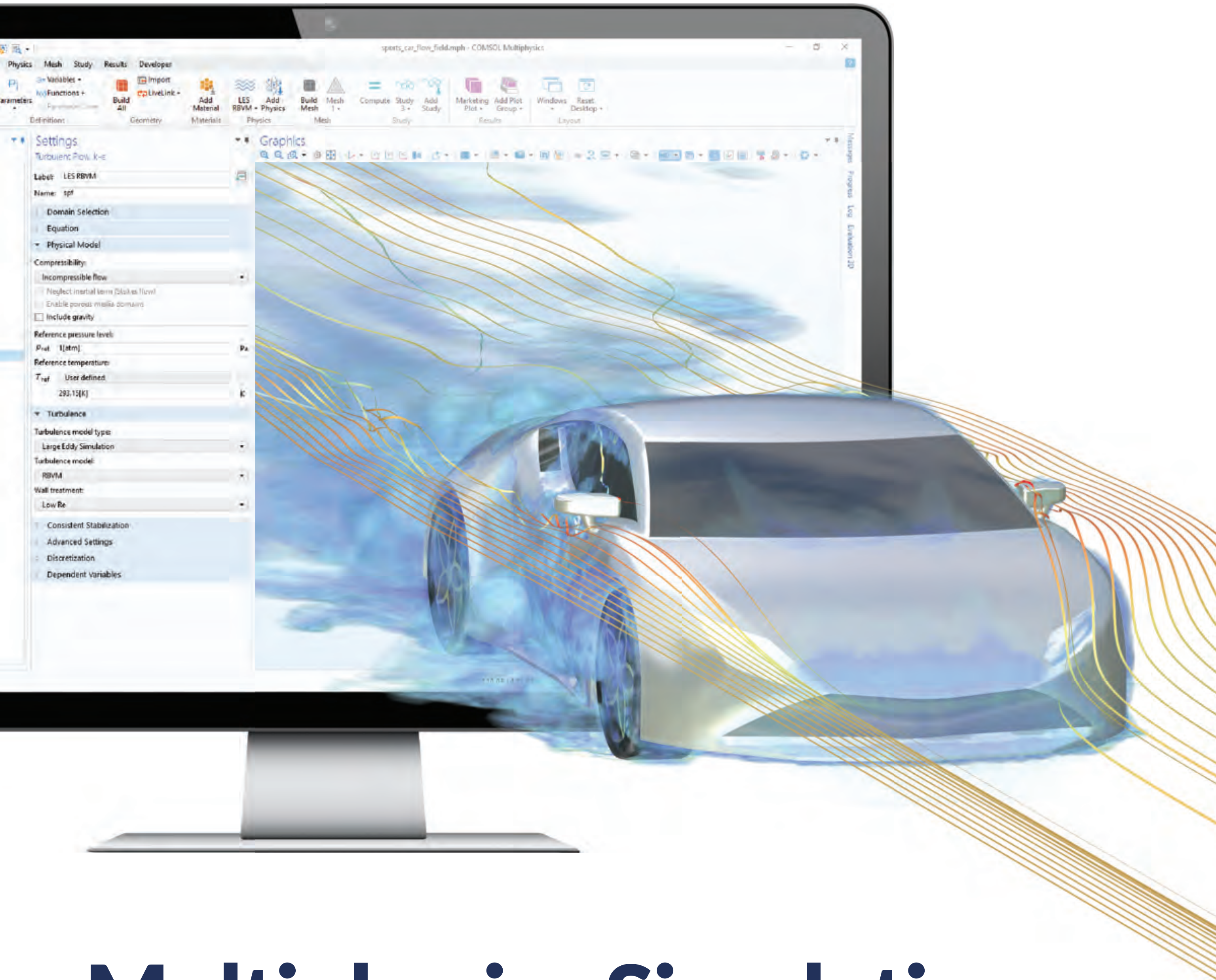
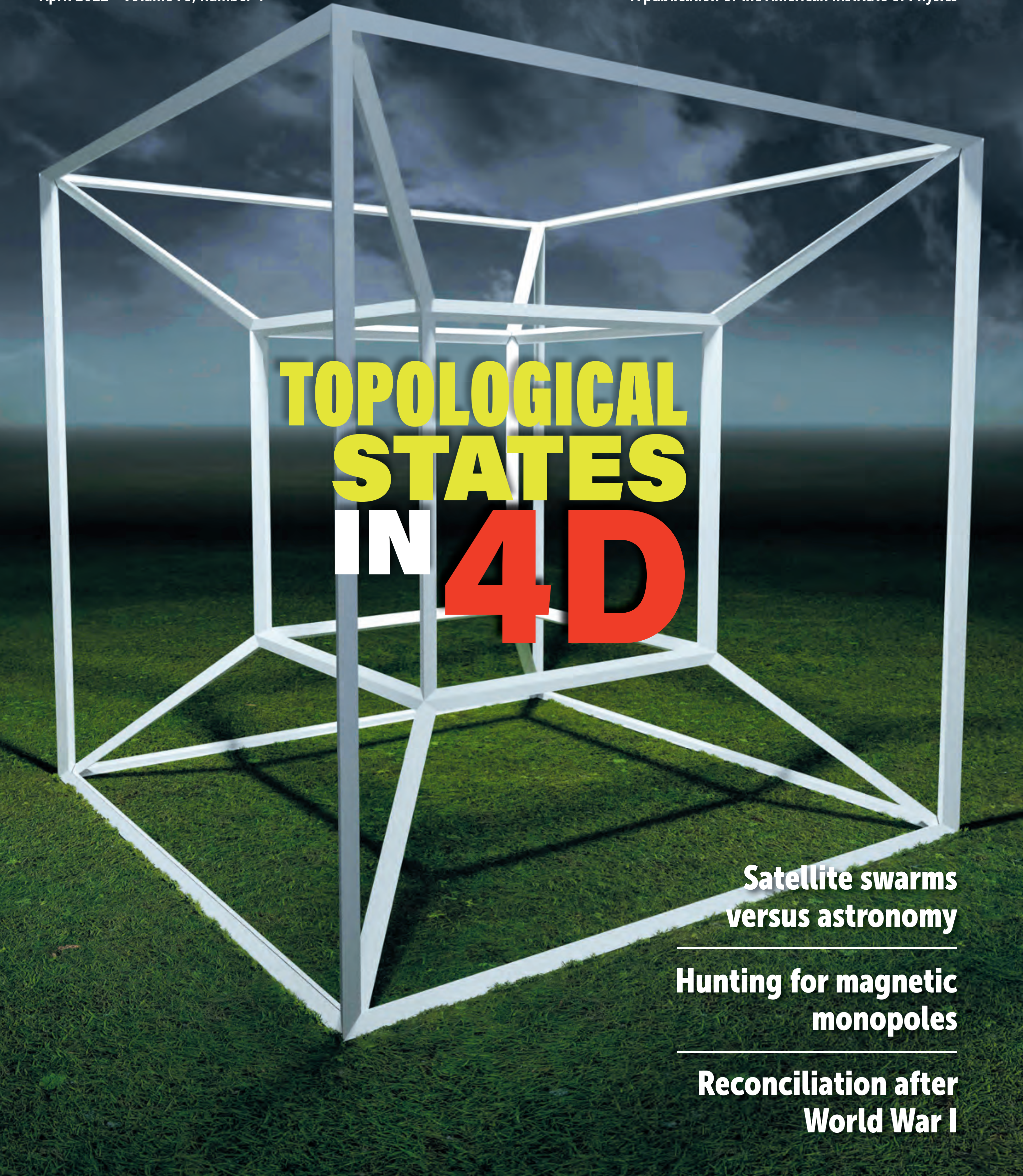

Settings
Graphics
Domain Selection
Equation
Physical Model
Turbulence
Large Eddy Simulation
Consistent Stabilization
Advanced Settings
Discretization
Dependent variables

PHYSICS TODAY

April 2022 • volume 75, number 4

A publication of the American Institute of Physics



TOPOLOGICAL STATES IN 4D

Satellite swarms versus astronomy

Hunting for magnetic monopoles

Reconciliation after World War I

Voltage Controlled Current Source

Your science. Our tools. SRS.

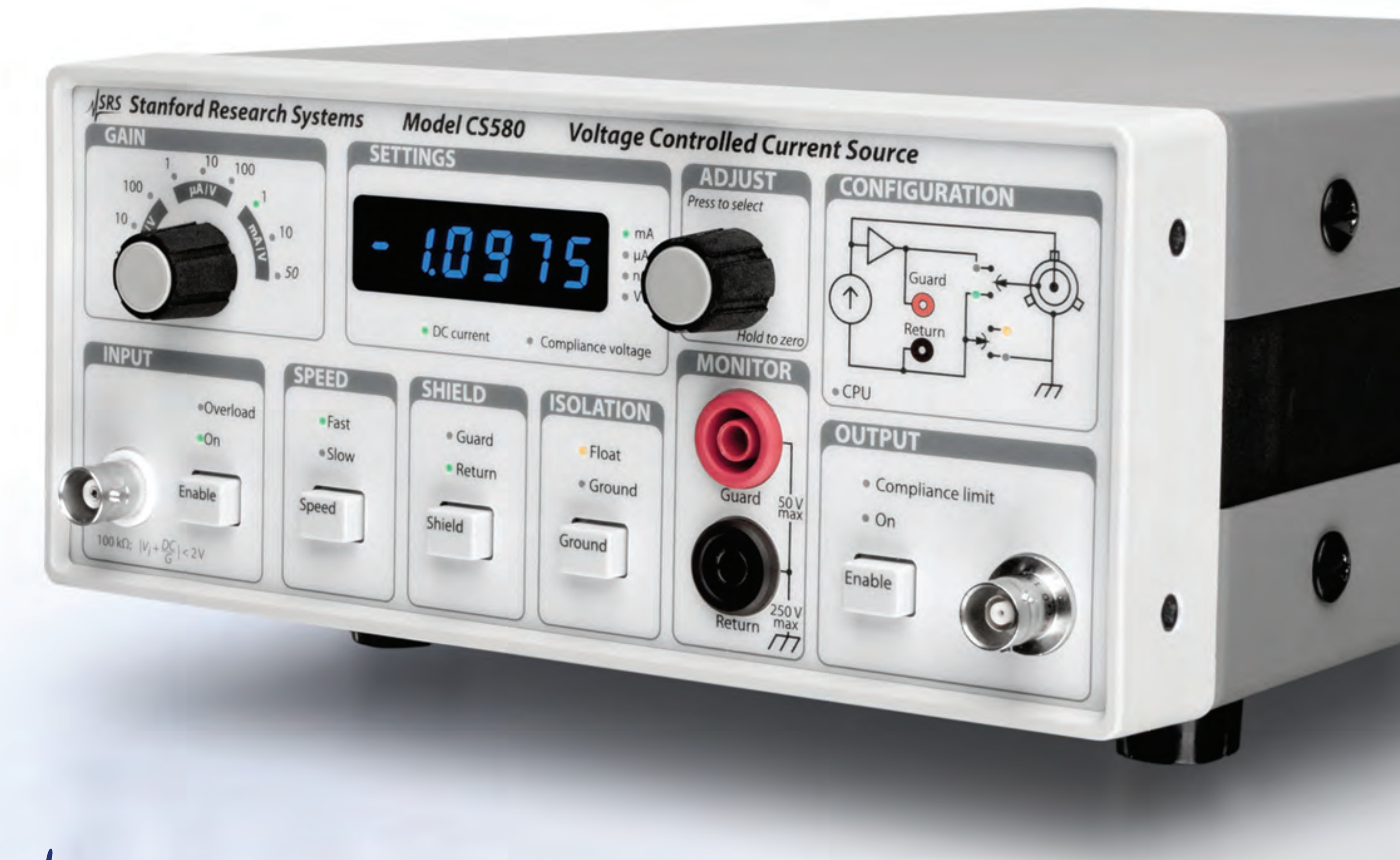
- Sources/sinks AC & DC current
- Analog voltage control input
- 1 nA/V to 50 mA/V gain
- 200 kHz bandwidth
- ±50 V compliance
- Ultra-low noise design
- RS-232 & optical fiber interfaces

The Model CS580 Voltage Controlled Current Source creates new capabilities for researchers needing an ultra-low noise current source in a flexible, easy to use instrument.

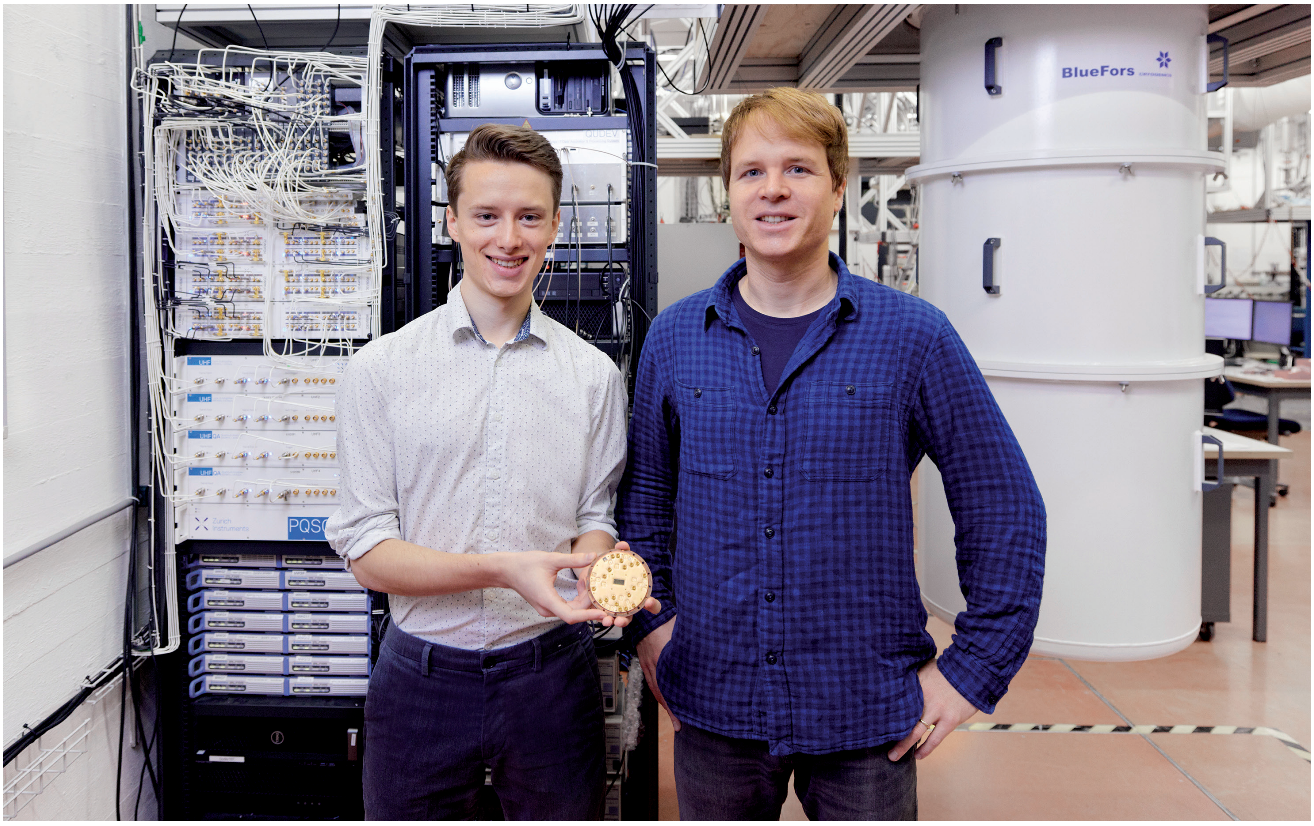
The CS580 is a natural companion to sensitive AC instruments such as lock-in amplifiers, providing a easy way of generating precision current directly from an AC or DC control voltage.

The CS580 is a welcome addition to any research lab studying semiconductors and transport phenomena, superconductivity, or nanotechnology, to name just a few.

Model CS580 ... $2995 (U.S. list)



SRS Stanford Research Systems
www.thinkSRS.com • Tel: 408-744-9040
www.thinksrs.com/products/cs580.html



Nathan Lacroix and Sebastian Krinner, ETH Zurich

Resilient quantum information processing

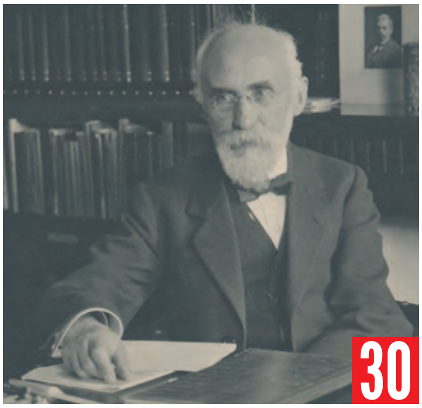
With the realization of a surface code of 17 qubits, the Quantum Device Lab at ETH Zurich takes a major step towards fault-tolerant quantum computing. For the first time, superconducting qubits were operated to form one stabilized logical qubit by continuously correcting for naturally occurring errors. The team demonstrated the protection of quantum states against unavoidable decoherence using a fast and precise correction scheme. Congratulations to Sebastian Krinner, Nathan Lacroix, and their colleagues on this impressive achievement!

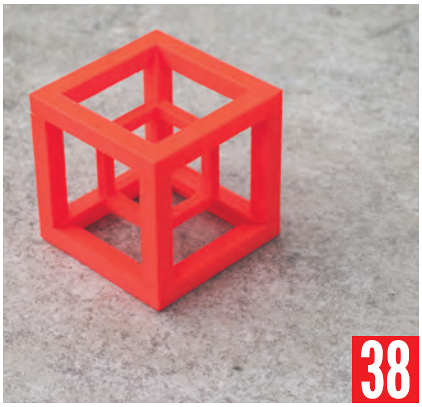
We are thrilled to support pioneering advances in quantum computing with the Zurich Instruments Quantum Computing Control System.

Zurich Instruments

PHYSICS TODAY

April 2022 | volume 75 number 4





FEATURES

ON THE COVER: The tesseract is the four-dimensional counterpart to the three-dimensional cube, much as the cube is the three-dimensional counterpart to the two-dimensional square. The pale gray structure on the cover is a common way to represent a tesseract in three dimensions. On **page 38**, Hannah Price describes both the rich theory of 4D lattices and the experimental tricks to simulate them in the 3D world. Inspired by the search for topological physics, the techniques work for ultracold atoms, photonics, acoustics, and electric circuits, among other systems. (Image by Robert Brook/Science Photo Library/Alamy Stock Photo.)

Recently on PHYSICS TODAY ONLINE

www.physicstoday.org

MARSTOURIST/PIXABAY

Mpemba effect

The idea of water freezing faster when it starts at a higher temperature, known as the Mpemba effect, is counterintuitive enough. Now researchers in Canada and France have demonstrated the inverse: that under certain conditions, a cold particle can heat up faster than a warmer counterpart.
physicstoday.org/Apr2022a

YONGKUN ZHANG

Fast radio bursts

The observed polarization of fast radio bursts increases and decreases with wave frequency, a recent study shows. The finding implies that the mysterious cosmic pulses originate in supernova remnants or other active regions where magnetic fields strip lower-energy waves of their initially strong polarization.
physicstoday.org/Apr2022b

LLNL

Fusion diagnostics

Companies in the fledgling nuclear fusion industry often make bold claims. Now, as part of a US government program, scientists are working with fusion firms to install diagnostic instruments and gather data on the performance of fusion plasmas and the experimental devices that manipulate them.
physicstoday.org/Apr2022c

PHYSICS TODAY (ISSN 0031-9228, coden PHTOAD) volume 75, number 4. Published monthly by the American Institute of Physics, 1305 Walt Whitman Rd, Suite 110, Melville, NY 11747-4300. Periodicals postage paid at Huntington Station, NY, and at additional mailing offices. POSTMASTER: Send address changes to PHYSICS TODAY, American Institute of Physics, 1305 Walt Whitman Rd, Suite 110, Melville, NY 11747-4300. Views expressed in PHYSICS TODAY and on its website are those of the authors and not necessarily those of AIP or any of its member societies.

PHYSICS TODAY

www.physicstoday.org

DEPARTMENTS

SUBSCRIPTION QUESTIONS? +1 800 344-6902 | +1 516 576-2270 | **ptsubs@aip.org**

Editor-in-chief
Charles Day *cday@aip.org*

Managing editor
Richard J. Fitzgerald *rjf@aip.org*

Art and production
Donna Padian, art director
Freddie A. Pagani, graphic designer
Cynthia B. Cummings, photographer
Nathan Cromer

Editors
Ryan Dahn *rdahn@aip.org*
Toni Feder *tf@aip.org*
Heather M. Hill *hhill@aip.org*
Abby Hunt *ahunt@aip.org*
David Kramer *dk@aip.org*
Alex Lopatka *alopatka@aip.org*
Christine Middleton *cmiddleton@aip.org*
Johanna L. Miller *jlm@aip.org*
Gayle G. Parraway *ggp@aip.org*
R. Mark Wilson *rmw@aip.org*

Online
Paul K. Guinnessy, director *pkg@aip.org*
Andrew Grant, editor *agrant@aip.org*
Angela Dombroski *atd@aip.org*
Greg Stasiewicz *gls@aip.org*

Assistant editor
Cynthia B. Cummings

Editorial assistant
Tonya Gary

Contributing editor
Andreas Mandelis

Sales and marketing
Christina Unger Ramos, director *cunger@aip.org*
Unique Carter
Krystal Amaya
Skye Haynes

Address
American Center for Physics
One Physics Ellipse
College Park, MD 20740-3842
+1 301 209-3100
pteditors@aip.org

PhysicsToday **@physicstoday**

AIP American Institute of Physics

Member societies
Acoustical Society of America
American Association of Physicists in Medicine
American Association of Physics Teachers
American Astronomical Society
American Crystallographic Association
American Meteorological Society
American Physical Society
AVS: Science & Technology of Materials, Interfaces, and Processing
Optica (formerly The Optical Society)
The Society of Rheology

Other member organizations
Sigma Pi Sigma Physics Honor Society
Society of Physics Students

FROM THE EDITOR

Physics and poetry

Charles Day

"The aim of science is to make difficult things understandable in a simpler way; the aim of poetry is to state simple things in an incomprehensible way. The two are incompatible."[1]

Paul Dirac's assertion to J. Robert Oppenheimer strikes me more as a witty aphorism than a heartfelt manifesto. Still, given that physicists and poets both strive for brevity and universality, it's worth exploring the similarities and differences of their respective vocations.

One commonality is youthful talent. The psychologist Dean Simonton has correlated the productivity of people in various intellectual occupations with their ages. Physicists, poets, and chess players tend to produce their best work early in their careers; novelists, biologists, and historians take longer to peak. A plausible explanation for the finding is that *everyone*'s intellectual ability is at its highest in their twenties. Practicing physics, poetry, and chess entails following a small set of clear rules, which can be acquired before one's peak is past. It takes longer to obtain biology's greater body of knowledge and to become comfortable with its fuzzier principles. Novelists need to observe a wide variety of life before they convincingly fictionalize it.

Another property that poets and physicists share is an appreciation for symmetry. The rhyme and meter of traditional poetry make lines symmetric and therefore delightful to digest. Symmetry is an important principle in physics, as David Gross observed in his feature article "Symmetry in physics: Wigner's legacy" (Physics Today, December 1995, page 46). Dirac and Subrahmanyan Chandrasekhar argued that mathematically neat formulations of physics were more likely to be valid than untidy ones.[2] Richard Feynman shared Chandrasekhar's appreciation for the beauty of mathematics—to a point. "If it disagrees with experiment it's wrong," he told students at his alma mater, Cornell University. "That is all there is to it."[3]

Favoring beautiful theories over ugly ones is, in my view, a prejudice. Conceivably, the theory of everything that unites the forces of nature could be, not a shapely, exotically dimensioned edifice, but a kludgy extension of the standard model of particle physics.

To research this essay, I trawled the Poetry Foundation's online collection of poems for instances in which physical phenomena are put to poetic use. I found barely a handful. Metaphors and similes work by tying an abstraction to a concrete image. To convey the pain of ingratitude, Shakespeare had King Lear say, "How sharper than a serpent's tooth it is to have a thankless child!" Physics is abstraction. Its use for metaphor and simile is limited.

The finest poem I found was "My Physics Teacher"[4] by David Wagoner (1926–2021), which was published in 1980. The first verse describes a comically clumsy teacher's failed physics demonstrations. The second and final verse continues:

> He believed in a World of Laws, where problems had
>   answers,
> Where tangible objects and intangible forces
> Acting thereon could be lettered, numbered, and crammed
> Through our tough skulls for lifetimes of homework.
> But his only uncontestable demonstration
> Came with our last class: he broke his chalk
> On a formula, stooped to catch it, knocked his forehead
> On the eraser-gutter, staggered slewfoot, and stuck
> One foot forever into the wastebasket.

Wagoner's poem recounts a futile attempt to teach physics to someone who became a professional poet. Are Physics for Poets courses also futile? Jon Miller, with the University of Michigan's Institute for Social Research, and his collaborators have compared scientific literacy in the European Union, the US, and other countries.[5] The US came second to Sweden, a high ranking that Miller attributes to compulsory science courses for nonscience majors.

References

1. H. S. Kragh, *Dirac: A Scientific Biography*, Cambridge U. Press (1990), p. 258.
2. S. Chandrasekhar, *Truth and Beauty: Aesthetics and Motivations in Science*, U. Chicago Press (1987).
3. R. Feynman, *The Character of Physical Law*, MIT Press (1967), p. 156.
4. D. Wagoner, *Poetry*, November 1980, p. 91.
5. For a summary of Miller's research, see A. Hobson, *Phys. Teach.* **46**, 404 (2008).

PT

READERS' FORUM

Some remarks about Rutherford

I read with interest Melinda Baldwin's article on Ernest Rutherford's publication strategies in relation to the journal *Nature* (PHYSICS TODAY, May 2021, page 26). She provides an excellent description of how, being in close competition with Pierre Curie and Marie Curie in France, Rutherford was conscious that working in Montreal put him—as he wrote in a letter to Otto Hahn on 6 January 1907—"on the periphery of the circle" and made it difficult to publish rapidly in Europe.

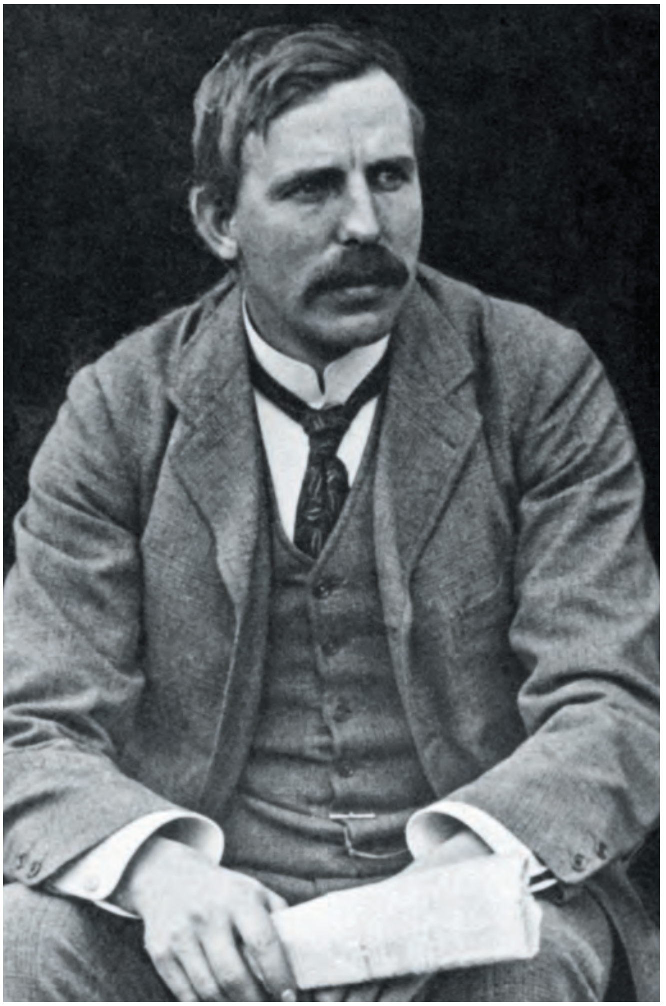
But there is more to the story. The discoveries of Rutherford and his colleague John McLennan, who was then also working on radioactivity at the University of Toronto, may have had a part in the creation of a means of rapid publication for the Royal Society of Canada (RSC).

