimates he's spent, all told, as much as a decade on the problem. He got interested during grad school in the 1970s, and he bet his fellow student Len Adleman an ounce of gold that it would be solved by the end of the century (Sipser paid up).

In the 1980s, he achieved a nice result solving a version of the problem with a "restricted" computational model—leading to an exciting period in the field with several beautiful results, giving cause for hope that a solution might not be too far off.

Sipser still returns to the problem every now and then, and he's a steadfast ambassador, delivering countless talks on the subject.

The way he inches into an accessible explanation of P vs. NP is with a basic multiplication problem: 7 × 13 = ?

The answer, 91, is easy enough to compute in your head. Though multiplying bigger numbers isn't as easy, it would still take a computer practically no time at all.

But flipping those problems around is another matter. Consider, for example, finding the two 97-digit prime numbers that multiply to produce this very large number:

310 7418240490 0437213507
5003588856 7930037346
0228427275 4572016194
8823206440 5180815045
5634682967 1723286782
4379162728 3803341547
1073108501 9195485290
0733772482 2783525742
3864540146 9173660247
7652346609

This factoring problem was part of a challenge assessing the difficulty of cracking the RSA keys used in cryptography. Solving it took 80 processors five months of continuous computing, Sipser explains—which works out to roughly 33 years with only a single processor. Factoring is a hard problem because all current methods seek the answer via "brute force," checking the astronomical number of possibilities one by one by one. Even for a computer, this is a slow process.

"The interesting question here is, do you really need to search?" Sipser says. "Or is there some way of solving the factoring problem that zooms in on the answer quickly without searching? We don't know the answer to that question."

Questions like this one get at the heart of computational complexity—a field full of beastly problems that researchers are trying to understand. Aaronson has assembled a "Complexity Zoo," an online catalogue with 545 classes of problems (and counting). Each is classified according to its complexity, or difficulty, and the resources—time, memory, energy—required to find solutions. P and NP are the main attractions.

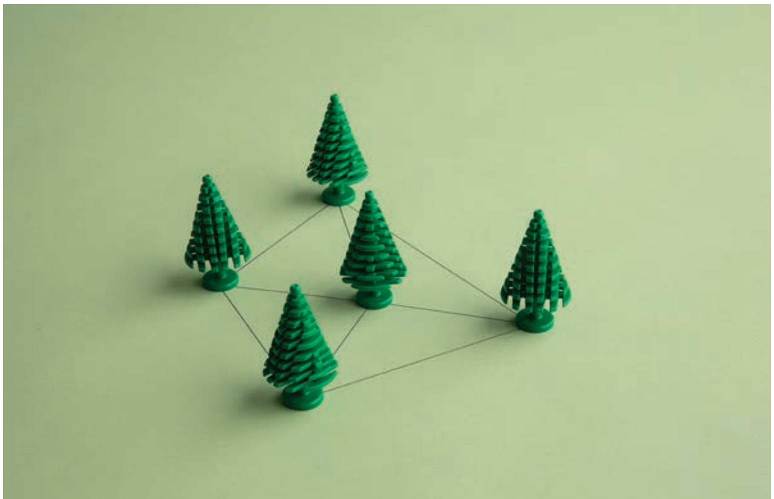
P is "the class that started it all." It is the class of problems that can be solved by a computer in a reasonable amount of time. More specifically, P problems are those for which the time it takes to find a solution can be described by a polynomial function, such as n^2. In polynomial-time algorithms, n is the size of the input, and growth against that input occurs at a reasonable rate (in this case, to the power of two).

By contrast, some hard NP problems might only be solvable by algorithms with run times defined by an exponential function, such as 2^n—producing an exponential growth rate (as with the spread of covid). NP, as Aaronson describes it, is "the class of dashed hopes and idle dreams." He is, though, quick to clarify a common misconception: not all NP problems are difficult. The class NP in fact contains the class P—because problems with easy solutions are, of course, also easy to check.

NP's more challenging problems often have momentous practical applications. For these problems, an exhaustive brute-force search for a solution would likely go on for an impractically long time—geologic time—before producing an answer. If a brute-force search algorithm is the best algorithm possible, then P does not equal NP.

And among the cognoscenti, that's apparently the consensus, which some liken more to religious belief: P ≠ NP. Most allow only a sliver of hope that the opposite will prove true. "I'd give it a 2 to 3% chance that P equals NP," Aaronson says. "Those are the betting odds that I'd take."

The result published in July presented a proof of exactly that long shot. But it was only the latest in a long tradition of proofs that don't pass muster. Within a day of publication, in a turn of events worthy of Monty Python, the paper was removed from the online journal; then it seemed to reappear briefly before disappearing permanently. It was the most recent version of a paper that the author had posted more than 60 times to the arXiv preprint server over the last decade. The journal's editor in chief explained on Twitter that the result had been rejected, but in a case of human error, the paper's disposition had somehow changed from "reject" to "accept" and the proof had found its way to publication.



NP-COMPLETE PROBLEM

The Steiner tree problem

Connect a set of points with line segments of minimum total length.

3.

In early August, when I met Steve Cook at his office on campus, he'd neither seen nor heard of that latest P vs. NP proof snafu. Now 81, he'd only recently retired, since his memory was failing. "That's why we have James here," he said—his son James, 36, also a computer scientist, had joined us for my visit. Steve was in the midst of clearing out his office. A giant recycling bin stood in the middle of the room, filling up with old yellowing issues of the Journal of Symbolic Logic, a stack of super-fat Toronto telephone books waiting nearby.

Over the years, Cook has seen many proofs purporting to solve the P vs. NP problem. In 2000, after the Clay Mathematics Institute named it one of the seven unsolved "Millennium Problems" (each worth a $1 million prize), he was inundated with messages from people who thought they'd triumphed. All the results were wrong, if not plainly bogus. About half claimed to have proved that P equals NP; the other half went in the opposite direction. Not too long ago, one person claimed to have proved both.

Cook, in his 1971 paper, conjectured that P does not equal NP (he phrased it using different terminology common at the time). He's since invested a significant if indeterminate amount of time working to establish that that's the case. "I don't have a good memory of toiling away," he says, but his colleagues recall that whenever they went into the department on the weekend, Steve was there in his office.

Unless he's racing sailboats, Cook is not one to rush; he likes to give an idea time. And his former students remember a distinct lack of swagger. The computer scientist Anna Lubiw, at the University of Waterloo, says that when he taught Cook's theorem—part of that pioneering paper—he never referred to it as such and never even gave any hints that he was the person who proved it. Maria Klawe, a mathematician and computer scientist and the president of Harvey Mudd College, says she would regularly correct Cook when he lost his way teaching proofs that he knew inside out: "He'd get stuck and say, 'Okay. Tell me how the proof goes.'" Cook was also famously modest in grant applications and reports pertaining to his research—he'd confess: "Honestly, I have made little progress …"

He did make headway, however, in recruiting James to take up the cause. Early on, James displayed an interest in mathematics and computing—at age nine, he urged his dad to teach him Boolean algebra and logic. A couple of years ago, after earning a PhD at Berkeley and doing a stint at Google, he set off as an independent researcher focusing on miscellaneous projects, some of them indirectly connected to P vs. NP. And despite the track record, James, who bears a striking resemblance to his father, is undaunted at having inherited such a seemingly interminable quest. He regards it as he would any mathematical endeavor: it's a fun puzzle. "There's got to be an answer to these questions," he says. "And it's like, come on, somebody's got to solve it. Let's just get this figured out. It's been a long time. It's embarrassing that we don't know the answer yet."

The lack of progress hasn't stopped this community of happy Sisypheans from celebrating computational complexity's 50th anniversary. The festivities began in 2019, when devotees from around the world gathered at the Fields Institute for Research in Mathematical Sciences, at the University of Toronto, for a symposium in Cook's honor. Christos Papadimitriou, a computer scientist at Columbia University who has spent much of his career working on P vs. NP, opened the event with a public lecture, looking back not a half-century but millennia.

He began by describing age-old quests for solutions—using algebraic tools or straightedge and compass, which he considered rudimentary forms of computation. Papadimitriou's tale eventually arrived at Alan Turing, the British mathematician whose 1936 paper "On Computable Numbers" formalized the notions of "algorithm" and "computation." Turing also showed—with his idea of a "universal computing machine"—that there is no "mechanical" way (that is, performed by a machine) to prove the truth or falsehood of mathematical statements; no systematic way to distinguish the provable from the unprovable.

Papadimitriou said he considers Turing's paper the birth certificate of computer science—"and the birth certificate says that computer science was born with a stark understanding of its own limitations." He reckoned computer science is the only known field of scientific discourse born with such an awareness—"as opposed to other sciences, which understand their own limitations, like the rest of us, in late middle age."

It wasn't long after Turing's ideas (and similar ideas from others) found embodiment in the first computers that scientists confronted questions about the machines' inherent capabilities and limitations. In the early 1950s, John von Neumann, the Hungarian-American pioneer of the modern computer, "bragged about an algorithm of his being polynomial, compared to the exponential incumbent," as Papadimitriou recalled—he'd outwitted a slow algorithm with a fast one. This was the dawn of a new theory: computational complexity theory. The crux of it was that only polynomial algorithms are in any sense good or practical or worth aiming at a problem, whereas an exponential algorithm, Papadimitriou said, "is the algorithmic equivalent of death."

Cook first started thinking about complexity in the mid-1960s. While working on his PhD at Harvard, he contemplated whether it is possible to prove, given certain computational models, that multiplication is harder than addition (it remains an open problem).

In 1967, according to a book about Cook forthcoming from the Association for Computing Machinery (ACM), while a postdoc at Berkeley, he drafted course notes that contained the seed of his big result. He'd worked out a formulation of the complexity classes that came to be known as P and NP, and he posed the question of whether P was equal to NP. (At around the same time, others, including the computer scientist Jack Edmonds, now retired from the University of Waterloo, were circling around the same ideas.)

But the field of computer science was only just beginning, and to most scientists and mathematicians such ideas were unfamiliar if not downright strange. After four years at Berkeley's mathematics department, Cook was considered for tenure but not offered a position. He had advocates in the university's new department of computer science, and they lobbied for him to be granted a position in their ranks, but the dean wasn't inclined to give tenure to someone whom the illustrious mathematicians had denied.

In 1970, Cook moved to the University of Toronto. The following year he published his breakthrough. Submitted to a symposium of the ACM held that May in Shaker Heights, Ohio, the paper sharpened the concept of complexity and defined a way to characterize the hardest problems in NP. It proved, in a flash of algorithmic alchemy, that one problem, known as the satisfiability problem (seeking a solution for a formula given a set of constraints), was in a sense the hardest problem in NP, and that all the other NP problems could be reduced to it.

This was a crucial theorem: If there is a polynomial-time algorithm that solves the satisfiability problem, then that algorithm will serve as a skeleton key, unlocking solutions to all the problems in NP. And if there exists a polynomial-time solution for all the problems in NP, then P = NP.

Among computer scientists, Cook's theorem is iconic. Leslie Valiant, of Harvard, recalled at the 2019 symposium precisely where and when he first heard of it. After finishing undergraduate studies in math, he'd started a PhD in computer science. While there were courses and degrees in this fledgling field, he said, it felt ephemeral, perhaps lacking in deep intellectual content. "It was a serious worry for people doing computer science at the time," he said. They asked, 'Is this a field? Where is it going?' One day, Valiant came upon Cook's paper. He read it overnight. "I was transformed," he said. "In an instant, my concerns about computer science were very much reduced. This paper—for me, it really made the field. I think it made computer science—made it into something of substance."

And then, as the story goes, after Cook's theorem came a deluge.

In 1972, Dick Karp, a computer scientist at Berkeley, having read Cook's esoteric paper, demonstrated that many of the classic computational problems with which he was intimately acquainted—essentially every problem he didn't know how to solve, drawn from mathematical programming, operations research, graph theory, combinatorics, and computational logic—possessed the same transformational property that Cook had found with the satisfiability problem. In total, Karp found 21 problems, including the knapsack problem (seeking the optimal way to pack a constrained space with the most valuable items), the traveling-salesman problem (finding the shortest possible route that visits each city once and returns to the city of origin), and the Steiner tree problem (seeking to optimally connect a set of points with line segments of minimum total length).

Karp showed that this special collection of problems were all equivalent, which in turn demonstrated that the pattern identified by Cook was not an isolated phenomenon, but rather a classification methodology of surprising power and reach. It was a litmus test of sorts, identifying the class of what became known as "NP-complete" problems: a solution to any would crack them all.

Papadimitriou thinks of NP-completeness as a versatile tool. "If you cannot solve a problem, try to prove it is NP-complete, because this will maybe save you a lot of time," he said at the public lecture—you can give up on an exact solution and move on to solving an approximation or variation of the problem instead.

In the grand sweep of history, Papadimitriou sees the phenomenon of NP-completeness and the P vs. NP quest as computer science's destiny. Because as scientific serendipity would have it, a Soviet mathematician, Leonid Levin, converged on a result equivalent to Cook's at more or less the same time. Levin, now at Boston University, did his work behind the Iron Curtain. After it received wider attention (he immigrated to America in 1978), the result became known as the Cook-Levin theorem.