Rutherford was elected a member of the RSC in 1900, and at the June 1904 meeting he was elected president of the RSC's section 3, devoted to mathematical, physical, and chemical sciences. As a section officer, Rutherford was a member of the RSC Council. In its annual report presented during the May 1905 general meeting, the RSC Council raised a problem related to the publication of the society's *Transactions*. The members pointed out that it was difficult to quickly publish a volume containing pieces from very different disciplines. In addition to Rutherford's section, the RSC also had members in the fields of biology and geology (section 4) and the humanities (sections 1 and 2). Though delays did not bother those in the humanities, it sometimes created problems for scientists.

The report insisted that "delay in the announcement of a scientific discovery may be very serious to original investigators, and, therefore, papers embodying important original results will not be sent to our volume of Transactions for publication" (reference 1, page II). Referring implicitly to Rutherford and McLennan's work on radioactivity, the report added the following:

> The revolution in scientific thought now in progress is fundamental, and some of our members are in the van of the movement. Conceptions of the constitution of matter which have been held for ages are now yielding to theories radically different, and laws established, even in recent times, are being profoundly affected. Under such conditions, and they have arisen very suddenly and recently, it might be well to inquire whether it would not be advisable to meet the emergency by issuing a bulletin. . . . In this way priority of discovery can be secured, and separate papers might be issued from the bulletin type. (reference 1, page II)

A committee, consisting of members of sections 2, 3, and 4, earlier had been formed to consider the idea of the bulletin. Presenting their report at the same 1905 meeting, they recommended a mechanism that would allow papers worthy of immediate publication, as judged by the secretary of the appropriate section, to be immediately printed. The author would receive a limited number of copies, which he could then distribute. Rutherford was absent, but Alexander Johnson, president of the society, moved the adoption of the report. McLennan seconded the motion, which was carried (reference 1, page XIII).



ERNEST RUTHERFORD did not make North American journals a major part of his publishing plans, but his discoveries may have motivated the Royal Society of Canada to create a rapid-publication mechanism. (Courtesy of the AIP Emilio Segrè Visual Archives, gift of Otto Hahn and Lawrence Badash.)

The first such bulletin was printed in June 1907. It contained two studies—one on radium and another on its emanation—that were carried out under Rutherford's direction. Notably, it appeared one month after Rutherford's departure for the University of Manchester in the UK. McLennan and William Kennedy's paper "On the radioactivity of potassium and other alkali metals" was later published in bulletin form as well.

So, even though "North American journals did not play a large role in Ruth-

CONTACT PHYSICS TODAY

Letters and commentary are encouraged and should be sent by email to **ptletters@aip.org** (using your surname as the Subject line), or by standard mail to Letters, PHYSICS TODAY, American Center for Physics, One Physics Ellipse, College Park, MD 20740-3842. Please include your name, work affiliation, mailing address, email address, and daytime phone number on your letter and attachments. You can also contact us online at **https://contact.physicstoday.org**. We reserve the right to edit submissions.

erford's publishing strategy," as Baldwin rightly notes, Rutherford nonetheless contributed to raising his colleagues' consciousness about the importance of rapid publication.[2] That led to improvement in the Canadian publication system, which then better served his Canadian colleagues and former students.

References

1. Council of the Royal Society of Canada, in *Proceedings and Transactions of the Royal Society of Canada*, 2nd ser., vol. 11 (1906).
2. For additional detail, see Y. Gingras, *Physics and the Rise of Scientific Research in Canada*, P. Keating, trans., McGill-Queen's University Press (1991), p. 86.

Yves Gingras
(gingras.yves@uqam.ca)
University of Quebec in Montreal
Montreal, Quebec, Canada

Melinda Baldwin's article in the May 2021 issue of Physics Today (page 26) makes a great contribution in demonstrating how Ernest Rutherford used *Nature* to get timely mention of his work in radioactivity and thus raised the weekly journal's profile in the new field. But it is disappointing that she refers to older works that introduced or propagated historical errors associated with Rutherford's early life and work. Such errors may be small, but more than 20 years after the publication of my book *Rutherford: Scientist Supreme* (1999), which is the only one to study original archives covering his early period, they occasionally make me wonder why I bothered.

Baldwin states that Rutherford "quickly distinguished himself as a talented student with a gift for physics and mathematics." But the records show that he was a normal kid who took two goes at each of the three scholarships he was awarded—for secondary school, university, and research overseas. In the last case, Rutherford was one of two candidates for the 1851 Exhibition Scholarship being offered to New Zealand for 1895. The other candidate, James Maclaurin, was judged better but declined the award.

But it is untrue, as Baldwin asserts, that Rutherford happened to graduate from Canterbury College (now the University of Canterbury) in the first year that students born in the British colonies could compete for the scholarship. The Royal Commission for the Exhibition of 1851 began offering the science research scholarships in 1891. They awarded one to a New Zealander every second year, starting in 1892 with David Jackson, a graduate of the Auckland University College. The scholarship for 1894 was pushed back a year, at the request of the University of New Zealand. If not for that delay, not only would Rutherford not have won the scholarship, but he would not have gone to Cambridge University because it did not accept non-Cambridge graduates for research until 1895.

The new policy is why Rutherford and other non-Cambridge graduates were treated as outsiders by the Cavendish Laboratory's Cambridge graduates. The usual progression to an academic position at Cambridge was receiving a Cambridge degree, then demonstrating in undergraduate laboratories and conducting research for a few years. Rutherford was the first non-Cambridge graduate accepted there for research, followed by John Townsend from Ireland. When Cambridge started accepting non-Cambridge graduates for research, the Cambridge graduates knew the path would now be much more competitive.

It is also untrue that Rutherford "developed a novel radio-wave detector back in New Zealand and brought it with him to Cambridge." His "wireless work" was done in England. In New Zealand he studied the magnetizing effect of very short current pulses ($^1/_{30\,000}$ of a second) during his first research year at Canterbury College in 1893. To generate pulses shorter than a fraction of a microsecond, in 1894 he used a Hertz oscillator to produce heavily damped oscillations lasting about one cycle that magnetized his detector needles. It was only when Rutherford got to England and placed his device in the receiving side of the Hertz oscillator to check its sensitivity that he carried out detection over a distance and in 1896 set a world record of half a mile.

Also, it is not exactly true to say that Frederick Soddy earned the 1921 Nobel Prize in Chemistry for his work with Rutherford. Officially, Soddy's prize was "for his contributions to our knowledge of the chemistry of radioactive substances, and his investigations into the origin and nature of isotopes." But it was mainly for the latter.

In 1901 the McGill Physical Society held a debate on the "existence of bodies smaller than an atom," with Rutherford chairing the meeting. Prior to the meeting, Rutherford wrote to J. J. Thomson that the goal was to "demolish the Chemists." Soddy, a temporary demonstrator in McGill University's chemistry department, spoke for the opposition and read a 32-page paper on "chemical evidence for the indivisibility of the atom." He condemned Thomson and Rutherford, who he said had "been known to give expression to opinions on chemistry in general and the atomic theory in particular which call for strong protest."[1] After the meeting, Rutherford, who had already determined that a radioactive gas emanated from thorium, asked Soddy to join in the chemical side of his research. In 1908 Rutherford received eight nominations for the Nobel Prize in Physics or Chemistry, but only one of those nominators proposed that he share the prize with Soddy.

Each of those small errors could have been corrected with just a minor change of wording if they had been known. There are many myths in circulation about Rutherford's life[2] (some of which I have recorded on the "Rutherford mythology" page of my website, www.rutherford.org.nz). He deserves to have his early life and work recorded accurately.

References

1. J. Campbell, *CERN Courier*, May/June 2019, p. 27.
2. J. Campbell, *AIP History Newsletter* **53**(2), 4 (2021).

John Campbell
(john.campbell@canterbury.ac.nz)
University of Canterbury
Christchurch, New Zealand

Quantum quibble

"Quantum" is a word with several definitions. As a noun—with the plural form "quanta"—it can refer to the smallest indivisible unit. It can also be an adjective meaning "quantum mechanical."

Yet I am here to argue against its use as an all-encompassing word referring to an alternatingly legitimate and overhyped morass of 21st-century research and the amusingly named "quantum industry." In addition to being vulgar, this usage is imprecise and leads to vague claims that cannot be evaluated—although that is part of its appeal.

Observe, if you like, the following sentence from page 26 of the December 2021 issue of Physics Today: "Israel's number of principal investigators 'at the core of quantum,' about 125, is low even for a small country, he says." Come again?

I am a skeptic of quantum technologies who has enjoyed years of research in quantum optics. But even zealous enthusiasts should seek a clear scientific language that aids in separating the wheat from the chaff.

The history of grammar is full of endless wars fought over trifles. But the precision of physics deserves protecting, and nothing is less precise than a catchall.

Paul Dieterle
(dieterle@g.harvard.edu)
Harvard University
Cambridge, Massachusetts

Another way to store energy

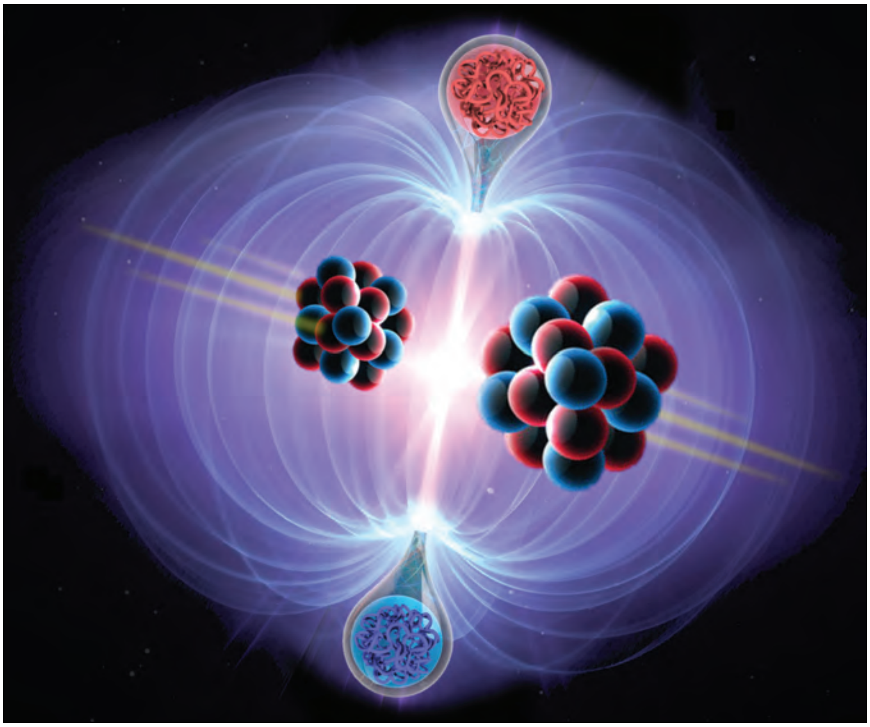
One method of energy storage that is notably absent from David Kramer's "Better ways to store energy are needed to attain Biden's carbon-free grid" (Physics Today, September 2021, page 20) is superconducting magnetic energy storage. SMES was studied extensively by Roger Boom and his group at the University of Wisconsin–Madison, beginning in the 1970s. Small SMES units have been built and operated successfully, which puts SMES ahead of some of the technologies mentioned in Kramer's item.

One of the advantages of superconducting magnets, as is well known, is that if they are sufficiently cooled, they can operate indefinitely. If properly designed, they also are not subject to the sort of catastrophic failure portrayed so colorfully in the James Bond film *The Man with the Golden Gun*.

John P. Heinrich
(jh@phoenixnmr.com)
PhoenixNMR LLC
Loveland, Colorado PT

A new search for magnetic monopoles

The latest results from CERN's Large Hadron Collider have established a lower mass limit for the elusive hypothesized particle.



THE COLLISION of two lead nuclei generates an exceptionally strong magnetic field. Magnetic monopoles and their antiparticles are theorized to be produced from the decay of that magnetic field, although they have yet to be observed. (Courtesy of James Pinfold, MoEDAL collaboration.)

A magnet always has a north and a south pole. But nothing in classical electrodynamic theory or quantum mechanics says that magnetic monopoles can't exist. They're the hypothetical analogues to electric charges in Maxwell's equations. In fact, their existence would make the equations more symmetrical: Electric terms could be transformed to magnetic ones, and vice versa (see the article by Arttu Rajantie, PHYSICS TODAY, October 2016, page 40).

Magnetic monopoles could be pointlike fundamental particles that carry magnetic charge, similar to electrons and described in 1931 by Paul Dirac.[1] Or they could be composite particles with a substructure, similar to neutrons or protons, as predicted by string theory, grand unified theories, and other explanations for physics beyond the standard model. Observing magnetic monopoles would provide evidence in support of such theories. And unlike the Higgs boson and other particles generated in collider facilities, monopoles are thought to be stable. Experimentalists could, therefore, track and manipulate them, potentially for specialized technologies.

The search for magnetic monopoles has so far come up empty. Although the now-defunct Tevatron, the Large Hadron Collider (LHC), and other particle accelerators were built mostly to study short-lived particles, researchers have used those facilities to search for magnetic monopoles. Elementary-particle collisions produce energies that are, theoretically, sufficient to produce monopoles with masses as high as a few trillion electron volts.

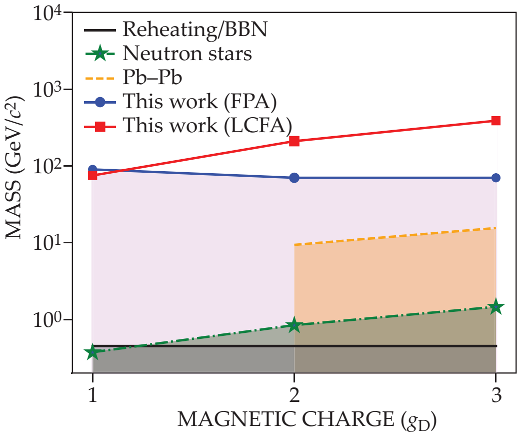
At collider facilities, the search has focused predominantly on monopole production via photon fusion or the Drell–Yan mechanism, in which the energy from the annihilation of a quark–antiquark pair is transformed to produce a point-like monopole and its antiparticle. Researchers might then detect a monopole by measuring the current it would generate in a superconducting ring, observing its strongly ionizing damage on a detector plate, or identifying signs of nucleon decay that would occur if a neutron or proton were to make contact with a monopole.

To make sense of collider measurements, researchers must have accurate estimates of the theoretical monopole production rate and momentum distribution. Otherwise, there's no concrete way to interpret whether the absence of a monopole signal is because of a low production rate or because monopoles don't exist.

The problem, however, is that pointlike and composite monopoles are expected to strongly couple to photons. That interaction has prevented researchers from using perturbation theory to reliably calculate their production cross sections, a measure of the likelihood of two particles interacting and subsequently generating the hypothesized monopoles (see PHYSICS TODAY, July 2006, page 16).

Another difficulty for detecting composite monopoles produced from elementary-particle collisions is that they're exponentially suppressed by a factor of $e^{-4/\alpha}$, where α is the electromagnetic fine-structure constant and has a value of about $1/137$. The suppression effectively makes monopoles unobservable and can be explained by an entropy argument: The probability of generating a coherent composite particle decreases dramatically as the number of objects in the system increases.

In 2018 the collaboration known as MoEDAL—the Monopole and Exotics Detector at the LHC—pursued a different detection strategy. The collaboration looked for the production of monopoles from heavy-ion collisions. That mechanism isn't exponentially suppressed, and the theoretical calculations of the monopole production rate are reliable. The team has now published its results.[2] Al-



MASS LIMITS have been placed for magnetic monopoles after the latest search at the Large Hadron Collider (LHC). Researchers used observations and two theoretical approximations of monopole production—free-particle approximation (FPA) and locally constant field approximation (LCFA)—to confirm at the 95% confidence level that they must have masses greater than 75 GeV. The new mass-exclusion region exceeds that from a previous LHC experiment (dashed orange line) and theoretical efforts that have considered cosmic monopole production from neutron stars (dashed green line) and Big Bang nucleosynthesis, or BBN (solid black line). (Adapted from ref. 2.)

though no magnetic monopoles were observed, the collaboration excluded the possibility of monopoles with masses smaller than 75 GeV, roughly 80 times as large as the mass of the proton.

Heavy ions

Smashing particles into one another at a collider facility isn't the only way to produce magnetic monopoles. The Big Bang would have had the conditions necessary to produce a lot of them. But if that's the case, where did they all go? If they exist, the question seems to be answered by cosmic inflation: The exponential expansion of space that occurred during the 10^{-36} of a second after the Big Bang would have greatly diluted their density and prevented them and their antiparticles from annihilating each other.

But so-called intermediate-mass monopoles could have conceivably been produced at the end of or shortly after cosmic inflation. That possibility has motivated researchers to look for monopoles among regular cosmic rays. Many observatories have hunted for evidence of cosmic monopoles in particle tracks and electric current fluctuations for the past few decades[3] (for more on those observatories, see the letters by Ken Frankel and Christopher Harrison, Physics Today, June 2017, page 13, and references therein). But those searches rely, of course, on a monopole traveling serendipitously through the facility's detector. The effort is akin to searching for a needle in a haystack.

Instead of elementary-particle collisions or astrophysical searches, the MoEDAL collaboration pursued a heavy-ion approach. The exceptionally strong magnetic fields created when heavy ions collide can give rise to the magnetic counterpart of the Schwinger mechanism, a vacuum-decay effect that produces electron–positron pairs in a decaying electric field.[4] Rather than electron–positron pairs, the decaying magnetic fields may create magnetic monopoles and their antiparticles. The particle production in both cases can be interpreted as quantum tunneling through the Coulomb-potential barrier. That mechanism, crucially, isn't limited by the exponential suppression of monopole production that plagues the elementary-particle-collision mechanism.

The second operational run of the LHC included a period of heavy-ion collisions in 2018. In anticipation of the MoEDAL experiment, Oliver Gould (now at the University of Nottingham), Arttu Rajantie (Imperial College London), and David Ho (now at MathWorks in Cambridge) developed a quantitative description of the production cross section and momentum distribution of monopoles that could be generated in lead–lead collisions at the LHC.[5]

Looking for monopoles produced via the Schwinger mechanism, rather than from elementary-particle collisions, meant that researchers could finally place mass limits on the hypothesized monopoles. Rajantie says that "the most important thing is that it's theoretically much easier to describe and to calculate how many monopoles you would actually expect."

Narrowing the search

In November 2018 a Pb–Pb collision at the LHC produced a magnetic field with a strength of 10^{16} T, the strongest ever observed in the universe. That's about 10 000 times as strong as magnetic fields found on the surfaces of neutron stars. To look for magnetic monopoles, the collaboration designed detector traps made from one ton of aluminum. The exceptionally large magnetic dipole moment of aluminum nuclei allows them to catch particles carrying a magnetic charge.

A DC superconducting quantum interference device (SQUID) then scanned the trapping volumes for the presence of magnetic charges. The signal for a monopole would be marked by the start of a steady current, whose value would depend on the magnetic charge of the monopole as it passed through the SQUID's coil. The current would continue to flow after the monopole had passed.

The researchers used two complementary methods with uncorrelated uncertainties to estimate the production cross section of magnetic monopoles in the LHC run. The first approach—free-particle approximation—calculates the spacetime dependence of the electromagnetic field produced in the heavy-ion collision but neglects self-interactions between monopoles. The second approach—locally constant field approximation—derives an exact solution for the magnetic monopole self-interactions but ignores the spacetime dependence of the magnetic field.

The work, however, wasn't finished with the newly calculated production cross sections. "We still had to translate them into the expected number of monopoles seen by the MoEDAL detectors," says Igor Ostrovskiy, a professor at the University of Alabama and a member of the collaboration. Ostrovskiy's graduate student Aditya Upreti worked on the challenging task of incorporating the new theoretical inputs into MoEDAL's Monte Carlo simulations that estimated the number of expected monopoles.

"At some point we were not sure if we would be able to calculate the trapping efficiency for all cases we needed—as simulations, ran with the help of the CERN's powerful computing infrastructure, were already taking weeks with no end in sight," says Ostrovskiy. After he, Upreti, and their colleagues carefully optimized the simulations, the calculations were completed for magnetic monopoles with Dirac charge of 1–3 g_D. (g_D is the minimum allowed magnetic charge and is equal to one-half the elementary charge e.)

The figure on this page shows the

estimated mass-exclusion region at the 95% confidence level for monopoles as a function of magnetic charge. Ostrovskiy says, "I, for one, was hoping to find monopoles! But we were still happy to produce the exclusion limits, as reliable limits help guide the theoretical development."

Complementary detectors

Although the MoEDAL collaboration didn't discover magnetic monopoles, the study's approach produced the most reliable calculation to date of the probable production rate of monopoles in strong magnetic fields. Furthermore, the negative result narrows the range in which future experiments will look for magnetic monopoles.

The search continues this spring: The LHC's third run will harness a beam with higher energy and five times the luminosity of that in the 2018 run. The MoEDAL experiment will use an updated detector to look for magnetic monopoles with higher mass and magnetic charge. Joining the aluminum trapping detectors will be nuclear tracking detectors consisting of stacked plastic sheets. When a highly ionizing particle rips through the sheets, the damage zone it leaves behind can be etched with a hot sodium hydroxide solution. Then an optical microscope identifies the precise path the particle traversed.

Other highly ionizing particles from beyond the standard model can emerge from heavy-ion collisions and may have strong electrical charges too. "If we do see something, it's going to be a real battle getting people to believe it," says James Pinfold, a physics professor at the University of Alberta and the MoEDAL spokesperson. "That's why we have the two methods."

The double-detector approach would use the nuclear tracker to reveal a monopole's path, and the trap would unambiguously identify the magnetic charge. Pinfold says, "If we do discover a monopole, it will be one of the most revolutionary discoveries of the century."

Alex Lopatka

References

1. P. A. M. Dirac, *Proc. R. Soc. A* **133**, 60 (1931).
2. B. Acharya et al., *Nature* **602**, 63 (2022).
3. A. Rajantie, *Contemp. Phys.* **53**, 195 (2012).
4. J. Schwinger, *Phys. Rev.* **82**, 664 (1951).
5. O. Gould, A. Rajantie, *Phys. Rev. Lett.* **119**, 241601 (2017).

Lawrence Livermore achieves a burning plasma in the lab

In that regime, fusion reactions are the plasma's primary source of heating.

Inertial fusion requires a thousand-fold compression of matter to ultrahigh densities and temperatures. The Sun and other stars use gravity to do the job and fuse hydrogen into helium. To mimic the effect on Earth, scientists at Lawrence Livermore National Laboratory's National Ignition Facility (NIF) use the world's most powerful bank of lasers to squeeze isotopes of hydrogen—deuterium and tritium—in a 2-mm-wide capsule.

The facility trains 192 laser beams into a 1-cm-tall, hollow, gold-lined cylinder known as a hohlraum, shown in figure 1, that suspends the capsule inside it. After absorbing UV-laser light, the hohlraum's interior wall reradiates a flux of soft x rays. Within 8 ns, those x rays accelerate and compress the hydrogen isotopes into a hot spot half the width of a human hair at a temperature of 60 million kelvin and a pressure of 350 GPa.

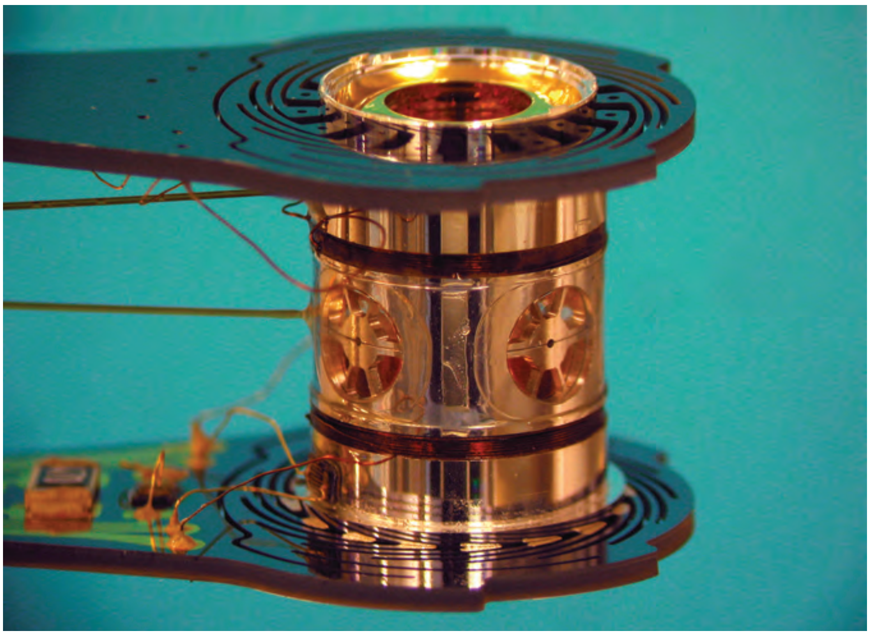


FIGURE 1. A GOLD CYLINDER known as a hohlraum holds a fuel capsule at its center for fusion experiments at the National Ignition Facility. Target-handling systems precisely position the capsule and cool it to cryogenic temperatures. (Courtesy of Lawrence Livermore National Laboratory.)

Under the capsule's surface the hydrogen fuel resides as a thin shell, cooled to 18 K prior to compression. The colder the fuel is initially, the more compressible it is—and hence the hotter and denser it becomes. The fuel's own inertia provides enough delay between the implosion and its sudden deceleration for the strong nuclear force to convert a small fraction of isotope pairs into neutrons and helium nuclei, or alpha particles.

Controlling those conditions is far from easy. Whenever a light fluid presses against a heavier one, the interface suffers Rayleigh–Taylor instabilities. Any imperfections on the capsule surface give rise to hydrodynamic fluctuations that rob the implosion of efficiency. Once the capsule starts to collapse, it can lose spherical symmetry and morph into a bumpy blob. Even worse, the imperfections can destabilize the implosion enough to mix compressed fuel with capsule material. Impurities in the fuel mixture radiate x rays away from the hot spot and rapidly cool it.

Nonetheless, researchers have been fusing hydrogen into helium, albeit inefficiently, for decades. Since 2009, NIF scientists have been striving to manage the Rayleigh–Taylor instabilities in the lab. The challenge is to heat the plasma hot spot faster than any cooling process, such as thermal conduction or bremsstrahlung radiation, can quench the fusion. That's been an elusive goal.

In three papers—one in *Nature*, one in *Nature Physics*, and a third on the arXiv eprint server—the NIF collaboration reports a more modest achievement: creating a burning plasma in four experiments that it conducted between late 2020 and early 2021.[1–3] In burning plasmas, the fusion reactions themselves—not the compression—are the primary source of heat for the plasma. Alpha particles produced by the reactions collide with electrons in the hot spot. Those electrons then thermalize and heat the fuel further.

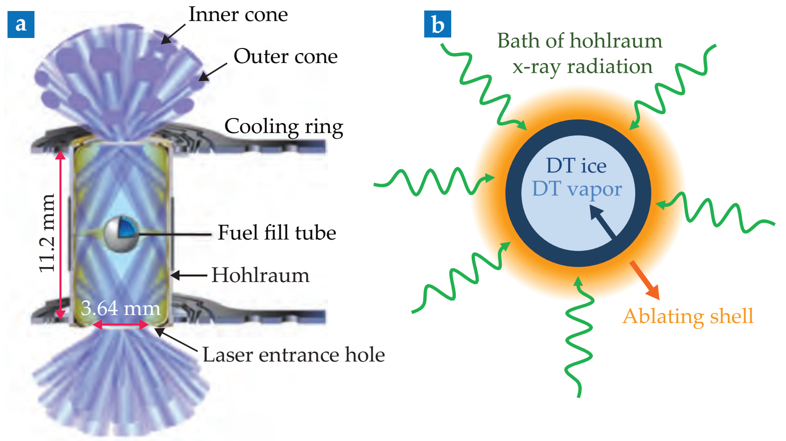


FIGURE 2. INERTIAL CONFINEMENT FUSION uses indirect laser excitation to spark a burning plasma. **(a)** Laser beams enter the hohlraum at various angles through top and bottom holes and heat its interior wall. **(b)** The flux of soft x rays reradiated by that wall expels the capsule's outer shell. By conservation of momentum, its inner shell of frozen and gaseous deuterium and tritium (DT) fuel is driven inward. (Adapted from ref. 1.)

Prelude to ignition

The process is an essential precursor to the ultimate goal of ignition, a regime in which the heat from those alpha particles exceeds all the heat losses from the system. The resulting thermal instability then triggers a nonlinear rise in temperature that sustains and propagates the burn deeper into surrounding fuel. The higher the temperature, the greater the fusion, the more alpha particles that collide in the hot spot, and the higher the temperature.

Reaching a burning-plasma state at NIF came from iterative optimization. No single measurement discloses the state's presence. Rather, a comprehensive suite of optical, x-ray, and nuclear diagnostics reveal key aspects of the implosion. Among the data is the neutron yield as a function of time and the size, volume, and energy of the hot spot. A simple metric for assessing the presence of a burning plasma is to evaluate whether the time integral of the fusion power—effectively, the energy gained by the hot spot from alpha particle heating—exceeds the total compressional work done on the hot spot. All the recent experiments satisfied that metric and other more rigorous ones.

"Having reached that regime," says Omar Hurricane, chief scientist of Lawrence Livermore's Inertial Confinement Fusion program, "we are now on the verge of ignition. The achievement not only opens access to interesting new physics, it fulfills NIF's central mission—supporting stockpile stewardship." As new data emerge, researchers there will be able to tune their computer codes to more accurately simulate what happens in a thermonuclear explosion.

Experiments and simulations have always worked hand in hand at Lawrence Livermore and NIF, and the inertial fusion program was an immense collaboration that took most of the past decade. More than 150 coauthors were involved in each of the three papers.

Holding a bomb

In the four experiments, the lasers unleashed 1.9 MJ in the form of an 8 ns pulse. Each shot roughly tripled the fusion energy achieved in previous record experiments—up to a maximum of 170 kJ. The collaboration stopped short of claiming ignition from those shots or from a fifth, record-making 1.3-MJ-yield shot it conducted a few months later in August. (See "Lawrence Livermore claims a milestone in laser fusion," PHYSICS TODAY online, 17 August 2021.)

To appreciate those numbers, keep in mind that 1 MJ is roughly the caloric energy of a candy bar. It's also the amount of explosive energy in a hand grenade. The difference lies in the amount of time each takes to release its energy. The 1.9 MJ energy of NIF's laser is fixed. So, to generate more powerful implosions the researchers had to increase the size of the fuel capsule by about 15% while keeping the hohlraum's dimensions nearly fixed.

That approach was complicated by the dynamics of the experiment. As the hohlraum heats up under irradiation, less room becomes available for the beams to propagate inside it. As shown in figure 2, an "outer cone" of laser beams reaches the hohlraum's wall close to its ends. Those beams produce a bubble of gold plasma that expands and can clip the inner beams aimed deeper, near the hohlraum's waist. The resulting nonuniformity in radiation temperature drives an aspheric implosion.

Years earlier, NIF scientists had partly resolved that problem by introducing helium gas to slow the expansion and forestall the clipping. The mere presence of gas, however, causes its own problems: laser–plasma instabilities that backscatter the beams and carry their energy out of the hohlraum. When scientists reduced the gas density, they found that those instabilities became much more manageable. But the reduction also increased the speed at which the plasma bubble expanded. Circumventing the problem, they realized, would take a faster implosion and hence more power to drive it.

The NIF scientists built a larger capsule to absorb more radiation and provide that extra power. Fortunately, beforehand they had also changed its composition—swapping out the capsule's plastic hydrocarbon shell for one made of microcrystalline diamond. Initially, the diamond capsules had many flaws that required some difficult engineering to solve, but eventually the replacements' outer surfaces were smoother and largely free of the pits and voids that had seeded instabilities and ruined implosions in earlier experiments.

With diamond's density triple that of plastic, the capsule became a better absorber of x rays and thus a more efficient compressor. Its shell was also thinner, which meant researchers could use a shorter laser pulse—down to 8 ns from 20 ns—to compress the capsule. That too sped the implosion.

Energy exchange

The new capsule design didn't entirely prevent the interception of the laser beams by an expanding plasma. To restore the uniformity of laser heating, the team tested two additional design tactics. One of them, an already well-established technique known as cross-beam energy transfer, was to shift the wavelength of the inner laser beams by just 1.5 Å relative to the outer ones. As the beams cross each other on entering the hohlraum, they scatter through an effective diffraction grating set up by laser–plasma interactions. The scattering transfers energy from the outer beams to the inner beams. And that transfer, in turn, delivers more heat to the hohlraum's waist and equalizes the x-ray flux on the capsule.

The second tactic was to add two pockets in the hohlraum near its poles. Those pockets provide space into which plasma may expand and thus delay the extent to which it occludes the inner beams. They were found to be insufficient for controlling the radiation symmetry. But they did reduce the wavelength shift needed to maintain that symmetry around the capsule.

Even if ignition is right around the corner, Hurricane cautions that converting the NIF experiment or any other fusion project into a clean, sustainable commercial energy source is a long way off. Still, "the House Science Committee seems keen on soon launching a federal fusion-based energy program," says Steven Cowley, director of the Princeton Plasma Physics Laboratory. The House-passed version of the Build Back Better bill includes $140 million over five years for the Department of Energy to carry out an inertial fusion R&D program. But the bill stalled in the Senate, where it doesn't have the votes required for passage.

Existing nuclear power plants use fission, the release of energy when uranium or other heavy elements are broken up into smaller nuclei. They also produce radioactive waste. Fusion, by contrast, produces only short-lived radioactivity induced in reactor components by the reactions' intense high-energy neutron flux. It's also safer because the reactions can be switched off by simply reducing the temperature.

As for what fusion approach—an upgraded and modified NIF reactor, tokamak, or some other system— eventually receives support, the jury is out. Cowley says, "When the time comes for a decision, it will be hard to choose."

R. Mark Wilson

References

1. A. B. Zylstra et al., *Nature* **601**, 542 (2022).
2. A. L. Kritcher et al., *Nat. Phys.* **18**, 251 (2022).
3. J. S. Ross et al., https://arxiv.org/abs/2111.04640v1.

Diamond-defect NMR monitors a surface reaction

Few other techniques can track adsorbed molecules in real time under ambient conditions.

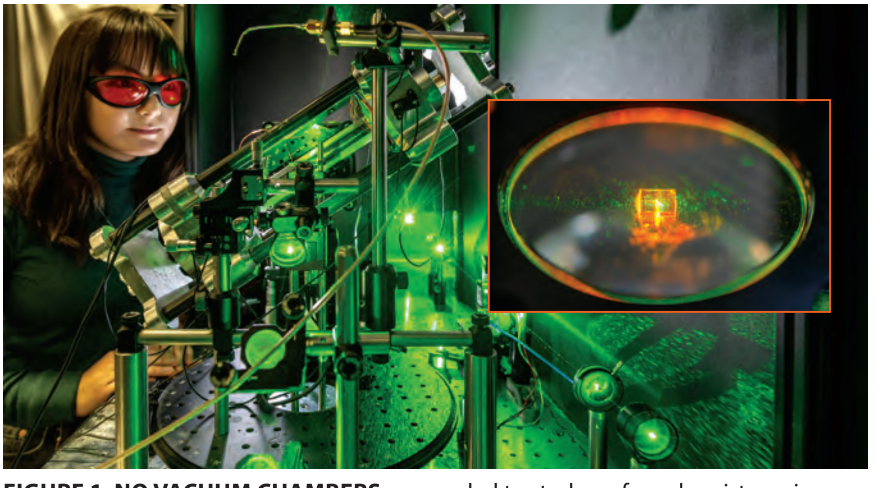


FIGURE 1. NO VACUUM CHAMBERS are needed to study surface chemistry using diamond NV-center NMR. Here, Kristina Liu of the Technical University of Munich operates the relatively simple experiment, which uses green light from an inexpensive solid-state laser to read the NV centers' spin states. The 2-mm-square diamond, not visible in the main image, is shown in the inset. (Photos by Andreas Heddergott, Technical University of Munich.)

The study of surface chemistry has always involved a bit of a paradox. Chemical processes at solid–liquid and solid–gas interfaces are ubiquitous in batteries, industrial reactors, biomedical devices, and many other systems. But despite some research at moderate pressures (see the article by Gabor Somorjai and Jeong Young Park, Physics Today, October 2007, page 48), most surface-science research tools, such as x-ray photoelectron spectroscopy and secondary-ion mass spectroscopy, work only under ultrahigh vacuum. Not only do they require bulky and expensive pumps and vacuum chambers, but they can't even access the conditions of greatest chemical and biological interest.

NMR spectroscopy is a time-honored tool for chemical analysis that works on bulk liquids, solids, and solid-like biomolecular systems. By measuring the precession frequency of spin-½ nuclei—for example, hydrogen-1, carbon-13, or fluorine-19—in a magnetic field, researchers can extract exquisite chemical information and even reaction dynamics. (See, for example, Physics Today, October 2019, page 21.) But because a two-dimensional surface contains fewer molecules than the three-dimensional bulk, conventional NMR isn't usually sensitive enough to study surface chemistry.

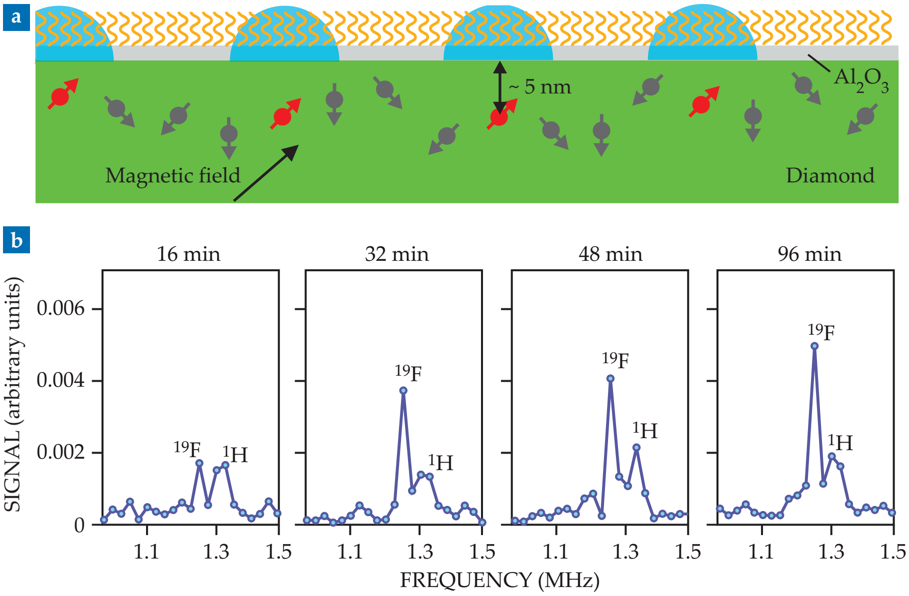


FIGURE 2. (a) A DIAMOND CHIP (green) coated with a thin film of aluminum oxide (pale gray) forms the basis for NV-center surface NMR. NV centers aligned with the magnetic field (red) are sensitive to the adsorbed molecules (orange) in the detection volumes (blue) on the surface just above them. **(b)** Real-time NMR data follow the formation of a self-assembled monolayer of fluorinated molecules on the Al_2O_3 surface. The growing fluorine-19 peak indicates the molecules' increasing surface coverage; the relatively constant hydrogen-1 peak comes from ^1H atoms from an unconstrained source, probably in the diamond itself. (Adapted from ref. 1.)

An NMR measurement can be made vastly more sensitive by swapping the induction coils of a conventional apparatus for nitrogen–vacancy (NV) centers, a type of point defect in diamond that's sensitive to magnetic fields. NV-center NMR has been under development for more than a decade, and although it hasn't yet emerged as a laboratory mainstay, researchers have collected rudimentary spectra from samples a millionth the size of those needed for conventional NMR.

Now Kristina Liu, her PhD adviser Dominik Bucher, and their colleagues at the Technical University of Munich have demonstrated surface-sensitive NV-center NMR.[1] Their experimental setup, shown in figure 1, is extremely simple by surface-science standards. It requires no ultrahigh vacuum, complicated optics, or sophisticated laser systems. Nevertheless, it can detect molecules on a surface and even monitor a simple surface reaction in real time.

In related research, Peter Maurer (University of Chicago), Nathalie de Leon (Princeton University), and colleagues are working toward using NV centers to detect and study single proteins and other biomolecules.[2] They haven't yet reached the stage of making NMR measurements themselves, but they've shown that with state-of-the-art biophysical techniques, they can tether biomolecules to NV-laden diamond without ruining either the biomolecular structure or the NV-center coherence. Combining their work with that of Bucher and colleagues could bring NV-center NMR into the realm of single-molecule biophysics.

Minuscule magnetometers

An NV center, as the name suggests, consists of a nitrogen atom and a vacancy at two adjacent sites in the diamond crystal lattice. The unpaired electrons bordering the vacancy form a spin-1 atom-like entity. Because the NV center is surrounded by an otherwise spinless sea of carbon, its spin is well shielded from its environment, and its quantum state retains its coherence as well as that of a trapped atom under vacuum. Moreover, NV-center spin states can be easily manipulated with microwave pulses and optically read with an inexpensive solid-state laser. The defects have been explored for applications in both quantum information and sensing. (See the article by Lilian Childress, Ronald Walsworth, and Mikhail Lukin, PHYSICS TODAY, October 2014, page 38.)

Under the combined influence of a static magnetic field and a series of RF pulses, spin-½ nuclei in a sample precess, and the oscillating magnetic fields they generate affect the spin-state evolution of an NV center a few nanometers away. With a suitably chosen measurement on the NV center, researchers can identify the precession frequency, which gives them information about the precessing atom's chemical identity and environment. That's the basis for NV-center NMR.

In the first proof-of-principle experiments, researchers used NV centers to detect NMR signals from nearby ^{13}C atoms in the diamond itself. Probing anything other than the inside of a diamond requires a careful balancing act: An NV center just below a diamond surface can pick up an NMR signal from a molecule just outside the diamond, but it may no longer fully benefit from the protective shielding of the carbon lattice.

The nature of the diamond surface, it turns out, matters a lot. When the dangling bonds at the edge of the lattice are capped with oxygen atoms, nearby NV centers retain their spin coherence, but when the surface is capped with hydrogen atoms, they don't. Moreover, surface chemists are mostly interested in the chemistry of surfaces other than diamond: To use NV-center NMR for surface chemistry, it's necessary to find a way to put an NV center in diamond in close enough proximity to the surface of some other solid—all without destroying the defects' delicate spin states.

Bucher and colleagues and the Maurer–de Leon collaboration both identified the same solution: coating the diamond surface with a nanometer or two of aluminum oxide. The coating is easily done with atomic layer deposition (ALD)—although as Bucher points out, the experimental capabilities for ALD and NV-center NMR aren't always present in the same lab. The diamonds, although synthetic, are expensive and in short supply (see PHYSICS TODAY, March 2022, page 22). Shuttling them back and forth between the materials science and quantum sensing labs added some logistical challenges to the work.

Both teams found that the ALD coating reduced the NV centers' coherence time, but only a little, and the defects were still capable of a sensitive NMR readout. And because Al_2O_3 is commonly used as a support material in surface-science experiments, Bucher anticipates that it would be easy to coat with a second layer of yet another material to study the chemistry of almost any surface.

Surface sensitivity

A schematic of the diamond sensors is shown in figure 2a. NV centers (red and dark gray) are implanted in the top several nanometers of a diamond, which is then coated with Al_2O_3. The defects can be oriented in a few different directions, depending on the relative positions of the N atom and the vacancy in the diamond lattice; only those NV centers parallel to the applied magnetic field yield NMR signals. Each one is sensitive to a detection volume above it and can pick up signals from any adsorbed molecules contained therein.

For their work on biomolecules, Maurer and de Leon were interested not in studying the chemistry of the protein–surface interaction but in exploiting it to hold the proteins in place long enough to study them with NV-center NMR. In a bulk solution, proteins diffuse around randomly, and they spend little time in the NV centers' detection volumes. A previous experiment on detecting single proteins with NV-center NMR immobilized the proteins by drying them onto the surface.[3] "That showed that the sensitivity is there," says Maurer, "but the proteins were completely denatured, and their structure was destroyed. The next step is to do the same thing on an intact protein."

Fortunately, biophysicists have developed a suite of chemical tools, drawing on a concept called click chemistry, for catching and holding biomolecules. A click reaction involves a pair of chemical groups that quickly "click," or bind together, whenever they're in close proximity. By placing one group on a biomolecule of interest and the other on a substrate, researchers can reliably join the biomolecule to the substrate.

Maurer, de Leon, and colleagues showed that they could attach half of a click-chemistry pair to an Al_2O_3-coated NV-laden diamond and use it to immobilize biomolecules on the surface. Using optical fluorescence, they showed that the surface-bound biomolecules retained their structure for several days—a promising step toward studying the molecules with NV-center NMR.

Bucher and colleagues, for their proof-of-principle surface-chemistry experiments, used fluorine-rich molecules and focused on detecting their ^{19}F signals, rather than the ^{1}H (proton) signals more typical of conventional NMR. "We're avoiding studying protons for now," says Bucher, "because protons are everywhere—even in the diamond—and it's not well controlled where the signal is coming from."

For example, when the researchers monitored the formation of a self-assembled monolayer on the Al_2O_3 surface, they saw a relatively unchanging ^{1}H peak, as shown in figure 2b, from the H atoms on and in the diamond. But the signal of interest was the ^{19}F peak, which steadily grew as more of the fluorinated molecules condensed out of solution and bound to the solid surface.

Regaining resolution

Collection of real-time data under chemically relevant conditions is a step forward for NV-center NMR, but Bucher and colleagues' spectral resolution is in a sense a step backward. In conventional NMR (and even some previous NV-center NMR experiments; see PHYSICS TODAY, May 2018, page 21), there's not just one peak per element. Rather, a series of fine spectral lines reveals differences in the atoms' chemical environments—for example, an atom that's bound to a benzene ring has a slightly different nuclear precession frequency than an atom of the same element that's not.

Those differences, called chemical shifts, convey highly detailed information about chemical structure, but they're unresolved in the new work. Says Bucher, "I know chemists are going to look at this and ask, 'Where's my chemical information?'"

Such poor resolution is a known issue for NMR on a solid or solid-like system. In an isotropic medium, such as a liquid, molecules tumble around rapidly, and the effect of their orientation with respect to the applied magnetic field is averaged out. But in an anisotropic environment, such as a solid or interface, molecular orientations are frozen in place, and the slight differences in precession frequency from molecule to molecule smear out the spectral lines into an unresolved blob.

There are ways of regaining the resolution in solid-state NMR. For example, spinning the sample at a so-called magic angle with respect to the magnetic field re-creates some of the effect of isotropic liquid tumbling. (See the article by Clare Grey and Robert Tycko, PHYSICS TODAY, September 2009, page 44.) But their implementation in NV-center surface NMR would greatly complicate the otherwise simple and inexpensive experiment.

Instead, Bucher seeks to look beyond spin-½ nuclei to quadrupolar isotopes with spin 1 or greater. In nuclear quadrupole resonance, or NQR, the precession frequencies are primarily determined not by interaction with an external field but by the electric field gradient produced by the surrounding atoms. That is, molecular orientation with respect to the field doesn't matter as much, and an applied field may not even be necessary. (See the article by Micah Ledbetter and Dmitry Budker, PHYSICS TODAY, April 2013, page 44.)

NV-center NQR, therefore, could convey detailed chemical information about a surface without the technical challenges of solid-state NMR. "These are very new ideas," says Bucher, "but I think that is the future of this technology."

Johanna Miller

References

1. K. S. Liu et al., *Proc. Natl. Acad. Sci. USA* **119**, e2111607119 (2022).
2. M. Xie et al., *Proc. Natl. Acad. Sci. USA* **119**, e2114186119 (2022).
3. I. Lovchinsky et al., *Science* **351**, 836 (2016).

PT

Electrification of cars and trucks likely won't disrupt the grid

The Biden administration's recently announced national charging network is a first step to enabling a fully electrified US vehicle fleet.

Matteo Muratori, who leads a research team at the National Renewable Energy Laboratory, is frustrated at social media posts warning that the rapid growth in electric vehicles will break the US electricity grid. The increase in electricity demand, or load, that will come as the US transportation system transforms to electric drive won't be any different from what occurred when air conditioning began to be widely adopted, he says.

"Utilities are excited. Selling more electricity is their business," Muratori says. "We build new industrial facilities, new hospitals, and new schools, and they make sure the electricity is there to support those needs." To balance supply with demand, the grid evolves on a daily basis as new load is added.

Michael Kintner-Meyer, an electrical engineer who leads mobility research at Pacific Northwest National Laboratory, agrees. "The lights will not go out," he says, and there isn't a tipping point that will overwhelm the grid.

There's no doubt that substantial growth in load will come from a wholesale conversion to electric vehicles (EVs) as the US, and the rest of the world, decarbonizes its transportation systems. Daniel Bowermaster, head of the electric vehicle research program at the Electric Power Research Institute, says US electricity demand would suddenly leap 25% if the nation's entire 290 million cars and trucks were converted to electric drive.

Muratori says EVs, which currently consume 0.2% of the grid's energy, will grow to 24% of demand when transportation becomes highly electrified. But Muratori, Bowermaster, and other experts agree that the shift to electric drive will be gradual and will provide utility and regulatory planners plenty of time to adapt.

President Biden has set a goal for EVs to make up half of all new vehicle sales by 2030. In February the administration unveiled its $5 billion, five-year plan for the National Electric Vehicle Infrastructure (NEVI) Formula Program. It will consist of 500 000 EV charging stations along interstate and other highways that are "designated alternative fuel vehicle corridors." NEVI is funded through the Bipartisan Infrastructure Law (BIL), which was enacted in November 2021.

States, which have to contribute 20% of NEVI infrastructure costs, are to submit their plans this summer on how they will spend the funding. The money is to be allocated under a formula based on the amount of funding that the states receive from the Federal Highway Administration.

The BIL also provides for a second, $2.5 billion competitive grant program to further increase EV charging access in locations throughout the country, particularly in rural and underserved communities. Guidelines for that program will be announced later this year, according to the Departments of Energy and Transportation, which are jointly administering NEVI.

In a 2020 DOE-commissioned study of the Western Interconnection, one of the three power grids in the US, the Pacific Northwest National Laboratory found that power demand from EVs could be accommodated through 2028. But Kintner-Meyer notes that data for the report were gathered in 2018, when only California projected an appreciable number of EVs. Demand from the other western states for EVs was "in the noise," he says.

Since 2018, planners at suppliers of bulk power have become much more attuned to the implications of an EV transition, Kintner-Meyer says. "The whole industry is aware new load is coming."

California's ambitions

California is far and away the leader in the adoption of EVs, and state resources devoted to the EV transition dwarf the federal effort. Hannon Rasool, a deputy director with the California Energy Commission (CEC), says spending, including incentives for purchasing EVs, is expected to total $10 billion over the next five to six years. Even before the promise of federal funds, the state had plans to install 250 000 EV chargers, including 10 000 DC fast chargers, by 2025.

NEVI funds will pay for some of those. According to the CEC's *2021–2023 Investment Update for the Clean Transportation Program*, for the state to reduce greenhouse gases, oil dependence, and air pollution, nearly 1.2 million public and private charging stations will be needed to support the roughly 8 million passenger EVs expected on Golden State roads by 2030. An additional 157 000 chargers will be needed to support the 180 000 medium- and heavy-duty vehi-

RAPID CHARGING stations servicing dozens of cars and trucks along busy highways will require megawatts of power and direct connections to the high-voltage transmission system.

cles anticipated on the same time scale.

Rasool says the CEC has been planning for the additional energy supply, in conjunction with the California Public Utility Commission, other agencies, and the California Independent System Operator, which manages the state's power grid. "Resiliency and reliability are really important to us," he says. "EVs create resiliency that combustion vehicles can't."

Modeling predicts that California will need 16% more electricity by 2030. Half of that will be for EVs, says Quentin Gee, a supervisor in CEC's energy assessments division. "We have a good sense on where the load is going to go, and a lot of EVs will start in urban regions," he says. "But more planning is needed to provide a more accurate geographical picture at the local level."

Electricity generation is one part of the electricity grid; the other two components are the high-voltage transmission and low-voltage distribution systems. The BIL stipulates that each NEVI station have a minimum capacity of 600 kW, enough to power four 150 kW DC fast-charging points. Fast chargers can recharge a battery to 80% in as little as 20 minutes. The last 20% can take as long as the first 80%. Level 2 chargers used for home and office require several hours to complete charging.

But charging hubs along interstate highways eventually will be expanded to accommodate perhaps 50 cars and 10 18-wheel trucks at a time, says Kintner-Meyer. That load might total 30 MW, well beyond the capacity of the low-voltage distribution system. That will require a direct connection to the transmission system, an expensive proposition that involves transformer substations and permitting, he says.

Distribution systems also will need upgrades to handle the extra load from charging EVs at homes, park-and-ride lots, and offices. "It's standard business for distribution planners to lay out electric infrastructure for a new subdivision with 100 homes," says Kintner-Meyer, "But now we have to consider how many EVs we will have in the next five years, when will they come, where will they charge, and how will they be charged."

In exchange for monopoly rights, utilities are obligated under both common law and state statutes to deliver power to new EV charging stations, just as they do to a new shopping center, industrial facility, or hospital, for example. Their costs could be recovered through a general rate increase approved by the state public utility commission, says Rasool. The new EV charging station owner might be billed for some parts of the power delivery infrastructure. Often utilities simply recoup the costs of upgrades from the sale of the additional power that is delivered.