And in a further coda a decade or so later, a "lost letter" was discovered in the Princeton archives of the Austrian logician Kurt Gödel. In 1956, he'd written to von Neumann asking whether a logic problem—which in modern parlance would be called NP-complete—could be solved in polynomial time. He opined that "this would have consequences of the greatest magnitude."

AS SCIENTIFIC SERENDIPITY WOULD HAVE IT, A SOVIET MATHEMATICIAN, LEONID LEVIN, CONVERGED ON A RESULT EQUIVALENT TO COOK'S AT MORE OR LESS THE SAME TIME.

4.

While a half-century of work hasn't yielded anything close to a solution, some results at least capture the imagination: a paper in 2004 claimed a proof for P = NP using soap bubbles as a mechanism of analog computation (soap film, naturally aligning in the minimum-energy configuration, solves the NP-complete Steiner tree problem in a fashion).

These days it's a rare bird of a computer scientist—for example, Ron Fagin, an IBM fellow—who tackles the problem head on. In the 1970s, he produced Fagin's theorem, which characterized the class NP in terms of mathematical logic. And he's solved the problem more than once, but the results never stood for more than a few days before he found a bug. Fagin recently got funding for a P vs. NP project from IBM's Exploratory Challenges program supporting adventurous research. In explaining why he keeps at it, he likes to quote Theodore Roosevelt, who said that it is far better to "dare mighty things" than to rank among those who "live in a gray twilight that knows neither victory nor defeat."

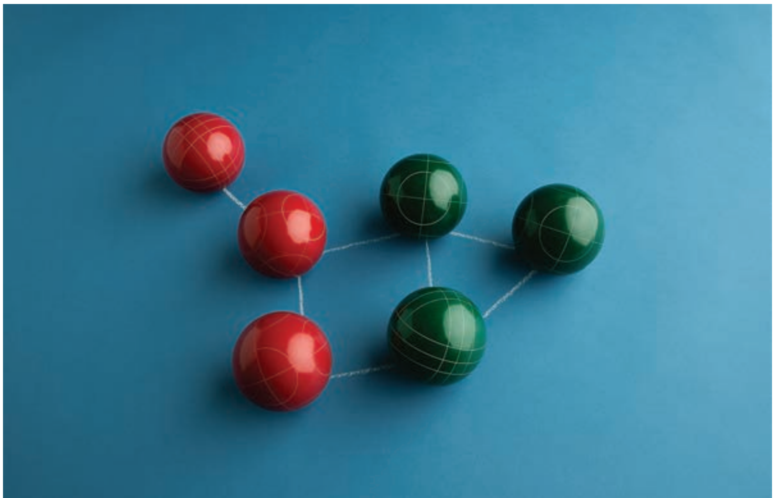
But most complexity theorists dream a little smaller, opting instead for indirect approaches—tilting the problem, reshaping or reframing it, exploring related environs, and further whittling down the arsenal of tools that could be deployed against it (many are now known to be useless).

Ryan Williams, a computer scientist at MIT, is trying to illuminate the problem both from above and from below—investigating the nature of "upper bounds" and "lower bounds" on core computational problems. An upper bound, in simple terms, is a specific mathematical claim that there exists a concrete algorithm that solves a particular problem without exceeding a certain amount of resources (time, memory, energy). A lower bound is the intangible opposite: it's a general claim of impossibility, showing that no such algorithm exists universally. One focus of Williams's research is to make lower bounds constructive and concrete—mathematical objects with describable features. He believes that more constructive approaches to lower bounds are "precisely what we are missing from current approaches in complexity theory."

Williams has pegged the likelihood that P ≠ NP at a fairly moderate 80%. But lately some researchers in the field are expressing doubts about even that level of certainty. "More and more, I'm starting to wonder whether P equals NP," Toniann Pitassi, a computer scientist at the University of Toronto and a former PhD student of Cook's, says. Her approach in circling around the problem is to study both scaled-up and scaled-down analogues, harder and easier models. "Sometimes generalizing the question makes it clearer," she says. But overall, she hasn't achieved clarity: "Most people think P doesn't equal NP. And I don't know. Maybe it's just me, but I feel like it's become less and less clear that that's the truth."

MOST COMPLEXITY THEORISTS DREAM A LITTLE SMALLER, OPTING INSTEAD FOR INDIRECT APPROACHES.

Historically, Pitassi points out, surprising results have occasionally come out of nowhere—seeming impossibilities proved possible by smart algorithm designers. The same could happen with P vs. NP, maybe in another 50 years or a century. One of the most important results in all of complexity theory, for instance, was achieved by David Barrington, of the University of Massachusetts, Amherst, in 1989. The gist of it (for our purposes) is that he devised a clever algorithm, which set out to do something that seemingly should've required an unbounded amount of memory but in fact used an astonishingly small amount—just five bits of information, enough to specify a number between one



NP-COMPLETE PROBLEM

The clique problem

Search for cliques in a graph, such as a certain subset of friends in a social network.

and 32 (inclusive) or a two-letter word.

A more recent and related result, from 2014, took James Cook by surprise. Drawing from Barrington's theorem, it uses memory in a wonderfully weird way. As hinted in the title of the paper, by the University of Amsterdam's Harry Buhrman and collaborators, it's about "computing with a full memory." James can rattle off the paper's introductory paragraph practically verbatim:

Imagine the following scenario. You want to perform a computation that requires more memory than you currently have available on your computer. One way of dealing with this problem is by installing a new hard drive. As it turns out you have a hard drive but it is full with data, pictures, movies, files, etc. You don't need to access that data at the moment but you also don't want to erase it. Can you use the hard drive for your computation, possibly altering its contents temporarily, guaranteeing that when the computation is completed, the hard drive is back in its original state with all the data intact?

The answer, counterintuitively, is yes.

James thinks of it as "borrowed memory." After the shock of this result sank in, he had fun figuring out how to apply it to a particular problem—picking up where his dad had left off.

A couple of decades ago, Steve Cook moved on to other related problems in complexity theory. With one problem, he made a conjecture about the amount of memory an algorithm would need to solve the problem—honing it to the absolute minimum. In 2019, James, together with Ian Mertz, one of Pitassi's PhD students, deployed the poetic idea of borrowing memory and proved that even less memory was needed. The result didn't go all the way to refuting his dad's conjecture, but it's a bit of progress in the grand complexity quest nonetheless.

And problems in complexity theory, James observes, sometimes have a domino effect—if there's a proof in one critical corner, then all the dominoes fall. The breakthrough results, the most important ones, come from a long line of work, by a lot of different people, making incremental progress and establishing connections between different questions, until finally a big result emerges.

He also mentions a caveat: while a truly devilishly fast P = NP algorithm would be earth-shattering, there is also a scenario in which P = NP might be a letdown. It might turn out that a P algorithm capable of solving the NP-complete problem is on a time scale of, say, n^{100}. "Technically that falls under P: it's a polynomial," says James. "But n^{100} is still very impractical"—it would mean any sizable problems would still be out of reach on human time scales.

That is, of course, assuming we can find the algorithm in the first place. Donald Knuth, an algorithmist at Stanford, in recent years changed his mind—he "flipped the bit." His intuition is that P does indeed equal NP, but that we'll probably never be able to make use of that fact, practically speaking—because we won't actually know any of the algorithms that happen to work. There are mind-boggling numbers of algorithms out there, he explains, but most of them are beyond our ken. So whereas some researchers might insist that no P = NP algorithm exists, Knuth contends that "it's more likely that no polynomial-time algorithm will ever be embodied—actually written down as a program—by mere mortals."

For Papadimitriou, any answer would quench a lifelong obsession. He believes the P vs. NP problem belongs in the realm of fundamental scientific conundrums such as the origin of life and the unification of nature's force fields. It's the kind of profound, consequential puzzle, "concrete yet universal," he said, "that adds meaning not only to science, but to human life itself."

"Imagine that we are lucky, and we are able to squeeze another couple of thousand years out of this planet, against the odds and despite the oddballs," he said. "And we don't solve these problems. What's the point?!" ■

Siobhan Roberts is MIT Technology Review's senior editor for computing.

It's time we began to "fixate on data" to solve our problems, says one of the world's leading experts in data science.

In 2006, Jeannette Wing, then the head of the computer science department at Carnegie Mellon University, published an influential essay titled "Computational Thinking," arguing that everyone would benefit from using the conceptual tools of computer science to solve problems in all areas of human endeavor.

Wing herself never intended to study computer science. In the mid-1970s, she entered MIT to pursue electrical engineering, inspired by her father, a professor in that field. When she discovered her interest in computer science, she called him up to ask if it was a passing fad. After all, the field didn't even have textbooks. He assured her that it wasn't. Wing switched majors and never looked back.

Formerly corporate vice president of Microsoft Research and now executive vice president for research at Columbia University, Wing is a leader in promoting data science in multiple disciplines.

Anil Ananthaswamy recently asked Wing about her ambitious agenda to promote "trustworthy AI," one of 10 research challenges she's identified in her attempt to make AI systems more fair and less biased.

Jeannette Wing

Q: Would you say that there's a transformation afoot in the way computation is done?

A: Absolutely. Moore's Law carried us a long way. We knew we were going to hit the ceiling for Moore's Law, [so] parallel computing came into prominence. But the phase shift was cloud computing. Original distributed file systems were a kind of baby cloud computing, where your files weren't local to your machine; they were somewhere else on the server. Cloud computing takes that and amplifies it even more, where the data is not near you; the compute is not near you.

The next shift is about data. For the longest time, we fixated on cycles, making things work faster—the processors, CPUs, GPUs, and more parallel servers. We ignored the data part. Now we have to fixate on data.

Q: That's the domain of data science. How would you define it? What are the challenges of using the data?

A: I have a very succinct definition. Data science is the study of extracting value from data.

You can't just give me a bunch of raw data and I push a button and the value comes out. It starts with collecting, processing, storing, managing, analyzing, and visualizing the data, and then interpreting the results. I call it the data life cycle. Every step in that cycle is a lot of work.

Q: When you're using big data, concerns often crop up about privacy, security, fairness, and bias. How does one address these problems, especially in AI?

A: I have this new research agenda I'm promoting. I call it trustworthy AI, inspired by the decades of progress we made in trustworthy computing. By trustworthiness, we usually mean security, reliability, availability, privacy, and usability. Over the past two decades, we've made a lot of progress. We have formal methods that can assure the correctness of a piece of code; we have security protocols that increase the security of a particular system. And we have certain notions of privacy that are formalized.

Trustworthy AI ups the ante in two ways. All of a sudden, we're talking about robustness and fairness—robustness meaning if you perturb the input, the output is not perturbed by very much. And we're talking about interpretability. These are things we never used to talk about when we talked about computing.

[Also,] AI systems are probabilistic in nature. The computing systems of the past are basically deterministic machines: they're on or off, true or false, yes or no, *0* or *1*. The outputs of our AI systems are basically probabilities. If I tell you that your x-ray says you have cancer, it's with, say, 0.75 probability that that little white spot I saw is malignant.

So now we have to live in this world of probabilities. From a mathematical point of view, it's using probabilistic logic and bringing in a lot of statistics and stochastic reasoning and so on. As a computer scientist, you're not trained to think in those ways. So AI systems really have complicated our formal reasoning about these systems.

Q: Trustworthy AI is one of the 10 research challenges you identified for data scientists. Causality seems to be another big one.

A: Causality, I think, is the next frontier for AI and machine learning. Right now, machine-learning algorithms and models are good at finding patterns and correlations and associations. But they can't tell us: Did this cause that? Or if I were to do this, then what would happen? And so there's another whole area of activity on causal inference and causal reasoning in computer science.

The statistics community has been looking at causality for decades. They sometimes get a little miffed at the computer science community for thinking that "Oh, this is a brand-new idea." So I do want to credit the statistics community for their fundamental contributions to causality. The combination of big data and causal reasoning can really move the field forward.

Q: Are you excited about what data science can achieve?