Demand-side management

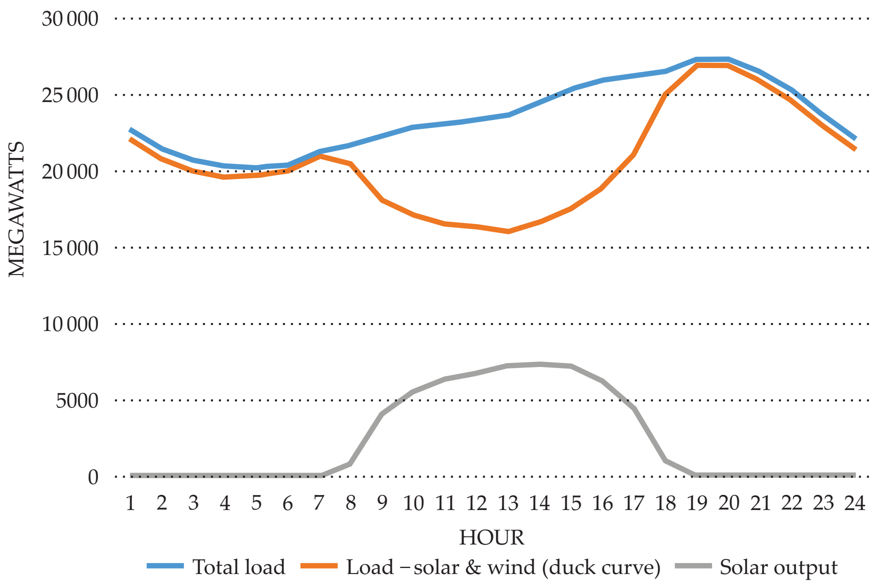
Lacking any incentives to do otherwise, EV owners would likely plug their cars in when they get home each evening, just as demand for grid power peaks. In California especially, distributed solar generation drastically lowers electricity demand from the grid during daytime hours. Once the Sun goes down, utilities must rapidly switch on generation. The phenomenon is known as the duck curve (see the illustration on page 24).

"If you lay on top of the peaks the added load from EV charging, you run out of capacity earlier than if you move [the load]," Kintner-Meyer says. So-called smart charging can shift much of the load to off-peak times. Utilities have long used variable rate structures to incentivize consumers to move their power usage to lower-demand periods. "You can charge in the daytime if you want to and pay more or wait until the middle of the night and pay less," says Muratori.

Since 80% of motorists drive 65 kilometers or less in a day, they may need to recharge only once or twice a week. Because the average US car is parked 96% of the time, there are plenty of off-peak hours in which to charge them. EVs, like most appliances, can easily be programmed to begin charging at an appointed hour. And, says Muratori, business models can be implemented to incentivize consumers and compensate them for using electricity during off-peak times.

Electrified long-haul trucks, delivery

ARNOLDREINHOLD/CC BY-SA 4.0



THE DUCK CURVE is the daily load pattern from natural gas power plants and other on-demand sources of electricity that occurs in states with large amounts of distributed solar energy. Electric vehicle charging in the evening will exacerbate the peak demand that already occurs after the Sun sets. These data are from California on 22 October 2016.

vans, and transit buses will have less flexibility on when to charge. Like a refrigerator that must run all day, those vehicles may require charging multiple times each day and may operate day and night.

"The concern is high-megawatt locations, such as a big fast-charging plaza or a fleet depot," says Bowermaster, pointing to places in Los Angeles where there may be 10 or 20 warehouses that are planning to move to all-electric delivery fleets. "How long will it take to upgrade those circuits and at what cost?" he says.

"When you decide to buy a truck, you usually don't think that you'll have to wait two years for my distribution system to be ready to charge it," says Muratori. "It's really important to plan ahead. Adopting these vehicles must be done in coordination with utilities."

Bowermaster is worried about timing. "Once the automotive supply chain gets going again, electric delivery vans and semitrucks are going to be coming off the assembly line far faster than the lead time for a substation," he warns. Regulatory bodies are pondering the amount of leeway they will give to utilities to build "ahead of the load" in anticipation of the increased demand.

But electrifying trucks provides "a lot of bang for your buck when it comes to greenhouse gas emissions and local air pollution," says Rasool. "We know that vehicle emissions impact low-income communities worse." Drayage trucks, which travel a well-defined route, such as from a port to a warehouse, are excellent candidates for electrification. The CEC is seeking proposals from companies operating 40–50 of those trucks to subsidize the cost of converting them to electric drive.

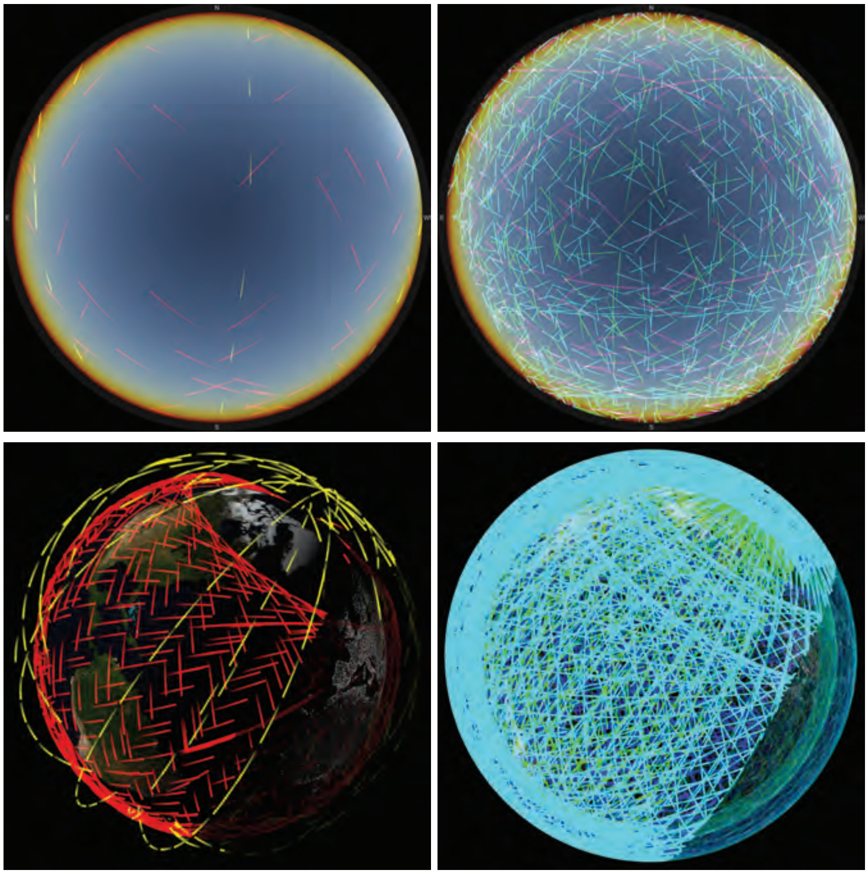
How much of the new electricity load from EVs will be supplied by renewable sources will likely vary from state to state. Many, if not most, states have renewable energy standards mandating that their total generation must be from nonpolluting sources. California aims to be 60% carbon-free by 2030 and is nearly there already, says the CEC's Gee. The state is targeting 2045 for full decarbonization of its power.

The electricity industry recognizes the need to retire fossil-fuel plants sooner than it had originally planned or to make further investments to clean them up with carbon capture and storage or the use of green natural gas, says Kintner-Meyer. "You need to keep the lights on," he adds, and gas turbines do provide a flexibility that wind and solar can't.

David Kramer

Ballooning satellite populations in low Earth orbit portend changes for science and society

Technological advances and business incentives far outpace space regulations.



HOW MANY SATELLITE SWARMS can low Earth orbit accommodate? The top images show a view of space from a latitude of 30° S—where the Vera C. Rubin Observatory will see first light in 2025—in stereographic projection at around sunset. The top left image includes existing satellites, and the top right incorporates the full systems of Starlink, OneWeb, and Project Kuiper. The two lower images show Earth from space with satellites buzzing around—again with existing numbers on the left and with the three planned megaconstellations on the right. The images were made with the freely available Mega Constellations app created by Hanno Rein of the University of Toronto Scarborough.

Several hundred thousand satellites could circle the globe in low Earth orbit by 2027, according to estimates based on license applications to national and international communications regulators. That's up from around 6300 active and defunct satellites today and 3300 in 2019.

As companies around the world rush to launch vast numbers of satellites into low Earth orbit, astronomers and others ponder the implications: How many satellites can space accommodate? What happens if that limit is exceeded? What are the effects of increasing launches, burn-up, and space debris on Earth's surface and atmospheric environments? How will celestial appreciation and discovery be affected? What can astronomers, satellite companies, and other stakeholders do to mitigate any ill effects?

Aaron Boley is a planetary astronomer at the University of British Columbia in Vancouver, Canada, who is interested in the sustainable use and exploration of space. Humanity increasingly depends on activities in space, he notes. Global positioning for navigation, time synchronization for banking and other purposes, monitoring activities for global security and disaster relief and recovery, understanding climate change, and providing internet access are examples.

"We want to use space because it can dramatically improve life on Earth," says Boley. But pointing to the depletion of the stratospheric ozone layer, accumulation of plastic in the oceans, and climate change, he cautions that "we humans have amazing capacity to underestimate our own influence on the environment" and that the wanton use of space has the "possibility of degrading the potential for its future use."

At the forefront among companies that plan to put up satellite swarms is Elon Musk's SpaceX. Around 2000 satellites in its Starlink program are currently in orbit, with the full system envisioned to have around 40 000. When SpaceX and others have put 65 000 satellites in orbit, "one out of every 15 points of light you see in the night sky will be a moving satellite," says Samantha Lawler, who studies the Kuiper belt at the University of Regina in Saskatchewan, Canada. OneWeb has launched 358 of a planned constellation of 6372 satellites—down from earlier plans for 48 000. Amazon's Project Kuiper is preparing to launch 7774 satellites. Other planned constellations include ones from China, with 12 992 satellites; Rwanda, with more than 327 000; and the Canadian company Kepler Communications, which has launched a few satellites and plans to create—or sell spots in—a constellation of 115 000.

"We are changing the sky," says Meredith Rawls, a University of Washington astronomer who works on data pipelines for the Vera C. Rubin Observatory under construction in north-central Chile. "We are at a turning point." Brighter skies from the sheer number of objects floating around reduce the number of stars anyone can see. And satellites and space debris complicate telescope observations. (For more on space debris and space environmentalism, see the interview with Moriba Jah at https://physicstoday.org/jah.)

Bright satellites

Satellites are visible when they reflect sunlight. The brightness depends on such factors as the materials, size, shape, position, orientation, and altitude of the satellite; the latitude, season, and time of night on the ground; the observing angle above the horizon; and the diameter of the telescope mirror. For an observer at 30° S—the latitude of the Rubin telescope—satellites at 600 km reflect sunlight for a few twilight hours at dusk and dawn.

So what's the harm of a satellite flying through the field of view of a telescope? Astronomers don't have the full answer yet. There are too many unknowns to accurately predict, says Jonathan McDowell of the Center for Astrophysics | Harvard & Smithsonian. How many satellites will actually be launched into low Earth orbit? How reflective will they be? What mitigation measures will be implemented?

Still, some satellite impacts are clear: For optical wavelengths, telescopes with a wide field of view and long exposure times will suffer most. Satellites are most visible at twilight, which will make it more difficult to detect near-Earth asteroids, and some could be missed. Transient events, such as supernovae explosions, could be obscured. And, says McDowell, "Faint things you don't know are there could be missed."

"You are hosed"

In May 2019, when SpaceX launched the first 60 Starlink satellites, the Rubin Observatory chief scientist, Anthony Tyson, began investigating their effect on astronomical observations. In lab experiments at the University of California, Davis, with the same CCD chips that the Rubin Observatory will use, he and colleagues discovered that satellite streaks are accompanied by parallel streaks caused by nonlinear cross talk in the electronics. Their calculations suggest that if the satellites are no brighter than seventh magnitude (coincidentally also the limit of what the naked eye can see in a dark, clear sky), the "ghost streaks" can be removed computationally.

"With 100 000 satellites, every Rubin exposure will have at least one satellite streak across the focal plane," says Tyson. Even if ghosts are suppressed, he says, "the main streaks will still need to be masked out, at the cost of lost survey area." (See "Dark-sky advocates confront threats from above and below," PHYSICS TODAY online, 9 February 2022.)

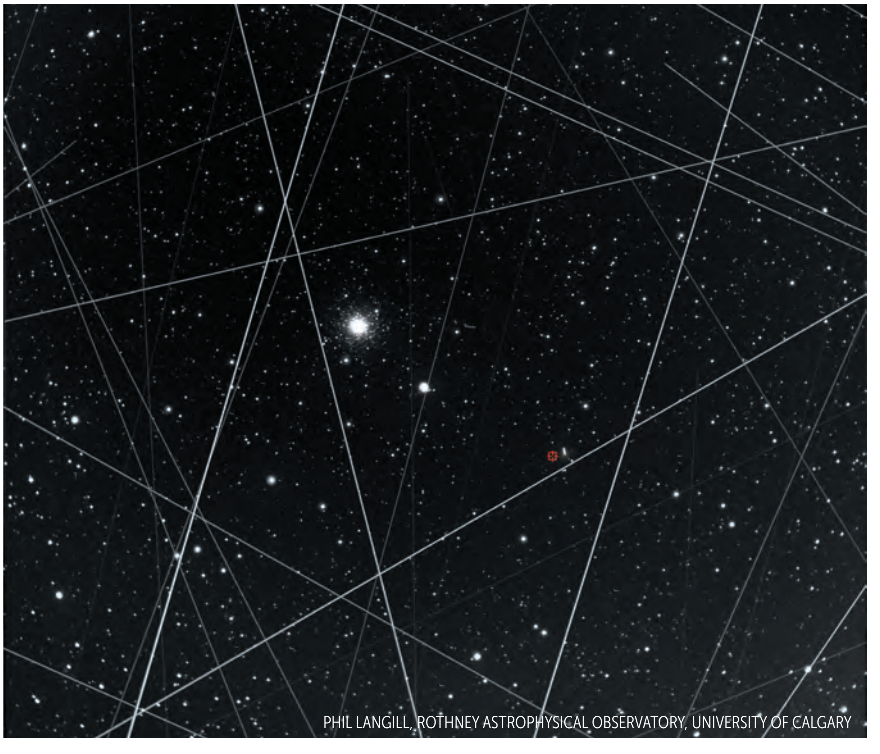
The European Southern Observatory's 40-meter Extremely Large Telescope under construction in Chile and other behemoths—the Giant Magellan Telescope and the Thirty Meter Telescope, which are not as far along—have smaller fields of view but will still be affected because of their typically long exposure times.

Multiobject spectroscopy is also vulnerable. Individual optical fibers positioned by small motors collect light from selected targets. "Some poor astronomer will have to figure out which fiber has been polluted by light reflected from a passing satellite," says Tyson. In terms of identifying interference, he adds, "their job is harder than ours at Rubin. We know immediately if we have a streak."

For a telescope with a wide field of view, "there is no dodging," says Rawls. But all telescopes will be affected—it's a matter of degree. If fewer than one streak per field of view is captured, corrupted images can be thrown out or cropped, says McDowell. But if a regime develops such that there are many streaks in every image, "you are hosed."

Radio astronomy is especially vulnerable because downlinks from satellites overpower the signals from celestial sources. Some parts of the spectrum are protected by International Telecommunication Union (ITU) rules. Harvey Liszt is spectrum manager for the US National Radio Astronomy Observatory and chair of IUCAF (the international Scientific Committee on Frequency Allocations for Radio Astronomy and Space Science). But ITU protection goes only so far, he says. Nowadays radio astronomers want to look at much more of the spectrum than decades ago when the regulations were adopted, he explains. "Because the universe is expanding, we receive redshifted signals from distant receding sources."

Moreover, special protections for radio telescopes in remote locations apply only to ground sources. "We have no protection from satellites," says Federico Di Vruno, spectrum manager at the Square Kilometre Array Observatory (SKAO), which is being built in Australia and South Africa. Radio telescopes may be able to partially adapt by recording shorter exposures to maximize their clean



SATELLITE STREAKS CRISSCROSS a two-hour sequence of observations collected in May 2021 at the Rothney Astrophysical Observatory outside of Calgary, Alberta, Canada. Near the center is a globular cluster and marked in red is a comet.

data, says Di Vruno. Interferometry also helps, he says, by "diluting the effects of signal from satellites."

"The new constellations pose a large change in the way we see the sky," says Di Vruno. "The difficulty is understanding how, and how to avoid both the immediate downlink and spillover noise." Liszt notes that radio astronomy has contended with satellites for decades. "The sky is falling, just not all of it, because satellites don't use the whole spectrum."

Astronomers are also contemplating hardware solutions. In radio astronomy, it may be possible to modify receivers to mask satellite signals. And a small wide-angle optical telescope could be placed near a much larger telescope to identify fibers in spectroscopic studies that get blasted with a satellite reflection. "If you pour in enough money, that might work," says Tyson. "You have to know the satellite is there, and how bright it is, to know how much solar spectrum to subtract."

Industry and regulatory measures

"The main problem for astronomy is that there is almost zero regulation," says Andrew Williams, who handles external relations at the European Southern Observatory. "The current regulatory landscape is not fit for these massive satellite projects." Astronomers, satellite companies, and others are working with national and international agencies to formulate policies.

At its annual meeting in February, the Scientific and Technical Subcommittee of the United Nations Committee on the Peaceful Uses of Outer Space (UNCOPUOS) discussed keeping skies quiet for science and society. "They added an agenda item for us for next year's meeting," says Connie Walker of NSF's NOIRLab. "That is a feat." Working at the policy level is slow and hierarchical, she says, "but eventually things get accomplished." And because the UN operates through consensus, once resolutions are adopted, they can be influential. "We hope the UNCOPUOS endorses best-practice guidelines," says Williams. "That would create a social norm to consider. We are also asking for recognition that astronomy is part of the space domain."

Richard Green is assistant director for government relations at the University of Arizona's Steward Observatory. He notes that an industry group is developing best practices for space use. Astronomy, he says, "will get the farthest and fastest by cooperating with the satellite companies." Eventually, he adds, "the mitigations should become a requirement for getting licenses to launch or operate."

Indeed, technology advances are far outpacing formation of legal frameworks, so for the short and medium term, astronomers are forced to rely on the goodwill of companies. The satellite companies "didn't set out to ruin astronomy," says the University of Washington's Rawls, "but it can be a side effect." Any mitigations by operators are voluntary, she notes. "There is nothing compelling them to take the needs of astronomy into account."

SpaceX painted satellites dark to make them less reflective. That made them too hot. Visors work better, but they block laser signals between satellites. Other approaches to reducing reflections include adjusting satellite orientation and optimizing surface materials. A spokesperson for Amazon says the company will test mitigation methods this fall with prototype missions for its Project Kuiper.

At higher altitudes, Earth provides less shadow, and satellites are visible for more of the night. They also move more slowly. So even though a satellite at higher altitude would be dimmer than one at lower altitude, the slower motion means more light would be collected by a given pixel in a telescope camera. "Due to the impacts on ground-based astronomy and concerns about debris generation and longevity," the top priority recommendation in a 2020 report by the government advisory body JASON is to "eliminate or highly regulate large satellite constellations in orbits higher than 600 km." OneWeb's satellites are at 1200 km, but most of the other planned constellations are below 600 km.

For some observations, astronomers say, accurate information about satellite positions and trajectories would be extremely helpful—they could time data collection. But trajectories are sometimes adjusted on short notice to, for example, avoid collisions with other satellites or debris.

"If some companies set a good example," says Rawls, "we hope others will adopt similar mitigation measures."

Coordinating protection efforts

Considerations for how to deal with a surging satellite population are coming under the umbrella of the new virtual Centre for the Protection of the Dark and Quiet Sky from Satellite Constellation Interference. Launched on 1 April by the International Astronomical Union (IAU) and cohosted by NSF's NOIRLab and the SKAO, the center has four hubs. They focus on policy, industry and technology, community engagement, and data and software. Walker, codirector with Di Vruno, says that the center "will try to coordinate efforts among all involved parties, unify voices across the astronomy community, and work with industry."

The center is starting with €135 000 ($150 000) for three years from the IAU. The center's data and software hub is tasked with creating a repository for information related to satellite interference, writing software to minimize the interference effects, and more. "Solutions won't be one-size-fits-all," says Walker, "but hopefully we can create tools that can be tweaked to accommodate different situations." A major challenge is money, she adds. "So far, people have been working pro bono. They can't do that forever. They have jobs."

"We are pretty good at finding solutions," says McDowell. "But we need resources. And working on this takes time away from doing science." Regina's Lawler agrees: Professional astronomy will lose data, she says. "And it will take more money and time to get discoveries."

The plan for the IAU hub on community engagement is to bring together environmentalists, Indigenous communities, astrotourism interests, astrophotographers, and planetarium professionals. Everyone has a stake in the skies, says Aparna Venkatesan, an astronomer at the University of San Francisco who is involved with the IAU hub. "Right now space is a free-for-all," she says. "The system is set up to benefit a few, and space is vulnerable to rogue action. We hope that the IAU will start an international dialog and lead to a more equitable and thoughtful use of space."

For astronomers, as well as for amateur sky gazers and for people who care about cultural heritage, the best outcome would be to keep the numbers of satellites to a minimum. But billions of dollars are "chasing these companies to launch as soon as possible," as Tyson puts it, "so that's not going to happen."

"I give SpaceX credit for stepping up

to the plate and trying to mitigate," says Tyson. Still, he adds, at the end of the day, they may not reach the seventh-magnitude goal.

"There is a huge fear of offending the satellite industry," says an astronomer who requested not to be identified. The concessions by industry are small, the astronomer says. The growing population of satellites "is an existential threat to astronomy."

Traffic jams

Even if concerns for astronomy don't lead to international regulations, orbital crowding might. "We have never operated in the environment we are about to create—congestion will be much higher than we have experienced before," says British Columbia's Boley. Collisions create debris, which can lead to more collisions. Such a runaway cascade could render future launches dangerous or impossible—and imperil satellites and the International Space Station. Estimates by JASON, NASA, and others show significant dangers at 1200 km with 8000 satellites. At lower altitudes, runaway cascades could also occur, but atmospheric drag slows debris.

BETTYMAYA FOOTT/INTERNATIONAL DARK-SKY ASSOCIATION

FAJADA BUTTE in Chaco Culture National Historic Park, New Mexico, shown here during an astronomy festival in 2018. The site is a registered International Dark Sky Park. Views of the sky will be compromised by the ballooning population of satellites.

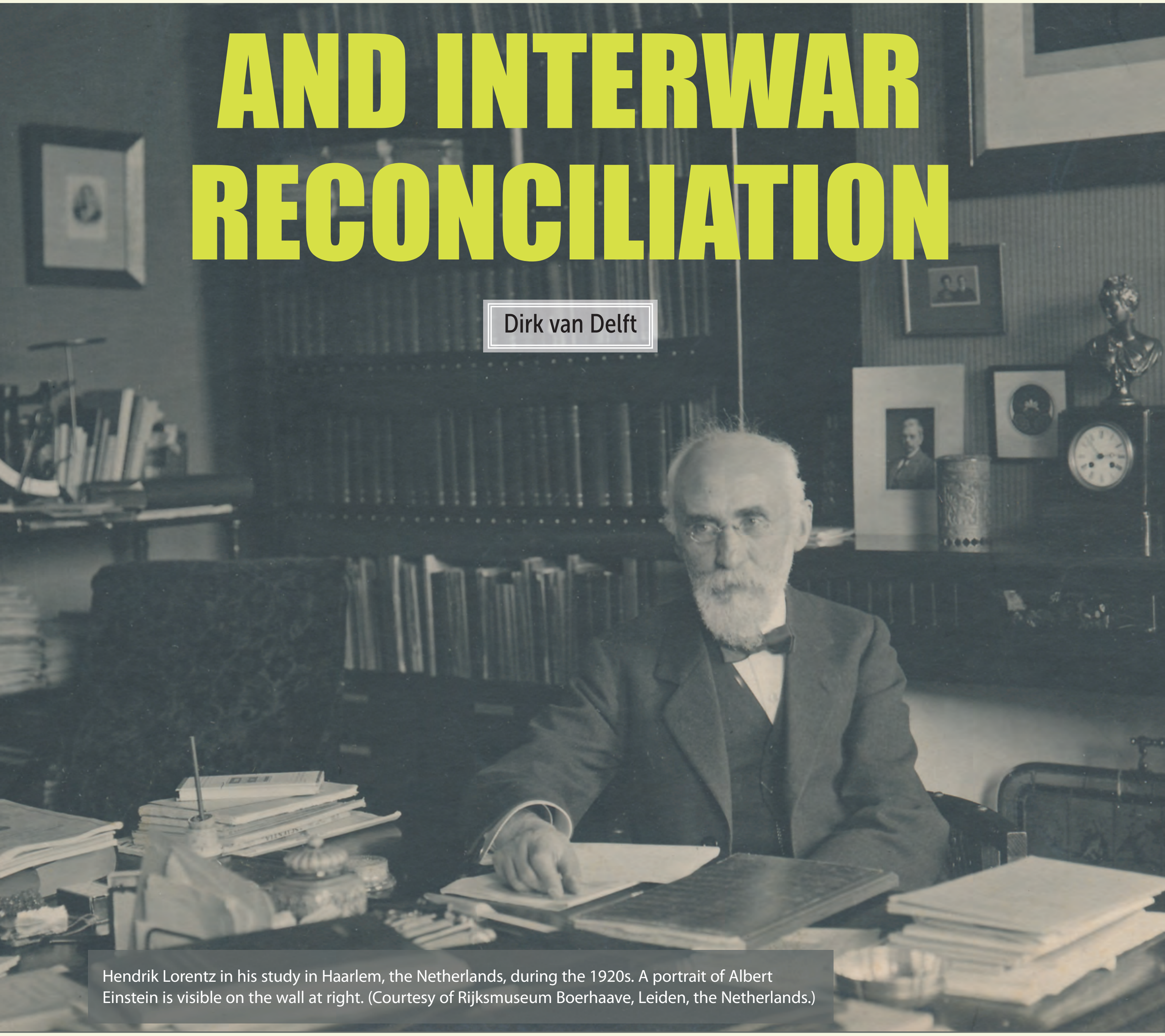
When satellites burn up on reentry, they spew aluminum and other materials into the atmosphere. Launches also pollute. The Starlink satellites will each be operational for five years. "If you do the math," says Lawler, "on average you deorbit 23 satellites a day. That comes to about six tons of aluminum added to the upper atmosphere." That would be more than from naturally occurring meteorites, she notes, which deposit more mass, but are made mostly of oxygen, silicon, magnesium, and iron.

Space is getting crowded, says Venkatesan. "The different stakeholders—astronomers, the military, storytellers, knowledge holders, cultural practitioners—all want access to space for the next few centuries, not just the next few years. That's where I think we can unite."

Toni Feder PT

Hendrik Lorentz, Robert Millikan, AND INTERWAR RECONCILIATION

Dirk van Delft



Hendrik Lorentz in his study in Haarlem, the Netherlands, during the 1920s. A portrait of Albert Einstein is visible on the wall at right. (Courtesy of Rijksmuseum Boerhaave, Leiden, the Netherlands.)

Dirk van Delft is a professor emeritus of the history of science at the Lorentz Institute for Theoretical Physics at Leiden University. He is the author of the 2007 book *Freezing Physics: Heike Kamerlingh Onnes and the Quest for Cold*.