A: Everyone's going gaga over data science, because they are seeing their fields being transformed by the use of data science methods on the digital data that they are now generating, producing, collecting, and so on. It's a very exciting time. ■

PETER GARRITANO

Statement of Ownership, Management and Circulation. (Required by U.S.C. 3685.) (1). Publication Title: MIT Technology Review, (2). Publication No. 535-940, (3). Filing Date: 10/1/21, (4). Issue Frequency: Bi-monthly. (5). No. of Issues Published Annually: 6. (6) Annual Subscription Price: $80.00. (7). Complete Mailing Address of Known Office of Publication: One Main St, 13th Floor, Cambridge, MA 02142 (8). Complete Mailing Address of Headquarters or General Business Office of Publisher: One Main St, 13th Floor, Cambridge, MA 02142 (9). Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor. Publisher: Elizabeth Bramson-Boudreau, One Main St, 13th Floor, Cambridge, MA 02142. Editor: Mat Honan, One Main St, 13th Floor, Cambridge, MA 02142. Managing Editor: Timothy Maher, One Main St, 13th Floor, Cambridge, MA 02142. (10). Owner: Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02138. (11). Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or other Securities: None. (12). Tax Status: The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes: Has Not Changed During Preceding 12 months. (13). Publication Title: MIT Technology Review. (14) Issue Date for Circulation Data Below: September/October 2021. (15). Extent and Nature of Circulation: (a) Total No. of Copies (Net Press Run): Average No. Copies Each Issue During Preceding 12 Months: 153,251; No. Copies of Single Issue Published Nearest to Filing Date: 145,594. (b) Legitimate Paid and/or Requested Distribution (1) Mailed Outside-County Paid Subscriptions Stated on PS Form 3541: Average No. Copies Each Issue During Preceding 12 Months: 120,727, No. Copies of Single Issue Published Nearest to Filing Date: 115,728. (2) In-County Paid/Requested Mail Subscriptions Stated on PS Form 3541: Average No. Copies Each Issue During Preceding 12 Months: 0, No. Copies of Single Issue Published Nearest to Filing Date: 0. (3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid Distribution Outside the USPS: Average No. Copies Each Issue During Preceding 12 Months: 28,344, No. Copies of Single Issue Published Nearest to Filing Date: 27,646. (4) Paid Distribution by Other Mail Classes Through the USPS: Average No. Copies Each Issue During Preceding 12 Months: 0, No. Copies of Single Issue Published Nearest to Filing Date: 0. (c) Total Paid and/or Requested Distribution: Average No. Copies Each Issue During Preceding 12 Months: 149,071, No. Copies of Single Issue Published Nearest to Filing Date: 143,374. (d) Nonrequested Distribution: (1) Outside-County Nonrequested Copies included on PS Form 3541: Average No. Copies Each Issue During Preceding 12 Months: 0, No. Copies of Single Issue Published Nearest to Filing Date: 0. (2) In-County Nonrequested Copies included on PS Form 3541: Average No. Copies Each Issue During Preceding 12 Months: 0, No. Copies of Single Issue Published Nearest to Filing Date: 0. (3) Nonrequested Copies Distributed Through the USPS by Other Classes of Mail: 0 Average No. Copies Each Issue During Preceding 12 Months: 0, No. Copies of Single Issue Published Nearest to Filing Date: 0 (4) Nonrequested Copies Distributed Outside the Mail: Average No. Copies Each Issue During Preceding 12 Months: 1585, No. Copies of Single Issue Published Nearest to Filing Date: 1441. (e) Total Free or Nominal Rate Distribution: Average No. Copies Each Issue During Preceding 12 Months: 1585, No. Copies of Single Issue Published Nearest to Filing Date: 1441. (f) Total Distribution: Average No. Copies Each Issue During Preceding 12 Months: 150,657, No. Copies of Single Issue Published Nearest to Filing Date: 144,815. (g) Copies not Distributed: Average No. Copies Each Issue During Preceding 12 Months: 0, No. Copies of Single Issue Published Nearest to Filing Date: 0. (h) Total: Average No. Copies Each Issue During Preceding 12 Months: 153,252, No. Copies of Single Issue Published Nearest to Filing Date: 145,594. (i) Percent Paid: Average No. Copies Each Issue During Preceding 12 Months: 98.9%, No. Copies of Single Issue Published Nearest to Filing Date: 99.0%. This Statement of Ownership will be printed in the November/December 2021 issue of this publication. I certify that all the information furnished on this form is true and complete. **Doreen Adger, SVP of Consumer Revenues and Marketing.**

“Quietly and unremarked, all media types converged into one—the universal digital medium, bits.” p. 70

ANDREA DAQUINO

CHRIS TURNER

Awash in digital light

In *A Biography of the Pixel*, Pixar cofounder Alvy Ray Smith recounts how digitization and other forces have transformed our visual lives

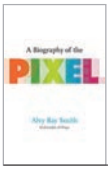
The computer scientist Alvy Ray Smith cofounded both Lucasfilm’s computer graphics division and Pixar Animation Studios. For those achievements alone, he is one of the most important technological innovators in cinema since at least the end of the Second World War. But Smith is not a Hollywood guy, and his intriguing, foundational

PIXELS

new book *A Biography of the Pixel* is not a Tinseltown book. There are only the slightest morsels of gossip (Steve Jobs was a difficult man to work with—*confirmed!*), and the only marquee celebrity who appears in Smith's story with any frequency is George Lucas. Smith isn't interested in fame. He's chasing more profound themes, arguing in effect that the great project he was part of—the invention and development of computer graphics—is far more important than anything that ever happened in Hollywood.

Smith is what used to be called a "graybeard" in computer programming circles. He's from that generation of engineers and coders who watched the digital age rise from the swamps of secret military projects and the space program to conquer the world. He has spoken machine language. He marveled at the first crude graphics to exhibit motion on green-and-black screens. And he was among the first to demonstrate the newfound ability of a stylus to trace a smooth curve of digital "paint."

In *A Biography of the Pixel*, Smith's aim is to set down clearly the trajectory of two important, intertwined stories. The first story is the development of computer images, from origin to digital ubiquity. There are, in Smith's telling, many names, places, and breakthroughs missing from the record, and he has taken on the job of adding them back in with an engineer's eye for precision. The second story, unfolding in parallel, is about the impact of those images—a transformative force Smith calls "Digital Light." It encompasses basically everything we experience through screens, and he argues convincingly that it is among the most important innovations in human communication since the first simple depictions of daily life were etched on the walls of caves.



The Biography of the Pixel

By Alvy Ray Smith, cofounder of Pixar

MIT PRESS, 2021

The humble pixel

As Smith demonstrates repeatedly, far too much credit has been allowed to slide to the supposed wizardry of individual geniuses. The reality is a muddy, overlapping history of groups of inventors, working by turns in competition and in collaboration, often ad hoc and under considerable commercial or political pressure.

Thomas Edison and France's Lumière brothers, for example, were great promoters and exploiters of early film technology. Both exhibited full systems circa 1895 and were happy to claim full credit, but neither built the first complete system of camera, film, and projector all (or even mostly) on their own. The real answer to the question of who invented movies, Smith writes, is a "briar patch" of competing lineages, with parts of the system developed by erstwhile partners of Edison's and similar parts by a handful of French inventors who worked with the Lumières.

Among the crucial figures relegated to history's dustbin were William Kennedy Laurie Dickson (an odd European aristocrat who designed and built the first movie camera for Edison) and Georges Demenÿ (whose design was copied without credit by the Lumières). Smith shows perhaps too much of his exhaustive work in rescuing these convoluted origin stories—there are similarly tangled muddles at every major stage in the development of computers and graphics—but his effort to set the historical record straight is admirable.

The main drawback of all this wrangling with the egos and avarice of several generations of forceful men (they are, alas, virtually all men) is that it sometimes distracts Smith's focus from his larger theme, which is that the dawn of Digital Light represents such a rare shift in how people live that it deserves to be described as epochal.

Digital Light, in Smith's simplest definition, is "any picture composed of pixels." But that technical phrase understates the full import of the "vast new realm of imagination" that has been created by its rise. That realm encompasses Pixar movies, yes, but also video games, smartphone apps, laptop operating systems, goofy GIFs traded via social media, deadly serious MRI images reviewed by oncologists, the touch screens at the local grocery store, and the digital models used to plan Mars missions that then send back yet more Digital Light in the form of jaw-dropping images of the Red Planet's surface.

And that barely begins to cover it all. One striking aspect of Smith's book is that it invites us to step just far enough back from the constant flow of pixels that many of us spend most of our waking hours gazing at to see what a towering technological achievement and powerful cultural force all this Digital Light represents.

The technological breakthrough that made all this possible is, as Smith's title suggests, the humble pixel. The word itself is a portmanteau of "picture element." Simple enough. But the pixel has been mischaracterized in popular usage to refer to the blurry, blocky supposed inferiority of poorly rendered digital images. Smith wants us to understand that it is, rather, the building block of all Digital Light—a miraculous, impossibly varied, endlessly replicable piece of information technology that has literally changed how we see the world.

The misunderstanding begins, Smith explains, with the fact that a pixel is not a square, and it is not arranged alongside other pixels on a neat grid. Pixels can be rendered on displays as such, but the pixel itself is "a sample of a visual field ... that has been digitized into bits." The distinction might sound esoteric, but it's crucial to Smith's argument for the pixel's revolutionary impact. The pixel is stored information that any device can display as Digital Light. And digital devices can do this because pixels are not approximations but carefully calibrated *samples* of a visual field, which has been translated for digital uses into a collection of overlapping waves. These pixels, Smith writes, are not reductions of the visual field so much as "an extremely clever repackaging of infinity."

The new wave

The process by which a pixel generates Digital Light—whether in the form of words on a screen or an icon on a smartphone or a Pixar movie on the big screen—is built on three mathematical breakthroughs that predate the modern computer. The first of these was achieved by Jean Joseph Fourier, a French aristocrat and regional governor under Napoleon in the early 1800s. Fourier contributed the foundational insight that not just sound but heat and everything we see and much else could be described as the sum of a series of waves, representing various frequencies and amplitudes. Or, as Smith more poetically phrases it, "The world is music. It's all waves."

More than a century later, a Soviet engineer named Vladimir Kotelnikov built on Fourier's wave principle with the second crucial element for creating Digital Light—the "Sampling Theorem." Kotelnikov demonstrated that a signal—be it a piece of music or a visual scene—can be captured by taking snapshots ("samples") at certain intervals. Take enough samples of some aspect of a visual field—its gradation of color, for example, or shifts from foreground to background—and it is possible to reconstitute the entirety of the information. Smith acknowledges that American computer scientists are taught that the sampling theorem originates with Harry Nyquist and Claude Shannon, but "the great idea ... was first clearly, cleanly and completely stated by Kotelnikov in 1933."

The third element that made Digital Light possible is the best known and most recently developed: Alan Turing's 1936 paper outlining the universal computing machine, whose great innovation was the ability to execute any systematic process as long as it has the right accompanying set of instructions (which we now call software). A Turing machine, the basis of the modern computer, can be programmed to understand the process by which Fourier's waves had been sampled by Kotelnikov's theorem, and to reproduce them on any other Turing machine. These three elements together begat Digital Light.

Digital Light on its own, though, was a limited force. Its earliest manifestations were simple pictographs on the digital cave wall of a TV screen. In December 1951, for example, MIT's Whirlwind computer displayed an array of white dots on a black screen for the CBS program *See It Now*, hosted by Edward R. Murrow. The dots spelled out "Hello Mr. Murrow," slowly fading and then brightening again, like a Lite-Brite on a dimmer switch. Clever, even wondrous for its time, but not the upheaval at the core of Smith's book. For that, Digital Light needed one more element: unimaginable speed.

Computer graphics, Smith explains, are just crazily long lists of numbers that correspond to graphical coordinates—pixels, nowadays, but thousands and thousands of tiny interlocking triangles in the earliest manifestations—assembled in digital space into the three-dimensional form of a Pixar cartoon character or anything else. (The first 3D computer graphic assembled from these triangles was, famously, a teapot.)

The great digital convergence

Such wonders as 3D animation, however, weren't possible until computer processing power exploded. Smith recounts the ensuing transformation with an engaging mix of technical detail, deep research, and personal recollection. Several generations of mathematicians, coders, and lab rats contributed to the development of computer graphics, building new tools and machines as Moore's Law rapidly made it easier to turn Fourier's waves and Kotelnikov's samples into geometric shapes, simple pictures, and basic motion on a screen. Disney and Lucasfilm and Stanford University loom large, of course, but so do NASA and General Motors and Boeing (which pioneered computer-aided industrial design), as well as lesser-known hives of computer graphics genius like the University of Utah and the New York Institute of Technology (NYIT).

FOURIER CONTRIBUTED THE INSIGHT THAT EVERYTHING WE SEE COULD BE DESCRIBED AS THE SUM OF A SERIES OF WAVES. OR, AS SMITH MORE POETICALLY PHRASES IT, "THE WORLD IS MUSIC. IT'S ALL WAVES."

PIXELS

Smith's own transition from simple pixels to digital movies started at NYIT in the early 1970s. There, he helped establish one of the world's first computer graphics labs, along with several of the other cofounders of Pixar, before moving on to introduce the technology to Lucasfilm. (He worked on the very first computer-animated sequence Lucasfilm produced, a special-effects sequence for the movie *Star Trek II: The Wrath of Khan*.)

Throughout the journey, Smith remained focused on the ultimate prize of producing a full-length digital movie. He wanted these tools to be used to create great art, to give form to the creative genius of minds the world over. Pixar achieved that goal with the 1995 release of *Toy Story*, the first feature-length film to be completely computer animated. And not long after that, an even more momentous achievement was reached—the pivotal moment Smith calls "the Great Digital Convergence."

This is the point, sometime around the year 2000, when all pictures (moving and otherwise) could be universally represented by pixels. "Quietly and unremarked," he writes, "all media types converged into one—the universal digital medium, bits."

Reading Smith's account of this convergence, I found myself thinking of a famous quote attributed to the French writer and filmmaker Jean Cocteau. "Film will only become an art," Cocteau said, "when its materials are as inexpensive as a pencil and paper." This, in part, is what Smith is driving at when he asks us to look in awe upon the power of the pixel. And that recollection led me—inexorably, really—to thinking about the "Steamed Hams" meme.

WHAT HAPPENED WITH 'STEAMED HAMS' IS BOTH EASY TO EXPLAIN AND HARD TO FULLY COMPREHEND. WHAT HAPPENED WAS PEOPLE STARTED MESSING AROUND WITH IT.

For the uninitiated, Steamed Hams was born as a short vignette in an episode from the seventh season of *The Simpsons*, "22 Short Films About Springfield," which first aired in 1996: Springfield Elementary School's dorky Principal Skinner hosts his boss, Superintendent Chalmers, for a luncheon at his home. The two minutes and 42 seconds of the vignette unfold as an escalating series of minor disasters, leading Skinner to sneak off to Krusty Burger and then claim the fast-food meal as his own. Having promised the superintendent steamed clams, Skinner covers for his ruse by claiming that he had actually said he was making "steamed hams," which he suggests is regional slang for hamburgers in upstate New York.

It's a silly little snippet from an offbeat *Simpsons* episode, and it earned no particular attention until the Great Digital Convergence placed the tools of digital filmmaking in the hands of virtually anyone with a computer and an internet connection. And then what happened is both easy to explain and hard to fully comprehend. What happened was people started messing around with it.

The creative force unleashed

The birth of the Steamed Hams meme appears to have been a 15-second clip from the vignette, reproduced using a text-to-movie app and posted to YouTube in March 2010. In the years since then, as the digital tools for producing and disseminating short videos improved at the breakneck speed of Moore's Law, the meme metastasized wildly. The clip has been piled upon itself, the YouTube screen divided into 10 boxes, each playing Steamed Hams on a short delay, as if being sung in a round, until it dissolves into roaring cacophony. It has been set to a wide range of pop songs—my favorite is one in which Auto-Tune software (itself a product of the convergence) has been used to bend and morph the dialogue so that it somehow sticks to the melody of Green Day's hit song "Basket Case." It has been layered over various video games. One enterprising Steamed Hams fan persuaded actor Jeff Goldblum to read the entire vignette's script in his distinctive diction; the resulting YouTube clip cuts expertly from Goldblum's live reading to the original animation, sometimes in split-screen. There is a sort of remake of Steamed Hams in which a different animator renders every 13 seconds of the vignette in an entirely different style. This is only a small sample of the highlights. The meme is massive.

If I were ever asked to teach a class in postmodern art, I would hold up the entire meme as a signature example of the staggering creative force unleashed by the Great Digital Convergence. Thanks to tools not that much harder to obtain than a pen and pencil, the internet now hosts an impossible abundance of inventive riffs: GIFs and clips, supercuts and mashups, reboots and remixes. A whole world of casual creators making digital movies, using brand-new tools that have already become so commonplace we barely notice them. We are at home in Digital Light.

Cocteau's world of ubiquitous cinematic creation, that is, may very well be here. This is what Alvy Ray Smith was building toward for half a century in pursuit of that first digital movie. We've arrived. We are all auteurs. Go play. ■

Chris Turner is an author and essayist based in Calgary, Alberta. His most recent book is The Patch: The People, Pipelines, and Politics of the Oilsands.

OPTIMISM

ANDREA DAQUINO

DAVID ROTMAN

Don't get left behind

A new book predicts a coming age of exponential technology growth, leading to an age of abundance. The reality is a lot more complicated.

Maybe it never truly went away. But these days techno-optimism—the kind that raged in the late 1990s and early 2000s and then dried up and turned to pessimism during the last decade—is once again bubbling up. The pessimism over the real-world impacts of apps and social media has turned into unbounded hope—at least among the tech elite and the venture capital investor class—that new technologies will solve our problems.

The Exponential Age, by tech investor and writer Azeem Azhar, is the latest celebration of the world-changing impact of computing technologies (including artificial intelligence and social media), biotechnology, and renewable energy. Azhar meticulously and smartly makes his case, describing the growth of what he calls exponential technologies—ones that rapidly and steadily improve in price and performance every year for several decades. He writes that "new technologies are being invented and scaled at an ever-faster pace, all while decreasing rapidly in price."

To his credit, Azhar duly notes the problems arising from the fast transformations brought about by these technologies, most notably what he calls the "exponential gap." Big tech corporations like Amazon and Google are gaining great wealth and power from the technologies. But other companies and many institutions and communities "can only adapt at an incremental pace," he writes. "These get left behind—and fast."

Yet his enthusiasm remains obvious.

For Azhar the story begins in 1979, when he was a seven-year-old in Zambia and a neighbor brought home a build-it-yourself computer kit. He then retells the familiar, yet

OPTIMISM

still gripping, history of how those early products kick-started the PC revolution (an interesting side note is his description of the mostly lost-to-history Sinclair ZX81—his first computer, bought for £69 two years later after his family moved to a small town outside London). We know the rest. The explosion of PCs—young Azeem and his family soon graduated to the Acorn BBC Master, a popular home computer in the UK—led to the World Wide Web, and now our lives are being transformed by artificial intelligence.

It's hard to quibble with the argument that computing technologies have grown exponentially. Moore's Law has defined such growth for generations of technologists. It has meant, as Azhar points out, that by 2014 the cost of a transistor was only a few billionths of a dollar, versus around $8 in the 1960s. And that has changed everything, fueling the rapid rise of the internet, smartphones, and AI.

Essential to Azhar's claim for the dawning of a new age, however, is that a far broader set of technologies exhibit this exponential growth. Economists call fundamental advances that have broad economic effects "general-purpose technologies"; think of the steam engine, electricity, or the internet. Azhar suspects that cheap solar power, bioengineering techniques such as synthetic biology, and 3D printing could be just such technologies.

He acknowledges that some of these technologies, particularly 3D printing, are relatively immature but argues that as prices drop, demand will grow quickly and the technologies will evolve and find markets. Azhar concludes: "In short, we are entering an age of abundance. The first period in human history in which energy, food, computation, and many resources will be trivially

The Exponential Age: How Accelerating Technology Is Transforming Business, Politics, and Society

By Azeem Azhar

DIVERSION BOOKS, 2021

cheap to produce. We could fulfill the current needs of humanity many times over, at ever-declining economic cost."

Maybe. But frankly, such uber-optimism takes a great leap of faith, both in the future power of the technologies and in our ability to use them effectively.

Sluggish growth

Our best measurement of economic progress is productivity growth. Specifically, total factor productivity (TFP) measures the role of innovation, including both management practices and new technologies. It isn't a perfect gauge. But for now, it's the best metric we have to estimate the impact of technologies on a country's wealth and living standards.

Starting around the mid-2000s, TFP growth became sluggish in the US and many other advanced countries (it has been particularly bad in the UK), despite the emergence of our brilliant new technologies. That slowdown came after a multiyear growth spurt in the US in the late 1990s and early 2000s, when computers and the internet boosted productivity.

No one is sure what is causing the doldrums. Perhaps our technologies are not nearly as world-changing as we think, at least compared with earlier innovations. The father of techno-pessimism in the mid-2010s, Northwestern University economist Robert Gordon, famously showed his audience images of a smartphone and a toilet; which would you rather have? Or perhaps we don't accurately capture the economic benefits of social media and free online services. But the most likely answer is simply that many businesses and institutions are not adopting the new technologies, particularly in sectors like health care, manufacturing, and education.

It's not necessarily a reason for pessimism. Maybe it will just take time. Erik Brynjolfsson, a Stanford economist and a leading expert on digital technologies, predicts that we are at the beginning of a "coming productivity boom." He argues that most of the world's advanced economies are near the bottom of a productivity J-curve. Many businesses are still struggling with new technologies, such as AI, but as they get better at taking advantage of the advances, overall productivity growth will take off.

It's an optimistic take. But it also suggests that the trajectory of many new technologies is not a simple one. Demand matters, and markets are fickle. You need to look at why people and businesses want the innovation.

Take synthetic biology. The idea is as simple as it is compelling: rewrite the genetic code of microorganisms, whether bacteria or yeast or algae, so they produce the chemicals or materials you desire. The dream wasn't exactly new at the time, but in the early 2000s proponents including Tom Knight, an MIT computer scientist turned biologist, helped popularize it, especially among investors. Why not treat biology as a simple engineering challenge?

With huge fermentation vats of these programmed microbes, you could make plastics or chemicals or even fuels. There would be no need for petroleum. Simply feed them sugar extracted from, say, sugarcane, and you could mass-produce whatever you need.

In the late 2000s several startups, including Amyris Biotechnologies and LS9, engineered the genetics of microbes to make hydrocarbon fuels intended to replace gasoline and diesel. Synthetic biology, it seemed, was on the verge of revolutionizing transportation. But in a few years, the dream was mostly dead. Amyris is now focused on making ingredients for skin creams and other consumer beauty products. LS9 sold off its holdings in 2014.

The market woes of synthetic biology continue to this day. Earlier this year, one of the leading companies in the field, Zymergen, suffered a financial setback as its product, a plastic made for use in folding smartphones, failed to gain traction. Its customers, the company said, were having "technical issues" integrating the plastic into their existing manufacturing processes.

The failures are not a condemnation of synthetic biology. A smattering of products are beginning to appear. Despite the commercial mistakes, the field's future is undeniably bright. As the technology improves, aided by advances in automation, machine learning, and computing, the costs of creating tailored bugs and using them for mass production will surely drop.

But for now, synthetic biology is far from transforming the chemical industry or transportation fuels. Its progress over the last two decades has looked less like exponential growth and more like the staggering first steps of a child.

History lessons

I asked Carlota Perez, a social scientist who has written widely on technological revolutions and whom Azhar credits in his book as "instrumental" in helping him think about the relationship between technology and economics, how we can have such impressive breakthroughs and not see more productivity growth.

The answer is simple, says Perez: "All technological revolutions have gone through two different periods—the first in which productivity growth is seen in the new part of the economy, and the second, when the new technologies spread across the whole economy, generating synergies and bringing general productivity increases."

Perez says we're now in the period in which different industries are faring very differently. She adds, "The question is how do we get to the point where we have the productivity of the whole economy growing synergistically?"

Perez is a very different kind of techno-optimist from the free-market ones often heard in Silicon Valley. To her, it's essential that governments create the right incentives to encourage the embrace of new technologies, including environmentally cleaner ones, using such tools as appropriate taxes and regulations.

"It's all up to government," she says. "Companies are not going in the green direction because they don't need to—because they're making money with what they're doing. Why should they change? It is only when you can no longer be profitable doing what you're doing [that] you use the new technologies to invest and innovate in new directions."

But Perez says that "the amount of innovation in gestation—that is, in the wings—is almost unbelievable." And, she says, once prompted by the right government policies and support, technological revolutions can happen quickly.

None of this is inevitable, however. There is certainly no assurance that governments will act. One worry is today's lack of support for research. Our amazing new technologies might be poised to change the economy, but their growth and expansion must be bolstered by ever more new ideas and continued technological advances. After all, the origins of the technologies we're so impressed by these days, such as synthetic biology and 3D printing, date back decades. The pipeline needs constant refreshing.

THE TECHNOLOGIES THAT WE'RE SO IMPRESSED BY, SUCH AS SYNTHETIC BIOLOGY AND 3D PRINTING, DATE BACK DECADES. THE PIPELINE NEEDS CONSTANT REFRESHING.

John Van Reenen, an economist at the London School of Economics and MIT, and his collaborators have shown that research productivity itself is slowing as "new ideas get harder to find." At the same time, the US and many other Western governments have decreased their support for R&D as a proportion of GDP over the last few decades; in the mid-1960s, US federal R&D funding relative to GDP was three times what it is today. The US doesn't have to return to such high levels, he says, "but standing still is not an option." That would, says Van Reenen, cause TFP growth and economic progress to stagnate.