World War I tore apart a close-knit international physics community. During the interwar period, two famous physicists attempted to repair those shattered bonds.

A photograph hangs in the stairwell of the east wing of Caltech's Norman Bridge Laboratory of Physics. Dating from 1922, it depicts 25 professors and students posing in their Sunday best with a dog lying at their feet (see figure 1). They are there to attend a special lecture series on light and matter. Seated in the middle of the front row, to the right of laboratory director Robert Millikan, sits the lecturer: Hendrik Lorentz, a theoretical physicist from Leiden University in the Netherlands.

Millikan and Lorentz are celebrated for their tremendous contributions to early-20th-century physics. But they also made their mark in the realm of international science policy. Based on the archives of Millikan and Lorentz in Pasadena, California, and Haarlem, the Netherlands, respectively, this article reconstructs a little-known aspect of that work: their joint effort to restore scientific cooperation between countries that had fought hard to destroy each other during World War I.

Four long years of bloody warfare had engendered tremendous hatred between scientists who cooperated amicably before the war but then ended up on opposite sides of the trenches. In their capacity as members of the League of Nations' International Committee on Intellectual Co-operation—the precursor of today's UNESCO—Lorentz and Millikan tried to make peace between the two sides.

Lorentz and Millikan

Born in 1853, Lorentz (see box 1) got along well with Millikan, who was 15 years his junior. The two physicists first met in 1900 at the International Congress of Physics in Paris—the first event of its kind. They rekindled their personal ties after World War I, when Lorentz, as the chair of the Solvay Conference on Physics, invited Millikan and his University of Chicago colleague Albert Michelson to the 1921 meeting (see figure 2). It was the first time that US scientists were invited to the prestigious gathering in Brussels.[1] On his way to the conference, Millikan visited Lorentz in the Dutch town of Haarlem in spring 1921. "I am going to keep you to your promise to visit America once more," Millikan wrote from London on his way back home from Europe, "and to suggest that the very best time will be sometime during the coming year."[2]

As figure 1 attests, Millikan (see box 2) indeed held Lorentz to his promise: The Dutch physicist spent the winter term in Pasadena during a lecture tour of the US in 1922. Lorentz's stay was part of the master plan developed by George Hale, an astronomer and the director of the Mount Wilson Observatory in Pasadena, to transform Caltech into a world-class research institution. The first step in the plan had been luring Millikan from the University of Chicago, which Hale did in 1921 with the help of a $250 000 donation (more than $3 million in today's dollars) from the physician and wealthy businessman Norman Bridge for the laboratory that still bears his name today.

The second step in Hale's plan was to invite famous foreign physicists to come to Pasadena as visiting professors. Lorentz was the first of Hale's invitees, which was no coincidence: He was the most highly respected theoretical physicist of the day. Other star scientists from Europe, including Albert Einstein, Paul Ehrenfest, Werner Heisenberg, Erwin Schrödinger, and Arnold Sommerfeld, would follow in his footsteps.

The newly finished Norman Bridge Laboratory was inaugurated during Lorentz's stay in Pasadena, and he was invited to speak at the opening ceremony on 28 January 1922. Tellingly, Lorentz's speech sang the praises of "physicists [who] form a kind of fraternity spread all over the world." Even though the world had recently emerged from a disastrous war, Lorentz said, "I am deeply convinced of this unity of science."[3]

International Research Council

Both Hale and Millikan were heavily involved in reorganizing international science after the Great War. Prior to the conflict, Hale had served as the foreign secretary of the US National Academy of Sciences. In that position, he became acquainted with the International Association of Academies, an umbrella organization that was intended to link scientific institutions across the globe. But that organization had achieved little since its founding in 1899. Hale believed that Allied countries should start something new.

To that end, he convened a meeting of scientists from Allied countries in London on 9 October 1918. Although it was preliminary in nature, the meeting foreshadowed what was to come: The US and UK delegates argued that Germany and the other Central powers should eventually be allowed to join the new international organization if they openly rejected the political actions that had led to the war. But that view touched a raw nerve with the Belgian and French delegates, who believed that the Central powers should be excluded.

FIGURE 1. CALTECH FACULTY, students, and staff pictured with Hendrik Lorentz during his 1922 visit. Behind the dog sit Lorentz (with the white beard) and Robert Millikan, to the left of him. A copy of this photo hangs in Caltech's Norman Bridge Laboratory of Physics. (Courtesy of the Caltech Archives.)

After another preparatory meeting in Paris, the new organization Hale envisioned—named the International Research Council (IRC)—was established on 28 July 1919, shortly after the signing of the Treaty of Versailles. Based in Brussels, the IRC served as an umbrella organization. International unions for specific disciplines, such as astronomy, chemistry, and geophysics, became its subsidiaries. The International Union of Pure and Applied Physics[4] joined the IRC in 1922.

In a concession to the French and Belgian delegates, the IRC bylaws explicitly excluded the former Central powers from membership. Yet in the early 1920s, many scientists from both formerly neutral and Allied nations began to campaign for admitting the defeated nations into the IRC. At the second IRC assembly in July 1922, the Swedish delegates proposed taking a more lenient attitude toward the Germans. The Belgian and French delegates refused to give way, which led to fierce clashes. Even a watered-down version of the Swedish proposal drove the French and Belgians into a frenzy. "Never!" they shouted.

Similar sentiments were in evidence at the 1924 meeting of

BOX 1. HENDRIK LORENTZ (1853–1928)

The son of a vegetable farmer, Hendrik Lorentz grew up in the provincial town of Arnhem in the eastern Netherlands.[15] He went to Leiden University in 1870 to study mathematics and physics. His doctoral work culminated in an 1875 PhD thesis on the reflection and refraction of light that extended Maxwell's theory of electromagnetism.

In 1877 Lorentz proposed the existence of minute, charged particles that were part of atoms and would later be called electrons. He formalized his electron theory in 1892. Published in its definitive form in 1904, the theory introduced the concepts now known as the Lorentz force, the Lorentz contraction, and the Lorentz transformation. Two years later Lorentz gave a series of lectures on that theory at Columbia University. They were published in 1909 as *The Theory of Electrons and Its Applications to the Phenomena of Light and Radiant Heat*.

In 1896 Lorentz used his electron theory to explain how spectral lines split under the influence of a magnetic field, which is commonly known as the Zeeman effect. Together with Pieter Zeeman, who experimentally discovered that behavior, Lorentz was awarded the 1902 Nobel Prize in Physics for that work. But he was unable to explain the anomalous Zeeman effect that Millikan and others soon discovered.

In 1912 Lorentz moved to Haarlem, where he led the Teylers Museum's research laboratory. That new position brought him relief from the heavy lecturing duties he had while he was teaching theoretical physics at Leiden University, although he continued to give lectures as a professor by special appointment. He primarily used his newly found spare time to further international scientific cooperation through work with the International Research Council, the International Committee on Intellectual Co-operation, and the International Solvay Institutes. Thanks to his tact, his command of languages, and the esteem with which he was held in the international physics community, Lorentz chaired the first five Solvay Conferences.

Lorentz left the lab in 1923 but continued to lecture at Leiden. He kept abreast of the new developments in relativity and quantum theory in the 1920s and even lectured on the latter topic at Cornell University during a 1926–27 US trip (see figure 4). Lorentz was good friends with Albert Einstein, who once called the Dutchman "a living work of art." In recognition of his contributions to Dutch science, the Royal Netherlands Academy of Arts and Sciences awards the Lorentz Medal every four years to an outstanding theoretical physicist (see figure 5).

the International Mathematical Union in Toronto. A huge commotion broke out when it became clear that delegates from Germany, the country of such mathematical giants as David Hilbert and Felix Klein, had not been invited because of the IRC bylaws. The US delegates proposed a motion to lift the ban. Although it was supported by the delegates from Denmark, the UK, Italy, the Netherlands, Norway, Sweden, and Switzerland, IRC regulations precluded such a move.

Hale, a member of the IRC's executive committee, was increasingly embarrassed by the continued ban on German participation. He feared that the organization was doomed if the ban was not scrapped at the next IRC assembly in 1925. As he wrote to the executive committee's British representative, the physicist Arthur Schuster, the IRC could not "afford" to let the "iron hand" of the French "wreck the whole organization." Germany needed to be admitted as soon as possible. Hale was not alone in that belief: He wrote to Schuster that Millikan felt similarly and "much more strongly than I do."[5]

FIGURE 2. A GROUP PORTRAIT from the Third Solvay Conference on Physics, held 1–6 April 1921. In the front row sit Hendrik Lorentz (fourth from right) and Robert Millikan (second from right). Between the two is Ernest Rutherford; Marie Curie sits to the right of Millikan. Apart from Albert Einstein, who was a dual German–Swiss citizen, German physicists were not invited to attend the conference because of lingering wartime resentment. (Photograph by Benjamin Couprie; courtesy of the International Solvay Institutes, Brussels.)

Millikan and the IRC

After receiving the Nobel Prize in Physics in December 1923—for which he had received warm congratulations from Lorentz—Millikan spent a few months in Europe. In May 1924 he gave his Nobel lecture in Stockholm, and in early June he and his wife visited Lorentz in Haarlem. Another item on Millikan's European itinerary was attending a meeting of the IRC executive committee, where he stood in for Hale, who was having health problems.

Believing that "all members" of the executive committee were ready to "move in broadening and bettering the conditions of international contact among scientific men," Millikan struck a positive note after the meeting. As he wrote to a colleague, "I have not the slightest doubt myself that Germany will be invited to full participation next July." Millikan believed that the IRC had proved its usefulness. It would only improve, he wrote in a biting aside referring to the US refusal to join the League of Nations, if "it is found possible to keep out of it the sort of spirit and attitude which has made the United States non co-operative since the war."[6]

At the third IRC assembly, held in Brussels in July 1925, the

BOX 2. ROBERT MILLIKAN (1868–1953)

The son of a minister, Robert Millikan spent his childhood in Morrison, Illinois.[16] In 1886 he began studying at Oberlin College in Ohio, where he obtained a BA in 1891 and an MA in 1893. He then went to Columbia University, where he became the institution's first PhD recipient in physics. After obtaining his doctorate in 1895, Millikan toured Europe for a year. Upon his return, Albert Michelson secured him an assistant professorship at the University of Chicago.

After 10 years at Chicago, Millikan decided to "get busy on some more serious research." That resulted in two brilliant proofs of his experimental talents. In 1910 he used falling oil droplets to prove the existence of an elementary charge, which he determined to be 1.6×10^{-19} coulombs. He then embarked on a series of ingenious investigations in which he experimentally confirmed Einstein's 1905 equation for the photoelectric effect and calculated the value of Planck's constant h with unparalleled accuracy. That work earned him the 1923 Nobel Prize in Physics.

The promise of a better salary and a new laboratory persuaded Millikan to move to Caltech in 1921. Under his leadership, the Norman Bridge Laboratory of Physics quickly developed into one of the world's most prestigious physics centers. After his retirement in 1945, he remained involved in experiments on cosmic radiation. Notably, Millikan maintained a lifelong aversion to government interference in science; rather, he argued that wealthy philanthropists, businessmen, and the public should support centers of scientific research, such as Caltech and other universities.[17]

Active in the public sphere, Millikan was often featured in the media and took part in local social and literary clubs. His public-facing activities extended to an ignominious aspect of his legacy—namely, his firm belief in eugenics. Many scientists of Millikan's era espoused such beliefs privately, but Millikan went much further: In 1937 he joined the board of trustees of the Human Betterment Foundation, a California-based organization that advocated eugenic sterilization of "mentally deficient" individuals.

Although Millikan did not have much influence on the organization's policies or strategy, a 2020 Caltech report concluded that his fame likely gave it "renewed legitimacy" in an era when Nazi sterilization policies were beginning to discredit eugenics.[18] As a result, Caltech decided in 2021 to remove his name from the university library, a professorship, and a fellowship fund. Other organizations with awards or buildings named after Millikan have followed suit.

FIGURE 3. ROBERT MILLIKAN (center) in conversation with Albert Einstein (left) and Marie Curie (right) at a meeting of the International Committee on Intellectual Co-operation in Geneva in 1924. (Courtesy of the Caltech Archives.)

Dutch and Danish delegates made a joint proposal to lift the ban on German membership. It was defended with great verve by Lorentz, who pleaded with Millikan beforehand to adopt "an attitude of benevolent neutrality."[7] Millikan, who was again standing in for Hale, told Lorentz that the pressure to admit Germany was "so strong" in the US that he was "very sure that the International Research Council itself will go to pieces unless some constructive action is taken in Brussels next summer."[8]

Although a majority of representatives at the meeting voted to lift the ban, the motion nevertheless failed because amendments to the IRC bylaws required the support of two-thirds of the total number of IRC member countries to pass. Because many IRC members did not attend the meeting, German admission would have been rejected even if all the countries present in Brussels had voted in favor. Fury reigned.

At an extraordinary meeting of the IRC in June 1926, the ban was finally lifted. Germany and its former allies Austria, Hungary, and Bulgaria were immediately invited to join the organization. But the Germans refused to eat humble pie. Against the express wishes of their own government, the German academies declined to join the IRC. It was not until 1952 that the West German academies would join the IRC's successor, the International Council of Scientific Unions, and 1961 when the East German academies entered.

International Committee on Intellectual Co-operation

In the meantime, Lorentz and Millikan had both joined the International Committee on Intellectual Co-operation (ICIC), an advisory committee to the League of Nations. The ICIC was a child of the league's idealism: It was based on the idea that cultural cooperation between intellectuals from formerly warring countries would build mutual trust between nations and therefore promote world peace.

Chaired by the famed French philosopher Henri Bergson, the ICIC consisted of 12 members drawn from the intellectual elite, including Marie Curie, Einstein, and Hale (see figure 3). The minutes of the ICIC's first meeting in January 1922 in Geneva illustrate how a dozen top intellectuals spent five days earnestly and diligently plowing through a huge agenda full of boring regulations, resolutions, and procedures. One of the topics discussed was the international exchange of publications and exchange programs for professors, lecturers, and students. They had to be implemented with great caution because of what were delicately described in the minutes as "feelings which were still extremely painful and went very deep."[9]

Ever the idealist, Bergson grandiloquently complimented the League of Nations on behalf of all intellectuals in the world for conceiving the "fine and noble idea" to found the ICIC.[10] Millikan, in contrast, offered sober realism: "The influence of the Committee on Intellectual Cooperation," he wrote in a letter to Bergson, "will be measured by the practicability and the success of the plans which it initiates." In comparison to the IRC and its unions, he argued that the ICIC would have to utilize existing initiatives developed by universities and libraries to promote the exchange of teachers, students, and scientific publications. As he concluded, "Money for scientific research is collected a hundred times more easily through the aid of a local appeal and a local interest than through the aid of any general appeal."[11]

Lorentz joined the ICIC in April 1923. Although he was motivated by a sense of humanitarian duty, he was clear-eyed

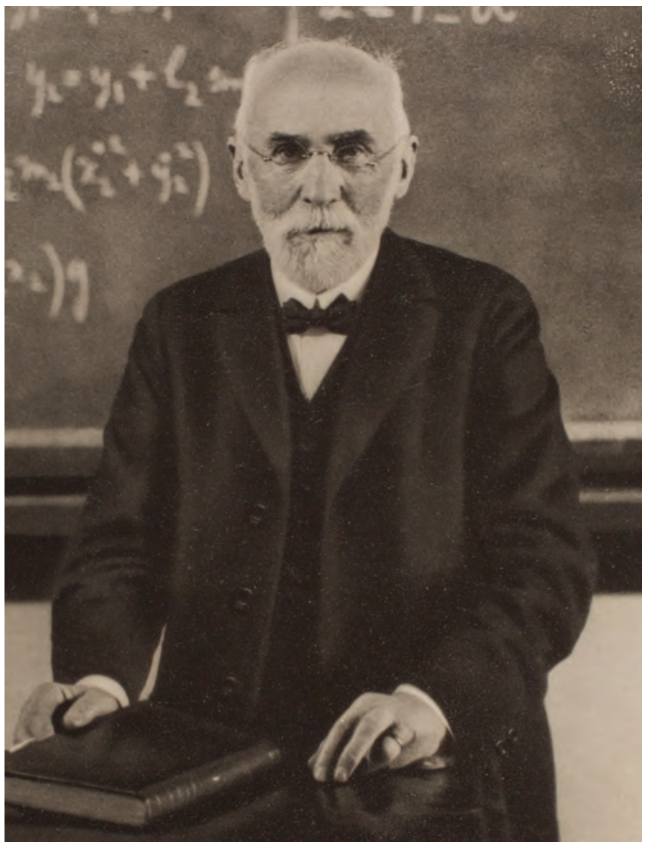


FIGURE 4. HENDRIK LORENTZ lecturing at Cornell University in 1926 during one of his US tours. (Courtesy of the Division of Rare and Manuscript Collections, Cornell University Library.)

FIGURE 5. THE LORENTZ MEDAL is awarded by the Royal Netherlands Academy of Arts and Sciences every four years to an outstanding theoretical physicist. First awarded to Max Planck in 1927, the medal was endowed in 1925 in honor of the 50th anniversary of Hendrik Lorentz receiving his doctorate.

about the ICIC's limitations. He quickly made a positive impression on those involved with the committee. As the League of Nations' Japanese undersecretary Inazō Nitobe informed Millikan, who was absent at the first ICIC meeting Lorentz attended in August 1923, "Your old friend, Dr. Lorentz, was a great acquisition. His wisdom and geniality were indispensable."[12] When Bergson was forced to step down from the chairmanship in January 1926 because of health problems, Lorentz was chosen as his successor.

The July 1924 ICIC meeting in Geneva brought a major surprise: The French government generously offered to fund an executive institute for the ICIC that would be housed in a Parisian palace. Although the offer was tempting, several committee members had reservations about aligning the ICIC with France. Millikan was worried about moving league matters to Paris. Einstein agreed and argued that having the headquarters in Geneva ensured neutrality. Millikan and Lorentz proposed integrating the tasks of the new institute with the activities of existing ICIC subcommittees. But certain matters could not be tampered with: As an advisory committee of the League of Nations, the ICIC needed to remain headquartered in Geneva.

A final attempt

Lorentz worked toward scientific reconciliation up to the last stroke of his pen. The very last letter he wrote, which he began drafting on 20 January 1928, was to Millikan. In that letter, which remained unfinished after Lorentz came down with a high fever, he brought Millikan up to speed on several ICIC initiatives.

One of those plans lay close to Lorentz's heart: convening a small-scale international meeting to persuade the reluctant Germans to join the IRC. A previous attempt to achieve that in summer 1926 had failed because the Germans had not yet joined the League of Nations. But Germany was finally admitted to the league later that fall. Millikan saw that as an auspicious sign and suggested to Lorentz in September 1927 that they try again.

Lorentz agreed and began planning the meeting. He envisioned that the attendees would be himself and Einstein on behalf of the ICIC, several presidents of German academies, several members of the IRC's executive committee, a representative of the Royal Society of London, and a League of Nations staff member.

Lorentz explained the initiative to Millikan as follows: "The task of the committee would be to consider in a friendly spirit the way in which we may best reach the really universal collaboration which we all desire. If for this purpose it should be felt desirable that the organization which exists now be more or less modified, we should have to examine the nature and the possibility of these modifications."[13]

Sadly, it again proved difficult to organize a reconciliatory meeting. The Germans refused to go to Geneva, so Haarlem—where Lorentz lived—was proposed as an alternative. Plans began to collapse after the German undersecretary to the League of Nations, Albert Dufour-Féronce, who had been involved with the planning process, announced that he would not attend the meeting. The League of Nations then asked Lorentz to serve as chairman, but he feared that little would result from a meeting that had lost its official character.

None of that mattered in the end, however, as Lorentz passed away "calmly and peacefully" on 4 February 1928. His student and successor, Adriaan Fokker, sent a copy of the unfinished letter to Millikan with an accompanying note stating that it was "a great pity Professor Lorentz has not lived to see the general reconcilement for which he so strongly longed."[14]

Tragically, Lorentz's dream of international scientific reconciliation would not be realized for many years. Even the small victories he and Millikan achieved were undone in the 1930s and 1940s when international relations were soured by the rise of nationalism, fascism, and Nazism. Only after World War II, in 1954, would the German Physical Society join the International Union of Pure and Applied Physics.

The author would like to thank the reviewers for their comments and Frits Berends for valuable discussions.

REFERENCES

1. Special issue, "The Early Solvay Councils and the Advent of the Quantum Era," *Eur. Phys. J. Spec. Top.* **224** (2015).
2. R. Millikan to H. Lorentz (15 April 1921), file 54, Hendrik Lorentz Papers, North Holland Archives, Haarlem, the Netherlands.
3. H. Lorentz, *Bull. Calif. Inst. Technol.* **31**(94), 33 (1922), p. 34.
4. F. Greenaway, *Science International: A History of the International Council of Scientific Unions*, Cambridge U. Press (1996).
5. G. E. Hale to A. Schuster (2 May 1924), box 47, folder 1, George Ellery Hale Papers, Caltech Archives.
6. R. Millikan to W. Willcox (17 October 1924; carbon copy), box 13, folder 19, Robert Andrews Millikan Papers, Caltech Archives.
7. H. Lorentz to R. Millikan (28 February 1925), box 41, folder 10, in ref. 6.
8. R. Millikan to H. Lorentz (4 February 1925), file 54, Lorentz papers, in ref. 2.
9. League of Nations Committee on Intellectual Co-operation, *Minutes of the First Session: Geneva, August 1st–5th, 1922* (11 October 1922), p. 21.
10. Ref. 9, p. 39.
11. R. Millikan to H. Bergson (2 November 1922; carbon copy), box 13, folder 17, Millikan papers, in ref. 6.
12. I. Nitobe to R. Millikan (10 August 1923), box 13, folder 17, Millikan papers, in ref. 6.
13. H. Lorentz to R. Millikan (20 January 1928; copy of unfinished letter made by A. Fokker), box 41, folder 10, Millikan papers, in ref. 6.
14. A. Fokker to R. Millikan (20 February 1928), box 41, folder 10, Millikan papers, in ref. 6.
15. F. Berends, D. van Delft, *Lorentz: gevierd fysicus, geboren verzoener* (*Lorentz: Celebrated Physicist, Born Mediator*), Prometheus (2019).
16. R. H. Kargon, *The Rise of Robert Millikan: Portrait of a Life in American Science*, Cornell U. Press (1982); L. A. du Bridge, P. S. Epstein, in *Biographical Memoirs*, vol. 33, National Academy of Sciences (1959), p. 241.
17. R. A. Millikan, *The Autobiography of Robert A. Millikan*, Prentice-Hall (1950), p. 262.
18. Committee on Naming and Recognition, *Final Report*, California Institute of Technology (17 December 2020; rev. 27 February 2021), p. 15.

PT

Originally Published in the Sept 2021 GradSchoolShopper Magazine

FINDING THE RIGHT PROGRAM FOR YOU

Samantha Pedek, graduate student, University of Iowa; co-chair, Physics Congress 2022 Planning Committee

Find Your People and Grad Program at the 2022 Physics Congress

Join hundreds of physics undergrads, grad school reps, and physics luminaries

Samantha Pedek

Samantha Pedek, 2022 Program Co-chair

Networking is one of the most important aspects of being a young professional. We've all heard the spiel about how networking can have positive impacts on future educational and career-related opportunities, but many of us struggle with making the initial contact that can lead to lasting connections.

In 2016 I attended the Physics Congress (PhysCon), the largest gathering of undergraduate physics students in the United States. Every few years, PhysCon brings together students, alumni, and faculty members for three days of frontier physics, interactive professional development workshops, and networking. It is hosted by Sigma Pi Sigma, the physics honor society, and anyone interested in physics can attend.

Networking at PhysCon was unlike any other professional development experience I had as an undergraduate physics student. The sheer number of like-minded people was daunting—hundreds of physics and astronomy undergraduates, representatives from graduate schools and summer research programs, employers from all over the country, and well-established professionals at the height of their careers were all under one roof for three days.

The Physics Congress is a high-energy, hands-on weekend designed explicitly for undergraduate physics students. Photo courtesy of SPS National.

PhysCon has continued growing in attendance, scope, and opportunities, and you won't want to miss the next one! In celebration of the 100th anniversary of Sigma Pi Sigma, an extra-special PhysCon is planned for October 6–8, 2022 in Washington, DC. With a little preparation, you'll have the chance to narrow down your graduate school search, meet potential employers, and make lasting connections with people heading down similar career paths.

The most direct opportunity to meet with representatives from physics and astronomy grad programs and potential employers occurs during the Expo, which encompasses both a grad school fair and a career fair. During the Expo, attendees can visit booths to learn more about a program, company, or undergraduate research experience as well as get tips and advice on applying. When I attended, seeing the wide variety of vendors enabled me to start thinking about my life after col-

NETWORKING TIPS

Before you attend a networking event, craft and practice your **elevator pitch**—a 30-second narration of who you are professionally, what you've accomplished, and where you hope to go in the future.

If you're attending an in-person event as a prospective student or employee, **business cards** (or contact cards) show that you're serious about your future and make it easy for new contacts to connect with you.

BE AN SPS INTERN

The Society of Physics Students summer internship program offers 10-week, paid positions for undergraduate physics students in science research, education, communication, and policy with various organizations in the Washington, DC, area.

www.spsnational.org/programs/internships.

lege, and I was blown away by the versatility that a degree in physics can provide.

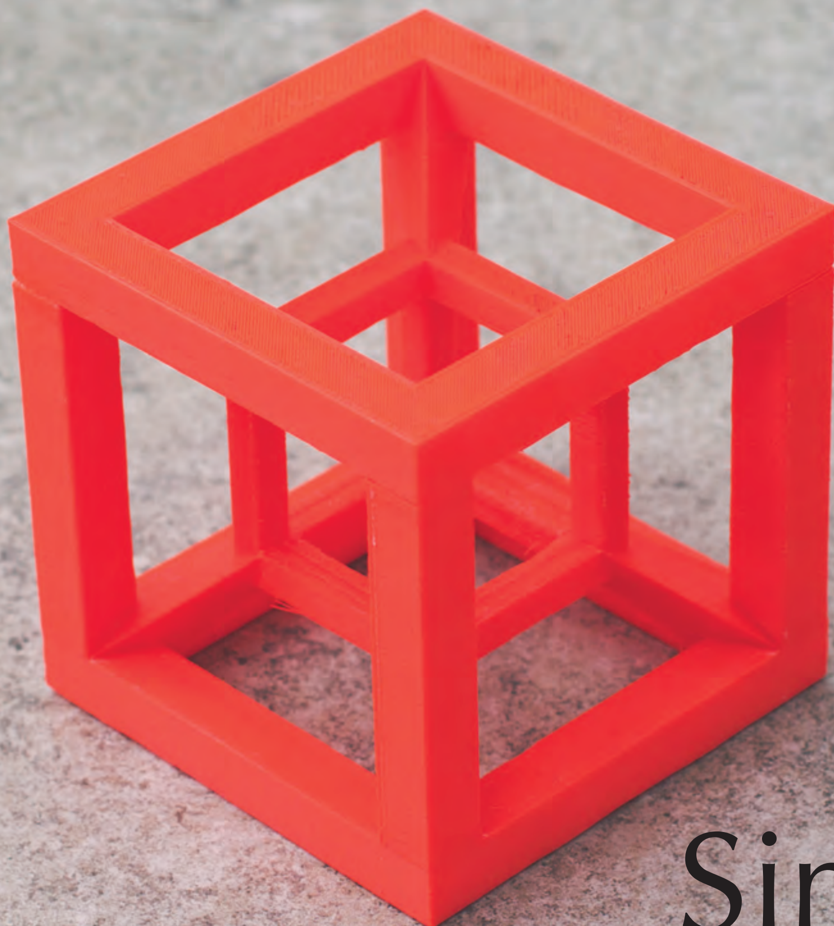
A more subtle opportunity to build your network as a young professional is to engage with attendees you don't already know, between events or at meals. Shuffling between workshops, plenaries, and banquets will be hundreds of people with lived experiences similar to yours. Be adventurous and sit at a meal or workshop table with strangers! You might find yourself next to a professor from a graduate school you're interested in, or even from a school you didn't realize you should be interested in. A quick conversation can leave a lasting impression.

A straightforward way to meet students and professionals is to go to the poster sessions, as a presenter or an attendee. These are excellent opportunities to have one-on-one interactions with others and to learn about new topics. Seeking out posters in subfields you're doing research in or interested in studying in grad school is a great way to form connections and learn about current research in the field. My favorite question to ask a presenter is "Can you tell me more about your research?" They likely have an answer prepared, which can be a bridge to more natural conversation.

2019 Physics Congress attendees visit one of the many graduate school booths in the exhibit hall to learn about the program and check out physics demonstrations. Photo courtesy of SPS National.

The physics and astronomy community is quite small, so if you meet people at PhysCon, you're likely to run into them again. Almost a year after I attended PhysCon 2016, I was a Society of Physics Students intern. Of the 14 of us, over half had met previously, largely at PhysCon. Having that shared experience helped me connect with the other interns right from the start. We even looked back at old PhysCon photos and tried to spot one another in the background, which was wildly entertaining.

Attending PhysCon is the networking gift that keeps giving. I have met others who attended in different years and we're still able to bond over our shared experiences. You are bound to find someone with similar interests and goals in a sea of over a thousand physics students, mentors, and advisers. Preparation is the key to successful networking, so practice your elevator pitch, make business cards, and I'll see you in 2022! GSS



Simulating FOUR-DIMENSIONAL PHYSICS in the laboratory

Hannah Price

Experimental methods to imitate extra spatial dimensions reveal new physical phenomena that emerge in a higher-dimensional world.

ISTOCK.COM/PIRANKA

Hannah Price is a Royal Society University research fellow in the theoretical physics group and a proleptic reader at the University of Birmingham in the UK.

What would the universe be like if it had four spatial dimensions instead of three? Experimentalists are starting to explore the physics of higher dimensions with the help of recently developed tricks that synthetically mimic an extra fourth dimension in platforms such as ultracold atoms, photonics, acoustics, and even classical electric circuits. Although any such trick necessarily has limitations, as the fourth spatial dimension is always artificial, those approaches have proven that they can simulate some four-dimensional effects in controlled experimental systems.

But what is a fourth spatial dimension? In nonrelativistic physics, in which space and time are distinct, a spatial dimension is simply a direction along which objects can move both forward and backward (unlike time, which always flows from past to future). The number of relevant spatial dimensions in a system is defined by the directions along which spatial motion can take place or, alternatively, the number of spatial coordinates—for example, (x, y, z)—that must be specified to define where an object is at a particular moment in time.

The number of spatial dimensions can be reduced by constraining a system. For example, threading a bead onto a long, straight wire limits the bead to move in only one spatial dimension: either forward or backward along the wire. A single coordinate gives the bead's position along the wire at any given moment.

What would happen, then, with an increase in the number of spatial dimensions to four or more? Theoretical physicists can simply extend familiar physical equations to an enlarged set of spatial coordinates—for example, (x, y, z, w). Often that extension leads to no new phenomena. But in certain fields of physics, new effects are predicted to emerge, such as so-called topological insulators, which are the primary source of inspiration for efforts to simulate 4D physics experimentally. This article delves into what 4D physics is and how experimental tricks to mimic 4D space work.

Topological insulators

The transfer of topological concepts from mathematics to physics has deepened researchers' understanding of states of matter and led to the discovery of a plethora of exotic topological materials. In mathematics, topology is most famously a framework to classify different surfaces. For example, donuts belong to the family of surfaces with one hole, whereas oranges belong to the family with no holes. If one smoothly squishes an orange, its shape deforms, but it cannot take the shape of a donut without tearing a new hole and thereby changing the topology, in that case quantified by an index known as the genus. Other mathematical problems have many other sorts of topological indices, such as the family of so-called Chern numbers, which are discussed later.

In physics, topological indices lie at the heart of electrical, optical, and other behaviors in many materials.[1] In particular, they often classify electronic energy bands in a crystal. When nontrivial, those indices guarantee special properties, such as the existence of currents circulating around the edge of a material despite the bulk remaining insulating—as in the aptly named topological insulator. Similar to the genus of a squishable orange, the indices are hard to change, so topological properties, such as those special edge currents, can be remarkably robust even in the face of disorder, as long as the bulk remains insulating.

Spatial dimensionality changes the nature of topological insulators and their edge currents. As depicted in figure 1, a 2D topological insulator has effectively 1D conducting edge channels, whereas a 3D topological insulator is covered with 2D conducting surfaces. Similarly, a 4D topological insulator should be an unusual material with robust 3D conducting surface volumes. What's more, not only the edge behavior but also the underlying physics and the definitions of topological indices depend on the spatial dimensionality and symmetries of the system.[1]

Quantum Hall effects

The story of 4D topological insulators starts with the 2D quantum Hall effect, discovered in 1980 by Klaus von Klitzing of the Max Planck Institute for Solid State Research in Stuttgart, Germany. That research earned him the 1985 Nobel Prize in Physics.

As the name suggests, the 2D quantum Hall effect is intrinsically a 2D phenomenon, first observed in an effectively 2D electron gas moving in a high-quality semiconductor heterostructure.[1] In his seminal experiment, von Klitzing exposed silicon-based heterostructures to low temperatures and high out-of-plane magnetic fields. He then flowed a current through his device and measured the voltage across it to find the Hall conductance. What he found was unexpected: The conductance exhibited robust plateaus that were precisely quantized by integer multiples of e^2/h, where e is the electron charge and h is Planck's constant. In fact, that quantization is so robust and precise that it became part of the 2019 redefinition of the

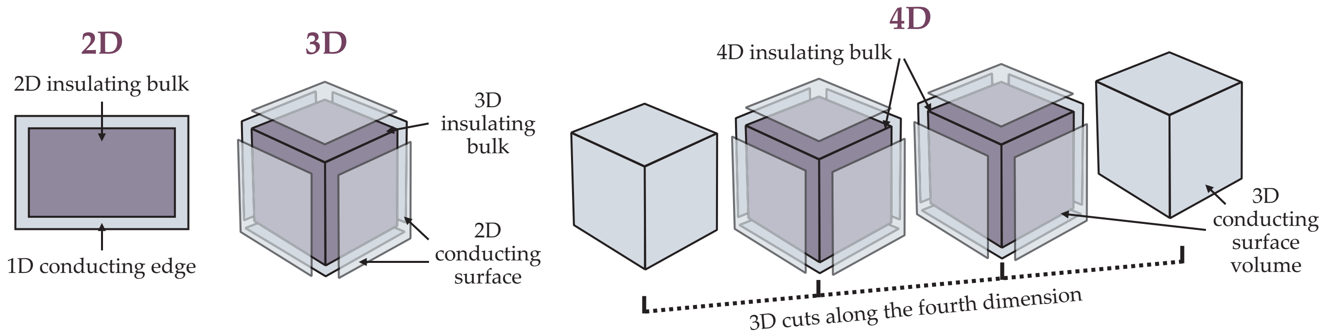


FIGURE 1. TOPOLOGICAL INSULATORS in two, three, and four dimensions conduct on their edges or surfaces (light gray) despite an insulating bulk (violet). That unusual behavior results from the topology of the electronic band structure. In 3D and 4D systems, the conducting surfaces are depicted lifted off the bulk to show it and the surface simultaneously. The 4D topological insulator is shown as several separate 3D cuts along the fourth dimension.

kilogram in SI units. (See the article by Wolfgang Ketterle and Alan Jamison, PHYSICS TODAY, May 2020, page 32.)

In 1982 David Thouless of the University of Washington in Seattle and his colleagues showed that the origins of the 2D quantum Hall effect lie in the topological nature of the electronic energy bands. That realization was, in part, why Thouless was awarded a share of the 2016 Nobel Prize in Physics. The integer in the Hall conductance is related to a 2D topological index called the first Chern number, which guarantees the existence of topological currents around the edge of the material[1] (see figure 1 and box 1). In other words, a 2D quantum Hall system is an example of what would now be called a topological insulator, with the robustness of the Hall conductance being one of its key experimental signatures.

After the discovery of the 2D quantum Hall effect, theorists suggested that certain 3D materials would also have bands characterized by first Chern numbers, except in that case a triad of them: one for each of the three Cartesian planes of the 3D material. The theorized 3D quantum Hall effect was indeed observed experimentally in 2019 in bulk zirconium pentatelluride crystals.[2] But the 3D quantum Hall effect is what's often referred to as a weak topological phenomenon because key properties, such as the first Chern numbers, remain essentially 2D concepts even though the system is 3D. The resulting topological behavior can thus sometimes be less robust.

In four spatial dimensions, however, a fundamentally different type of quantum Hall effect was proposed in the early 2000s by Jürg Fröhlich and Bill Pedrini from ETH Zürich in Switzerland and independently by Shou-Cheng Zhang and Jiangping Hu of Stanford University.[3] That 4D quantum Hall effect has a different form of quantized Hall conductance from its 2D cousin and is instead related to a 4D topological invariant called the second Chern number, which creates 3D conducting surface volumes, as shown in figure 1.

To date, various 4D quantum Hall models have been proposed.[3–5] Some, similar to the 2D quantum Hall effect, describe charged particles in magnetic fields. Others, such as that of Zhang and Hu, exploit the physics of a Yang–Mills gauge field, as explained in box 2, and take inspiration from particle physics.

The 4D quantum Hall effect is not the end of the story. Over the past 20 years, other quantum Hall effects have been predicted in 6D and 8D systems, while many other families of 2D and 3D topological insulators have been discovered that require topological invariants other than Chern numbers.[1] Mathematical classifications categorizing topological phases of matter up to arbitrary numbers of spatial dimensions also suggest other higher-dimensional phenomena waiting to be uncovered.[6]

Not just electrons

Bringing the physics of higher dimensions into the laboratory requires thinking beyond solid-state materials—where the 2D and 3D quantum Hall effects were observed—to other more controllable platforms.

Although originally associated with electronic transport, many topological properties are now understood instead to stem from band theory and the general physics of waves.[5] In other words, a topological index, such as the first Chern number, also applies to ultracold atoms, classical waves of light, mechanical oscillations, and waves on the ocean surface, to name just a few possibilities.

Intuitively, classical waves or noninteracting bosons shouldn't be called topological insulators, because without the Pauli exclusion principle or other effects to fill up the states in an energy band, those systems will not be insulating in the usual sense. The current convention, however, is to use the term topological insulator whenever the physics derives from energy bands with well-defined topological indices.[5]

Probing the topological physics of nonelectronic systems requires different experimental methods because those systems no longer have robust quantized plateaus in the Hall conductance. For wave-based systems, the most important experimental signature is instead typically the existence of robust modes localized on the system's surface at frequencies forbidden to penetrate the bulk. In those cases for a given frequency, waves can propagate on the surfaces but not in the bulk, as sketched in figure 1. Such topological protection may someday be useful for applications such as photonics devices because it provides a way to robustly guide light around any disorder and imperfections introduced during device fabrication.[5]

The expansion into nonelectronic platforms has also been advantageous for the study of topological phenomena (see "Topological insulators: from graphene to gyroscopes," PHYSICS TODAY online, 27 Nov 2018). Many of those platforms are easier to engineer than real materials and have thus allowed scientists to explore beyond what is currently accessible in solid-state physics.[5] As part of the push, researchers have developed ex-

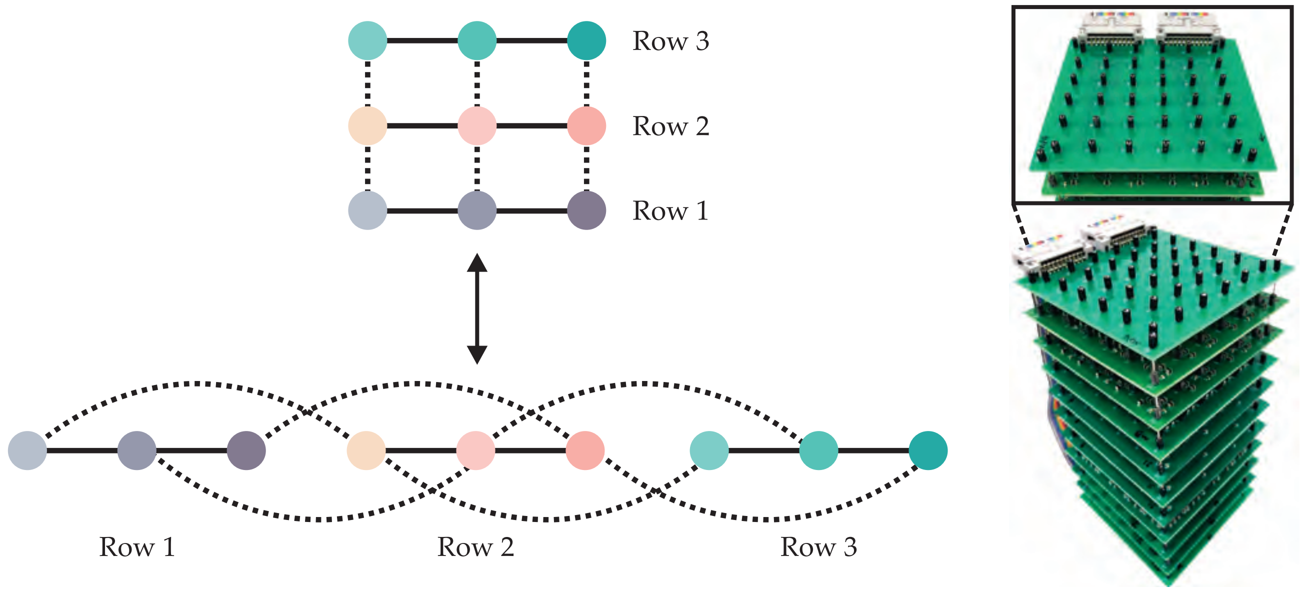


FIGURE 2. HIGHER-DIMENSIONAL LATTICES can be constructed in lower-dimensional systems. On the left, a two-dimensional discrete lattice model is composed of lattice sites (circles) with connections (lines). That same lattice can be effectively embedded into one dimension provided the same connectivity is maintained. On the right, that embedding trick was used to encode a 4D lattice into this 3D stack of circuit boards. (Photo from ref. 9.)

perimental tricks to mimic extra dimensions, in part to probe higher-dimensional topological insulators. Three main approaches are topological pumping, connectivity, and synthetic dimensions, although other schemes are also under development.

Topological pumping

One of the earliest but perhaps most abstract tricks to mimic higher dimensions is topological pumping, which Thouless first proposed in 1981 as a method to realize the 2D quantum Hall effect. He predicted that slowly tuning the parameters of certain types of 1D quantum systems could robustly pump particles across the system.[1]

The simplest example starts with an insulator in which particles occupy every minimum of a 1D chain of periodic potential wells. If the overall potential's location in space is then slowly tuned such that the entire crystal slides along the chain, the resulting motion of the minima drags the particles along with it. Thouless calculated not only that such robust particle transfer was a product of a topological invariant but that the invariant was the same 2D index—the first Chern number—as in the 2D quantum Hall effect. The result suggested that, in a sense, a 1D topological pump is a dynamic version of the 2D quantum Hall effect, as has since been explored experimentally.

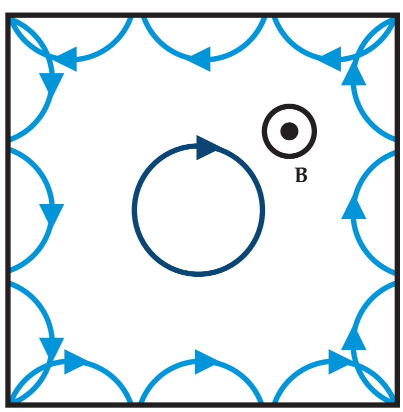
Going from one dimension to two dimensions may seem quite far from higher-dimensional physics. But in 2013 Yaacov Kraus of the Weizmann Institute of Science in Israel, Zohar Ringel of Oxford University in the UK, and Oded Zilberberg of ETH Zürich predicted that a 2D topological pump would be related to the 4D topological index—the second Chern number—of the 4D quantum Hall effect.[4]

The prediction proved correct in 2018 in two complementary experiments led by Zilberberg. One he conducted in photonics with the team of Mikael Rechtsman at the Pennsylvania State University; the other was in cold atoms with the team of Immanuel Bloch at the Max Planck Institute of Quantum Optics and the Ludwig-Maximilians University Munich in Germany and my collaboration at the University of Birmingham.[7] Those experiments identified signatures of the 4D quantum Hall effect in the propagation of light around the edge of a waveguide array and in the net motion of atoms across a system, respectively, and have since been extended by other groups to acoustic platforms.

Topological pumping has many intrinsic limitations because it is essentially a mathematical trick based on slicing up a higher-dimensional model in a clever way. In reality, the particles are only ever able to move in the lower-dimensional system and do not have the full

BOX 1. SKIPPING ALONG THE EDGE

What is the origin of the two-dimensional quantum Hall edge current? Classically, when a charged particle confined to 2D motion experiences an out-of-plane magnetic field **B**, it executes closed cyclotron orbits in the bulk (dark blue circle), but one-way skipping orbits along the boundary of the box (light blue arrows). Even if the boundary is deformed, those skipping orbits keep moving in the direction dictated by the orientation of the magnetic field. Quantum mechanically, that behavior translates to the characteristic insulating bulk energy bands and robust conducting edge states of a topological insulator.



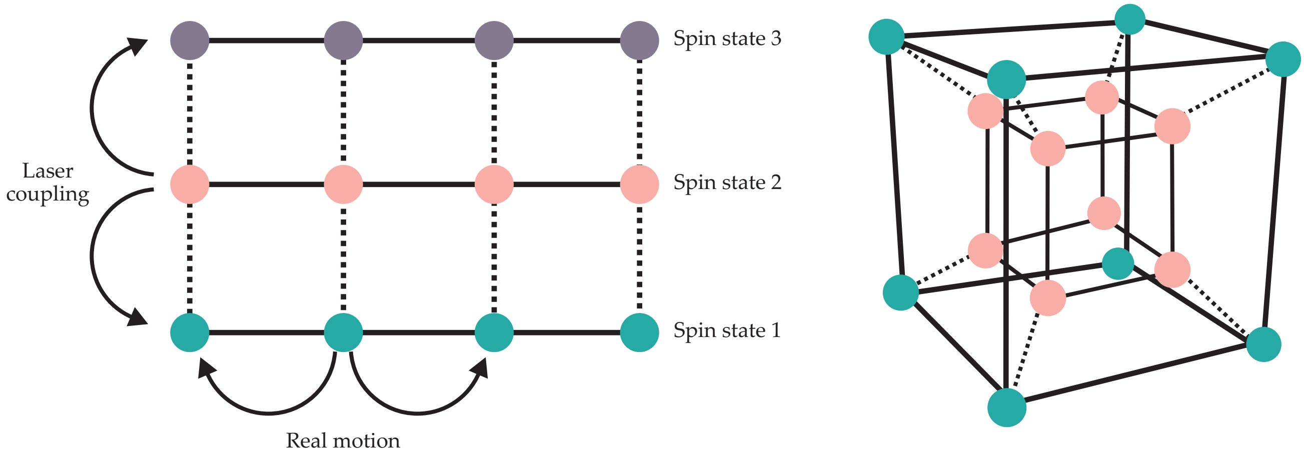


FIGURE 3. SYNTHETIC DIMENSIONS turn atomic spin states—or other internal states or intrinsic properties—into something similar to a spatial dimension. A two-dimensional discrete lattice model **(left)** comprises one real spatial dimension and one synthetic dimension composed of atomic spin states. Hopping along the real dimension (solid lines) corresponds to real atomic motion, whereas hopping along the synthetic dimension (dashed lines) corresponds to laser-induced transitions between spin states. The unit cell of a 4D hypercubic lattice **(right)** is a tesseract. Such shapes can be crafted with a suitable combination of real and synthetic spatial dimensions.

freedom of higher dimensions. Something closer to real higher dimensions may be possible through other types of experimental schemes.

Electrical circuit connectivity

The second method to simulate higher spatial dimensions is based on the idea of connectivity, which can be understood by starting with discrete lattice models. In those models, particles can exist only on a set of lattice sites. Those sites can be represented as a set of discrete points distributed in space, as shown in figure 2. Depending on the model specifics, particles can hop between pairs of lattice sites, as indicated by the dashed lines. Such discrete lattice models are common approximations for real systems, including electrons moving through a solid-state material and an electrical current moving around a circuit. They can also identify and isolate the essential ingredients of phenomena.

The key point for understanding higher-dimensional simulations is that a discrete lattice model is essentially a network of nodes, as in lattice sites, and connections, as in allowed hops. That perspective reveals that it does not matter where the nodes are physically located in real space, provided that all the connections are the same.

For example, the 2D square lattice in figure 2 can transform to a 1D chain if each row of sites is laid out end to end. So long as the same types of connections exist between sites, the system will obey the same mathematical equations as before. In a sense, the process embeds the 2D model into a 1D scheme—albeit a strange 1D scheme, in which some short-range connections are absent, while other long-range connections appear.

The same idea extends to higher-dimensional lattices too—for example, creating a 4D lattice model with a 3D or 2D scheme. The embedding trick therefore offers a recipe for realizing a 4D lattice model in a real physical system, but with the challenge of engineering complicated connections between sites.

In an early proposal from 2013, Dario Jukić and Hrvoje Buljan from the University of Zagreb in Croatia envisioned simulating a discrete 4D lattice with photonic waveguides.[8] Since then, research interest has focused on more flexible systems, such as electrical circuits, with various proposals for how lattice sites composed of inductors, capacitors, and resistors can be wired together to realize 4D topological models.

In 2020 You Wang, Baile Zhang, and Yidong Chong of the Nanyang Technological University in Singapore and I applied the approach for the first time experimentally, as shown in figure 2. We created a small 4D topological lattice of 144 sites embedded in an electrical circuit.[9] In the experiment, we designed a stack of 3D circuit boards and wired them together to match a 4D discrete lattice model for the 4D quantum Hall effect. As predicted for a 4D topological insulator, we observed that currents flowed through the sites that would be on the surface of the 4D topological insulator but not through the bulk.

Those electrical circuit experiments do have limitations be-

BOX 2. EXOTIC MONOPOLES

One way to think about topological pumping is that it replaces some of the real spatial dimensions in the Hamiltonian with externally controlled parameters. But if all the spatial dimensions are swapped for externally controlled parameters, then no real spatial degrees of freedom are required to simulate higher dimensions.

In 2018 Seiji Sugawa, Ian Spielman, and their colleagues at the Joint Quantum Institute and the University of Maryland in College Park used that type of approach. Inspired by the work of Shou-Cheng Zhang and Jiangping Hu on the 4D quantum Hall effect,[4] the researchers experimentally simulated what's known as a Yang monopole in an effective 5D parameter space created by coupling four internal states of an atomic quantum gas.[16] Similar to how Paul Dirac postulated the hypothetical magnetic monopole as a source for the magnetic field, the Yang monopole is proposed as the source of a Yang–Mills gauge field in five dimensions. Sugawa, Spielman, and their colleagues mapped out the properties of the simulated monopole and verified that it is characterized by the second Chern number, as predicted.

More recently, in 2020, similar experimental approaches have simulated so-called 4D tensor monopoles, which are postulated as the sources of tensor gauge fields and are characterized by an exotic topological index called the Dixmier–Douady invariant.[14]

cause they typically cannot access the whole energy spectrum of states at once. They are also classical systems, which cannot exhibit quantum effects. Nevertheless, the simplicity of manufacturing electrical circuits and their flexibility make them a fruitful avenue to explore 4D physics.

Synthetic dimensions

The final trick—synthetic dimensions—gets closest to genuinely simulating particles moving in four dimensions. The method interprets some set of a system's internal states or intrinsic properties as lattice sites along an imaginary extra dimension.[5] By combining that strategy with other real or synthetic dimensions, it has the potential to realize high-dimensional lattice models.

To get a feel for how the trick works, consider the example of a gas of identical atoms trapped in a vacuum chamber and cooled close to absolute zero. Each atom has various possible internal atomic spin states, which correspond to different configurations of its constituent electrons and nucleus. Shining suitable lasers onto an atom can stimulate a sequential transition between those internal states, as sketched in figure 3. As those transitions happen, the atom's spin-state label changes step-by-step, similar to how a discrete spatial coordinate changes when particles hop between lattice sites. That analogy is powerful and effective, and it reframes different spin states as spanning a synthetic dimension.

The idea of synthetic dimensions of atomic spin states originated in 2012 in work by Octavi Boada and José Ignacio Latorre at the University of Barcelona in Spain and Alessio Celi and Maciej Lewenstein at the Institute of Photonic Sciences in Barcelona.[10] The same idea was extended three years later to a discrete 2D quantum Hall lattice model with one real dimension and one synthetic dimension realized in cold-atom experiments, explained in box 3. In the future, the approach may be pushed even further to realize 4D topological models.[5]

Since 2015 the field of synthetic dimensions has expanded dramatically. One prominent innovation was to swap spin states for atomic momentum states in cold atoms. The momentum states can be coupled into a synthetic dimension by pulsing a standing wave of light, which kicks the atoms and changes their momenta by quantized amounts along the wave's direction.[5] Ulrich Schneider's group at the University of Cambridge in the UK recently extended that approach to four separate standing waves of light at once, each one pointing along a different direction in the 2D plane. The feat engineered up to four synthetic dimensions simultaneously.[11] Although not yet topological, the results of the experiment could be interpreted in terms of atoms hopping on a 4D hypercubic lattice, as shown in figure 3, that is composed of momentum states.

Photonics has also undergone significant recent developments in synthetic dimensions. Most notable are two schemes: one in which the synthetic dimension is formed from the frequency modes of a ring cavity and the other in which it's formed from the lattice modes of a waveguide array. Shanhui Fan at Stanford University and his colleagues demonstrated two simultaneous independent synthetic dimensions based on frequency modes in a single photonic cavity.[12] Mordechai Segev's group at the Technion–Israel Institute of Technology in Haifa proposed and developed experiments based on lattice modes, which have already revealed both 2D and 3D topological edge physics with a synthetic dimension.[13] Both approaches may someday lead to realizations of 4D topological insulators.