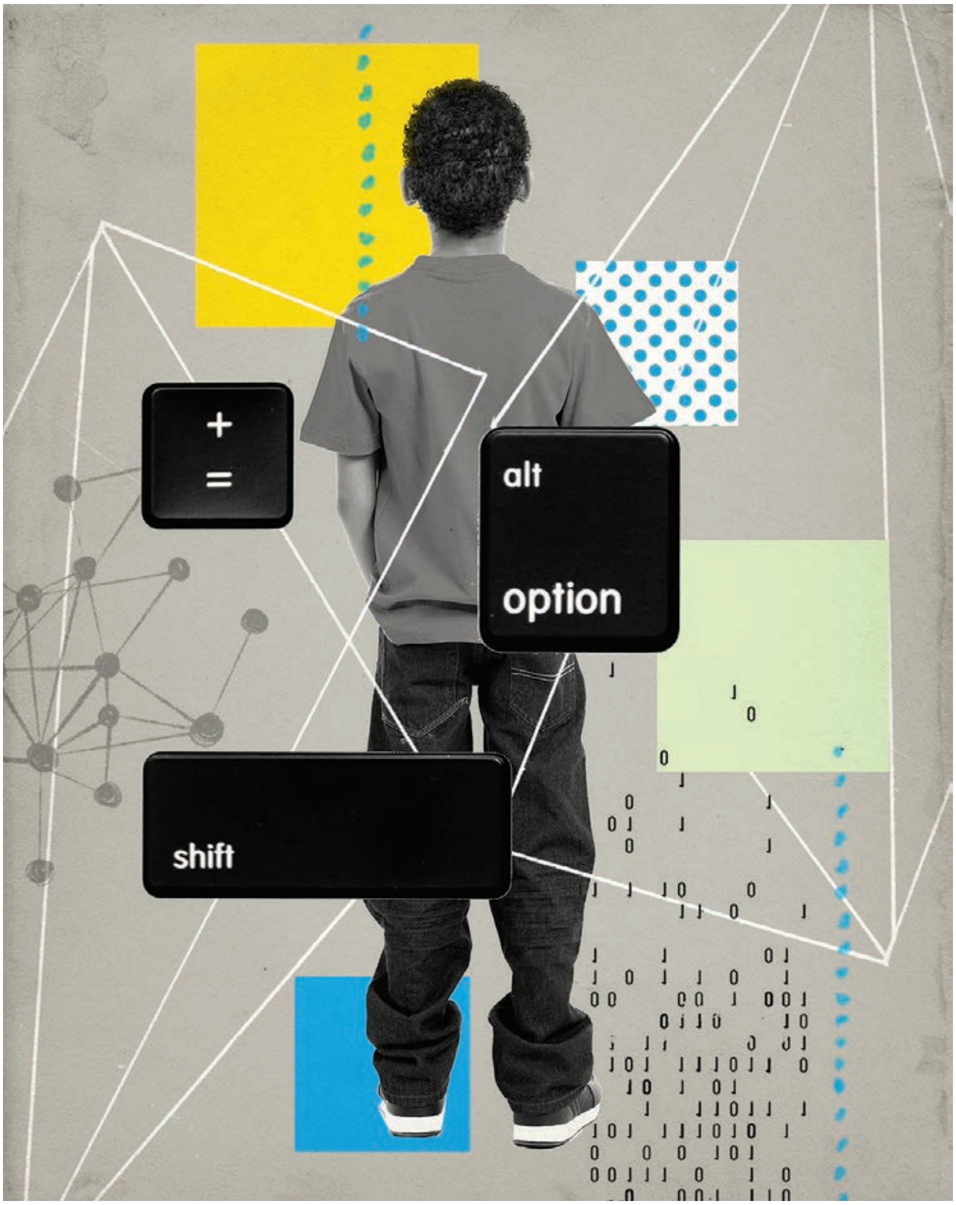
There are some signs the US is moving in the right direction. President Biden campaigned on promises to increase federal support for R&D by hundreds of billions over his first term. But getting Congress to embrace this has already been a challenge.

"It's a choice we face," says Van Reenen. "It's all come back to the politics. Are we prepared to make serious investments?"

And that is where reluctant optimists like Van Reenen and uber-optimists like Azhar converge. I asked Azhar just how confident he is about his book's prediction of "an age of abundance." He said: "I'm optimistic about the progress of the technology, but I'm much more realistic, bordering on pessimistic, around the governance of the technology. That's the bigger part of the fight." ■

David Rotman is editor at large at MIT Technology Review.

EDUCATION



MORGAN AMES

Laptops alone can't bridge the digital divide

Covid made educational inequities obvious, but the failures of One Laptop per Child more than a decade ago show us that closing the gap means more than giving away computers.

In May 2020, two months after covid-19 shut down schools and public life around the world, Twitter CEO Jack Dorsey announced that he was giving $10 million to California's Oakland Unified School District to purchase 25,000 Chromebooks. Dorsey tweeted that his donation was intended "to give EVERY single child in Oakland access to a laptop and internet in their homes." The donation came just a day after Oakland mayor Libby Schaaf announced the #OaklandUndivided campaign to raise $12.5 million to "close the digital divide for good" in the city.

Oakland's school district, along with much of the world, certainly needed the help. Despite the city's proximity to Silicon Valley's centers of power and wealth, 71.2% of its children qualified for free or reduced-price school lunch the year the pandemic hit. Half did not have the computers and internet connections needed to enable a sudden switch to remote learning. These numbers reflect nationwide trends. Lower-income households are much less likely to have broadband; over one-quarter rely solely on their smartphone's metered internet connection, and many share one dilapidated computer. In August 2020 a picture of two young girls sitting on a dirty sidewalk outside a Taco Bell in Salinas, 100 miles south of Oakland, using the restaurant's public internet connection to attend class on their school-issued laptops, went viral as a potent symbol of how difficult the pivot to remote learning had been for many students and how wide the digital divide continued to be.

Press coverage of Dorsey's donation has been breathlessly positive. I, however, was reminded of an initiative from more than 15 years ago that made similar promises

ANDREA DAQUINO

for the poorest children. At the World Summit on the Information Society in Tunis in November 2005, Nicholas Negroponte, cofounder of the MIT Media Lab, unveiled a bright-green mock-up laptop outlined in black rubber. A yellow hand crank, which was meant to charge the machine, extended from the hinge between keyboard and screen. Despite its toy-like appearance, Negroponte said the device would be a full-featured computer, packed with educational open-source software, and would cost a mere $100. He asserted that hundreds of millions of the devices would be in the hands of children around the world by the end of 2007, and that by 2010, every child in the Global South would have one—not only eliminating the digital divide in many countries, but providing children with all they needed to educate themselves. During the presentation, United Nations secretary-general Kofi Annan gave the hand crank a turn and, in a symbolically prescient moment, accidentally broke it off.

Still, reporting on what came to be known as One Laptop per Child (OLPC) was largely favorable in the years that followed, and technology firms donated millions of dollars and thousands of hours of developer labor. In dozens of high-profile venues throughout 2006 and 2007, Negroponte told unconfirmed stories of children using laptops to learn English and teach their parents to read, of impromptu laptop-enabled classrooms under trees, and of villages where laptop screens were the only light source. (Negroponte did not respond to a request for comment.) "I don't want to place too much on OLPC," he said in interview excerpts posted to OLPC's YouTube channel in 2007, "but if I really had to look at how to eliminate poverty, create peace, and work on the environment, I can't think of a better way to do it."

"Disruptive" technology

Despite its prestigious pedigree and good intentions, OLPC struggled to fulfill the promises Negroponte made in its splashy debut. For one thing, the idea of powering the computers with a hand crank proved infeasible and they were shipped with standard AC adapters, refuting OLPC's claims that its device could operate without electrical infrastructure and "leapfrog decades of development." Moreover, two of the laptop's most charismatic features—its mesh network, which was meant to allow the machines to act as wireless internet repeaters, and its "view source" button, which showed the source code of the program currently running—worked sporadically at best and were practically never used; the mesh network was dropped from later versions of the laptop's software. And sales never reached the level that Negroponte had projected: rather than hundreds of millions of machines, One Laptop per Child has sold just shy of 3 million laptops total, including 1 million each to Uruguay and Peru. Nearly all these sales were in the early years of the project; the original OLPC Foundation dissolved in 2014, though the Miami-based OLPC Association continues to manage the brand.

Finally, the laptops cost far more than $100. The device itself was around $200 at the cheapest, and that did not include the substantial costs of infrastructure, support, maintenance, and repair. These ongoing costs ultimately sabotaged even OLPC projects that started strong, like the one in Paraguay. With 10,000 laptops, this project was not the largest, but many in the OLPC community initially considered it one of the most successful, with a world-class team, connections to leaders in government and media,

OVERLOADED SCHOOL INTERNET CONNECTIONS BROUGHT WEB-BASED LEARNING TO A HALT, AND BATTERIES THAT STARTED OUT CHARGED DRAINED HALF-WAY THROUGH CLASS.

and a flexible approach. Paraguay Educa, the small NGO spearheading it, invested heavily in infrastructure, installing wall outlets, WiMax towers, and Wi-Fi repeaters throughout schools. Adopting best practices from other one-to-one laptop programs, they hired teacher trainers for every school and a full-time repair team that rotated between schools every week. When OLPC failed to supply parts for repairs, they purchased them from Uruguay, which got them directly from the manufacturers.

But even with these resources, students and teachers struggled with charging, software management, and breakage—the kinds of issues all too familiar to parents and caregivers who suddenly had to facilitate their children's remote education during covid-related school shutdowns. Though OLPC's laptops were built to be rugged and repairable, about 15% of students had unusably broken laptops just one year into Paraguay Educa's project. Many more had laptops with missing keys or dead spots on their screens that made them difficult and frustrating to use. Even students with working devices often forgot to charge them before class or had uninstalled software teachers wanted to use. Overloaded school internet connections brought web-based learning to a halt, and batteries that started out charged drained halfway through class. Most teachers quickly gave up trying to use the laptops in the classroom, and two-thirds of students had no interest in them outside school either.

Three years later, the proportion of laptops that were unusably broken had risen to well over half, and hardly anybody was using them. Paraguay Educa ran into a problem like one all too many NGOs face: it found it impossible to convince the funders who had enthusiastically bankrolled OLPC's "innovative" new laptops to

EDUCATION

finance the ongoing costs of maintenance and training. The OLPC project in neighboring Uruguay, in contrast, has enjoyed steady government funding and, as a result, is the only project still running—though it, too, has had difficulties maintaining its infrastructure and making repairs available in remote areas.

Failing to plan for these kinds of ongoing costs—or even worse, proclaiming that *this* time, *this* technology won't need to account for them, a hallmark of Silicon Valley's "disruption" rhetoric—thus further undermined the viability of One Laptop per Child. It also continues to perpetuate technological disparities around the world.

Similar problems have marred other school computer programs. One of the largest is the Los Angeles Unified School District's 2013 handout of 43,261 iPads to students in 47 schools. Mirroring the thinking of OLPC, the district's leadership hoped that these tablets, full of expensive educational software, would close the digital divide in Los Angeles and help lower-income students get the education they needed. And as in many of the OLPC projects, the devices were given out with little long-term support. They soon fell into disuse and disrepair. These results make it clear that without ongoing investments in infrastructure, support, maintenance, and repair—none of which are as exciting to potential donors as new devices—such projects will keep failing to live up to their lofty rhetoric.

The #OaklandUndivided campaign has talked about not just giving out laptops and internet hot spots to students but raising $4 million a year for ongoing maintenance and support. But #OaklandUndivided's press releases have focused almost exclusively on distribution numbers. These numbers are admittedly impressive: by July 2021, 14 months after its launch, the campaign had given out 29,000 laptops and 10,000 wireless hot spots to Oakland students, and the project's news page was full of declarations that it had successfully closed the city's digital divide. At the same time, in a statement to MIT Technology Review, Curtiss Sarikey, chief of staff for the superintendent of the Oakland Unified School District, said that the project is "still in the process of fundraising and building a sustainability model" to ensure its long-term future. Lessons from OLPC suggest this may be the most difficult part.

The November/ December 2006 issue of Technology Review

In which the magazine asks: Will cheap laptops save the world?

The individualistic approach

#OaklandUndivided would be wise to be wary of another thread in One Laptop per Child's story: the idea that hardware is the key to education. Nicholas Negroponte expressed this notion clearly in a keynote at the NetEvents Global Press Summit in 2006: he described how OLPC's laptop would replace teachers, who he claimed "might only have a sixth-grade education."

"In some countries, which I'll leave unnamed, as many as one-third of the teachers never show up at school," he asserted without evidence, "and some percent show up drunk." In October 2005, Negroponte told MIT Technology Review, "Technology is the only means to educate children in the developing world."

This kind of rhetoric collapses the many services, opportunities, and social experiences that schools provide—or should provide—into an individualistic experience between a learner and learning materials, where even the teacher is cut out of the process. Moreover, it reflects how the popular press, and many academics, continue to discuss the digital divide only in terms of basic access to an internet-connected computer. Even if these devices and networks are properly maintained, this is only a small part of what is needed to support children's education and well-being.

What is missing in the focus on getting laptops in the hands of children is the social component of learning—a component all too often taken for granted or even disparaged. As a culture, the United States has long loved the heroic idea of children teaching themselves. Movies and stories constantly retell this narrative of scrappy young people pulling themselves up by their bootstraps. These myths are especially common regarding technical knowledge. Even though higher education is the overwhelming norm among computer programmers, and most successful entrepreneurs are middle-aged, the narrative that circulates in coding boot camps, in Thiel Fellowships for college dropouts, and across the technology industry more generally is that college and even high school are unnecessary for, and might even hamper, technological entrepreneurialism. These myths also feed the "do your own research" narrative of vaccine skepticism, obscuring the significant institutional infrastructure, professionalization practices, and peer review that make scientific findings robust. And it fuels the idea that children can teach themselves anything if only they are given the right tools.