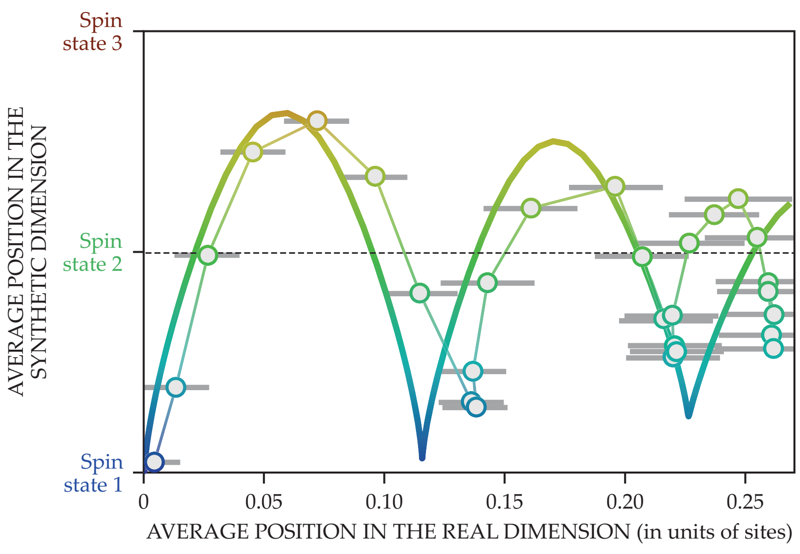
Opening up higher dimensions

Despite so much progress over the past few years, experiments simulating 4D physics are still in their early stages. Topological pumps have successfully employed mathematical tricks to observe signatures of 4D effects, but they cannot completely capture 4D dynamics. Electrical circuits can capture the full connectivity of a 4D topological lattice, but they have not yet provided full access to 4D physics. In the future, all those limitations will hopefully be overcome by synthetic dimensions, in which particles may be able to move as if in 4D space.

Synthetic dimensions may also reveal new ways to think about the 3D world. After all, a synthetic dimension consists of coupling together existing physical degrees of freedom. For example, creating a synthetic dimension of optical frequency modes involves controlling the frequency of light, whereas finding topological edge currents in such a setup is about identifying a new mechanism to robustly channel light or convert

BOX 3. SKIPPING IN A SYNTHETIC DIMENSION

In 2015 the groups of Leonardo Fallani and Massimo Inguscio at LENS (the European Laboratory for Nonlinear Spectroscopy) and the University of Florence and of Ian Spielman at the Joint Quantum Institute and the University of Maryland in College Park both realized a two-dimensional quantum Hall system made of a real spatial dimension and a synthetic dimension of three atomic spin states,[17,18] similar to those in figure 3. As shown here (adapted from reference 17), the systems exhibited the key signature of Hall physics: skipping orbits along the edge of the system, analogous to those of a charged particle in a magnetic field, as explained in box 1.



its frequency. By giving an alternative viewpoint for understanding and designing complex systems, synthetic dimensions may, in the long term, lead to applications in optical isolators or the spectral manipulation of light, for example.[5,12]

In terms of fundamental science, much more 4D physics is left to explore. The topics in this article are all single-particle physics, in that particle–particle interactions are negligible. Only a few steps have been taken in the theoretical understanding of 4D phenomena, such as Zhang and Hu's proposed generalization[3] of the 2D fractional quantum Hall effect to four dimensions. Understanding what many-body physics might emerge in higher dimensions and whether those phenomena can be accessed with current experimental tricks requires further work.

From the experimental point of view, a future challenge is that particle–particle interactions naturally depend on the particle separation in the real 3D world rather than in the synthetic 4D system.[5] In the case of synthetic dimensions, for example, two atoms in different spin states often interact strongly so long as they occupy the same physical location. Those interactions correspond to strange nonlocal interactions along the synthetic dimension. Researchers are developing various approaches to understand and tackle such problems.

Finally, although the simulation of 4D physics started with the 4D quantum Hall effect, the field should flourish far beyond that effect in the future. Recent experiments have already shown other topological effects, such as the exotic 4D tensor monopoles[14] described in box 2. Other experimental tricks are also in development, including schemes based on using multiterminal Josephson junctions to replace spatial degrees of freedom with superconducting phases.[15] In the near future, more 4D physics will be simulated in the laboratory.

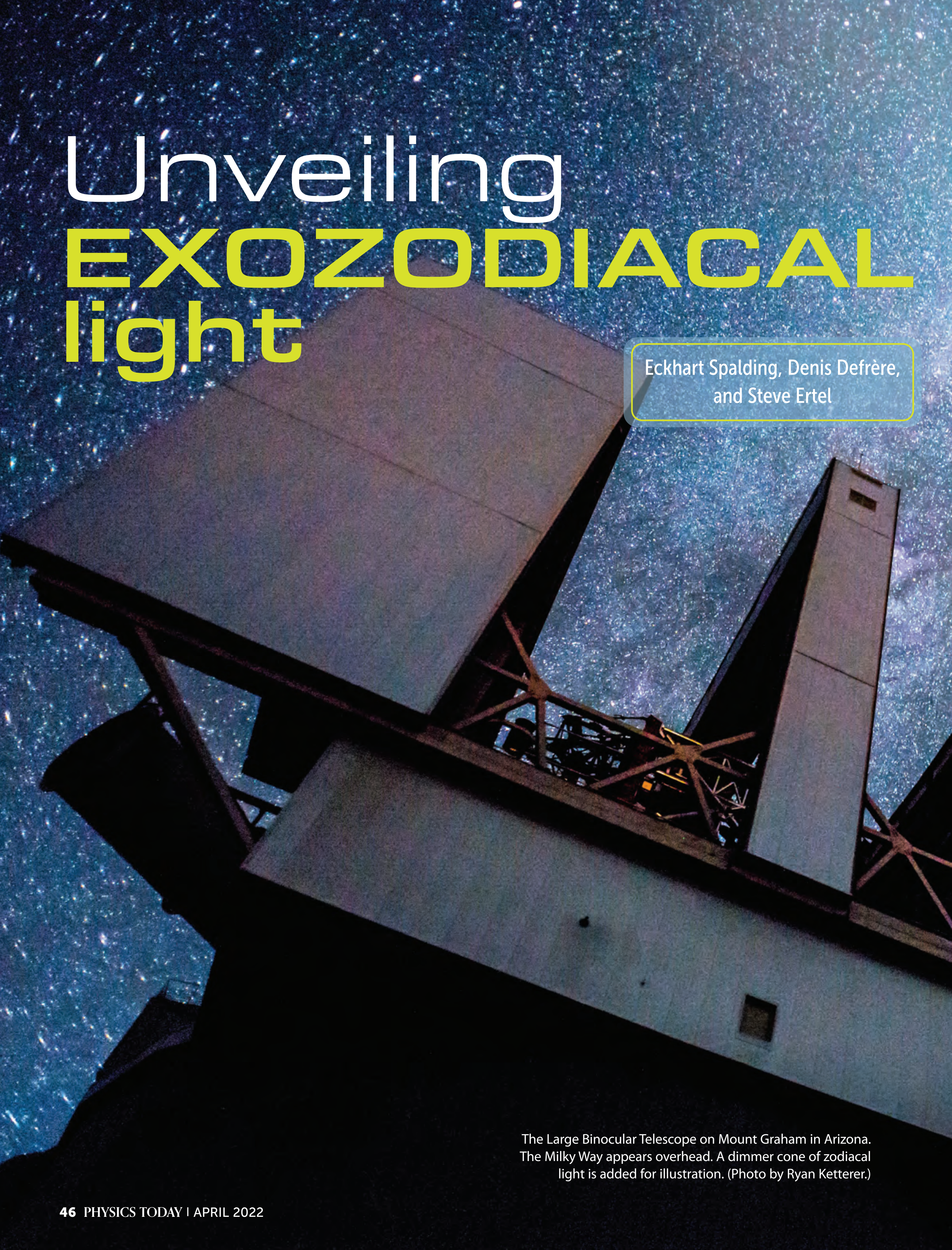
REFERENCES

1. B. A. Bernevig, T. L. Hughes, *Topological Insulators and Topological Superconductors*, Princeton U. Press (2013).
2. F. Tang et al., *Nature* **569**, 537 (2019).
3. J. Fröhlich, B. Pedrini, in *Mathematical Physics 2000*, A. Fokas et al., eds., Imperial College Press (2000), p. 9; S.-C. Zhang, J. Hu, *Science* **294**, 823 (2001); D. Karabali, V. P. Nair, *Nucl. Phys. B* **641**, 533 (2002).
4. Y. E. Kraus, Z. Ringel, O. Zilberberg, *Phys. Rev. Lett.* **111**, 226401 (2013).
5. N. R. Cooper, J. Dalibard, I. B. Spielman, *Rev. Mod. Phys.* **91**, 015005 (2019); T. Ozawa et al., *Rev. Mod. Phys.* **91**, 015006 (2019); T. Ozawa, H. M. Price, *Nat. Rev. Phys.* **1**, 349 (2019).
6. C.-K. Chiu et al., *Rev. Mod. Phys.* **88**, 035005 (2016).
7. O. Zilberberg et al., *Nature* **553**, 59 (2018); M. Lohse et al., *Nature* **553**, 55 (2018).
8. D. Jukić, H. Buljan, *Phys. Rev. A* **87**, 013814 (2013).
9. Y. Wang et al., *Nat. Commun.* **11**, 2356 (2020).
10. O. Boada et al., *Phys. Rev. Lett.* **108**, 133001 (2012).
11. K. Viebahn et al., *Phys. Rev. Lett.* **122**, 110404 (2019).
12. A. Dutt et al., *Science* **367**, 59 (2019).
13. E. Lustig et al., *Nature* **567**, 356 (2019); E. Lustig et al., in *Conference on Lasers and Electro-Optics*, Optica (2020), paper FW3A.2.
14. X. Tan et al., *Phys. Rev. Lett.* **126**, 017702 (2021); M. Chen et al., http://arxiv.org/abs/2008.00596.
15. R.-P. Riwar et al., *Nat. Commun.* **7**, 11167 (2016).
16. S. Sugawa et al., *Science* **360**, 1429 (2018).
17. M. Mancini et al., *Science* **349**, 1510 (2015).
18. B. K. Stuhl et al., *Science* **349**, 1514 (2015).

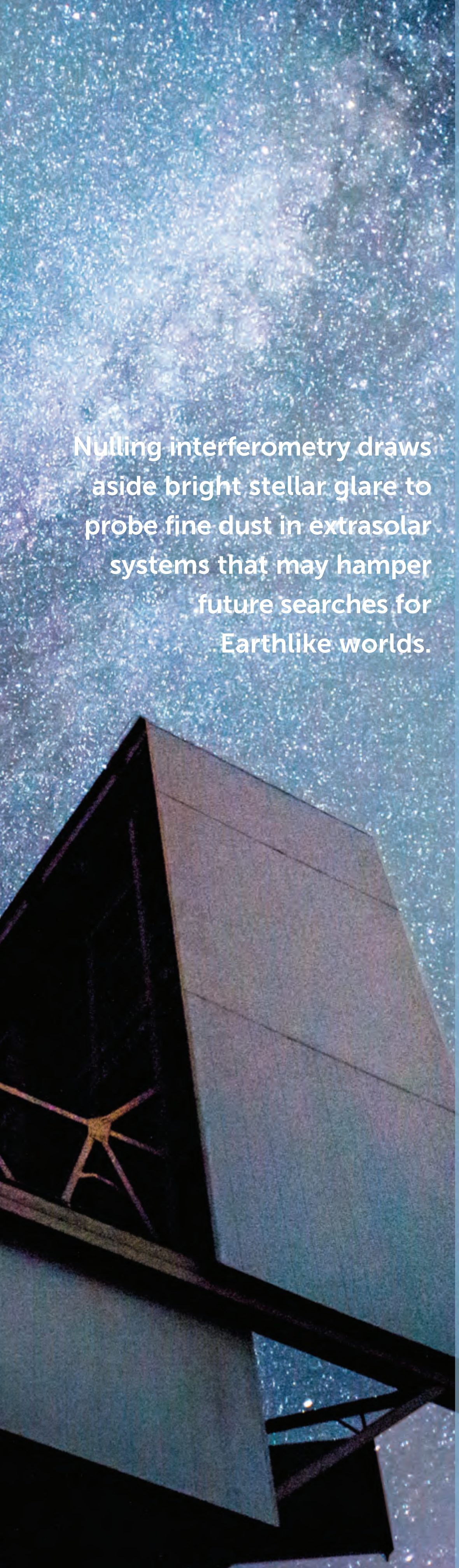
PT

Unveiling EXOZODIACAL light

Eckhart Spalding, Denis Defrère, and Steve Ertel



The Large Binocular Telescope on Mount Graham in Arizona. The Milky Way appears overhead. A dimmer cone of zodiacal light is added for illustration. (Photo by Ryan Ketterer.)



Nulling interferometry draws aside bright stellar glare to probe fine dust in extrasolar systems that may hamper future searches for Earthlike worlds.

Eckhart Spalding is a physics postdoctoral research associate at the University of Notre Dame in Indiana. **Denis Defrère** is an associate professor of astronomy and instrumentation at KU Leuven in Belgium and currently leads a project to build the first nulling instrument for the Very Large Telescope Interferometer in Chile. **Steve Ertel**, an associate researcher with the department of astronomy and Steward Observatory at the University of Arizona in Tucson, is the lead scientist of the Large Binocular Telescope Interferometer on Mount Graham.

Under a dark sky and with the proper celestial orientation, stargazers can see the plane of the Milky Way as a soaring thoroughfare of glowing stars and the silhouettes of dust clumps and tendrils. If the season is right and the eyes adjust enough, a dimmer cone of light emerges that intersects the Milky Way at an angle of about 60 degrees. That cone is zodiacal light: sunlight scattered by small dust grains orbiting the Sun in the ecliptic plane. The dust begins a few solar radii from the Sun and stretches out to the asteroid belt. It is optically thin, with a total mass about 10^{-9}–10^{-8} times that of Earth and originates mostly from the residue of comet tails and asteroid smashups.

Until recently, the demography of equivalent dust populations around other stars—exozodiacal disks—was unknown. Yet such disks are of paramount importance in understanding exoplanetary systems whose architectures and dynamics leave imprints in the disks' shapes, thicknesses, and compositions. Exozodiacal disks also carry environmental signatures of a star's habitable zone (HZ), commonly defined as the region around the star where liquid water can be expected to exist on the surfaces of Earth-sized exoplanets with atmospheres. The HZ environment influences a rocky planet's ultimate fate: It can cause the planet to remain an uninhabitable rocky orb, evolve into a steamy water world with a punishing greenhouse effect, or end up a balmy habitable middle ground, among other possibilities.

Roughly 5000 exoplanets are known to date from various detection methods. The transit method has uncovered most of them, but it requires a planet to pass in front of the host star and is best suited for planets with tight, short-period orbits that enable repeated observations. Direct imaging has uncovered just a few tens of exoplanets. The method's strength lies in its ability to characterize the atmospheres of exoplanets, particularly those with wide orbits, without requiring the planetary system to be close to an edge-on orientation relative to Earth's line of sight.

Direct imaging can also peer deeper into exoplanet atmospheres and pick out absorption lines of species that include water, oxygen, and methane. In certain combinations, high signal-to-noise detections of such chemical species could provide strong evidence of a biosphere.[1] Although an exozodiacal disk can strongly influence a planet's habitability, it can also snarl observations by acting as a noise source.

Probing extrasolar systems

The greatest obstacle for directly imaging an exoplanet is not the planet's dimness but the host star's blinding glare, which is particularly challenging for ground-based instruments that collect light waves after they have been aberrated by Earth's turbulent atmosphere. Although ground-based instruments continue to

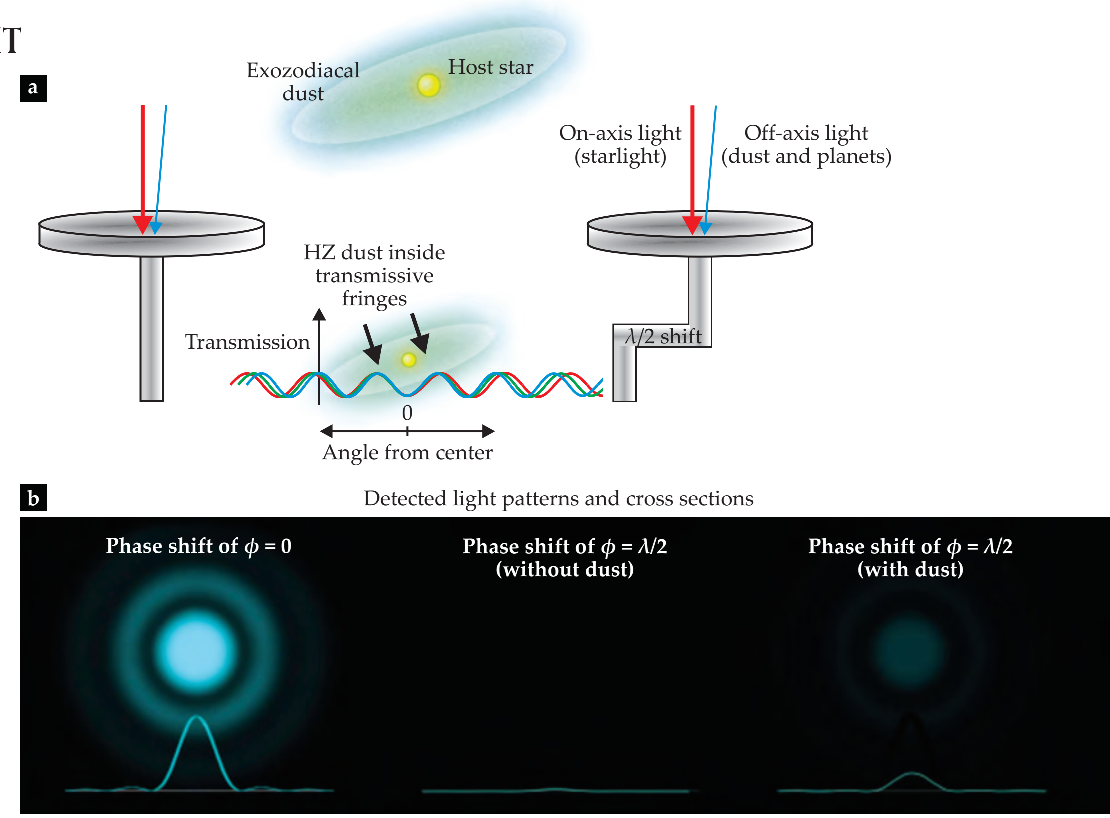


FIGURE 1. INTERFEROMETRIC NULLING separates a star's overpowering light from the dim glow of circumstellar material. **(a)** Light from a star and its surrounding material is collected by two apertures. One beam undergoes a half-wavelength phase shift, and combining the beams produces a destructive transmission fringe over the on-axis host star. Light from circumstellar material remains because it is off-axis and seeps through constructive transmission fringes. (Adapted from ESA 2002/Medialab.) **(b)** Imaging a star with a telescope produces a circular Airy pattern (left). If the star were bare, nulling would remove the entire signal except for trace leakage from phase noise or the star's nonzero angular diameter (center). But if the star has circumstellar material in its habitable zone (HZ), a dim signal remains (right).

reach ever-greater sensitivities and produce valuable results, the atmosphere imposes so much noise that large space telescopes remain the most promising route to directly imaging and characterizing Earth-sized exoplanets.

Space telescopes are colossally expensive, so mission planning must maximize the scientific return to justify the price. The best-informed mission plans inescapably require some foreknowledge of exozodiacal dust because such dust could well be the greatest obstacle to the mission's sensitivity. For a given amount of smoothly distributed dust, an observation's signal-to-noise ratio will depend on the integration time and the diameter of the telescope's primary mirror. A clumpy dust distribution can introduce additional noise and further complicate the analysis.[2] Even if exozodiacal dust is optically thin, as zodiacal dust is, the total light from such a disk may make it orders of magnitude brighter than a planet. Those factors have significant consequences for the size, lifetime, operational cadence, and cost of a mission.

Whatever the precise nature of exozodiacal light, one contribution is almost certainly Poynting–Robertson drag, whereby an orbiting particle's angular momentum diminishes as it absorbs stellar photons and reemits them. Over time it causes dust particles to creep inward from cold primordial outer rings, such as the solar system's Kuiper belt. As the dust spirals inward, collisions grind the grains down to sizes small enough for radiation pressure from the host star to blow them out of the system.

Minor bodies can also produce exozodiacal dust. A planet's gravitational pull can scatter an object from a distant debris belt onto an orbit that brings it sufficiently close to the host star. Increased exposure to the star's light causes volatiles to outgas, thereby depositing dust in the inner system.

Scientists face the same fundamental challenge when imaging exozodiacal disks and exoplanets: detecting a faint signal separated by a tiny angle from a blazingly bright star. In a copy of the solar system 10 parsecs (approximately 3.26 light-years) away, Earth would sit just 0.1 arcsecond from the Sun—roughly the diameter of a quarter held 30 miles away. And the Sun would be 10 million times brighter than Earth at IR wavelengths. In principle, a stellar interferometer with a baseline length of 15–20 m can obtain that spatial resolution at an observing wavelength of around 10 μm—the wavelength at which HZ dust at around 300 K is expected to be most emissive. From the ground, however, the atmospherically induced perturbations that make stars twinkle are essentially prohibitive for imaging an exo-Earth.

The first IR nulling

In 1978, when exoplanets had yet to be discovered and the *Hubble Space Telescope* was little more than a vision, Roland Bracewell at Stanford University proposed that a space-based interferometric array twirling about an axis could detect exoplanets in the IR.[3] A schematic of the technique, known as nulling interferometry, is shown in figure 1.

In nulling interferometry, light is collected by two apertures at the ends of the interferometer arms. A half-wavelength phase shift is applied to one arm, and the two beams are combined to create a transmission pattern composed of alternating constructive and destructive interference fringes. On-axis light from the star is multiplied by a destructive transmission fringe, thereby effectively subtracting it from the combined signal. But light from an exoplanet or circumstellar dust, which enters the apertures at an angle relative to the starlight, is multiplied by a constructive transmission fringe and is therefore not sieved out with the starlight. The amplitude of the residual light rises and falls with time, either because the optical path length changes or because an asymmetric structure around the star flits in and out of view as the sky rotates overhead. An exoplanet's signature would appear as a characteristic ripple in the observed brightness.

In the 1990s Roger Angel and Nick Woolf at the University of Arizona in Tucson were aware that exozodiacal dust could

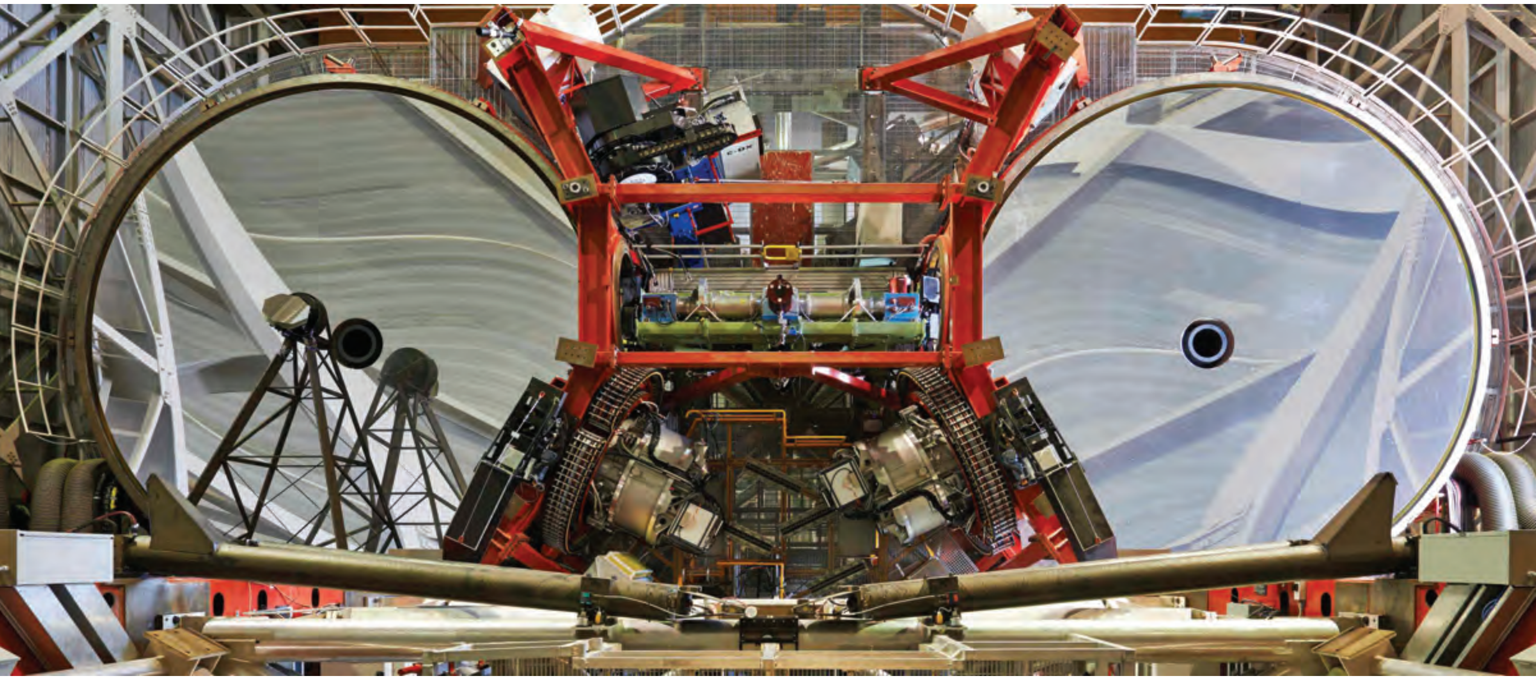


FIGURE 2. THE TWIN APERTURES of the Large Binocular Telescope sit 14.4 m apart, center to center. The green structure between them is an interferometer that enables the telescope to be used for nulling. (Courtesy of LBTO–Enrico Sacchetti.)

be a limiting source of noise for space missions attempting to image exo-Earths. They also knew that the planned Large Binocular Telescope would be able to characterize the dust from Earth using nulling interferometry. The LBT, shown in this article's opening image, was slated to have a unique configuration of large twin primary mirrors on a single mount,[4] shown in figure 2. (Angel also innovated a new mirror casting technique that used a rotating furnace to produce the 8.4-m-diameter mirrors—the largest monolithic mirrors for astronomy, then and now. He and coauthors Buddy Martin and John Hill describe the advances in mirror design in an article in PHYSICS TODAY, March 1991, page 22.)

While the LBT was under construction, Angel's then graduate student Philip Hinz, with assistance from William Hoffmann, Donald McCarthy, and others in the department, built a nulling instrument with a 5 m baseline between two mirrors of the Multiple Mirror Telescope (MMT). The detector was sensitive to light in the 8–28 μm range, which includes HZ dust emission.

The researchers placed a narrow null over the star Betelgeuse and measured the transmitted light from its dust nebula down to within 0.2 arcsec of the star. The resolution was better than that from the full MMT aperture; in that case, a perfect wavefront would have delivered a characteristic resolution of 0.3 arcsec. Although the turbulent atmosphere induced rapid changes in the optical path length between the two arms of the interferometer, the researchers used a series of rapid readouts to quantify the null depth—the excess light above that expected from perfect stellar light suppression, which originates from circumstellar emission—to within the limits imposed by phase noise.

Hinz and coworkers produced the first nulling constraints on circumstellar structure[5] in 1998. The project's primary objective was not to complete a full exozodiacal survey—the sensitivity was not yet good enough—but to provide a proof of concept for using null measurements to fit physical models of dusty disks around young stars.