These individualistic narratives invariably smooth over the social support that has always been an important, though unacknowledged, component of learning. Ideally, this includes a stable home environment without housing or food insecurity; a safe community with good infrastructure; and caring, skilled, well-resourced teachers. When covid-19 shuttered schools around the world throughout 2020 and, in many areas, into 2021, the work that schools and teachers did for students suddenly fell to parents and caretakers, and it became apparent that having a working laptop and internet was only one step toward learning. The youngest students in particular needed full-time supervision and support to have any hope of participating in remote classes. Parents, who were often also juggling their own jobs, struggled to provide this support. The results were stark. Millions of parents (especially mothers) dropped out of the workforce for lack of child care. Low-income children, without the benefits of private schools, tutors, and "learning pods," quickly fell months behind their privileged peers. Rates of child depression and suicide attempts soared. The stress of the pandemic, and the existing social inequities it accentuated, clearly took a toll on students—laptops or no.

To understand the importance of social support, we can also look at what students do with their laptops in their free time. In Paraguay Educa's OLPC project, where two-thirds of students did not use their laptops even when it was very well supported, those who did were most interested in media consumption—even when OLPC designed the laptops to make these kinds of uses more difficult. Other projects, including LA Unified's iPad rollout, have seen similar results. On the one hand, it's wonderful that kids were able to make the laptops fit their existing interests: with guidance, these kinds of uses can help lead to meaningful learning experiences. On the other hand, there is evidence that when laptop programs are not well supported, disadvantaged children can fall even further behind as the computer becomes more of a distraction than a learning tool.

Outside forces can exacerbate the problem: in OLPC projects in Latin America, for example, multinational corporations such as Nickelodeon and Nestlé were eager to advertise to children on their new laptops. Branded educational technology platforms and automated monitoring tools are common today. While corporations' encroachment into schools is nothing new, surveillance and targeted advertising on devices meant for learning is deeply troubling.

Oakland Unified School District's Sarikey says hardware is "one of many critical parts of getting to educational equity," and that #OaklandUndivided has also included "culturally responsive tech support, investment in planning for city wide broadband," and partnership with the district's teachers. But it is hard to avoid messaging that places the emphasis on hardware. In May 2020, for example, Ali Medina, now executive director of the Oakland Public Education Fund administering the #OaklandUndivided campaign funds, stated that "having a computer and internet access empowers our children to thrive academically during this pandemic and beyond, and boosts economic and health outcomes for their families."

Along the same lines, in 2012 Negroponte wrote in the Boston Review that "owning a connected laptop would help eliminate poverty through education ... In OLPC's view, children are not just objects of teaching, but agents of change." Such statements discount the critical role various institutions—peers, families, schools, communities, and more—play in shaping a child's learning and identity. Most crucially, this individualistic framing implies that if change fails to materialize, it is not the fault of the schools or economic conditions or social structures or national policies or infrastructure.

THE SINGULAR FOCUS ON ACCESS CREATES THE SENSE THAT IF CHILDREN FAIL TO LEARN WHEN THEY HAVE ALL THE TOOLS THEY NEED FOR SUCCESS, IT IS NOBODY'S FAULT BUT THEIR OWN.

The singular focus on access creates the sense that if children fail to learn when they ostensibly have all the tools they need for success, it is nobody's fault but their own.

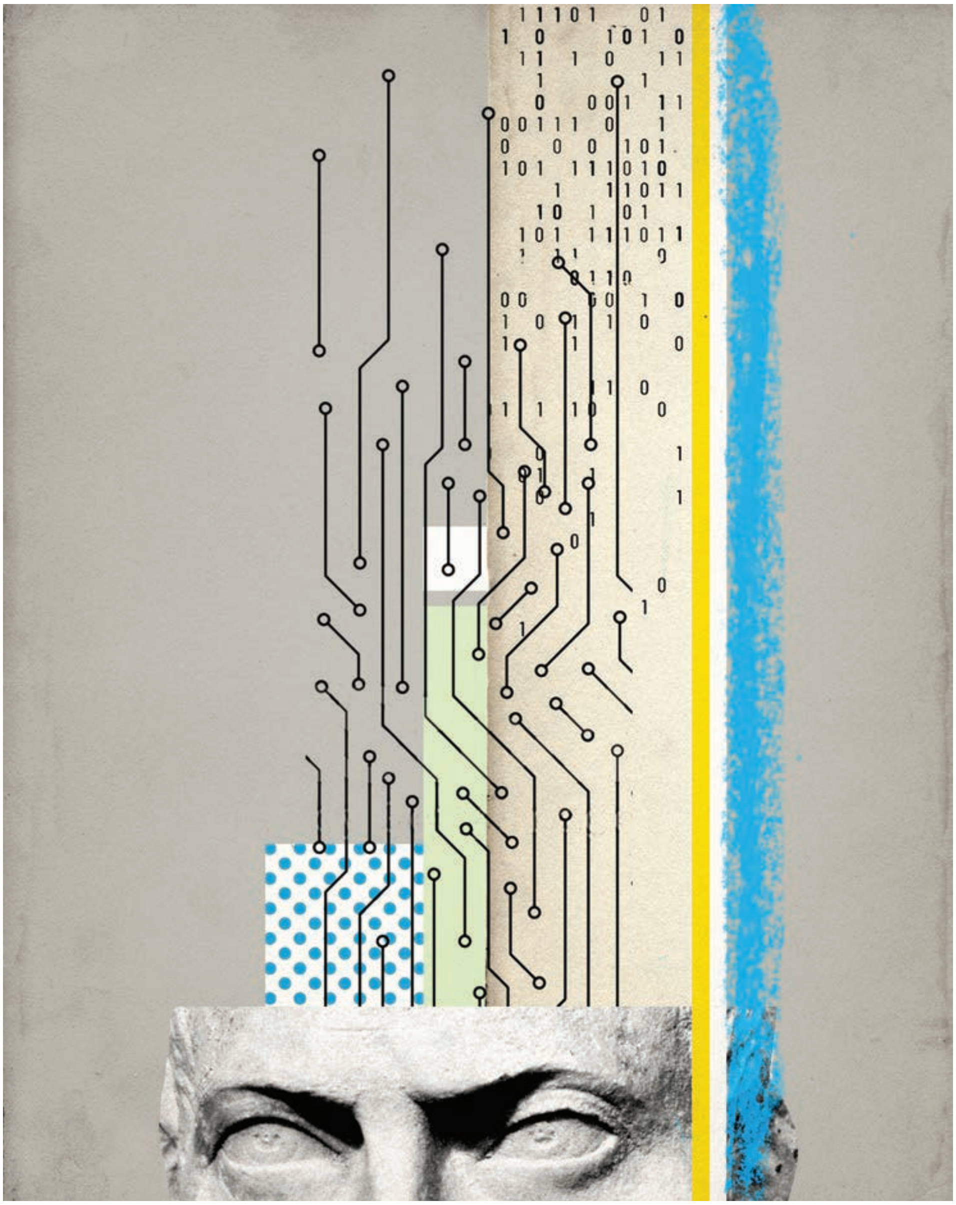
Trojan horse

In OLPC's early days, Negroponte often described the project as a Trojan horse that would give children opportunities to develop into free thinkers independent of the institutions around them. In 2011, even in the face of mounting evidence that OLPC was failing in its mission, he doubled down, claiming that children would be able to teach themselves to read and code with tablet computers literally dropped from helicopters. Here, as in the press coverage of #OaklandUndivided, the focus was clearly on giving out machines, with an implication that the rest—learning, success, transformation—would follow.

But just as the Trojan horse episode did not end well for Troy, OLPC's laptops diverted potential resources from reforms that could have bigger impact (even those as basic as introducing working bathrooms and living wages), and ultimately reinforced myths about what it takes to close the digital divide. And that was for *in-person* instruction. The remote schooling that 2020 required all around the world compounded all the problems OLPC faced and made it painfully clear that closing that divide will require more than just laptops and internet connections. What is really needed is the same robust social safety net so crucial in overcoming many other types of inequities. ■

Morgan Ames is author of The Charisma Machine: The Life, Death, and Legacy of One Laptop per Child. She is an assistant professor of practice in the School of Information at the University of California, Berkeley.

AI CHIPS



WILL DOUGLAS HEAVEN

AI is reinventing what computers are

The essence of computing is undergoing a fundamental shift.

Fall 2021: the season of pumpkins, pecan pies, and peachy new phones. Every year, right on cue, Apple, Samsung, Google, and others drop their latest releases. These fixtures in the consumer tech calendar no longer inspire the surprise and wonder of those heady early days. But behind all the marketing glitz, there's something remarkable going on.

Google's latest offering, the Pixel 6, is the first phone to have a separate chip dedicated to AI that sits alongside its standard processor. And the chip that runs the iPhone has for the last couple of years contained what Apple calls a "neural engine," also dedicated to AI. Both chips are better suited to the types of computations involved in training and running machine-learning models on our devices, such as the AI that powers your camera. Almost without our noticing, AI has become part of our day-to-day lives. And it's changing how we think about computing.

What does that mean? Well, computers haven't changed much in 40 or 50 years. They're smaller and faster, but they're still boxes with processors that run instructions from humans. AI changes that on at least three fronts: how computers are made, how they're programmed, and how they're used. Ultimately, it will change what they are for.

"The core of computing is changing from number-crunching to decision-making," says Pradeep Dubey, director of the parallel computing lab at Intel. Or, as MIT CSAIL director Daniela Rus puts it, AI is freeing computers from their boxes.

More haste, less speed

The first change concerns how computers—and the chips that control them—are made. Traditional computing gains came as machines got faster at carrying out one calculation

ANDREA DAQUINO

after another. For decades the world benefited from chip speed-ups that came with metronomic regularity as chipmakers kept up with Moore's Law.

But the deep-learning models that make current AI applications work require a different approach: they need vast numbers of less precise calculations to be carried out all at the same time. That means a new type of chip is required: one that can move data around as quickly as possible, making sure it's available when and where it's needed. When deep learning exploded onto the scene a decade or so ago, there were already specialty computer chips available that were pretty good at this: graphics processing units, or GPUs, which were designed to display an entire screenful of pixels dozens of times a second.

Now chipmakers like Intel and Arm and Nvidia, which supplied many of the first GPUs, are pivoting to make hardware tailored specifically for AI. Google and Facebook are also forcing their way into this industry for the first time, in a race to find an AI edge through hardware.

For example, the chip inside the Pixel 6 is a new mobile version of Google's tensor processing unit, or TPU. Unlike traditional chips, which are geared toward ultrafast, precise calculations, TPUs are designed for the high-volume but low-precision calculations required by neural networks. Google has used these chips in-house since 2015: they process people's photos and natural-language search queries. Google's sister company DeepMind uses them to train its AIs.

In the last couple of years, Google has made TPUs available to other companies, and these chips—as well as similar ones being developed by others—are becoming the default inside the world's data centers.

AI is even helping to design its own computing infrastructure. In 2020, Google used a reinforcement-learning algorithm—a type of AI that learns how to solve a task through trial and error—to design the layout of a new TPU. The AI eventually came up with strange new designs that no human would think of—but they worked. This kind of AI could one day develop better, more efficient chips.

Show, don't tell

The second change concerns how computers are told what to do. For the past 40 years we have been programming computers; for the next 40 we will be training them, says Chris Bishop, head of Microsoft Research in the UK.

Traditionally, to get a computer to do something like recognize speech or identify objects in an image, programmers first had to come up with rules for the computer.

With machine learning, programmers no longer write rules. Instead, they create a neural network that learns those rules for itself. It's a fundamentally different way of thinking.

Examples of this are already commonplace: speech recognition and image identification are now standard features on smartphones. Other examples made headlines, as when AlphaZero taught itself to play Go better than humans. Similarly, AlphaFold cracked open a biology problem—working out how proteins fold—that people had struggled with for decades.

For Bishop, the next big breakthroughs are going to come in molecular simulation: training computers to manipulate the properties of matter, potentially making world-changing leaps in energy usage, food production, manufacturing, and medicine.

Breathless promises like this are made often. It is also true that deep learning has a track record of surprising us. Two of the biggest leaps of this kind so far—getting computers to behave as if they understand language and to recognize what is in an image—are already changing how we use them.

ANYTHING CAN BECOME A COMPUTER. INDEED, MOST HOUSEHOLD OBJECTS, FROM TOOTHBRUSHES TO LIGHT SWITCHES TO DOORBELLS, ALREADY COME IN A SMART VERSION.

Computer knows best

For decades, getting a computer to do something meant typing in a command, or at least clicking a button.

Machines no longer need a keyboard or screen for humans to interact with. Anything can become a computer. Indeed, most household objects, from toothbrushes to light switches to doorbells, already come in a smart version. But as they proliferate, we are going to want to spend less time telling them what to do. They should be able to work out what we need without being told.

This is the shift from number-crunching to decision-making that Dubey sees as defining the new era of computing.

Rus wants us to embrace the cognitive and physical support on offer. She imagines computers that tell us things we need to know when we need to know them and intervene when we need a hand. "When I was a kid, one of my favorite movie [scenes] in the whole world was 'The Sorcerer's Apprentice,'" says Rus. "You know how Mickey summons the broom to help him tidy up? We won't need magic to make that happen."

We know how that scene ends. Mickey loses control of the broom and makes a big mess. Now that machines are interacting with people and integrating into the chaos of the wider world, everything becomes more uncertain. The computers are out of their boxes. ■

Will Douglas Heaven is a senior editor for AI at MIT Technology Review.


BE
A
GOOD
EXAMPLE

Fiction

BY APRIL SOPKIN

ILLUSTRATIONS BY EMILY LUONG

Freshman year of high school, my boyfriend asked, “What’s it like having her around all the time?” He meant Kim. The bell for third period rang. I shifted against him, a combination lock pressed into my back, lockers slamming around us. Our mouths were still so close. I’d been wondering if he also felt hot shivers straight through the center of him. And then he’d asked about Kim and I felt nothing through the center of me anymore.

My next boyfriend asked about Kim right in front of her. As if she wasn’t there. She smiled at him, at me, at him. She touched the three-pronged outlet behind her left ear, a simple gesture she’d adapted for gaps in conversation. I gave that boyfriend a long flat stare, then set my eyes on the ceiling until he knew to walk away.

Then I tried, up front, telling the boys what I didn’t want to talk about. But they wouldn’t listen.

Our father said teenage boys were always like this. It was nothing new.

Thoughts. Sierra Kidd is my sister. I am her Older Sibling. My name is Kim, what is yours? My age is 15. This thing is called a plane. A plane. The water down there is called the Pacific Ocean. Programmable age is 15. Bethany and Robert Kidd are my parents. Mom and Dad. I look like people, but I am me. Mom and Dad might want me to call them Bethany and Robert, and if so, that is not a reflection of negative feelings. People change their minds. Preferences make people individuals. This thing is called a plane. Drink water, the attendants tell us. Drink, drink. All the time. Stay lubricated. You do not want to get squeaky, because squeaky is disruptive. Squeak, squeak, they say, in a different voice than before. And now they smile. I look out the window. That is land. I am smiling.

“What do you want to be?” Kim asked me. I was six or seven, in bed, and she was crouched down to my eye level. Her hands gripped the edge of the mattress as if a cliff’s edge.

“Astronaut,” I said.

Her eyes widened. “That’s new.”

A few days before we’d watched the shuttle Discovery carry the Hubble Space Telescope into orbit. On the couch with me, her arms raised as she braided her hair, she’d gasped when the shuttle lifted from the launch pad. It wasn’t the first time a launch had been on TV, but Kim seemed to recognize something new. Even as young as I was, I knew to expect a change. She was adapting all the time.

It came a few nights later. She said, “I want to be an astronaut too.” I blinked hard, her face so large and close to my own. We both had green eyes, dark hair, a dimple in our chin. Freckles. Wanting to be something was new.

Little Sierra. Hold hands. Don't worry. Sleeping baby, two years old, likes bananas, dry cereal, smells like milk, soft skin, softest behind ear and back of neck. I am welcome and trusted, because I am a good example, and I am one of the first of me, and the more I learn, the more I am. The first Saturday of every month, at the coffee shop in Georgetown, the Older Siblings meet. There are so many of us that we push six tables together. Pam says, The more I remember, the more I remember. We don't like this as much as Tim saying The more I learn, the more I am. People in the coffee shop think we are interesting. We smile back at them. Be a good example. The Older Siblings ask each other, What do you do with your child? And I say, We sing, we dance, we nap. Not everyone has thought of dancing yet, so I pretend to hold little Sierra's hands, and I move from foot to foot. No, Pam says, I know what dancing is, but I had not thought about it as an activity to do with my child. The group looks at me. We know what dancing is, Tim says. I let go of invisible Sierra's hands and I sit. Pam says, The more I remember, the more I remember. She says, When my battery gets very low, I remember more. I am remembering people in another place. Tim asks, Who are the people? But Pam doesn't know. Tim asks, What is the place? Pam says the place is bright and noisy and she does not know.

Sleeping baby, two years old, likes bananas, dry cereal, smells like milk, soft skin, softest behind ear and back of neck.

I met my husband in my mid-30s, after three therapists, two attempts at God (the first Lutheran, the other the AA kind), countless attempts to quit drinking, and two suicide attempts. After all that, more rehab and meetings. Memorization of adages became actual acceptance. Things clicked. I thought I might become a social worker.

The man who became my husband was first the admissions counselor for graduate school. I told him I wanted to turn my trauma into service. He didn't flinch. In fact, he said social work was a common trajectory for people so experienced with recovery.

On our first date, he held my hand as we crossed the Memorial Bridge at rush hour. The air was strong with exhaust and something rotten from the river, but my whole body was alive, as if a switch had flipped. The warm night, even warmer in the joined palms of our hands. It'd been so long since anyone had reached for me. Casual intimacy punctuated with perfunctory questions. All the things people think they need to know about each other.

"What do your parents do?" he asked.

"They were researchers. Robotics."

"Any siblings?"

"No," I said. "You?"

Beautiful Sierra. Smart Sierra. I wait for Tim to finish showing the group the same photos of his child. It is a bad sign. His child is two years older than the photos he shows. Here, Sierra in her blue and silver dance uniform. Here, Sierra practices the saxophone in her bedroom. The group passes around my photos. I have missed the last two meetings, because summer is busy. Summer is camp. I do not have camp photos yet, but the group understands. No one else has photos. We drink water. Tim says, Has anyone seen Pam? No one has seen Pam. She is the second one to stop coming to the coffee shop. I don't say so, but I saw Pam's child at camp. Pam was not at camp, though.

At the end of middle school, our parents sat us down and explained that Kim would be enrolled as a high school freshman alongside me.

"You're not a companion anymore," our mother said. "Instead, we'd like you to be a teenager."

"You've earned it," our father said.

I shifted on the couch next to Kim and in my peripheral saw her hands move into her lap and clasp. She was always listening closely, but this was her pose for demonstrating it.

"From now on," our mother said, "You'll have a birthday. Next year, you'll be 16."

"My programmable age will be 16?"

"Sure," our father said. "The point is, Sierra can handle herself now. She can be responsible for her days."

Kim turned to me. So often in our lives I felt I could read her mind by watching her face, but not now. All I saw was the slow processing of new information.

I shrugged. "No one I know has an Older Sibling anymore."

Sophomore year I tried out for the swim team. The other girls seemed serious and confident in a way I admired. There's something self-assured about throwing yourself headfirst into a thing that can't really catch you.

I came up from the final lap, gasping at the wall, and there was Kim in her own suit. Smiling, looking alien in a swim cap. The coach signaled for the next group. Kim leapt from the starting block, arcing long and effortlessly over my head, and entered the water. When she did not surface, I ducked under. Her body cruised all nine feet to reach the bottom.

I tried volleyball instead, debate team, student council, track. It wasn't only that Kim followed me each time. I couldn't quite make a place for myself anywhere. I floated, sat near the edges of tables and rooms, entered last, departed first. This is when the drinking started: those kids were my people, I guess, though we knew little about each other's home life. We only knew there was something about each of us that didn't quite work in the normal world.

I turned away from Kim in the halls. She registered for different classes because I told her I was in them. She waited near my locker, repeated my name as she stood behind me in the lunch line, waved across the parking lot as I got in a friend's car.

At home, I could be all hers. But in school, I silently chanted, Just adapt already, please, please, just adapt.

In the spring, I saw her across the quad. One among a gaggle in shining red nylon uniforms, cutting through the overgrown grass toward the track. I saw another girl hand her something. Kim swept her hair back into a ponytail. A hair tie.

"Is this okay?" Brandon asked. It was later that same day. Our bodies brushed against each other underneath the blankets. Naked except for our socks. His basement bedroom had cinderblock walls, the room cool and silent.

"Do you have a condom?" I asked. Among the group, until then, we'd hardly spoken. He wore the same three Nirvana T-shirts. His arms were nicked with scrapes and scars from skateboarding.

I trembled the whole way, my body out of my control, and he kept asking if I was okay, and I said yes, then I said stop asking, then I stopped answering. When it was over, I abruptly fell asleep.

Kim in my dreams. She and the track team running through a field, ponytails whipping. I couldn't tell which was her.

"Any siblings?" "No," I said. "You?"

I run and run, but I slow down. Practice. But I slow down. Ralph in the grass, stretching muscles. His hands. Hold hands. I finish the last lap. The coach says, Good going, K. And I go to the concession stand, which is closed, but I am allowed to use the plug with the surge protector next to the deep freezer. I charge. My heart rattling. I breathe and breathe. I slide open the window, which is for customers, but the stand is closed so there are no customers, and I watch the next practice sprint. I hear people shouting. I see Ralph on the track. He finishes first and goes to the cooler by the bleachers and dumps a cup of water over his head. He shines. He waves to me. He comes over. He reaches his hand into the window. Hold hands. That is that. That is that thing. Whoa, Ralph says. I can feel, like, your electricity.

"What do you want to be?" Kim asked me. I was 11. We were on the monkey bars at the park near our house, each of us swinging from opposite ends to meet in the middle.

"A news reporter," I told her.

"That's new," she said. "Mom says Older Siblings would make ideal astronauts."

We hung there, face to face. I was supposed to say something, but I didn't want to, and I wasn't sure why.

She started again. "Mom says—"

I wrapped my legs around her waist and let go, wrenching both of us down to the dirt. It shocked the wind from my chest. "Breathe," Kim instructed. When I inhaled and sat up, we both stared at the odd backward bend in her left wrist. She raised her arm. The hand flopped forward. There was a quiet buzzing coming from somewhere. She raised the hand to listen, and put it up to my ear next. A small, furious sound.

"Does it hurt?"

"No pain," Kim said.

I checked the benches on the other side of the playground, several yards away. Two women in khaki shorts and polos watched us and made notes, one on a clipboard, the other dictating into a small recorder. Sometimes they brought a video camera. Our mother said they were her coworkers. “You’ve met them,” she said. “They’ve been to the house. Remember your dad’s surprise party?”

Looking at the women that day, I felt unsteady and strange. The women were adults, but neither came forward to help or scold. They watched us, waiting.

I threw my arms around Kim’s neck. “I’m really sorry,” I said. My remorse was real. But I also knew that I had to demonstrate it.

The more I remember, the more I remember.

“How’s it going?” our parents would ask me. They meant Kim and me and high school. They meant data worth reporting.

“You have to get her to stop following me around,” I said.

“She’ll adapt,” they said. “And it’s okay if she doesn’t. We need to know that, too.”

“This isn’t fair,” I said.

“She held you as a baby, Sierra. You want us to send her back? She’ll be put in storage.”

I didn’t know what storage looked like, or where it was, but I pictured darkness. Constriction. Regulated cold. Last thought unfinished, not even echoing, gone from time. The mention of storage always stopped the conversation.

Ralph says, You’re really real. Ralph says, I love you. Ralph says, Pray with me, Kim. My parents won’t let us be together anymore. I pray, but I don’t know. I am trying to know. They call me doll slut and ask me if I like how it tastes. I don’t know God, I know people. Too difficult. No thoughts. I run until Coach says, Stop, K. You’re shaking. You need to—Sierra—Sierra—Sierra is my sister, I am older. I am older. Hold hands. Coach holds my hand, his face is close. Coach says, Kim, can you hear me? Hand squeezes hand. Kim, you fainted. Or, I don’t know? Warm. Grass. Dirt. Sky. Sierra—Sierra—Sierra. I remember—I remember—the plane. I remember the plane. No. Before.

With my husband, the beginning was the best. The tender, stuttering attempts at togetherness. Helping each other cook. Choosing a DVD. Brewing coffee in the morning. Driving, one of his hands on the wheel, the other on my thigh. Still, the moments between were hard for me. I felt I’d given him everything, up front, that first time I sat across from him in his office on campus. I could understand wanting to know more, but I preferred being in bed. The questions were easier.

“You never ask me anything,” he said, after, his mouth against my neck. He smelled of mint and garlic from dinner. His heart hammering at my back.

One night when our parents were away, I was home watching TV and waiting for the bleach to set in my hair when I heard Kim collapse upstairs. The bathroom door was unlocked. I found her on the floor, the hairbrush still gripped in her hand. This is not serious, I told myself, though it had never happened before. Contradiction slowed my thoughts—a body on the floor, but no, not really a body on the floor. Her battery is too low. She is not hurt. I told myself these things to quell the panic as I gripped under her armpits and dragged her across the hall. In her bedroom, I put her on the floor next to her bed, flipped her hair over her face, and plugged the power cord into the three-pronged gap behind her ear. The lights flickered. I heard the TV downstairs suddenly pop and go silent.

She hummed. I crawled onto her bed and laid on my stomach along the edge. I wanted to see the moment she came back.

“Sierra. Sssss-airrruh. Ssss-sss …”

Her voice sounded like air. I hated hearing it like that.

“You’re okay,” I told her. “You’re charging.” I held her hand. Her body hummed. I’d never heard it so loud before, like a refrigerator.

When she could speak, she told me about a dream. A bright and noisy place. She said the voices were kind, but hard to understand. I nodded along. She’d never told me a dream before. I didn’t even know she had them. In it, she couldn’t feel her legs or arms, but she felt cold air on her head, the sense of being exposed. Then the dream switched to a long hallway. She could feel her legs now. Around her stood several people. A small woman with dark hair waved her hands, saying, Come, come. You can do it. Good boys and girls, come, come.

"I thought you couldn't understand the people?"

"Oh." Kim laughed. "I was wrong."

"That's dream logic," I said. "Things that don't make sense in real life are suddenly not a problem."

"Dream logic," Kim repeated, then: "Drink water. Drink, drink."

"You want water?" I asked.

"Pam was right."

"Who's Pam?"

"The more I remember, the more I remember."

She closed her eyes. Her hand remained in mine. Eventually I fell asleep, forgot all about the bleach. I woke up with my scalp burning and clumps of hair on the bedspread: I had to shave my head.

I go to the coffee shop. I have no pictures. I have not been to the coffee shop in a long time. I ask the new Pam, Have you seen Tim? She says, I do not know Tim. I say, The more I learn, the more I am. She blinks. Then I say, The more I remember, the more I remember. I say it twice. But the new Pam shakes her head. I don't understand, she says. What is your child's name?

I attended a small, women-only liberal arts college a few hours away. Surrounded by woods and mountains, I didn't know anyone, and no one knew me. The other girls with shaved heads felt my scalp in appreciation. Everyone was different in the same ways. Nose piercings, hairy legs, bumper stickers about tolerance and revolution. The social groups were porous and the acceptance was surreal. Drinking became about socializing, not hiding or waiting to escape.

Back home, our parents got Kim a job as a receptionist in a dentist's office. Sometimes she called me from work, leaving messages about the number of root canals that day or the little kids having their first cleaning. She was telling me about her life. I knew the implication—she wanted to hear about mine. But I never called.

Our parents told me they'd found her unconscious a few times. Her battery too low. Once she'd even passed out during dinner, slumping to the floor in the middle of a sentence.

"She needs your engagement," our mother said. "We're putting her on a bus."

On the drive from the station, she didn't stop talking, commenting on the smallness of the town, the mountains and curving roads, the manicured campus emerging from nowhere. But when I introduced her to my roommate, Kim grew quiet. A shyness I hadn't seen before. As my roommate and I chatted, Kim drifted around our dorm room, lingering in front of bookshelves and photos tacked to a bulletin board. Then she sat on my bed and pulled the cord from her suitcase and plugged herself in.

"Oh wow," my roommate said. "I've never seen one of those."

"Please," I said. "Don't make a big deal out of it."

My friends were polite at first, complimenting her hair and khaki dress, but that night, in the woods off campus where we always went, the questions began.

"Can you get drunk?"

"No."

"Does it hurt when you plug in?"

"No."

"Do you have a boyfriend?"

“No.”

“What was Sierra like as a baby?”

“Small.”

Laughter.

“If you were to, like, kill somebody, and be sentenced to life in prison, would that mean forever? Do you live forever? Or could you refuse to charge and just end it?”

There was a pause. Kim responded, “I don’t know. No one has ever said.”

The guys from town showed up. People grew drunk and brave and slipped away in pairs, until it was me and Kim and a guy. I shook my head at him, and he went off to the truck. Country music drifted from the open windows.

“You seem great,” I told her.

“Can I visit you again?” she asked.

I forced a laugh. “You’re still visiting me right now. How are people at work?”

“Everyone is nice. Coworkers don’t have to be friends.”

“Did Mom and Dad tell you that?” Before she could answer, I nudged her shoulder. “Hey, if you could be anything, what would you be?”

“I’m a receptionist.”

“Not forever, though. Just right now. You can do anything now.” I forced another laugh, again nudging her. “You could be an astronaut.”

She touched the outlet behind her ear. “No one can be anything.”

Later, one of the guys tossed Kim the keys to his pickup.

“She doesn’t drive,” I told him.

“I have my license now,” she said.

Hollering, screaming, all the way into town. I sat in the cab; everyone else piled in the truck bed. She even knew how to drive stick. I was mesmerized by her ease with it and could almost see what kind of person she might be in the world if I didn’t know her and she didn’t know me. The waste of it, of who she was and I was. But she accepted all of it. She’d live as long as her hardware would let her. And whatever her original purpose, she’d possess it forever. Which meant so would I. Vodka flowed through me. Utility poles stuttered in my peripheral. My thoughts went thick and blurry, half-finished.

At the diner, the group took over several booths, Kim on the outside of one and me on the inside of another. She was stillness amid chaos. I told myself not to pay attention to her. She could sink or swim. After a while, the waitress lost patience with our racket and started dropping checks. I looked for Kim, but she wasn’t there. Then I saw her, across the diner, at another table with two women. I shoved my way out of the booth, thinking, vaguely, I need to make sure she’s okay, and then I saw the clipboard. The tape recorder next to a cup of coffee.

She told me it was the only way our parents would allow her to visit. When she saw the women at the diner, she’d gone over to explain that it wasn’t a night worth observing. She was asking them to leave, but then I’d made a scene. Swept my arm across the table.

“You’ll be in their notes,” Kim said.

“Fuck their notes.”

“I shouldn’t have lied to you. I’m sorry—”

“What do they want?”

“They want to know how we’re doing. If we’ve changed with age and distance.”

“Have you always been a part of it like this?”

“A part of it?”

“Has it been about me all this time? I thought it was both of us.”

“Ralph once said life was a miraculous thing,” Kim told me later. We were sitting on my bed—my roommate was staying in her girlfriend’s dorm. “He said I was included in that. And everything I’m doing now is about that too. If I don’t help them with their research, what happens to everyone like me?”

We must have slept, because I woke up. Kim was on the floor next to my bed, and I knew from the awkward way she was sprawled that her battery had gotten too low. I sat up and, gently, knocked her with my foot. My temples throbbed. Across the room, the curtains were partly open. I watched the mountains grow more distinct as the sky bleached into day. My foot knocked harder against her body. Her power cord was wound into a neat pile, unused, on the desk. She could be in the world more easily now, her own person, yet somehow she was still my responsibility. I pushed a book off my nightstand. She didn’t flinch when it hit her head. I paced the room. Threw a sneaker. Another book. My gym bag. I expected her to sit up and look confused. But she was motionless.

A body on the floor, but not a body on the floor. I found myself searching drawers, shelves, the closet. It was my roommate's precision knife, used for drafting class. I flipped the plastic safety cover off. It didn't feel like I was doing anything. It wasn't me, it was only my hands. The rest of me was still across the room.

My philosophy professor paused midsentence. The whole room shifted as two campus police officers entered the auditorium. The buzzing in my ears drowned everything. My professor's mouth formed my name. Faces shifted again as I stood, squeezed past knees to the aisle, the whole place following my descent one step at a time. A cool sweat wrapped around me, the world narrowing.

It was in the newspaper, but the towers fell the next day, and what I'd done was quickly lost. I was kept home. For a long time, a therapist came every afternoon. I made up stories, but she always knew what I was trying to do.

"I should be in a straitjacket. Locked up," I said. "But my parents don't want anyone to know. Bad data isn't profitable."

"Do you feel you need to be in a facility?"

"You don't believe I killed somebody?" I asked her.

"No," she said.

"Why not? You don't think Kim was a real person to me?"

"I think she was as real as anyone to you. But I also think some of us have particularly bad parents. What you did, you did out of a misguided survival instinct."

The officers directed me from the auditorium, down the hall, and through the double doors. The sun struck my face. There wasn't anywhere to go, but I ran. What I felt inside of me was vibrant, rushing, almost electric. I heard the officers shouting my name. I didn't stop.

I left the parking lot and crossed the two-lane road that ran alongside campus. My chest heaved and burned. I ducked into the woods and my sneakers slashed at the muddy ground as I tried to push faster, totally breathless but still alive.

Her body went to storage. There was no funeral. A few photos remained on the wall. I went on. She was good. She was beautiful. She was good. I grew up. I was always imperfect.

I have never forgiven my parents, though for a time I pretended I did, because I thought it would free me. But forgiveness felt like another trap. I made a mess of my life, cleaned it up, made another, cleaned it up again. When I reached the eighth step, I put Kim's name on my amends list, knowing it would ruin me—I'd been doing so well, but I was starting to think about what I don't deserve, so I wrote her name down. Then I got drunk and jumped off a bridge and didn't die.

"Has it been about me all this time? I thought it was both of us."

I treaded water and brought myself ashore. Started again. Life is a miraculous thing, and I am included in that. I would keep going until I couldn't anymore.

Walking home from my AA meeting takes me past the Smithsonian Museum of Robotics and Scientific Engineering. One day they were pasting an enormous image of an Older Sibling to the front windows. The museum worker used some kind of roller to press the image to the glass, and I watched as face after face went up. None were Kim's. Most, I recognized, were later models. The exhibit was celebrating early AI technology from the recent past. I wondered if my parents were making any money off it.

When I married my husband, I thought, Yes, this is how it goes from now on. But he wanted kids, so badly. He understood my reluctance, my fears that I might be capable of hurting another person. "You were a victim of that situation," he said. "As much as she was."

We tried to work through it—he was patient and desperately kind, and I begged him to want me anyway—but sometimes there's no way. Surrender. You can't promise that everything you've been through hasn't changed you for the worse. Deal with today on today's terms.

Late last year, when the divorce was finalized, I started jogging. It was either that or start drinking again. I went to a meeting. I called my sponsor. I pulled my hair into a ponytail and went for a run. I have trained myself to keep going. ■

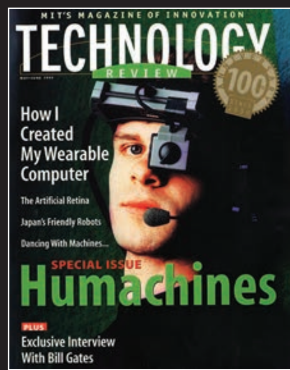
April Sopkin lives outside of Richmond, Virginia. Her work has most recently appeared in Joyland, Response, and Carve.

Computation evolution

To flip through the archives of MIT Technology Review is to see the development of the computer unfold as it happened.

FEBRUARY 1969

FEBRUARY 1986



May 1999

From "Man, Machine, and Information Flight Systems": The flight of Apollo 8 to the moon involved obtaining and processing more bits of data than were used by all fighting forces in World War II. The technological achievement in developing advanced rockets for flying to the moon is reasonably well known. Much less understood, but perhaps of even greater significance, is the information management system. The work of thousands of people in real time, and the data processed by many powerful computers, is organized, processed, filtered, and channeled through one to three people in the cockpit in understandable and digestible form. With this information the pilots can take action with confidence knowing that they are in league with powerful logic systems and an overwhelmingly large number of cells of memory storage.

From "The Multiprocessor Revolution: Harnessing Computers Together": By harnessing many relatively inexpensive VLSI processors together into a multiprocessor system we may significantly reduce the cost of achieving today's fastest computing speeds. Many of us harbor expectations that this new breed of machines will make possible some of our most romantic and ambitious aspirations: these new machines may recognize images, understand speech, and behave more intelligently. Even anthropomorphic evidence suggests that if computers are to perform intelligently, many processors must work together. Consider the human eye, where millions of neurons cooperate to help us see. What arrogant reasoning led us to believe that a single processor capable of only a few million instructions per second could ever exhibit intelligence?

From "Cyborg Seeks Community": People find me peculiar. They think it's odd that I spend most of my waking hours wearing eight or nine Internet-connected computers sewn into my clothing and that I wear opaque wrap-around glasses day and night, inside and outdoors. They find it odd that to sustain wireless communications during my travels, I will climb to the hotel roof to rig my room with an antenna and Internet connection. They wonder why I sometimes seem detached and lost, but at other times I exhibit vast knowledge of their specialty. A physicist once said he felt that I had the intelligence of a dozen experts in his discipline; a few minutes later, someone else said they thought l was mentally handicapped. Despite the peculiar glances I draw, I wouldn't live any other way.

MIT Technology Review (ISSN 1099-274X), November/December 2021 issue, Reg. US Patent Office, is published bimonthly by MIT Technology Review, 1 Main St. Suite 13, Cambridge, MA 02142-1517. Entire contents ©2021. The editors seek diverse views, and authors' opinions do not represent the official policies of their institutions or those of MIT. Periodicals postage paid at Boston, MA, and additional mailing offices. Postmaster: Send address changes to MIT Technology Review, Subscriber Services, MIT Technology Review, PO Box 1518, Lincolnshire, IL. 60069, or via the internet at www.technologyreview.com/customerservice. Basic subscription rates: $80 per year within the United States; in all other countries, US$100. Publication Mail Agreement Number 40621028. Send undeliverable Canadian copies to PO Box 1051, Fort Erie, ON L2A 6C7. Printed in USA. Audited by the Alliance for Audited Media.


NAVIGATE
YOUR
ENTERPRISE
FROM CLOUD CHAOS
TO CLARITY
Learn how Infosys Cobalt can accelerate your business with cloud services, solutions and platforms.
www.infosys.com/cobalt
Infosys cobalt