Scaling up sensitivity

The same atmospheric turbulence that causes stars to twinkle also degrades the resolution of any large rigid, single-aperture ground-based telescope to that of a much smaller telescope just a few centimeters across. Increasing the sensitivity of nulling interferometry at the MMT therefore required at least partial correction for Earth's roiling atmosphere. To overcome atmospheric effects, scientists built adaptive optics (AO) systems with sensors to measure aberrations in the wavefront, compute a full wavefront reconstruction, and deform mirrors to cancel the aberration, typically at hundreds of Hz to roughly 2 kHz.

Serious development of AO began in the 1970s behind the curtain of classified US defense research, but it was still a relatively new technology in the civilian astronomical community[6] in the 1990s. The first telescopes to be upgraded with AO had to be retrofitted with additional optics between the telescopes' optical trains and their detectors.

When the MMT was outfitted with an AO system in 2002, the implementation was based on a novel design funded by the US Air Force: a deformable secondary mirror, integrated with the telescope's optics, that removed the need for additional optics and thereby minimized thermal noise.[7] The secondary mirror also benefited nulling observations by consolidating photons from an object onto a smaller area of the detector, thereby further reducing background noise, and by deblurring the object and straightening transmissive fringes.

Even with substantial wavefront correction, there was still enough atmospherically induced slope in the wavefront phase to cause the two beams from the subapertures to retain differential phase noise. The noise caused the on-sky transmission pattern to jitter, which in turn caused contaminant stellar light to flicker in the data.

For a time, the team at the MMT resorted to a slow, makeshift form of phase correction: They manually tuned an electric field over a piezoelectric ceramic mount to move an internal mirror and change the optical path length. By doing so, Hinz, with then graduate students Wilson Liu and Nathan Stock, resolved dust around young stars ensconced in gaseous envelopes and protostellar disks. They derived upper limits for the amount of HZ exozodiacal dust around a few older main-sequence stars, and, by incorporating complementary data from elsewhere, they determined that a bright debris disk around the star β Leonis consists of multiple rings of material.

As IR nulling ramped up at the MMT and construction of the LBT continued, scientists began using other facilities to observe extremely dusty systems so they could characterize the brightest end of the exozodiacal luminosity function. Space-based data came from the *Spitzer Space Telescope*'s infrared spectrograph in the 5–37 μm band, from its multiband imaging photometer in the 24 μm and 70 μm bands, and from the *Wide-Field Infrared Survey Explorer* (*WISE*) in the 12 μm band. Data from 1.6 μm to 2.2 μm came from ground-based interferometers, including the Infrared Optical Telescope Array on Mount Hopkins in Arizona, the array and instruments at the Center for High Angular Resolution Astronomy at Palomar Observatory on Mount Wilson in California, and the Very Large Telescope Interferometer on Cerro Paranal in Chile.

Although those interferometric observations did not involve nulling, they exploited the instruments' high angular resolutions to detect emission from thick, hot dust clouds close to the clouds' host stars. Models suggest that such dust comes predominantly from continual replenishment by a rain of infalling comets.[8] The Sun's F-corona scatters light from small amounts of dust at a similarly close separation, and Sun-grazing comets deliver small amounts of dust, but our solar system has no direct analogue to that thick, hot dust.

Those ground-based interferometric observations were not sensitive enough and were at wavelengths that were too short to detect the more tenuous dust near a star's HZ. The space-based photometry, for its part, could not suppress the target star's light enough to reveal faint circumstellar emission. Such measurements need ground-based mid-IR nulling interferometry.

For future space missions to image exoplanets, HZ exozodiacal dust would have to be characterized down to a sensitivity approaching the level of a few zodis. (One zodi equals the brightness of the zodiacal disk.) In 1997 NASA began funding the project to optically connect the twin Keck telescopes on Mauna Kea in Hawaii, thereby creating an interferometer. Operating in a nulling mode, the telescope became the Keck Interferometer Nuller (KIN) and was used to measure the exozodiacal luminosity function—that is, the number of exozodiacal systems per luminosity interval.

The KIN had multiple baselines: 4 m across the two halves of each individual telescope and an 85 m baseline between the two telescopes.[9] Between 2008 and 2011, the KIN collected a data set of 44 systems in multiple wavelength channels ranging from 8 μm to 13 μm. Based on the 8–9 μm measurements, the survey put an upper limit on the median brightness of exozodiacal disks around Sun-like stars at 60 times the thickness of the zodiacal disk,[10] or 60 zodis.

As the KIN survey came to an end, the LBT was getting up and running. Its twin monolith telescopes have center-to-center separation of 14.4 m, a number chosen so that baselines up to the nearly 23 m distance between the outer edges of the mirrors can be continuously sampled. The LBT's two sub-telescopes are each 8.4 m wide, and either one would be a large research telescope by today's standards. Both are crammed with a twin menagerie of instruments built and operated by a consortium of institutions and use adaptive secondary mirrors based on the pioneering tests conducted at the MMT.

The HOSTS survey

Because the LBT has a squat configuration, a compact baseline, and adaptive secondary mirrors, it can avoid the added complexity and thermal emission of long optical trains when it combines the beams from the sub-telescopes. The LBT Interferometer (LBTI) thus has less instrumental noise and achieves greater sensitivities than other facilities with longer baselines.

The LBTI group, headed by Hinz, who once led the early nulling experiments at the MMT, embarked on a new exozodiacal disk survey in the spring of 2014. The Hunt for Observable Signatures of Terrestrial Planetary Systems (HOSTS) survey targeted A-, F-, G-, and K-type stars. Those stars were subdivided into two subsamples: One contained A- and F-type stars, which are particularly hot and bright and consequently have HZs at larger radii from the stars. The LBTI would have the greatest sensitivity to dust in those systems. The other subsample included close Sun-like analogues that would be of greatest interest for future space-based missions. The full target list comprised 68 stars within 30 parsecs, close enough that emission from the inner HZ should be detectable.[11]

For their observations, the researchers

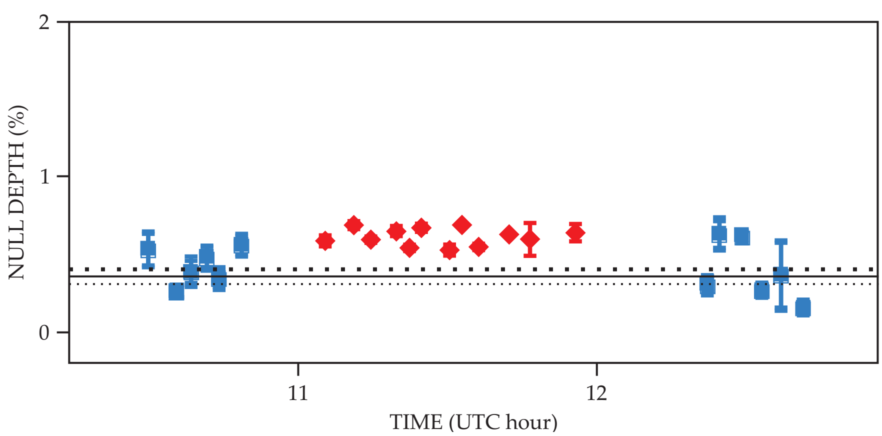


FIGURE 3. THE NULL DEPTH embodies the amount of circumstellar emission detected through nulling interferometry. The offset between the data for Vega (red), a star from the Hunt for Observable Signatures of Terrestrial Planetary Systems survey, and those for calibration stars (blue) indicates a glow from exozodiacal dust in Vega's habitable zone. The dotted lines represent the one-sigma uncertainty of the null floor.

alternated placing the central null over science stars—the targets of the investigation—and calibration stars. The offset in the null level between the two types of targets indicated light from circumstellar dust (see figure 3). The interferometer's tiny field of view minimized the probability of background sources, such as distant galaxies, masquerading as HZ dust.

The star η Corvi was chosen as an early proof-of-concept target because it was already known to be particularly dusty. But when the LBTI observed the star in early 2014, it found that the null depth corresponded to excess light of only about 4%, smaller than expected, given the 23% total disk-to-star flux ratio measured by *Spitzer*. Subsequent modeling suggested that much of the dust must be close to the star,[12] constrained to within 0.5–1.0 AU.

The LBTI group had developed a software loop to track the differential phase between the two sub-telescopes and send commands to an internal mirror to correct one of the beams. The team members improved their phase-tracking system after the η Corvi observation, but the improvements did not translate to better observational sensitivity when they looked at another system, β Leonis. They finally solved the problem with a data-reduction technique, nulling self-calibration, that was developed for the Palomar Fiber Nuller by Bertrand Mennesson of NASA's Jet Propulsion Laboratory in California and Charles Hanot of the University of Liège in Belgium. The algorithm uses the statistical distribution of the interferometric output signal to calculate the true astrophysical null. Applied to the LBTI data set, nulling self-calibration provided 5–10 times better null accuracy.

The high luminosity of β Leonis and its proximity to Earth have enabled dust in its HZ to be particularly well resolved, and the use of different aperture radii in the data analysis provides details about the dust's structure. The data are most consistent with a composition of silicates and organics, and with a modeled grain-size distribution, they suggest the existence of a two-component outer dust disk. The material that makes up the inner component was found to be so dense that collisions between dust and rocks would have ground down most of the particulates over the age of the star. The dust must therefore have been replenished from an external source, such as disintegrating comets.[13]

The survey's criterion for success was to observe 35 stars, and by the spring of 2018, the LBTI had observed 38. Figure 4 shows the measured exozodi levels around the stars. Ten had detections of dust, ranging in thickness from 38 to 2600 zodis, with typical uncertainties of a few tens of zodis. Particularly dusty systems, such as η Corvi, were already known to be the exception and not the rule, but analysis of the HOSTS results mercifully revealed that even systems at tens of zodis are unusual.

Figure 5 shows the constraints that the LBTI and *WISE* placed on the exozodiacal luminosity function. To within 95% confidence, the analysis shows that the median surface brightness of exozodiacal disks around Sun-like stars is 27 zodis or less, with a best fit to the data of only 3 zodis, which suggests that our solar system may actually be typical.[14] Exozodiacal dust should therefore not pose a serious challenge for directly imaging Earth-like planets from space, as it would if typical dust levels were beyond 20 zodis or so.

Next steps for HOSTS

The LBTI currently has the best sensitivity to HZ exozodiacal dust of any instrument, and the HOSTS results will form the basis for planet-yield estimates when planning future space missions for imaging exo-Earths. The possible presence of HZ dust still poses a challenge for detecting faint planets, but the exozodiacal luminosity function traced out by HOSTS indicates that space-based designs will not need to have significantly expanded apertures, as would have been necessary for dustier scenarios. Such modifications could have easily incurred additional costs on the order of billions of dollars for a single mission.

Even after the success of HOSTS, a better understanding of exozodiacal dust would still be valuable. More data could help elucidate disk morphologies and constrain the dust's composition and dynamics—and, thus, its origin. Clumps or disk-shape asymmetries could suggest perturbations from planets. The shape of the dust's radial surface density could also indicate the clearing out of planetary material, a smooth funneling of dust from outer belts, or the fitful deposition of material from bodies such as comets and asteroids. Such processes will provide critical context once rocky HZ planets can be imaged and studied directly. In addition, if tighter constraints can be placed on dust levels around stars with HOSTS nondetections, smaller and cheaper space-based apertures could potentially be deployed and obtain a sufficient signal-to-noise ratio to image planets.

The LBTI group has plans in place to use the interferometer for a deeper study of detected disks, in which observations will be taken at multiple wavelengths and use more on-sky rotation as targets move overhead. The sensitivity of the LBTI is currently limited by detector

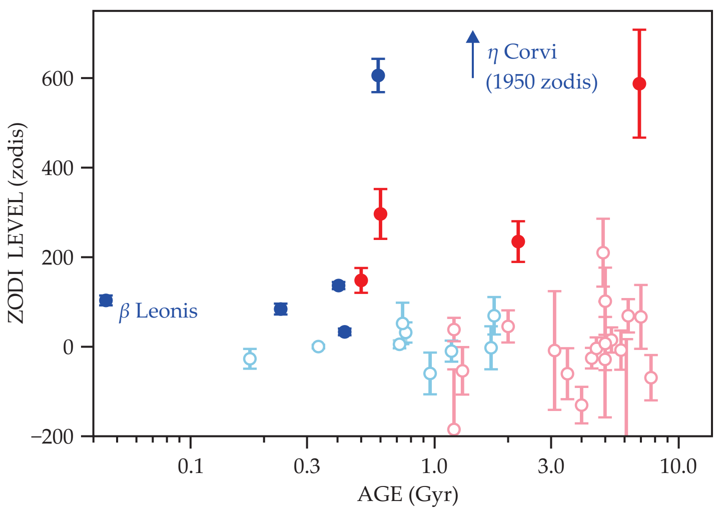


FIGURE 4. THE BRIGHTNESS of exozodiacal disks observed in the Hunt for Observable Signatures of Terrestrial Planetary Systems survey is measured in zodis, where one zodi is the brightness of the zodiacal disk. Filled data points indicate detections of exozodiacal light, and unfilled data points are nondetections. The majority of systems observed were nondetections, which suggests that the amount of light from exozodiacal dust in most stellar systems is similar to that in our solar system. (Adapted from ref. 14.)

noise; imperfect background subtraction; and instrumental vibrations, particularly from the telescope swing arms that hold the secondary mirrors. The group is currently working to improve all three factors. NASA has allocated funding for the LBTI's efforts over the next three years through an Exoplanets Research Program grant.

Nulling's future

Nulling will continue to be used in exoplanet science. It has the potential to find planets at smaller angles to their host stars than coronagraphy, the technique of successive optical masking to remove on-axis stellar light. In recent years, research groups have tested new nulling techniques with simulations, in labs, and on telescopes. Recently the Palomar Fiber Nuller was used to make the first detection of a faint companion star by rotating a null fringe around the primary star.[15]

An upcoming project, Hi-5, will turn the Very Large Telescope Interferometer into a giant four-aperture nuller.[16] It will benefit from the observatory's state-of-the-art infrastructure, the scientists' experience from past nulling projects, and improved software control and data processing. Hi-5 will also have integrated optics—miniaturized optical components in photonic analogues of integrated circuits—to provide increased optical stability and beam-combination flexibility. In addition to its main goal of detecting young giant planets in the HZs of nearby planetary systems, Hi-5 has an ambitious slate of science objectives that includes studying the multiplicities of stars in different evolutionary states to better constrain star formation models and detecting exozodiacal disks at wavelengths between the sweet spots of other nulling instruments.

Atmospherically induced noise remains prohibitive for observing habitable rocky planets from the ground for all but a handful of the closest stars, even with future 30-m-class telescopes. Nulling observations of such planets must wait for space-based missions in the decades ahead. Ambitious space nulling projects from years ago, such as the *Terrestrial Planet Finder*, never advanced beyond the development stage. But new designs have recently emerged with the benefit of an additional decade's worth of progress. A European team is proceeding with the study phase of the space-based *Large Interferometer For Exoplanets*, a proposed nuller with up to four mirrors on long baselines.[17] The researchers predict that it could detect hundreds of small rocky planets around nearby stars, including possibly dozens of Earth-sized planets in their stars' HZs.

Even if planets in HZs are common, the distances between the nearest Sun-like stars remain staggeringly vast on any human scale. Still, there is no physical reason why the distances to the nearest stars cannot be traversed in future centuries, with technological analogues of the outrigger sailing craft of Pacific Islanders who millennia ago set out on expeditions to the remotest of islands and navigated by the stars. Should that come to pass, nulling observations will help furnish the navigational charts to the most suitable destinations.

Or perhaps Earth-like worlds, glassy with ocean glint and speckled with clouds, are truly vanishingly rare. The evolution of life may also be so unusual that atmospheric studies of habitable exoplanets in our own century will provide no reason to suspect they are anything more than serene but deserted outposts, with nothing and no one to hear the wind and the waves. With time we will begin to know.

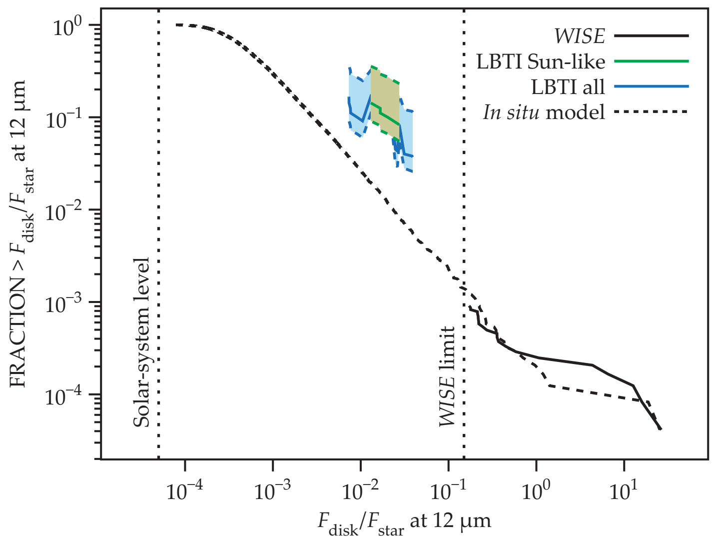


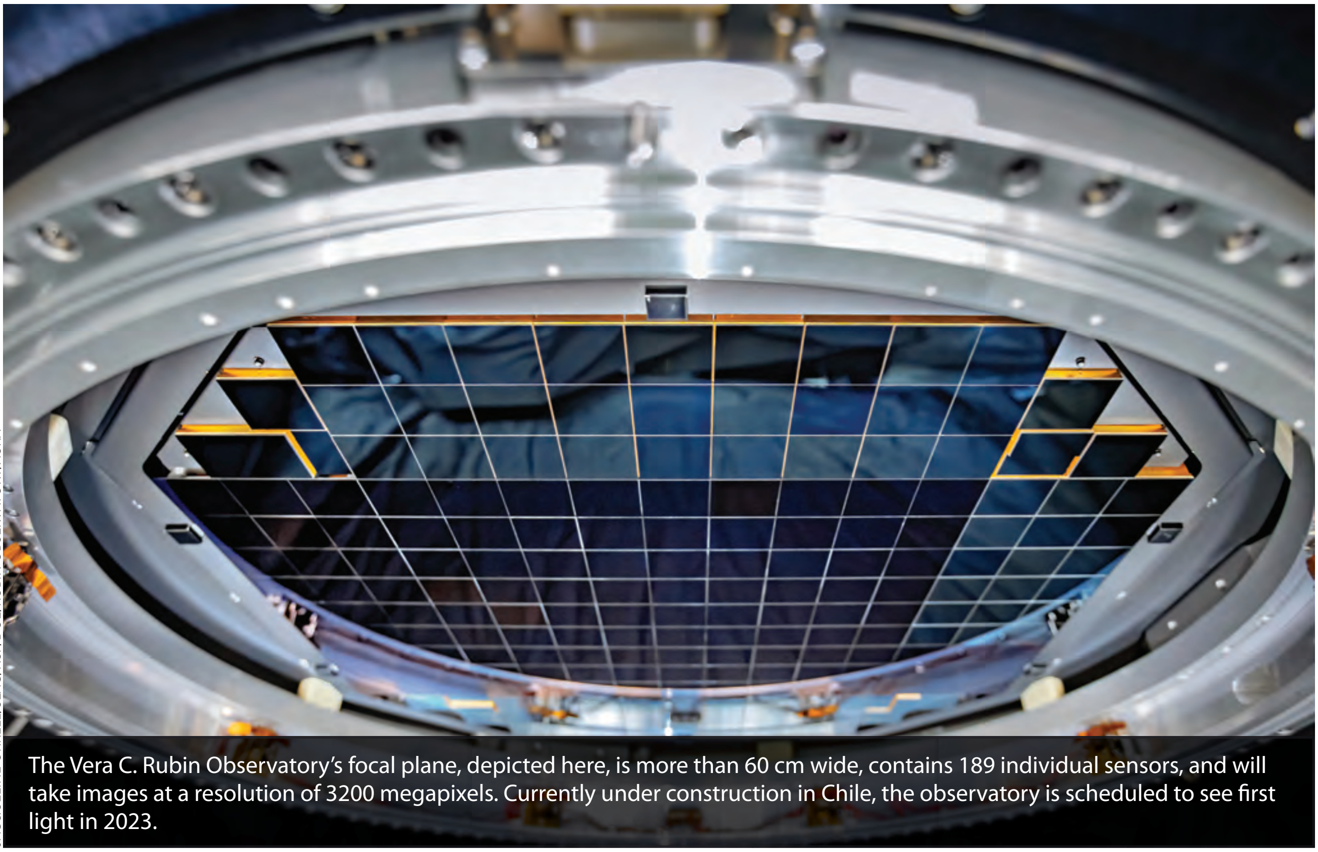
FIGURE 5. THE EXOZODIACAL LUMINOSITY function is constrained by data from the Large Binocular Telescope Interferometer (LBTI; blue and green lines) and the *Wide-Field Infrared Survey Explorer* (*WISE*; solid black line). The shaded regions around the LBTI data indicate uncertainties. The dashed line shows a model extrapolation of the *WISE* results. This plot is a reverse- cumulative distribution of the ratio of flux from the disk to flux from the star. Like luminosity, the flux ratio is independent of distance. It's measured at 12 μm, where habitable-zone dust is especially emissive. The green color shows the LBTI constraint from only Sun-like stars (F, G, and K types), and the blue is from all observed stars, including hot A types. The finding that the LBTI data have a shallower slope than the *WISE* extrapolation suggests the presence of additional physical processes that produce small levels of dust. (Adapted from S. Ertel et al., *Astron. J.* **155**, 194, 2018.)

We thank Philip Hinz for sharing his experiences and documentation. We also thank the current and past members of the LBTI, the support staff at the MMT and LBT, and funding sources, which have included the US Air Force and NASA. We also acknowledge the significant cultural role of the summit of Mount Graham—in Apache, Dził Nchaa Si'an, or "big seated mountain"—and reverence with which it is held in the Indigenous White Mountain and San Carlos Apache communities.

REFERENCES

1. V. S. Meadows et al., eds., *Planetary Astrobiology*, U. Arizona Press (2020).
2. A. Roberge et al., *Publ. Astron. Soc. Pac.* **124**, 799 (2012).
3. R. N. Bracewell, *Nature* **274**, 780 (1978).
4. J. R. P. Angel, N. J. Woolf, *Astrophys. J.* **475**, 373 (1997).
5. P. M. Hinz et al. *Nature* **395**, 251 (1998).
6. R. W. Duffner, *The Adaptive Optics Revolution: A History*, U. New Mexico Press (2009).
7. M. Lloyd-Hart, *Publ. Astron. Soc. Pac.* **112,** 264 (2000).
8. É. Sezestre, J.-C. Augereau, P. Thébault, *Astron. Astrophys.* **626**, A2 (2019).
9. E. Serabyn et al., *Astrophys. J.* **748**, 55 (2012).
10. B. Mennesson et al., *Astrophys. J.* **797**, 119 (2014).
11. A. J. Weinberger et al., *Astrophys. J., Suppl. Ser.* **216**, 24 (2015).
12. D. Defrère et al., *Astrophys. J.* **799**, 42 (2015).
13. D. Defrère et al., *Astron. J.* **161**, 186 (2021).
14. S. Ertel et al., *Astron. J.* **159**, 177 (2020).
15. E. Serabyn et al., *Mon. Not. R. Astron. Soc.* **489**, 1291 (2019).
16. D. Defrère et al., *Exp. Astron.* **46**, 475 (2018).
17. S. P. Quanz et al., https://arxiv.org/abs/2101.07500.

PT

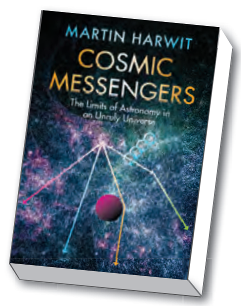


JACQUELINE ORRELL/SLAC/NSF/DOE/RUBIN OBSERVATORY/AURA

The Vera C. Rubin Observatory's focal plane, depicted here, is more than 60 cm wide, contains 189 individual sensors, and will take images at a resolution of 3200 megapixels. Currently under construction in Chile, the observatory is scheduled to see first light in 2023.

Astronomy's upper bounds

Cosmic Messengers
The Limits of Astronomy in an Unruly Universe
Martin Harwit
Cambridge U. Press, 2021. $39.99



Fueled by dramatic technological advances, astronomy has progressed rapidly in recent years. Over the past several decades, astronomers have detected the shadows of supermassive black holes, gravitational waves from merging black holes, fast radio bursts of uncertain extragalactic origin, dark energy, and exoplanets. The *James Webb Space Telescope*, with its 6.5-meter-diameter primary mirror, recently launched; the Vera C. Rubin Observatory in Chile will soon be imaging the entire available sky every few days; and 30-meter ground-based optical telescopes will be operational before 2030. The detection of signs of extraterrestrial life seems within reach. Multimessenger astronomy is thriving.

So it may come as a surprise that Martin Harwit's new book, *Cosmic Messengers: The Limits of Astronomy in an Unruly Universe*, provocatively argues that there may be ultimate bounds to astronomical knowledge. *Cosmic Messengers* is the final book in a trilogy. The first book, *Cosmic Discovery: The Search, Scope, and Heritage of Astronomy* (1981), enumerates the number of distinct breakthroughs in astronomical research and attempts to project the total number of such events (see the article by Harwit, Physics Today, November 1981, page 172). In the second volume, *In Search of the True Universe: The Tools, Shaping, and Cost of Cosmological Thought* (2013), Harwit discusses how new ideas in theoretical physics have led to the discovery of new phenomena in the universe (see Physics Today, October 2014, page 54).

In *Cosmic Messengers*, Harwit synthesizes his earlier efforts. He continues to argue that there may be a finite number of astronomical phenomena to be discovered and that we may someday approach the point of diminishing returns. But not to worry: He projects it will take several more centuries to get there.

Cosmic Messengers is important and thought-provoking because it revisits Harwit's basic idea almost exactly 40 years after the publication of *Cosmic Discovery*. He demonstrates how advances in the field since his first book appeared support his conclusion that there are a total of about 100 distinct phenomena to be discovered. Harwit's historical perspective is fascinating. In *Cosmic Discovery*, for example, he recommends looking into the feasibility and cost of large instruments that would directly detect gravitational waves or neutrinos from cosmic sources. Both types of instruments are now a reality and have proved spectacularly successful.

Harwit is well qualified to take us through such a broad-brush assessment of astronomy. He was a pioneer in the emerging field of IR astronomy in the 1960s and 1970s, a professor of astrophysics at Cornell University for many years, and the author of the 1973 textbook *Astrophysical Concepts*. He also led key review committees for NASA as it built its four highly successful space-based Great Observatories, which were launched in the 1990s and 2000s.

In his new book, Harwit makes a careful distinction between cosmic messengers, such as electromagnetic radiation, cosmic rays, gravitational radiation, and neutrinos, and the fundamentally distinct phenomena discovered by decoding their messages, such as exoplanets, fast radio bursts, and merging black holes. He makes a compelling case for why our knowledge may be ultimately limited by explaining how and why the universe blocks many of its messengers: The electromagnetic spectrum is bounded at longer wavelengths at about 10^6 cm because of the ionized interstellar medium; at shorter wavelengths, the bound is about 10^{-19} cm, below which photons are destroyed by their interaction with microwave photons from the cosmic microwave background.

A more basic problem, however, is that as the universe is examined with ever-finer temporal, spatial, and energy resolutions, the messages become unreliable because of propagation effects. Perhaps the most fundamental example is the way gravitational lensing obscures the arrival time and direction of all conceivable messengers.

The first half of the book is devoted to describing the known cosmic messengers, the instruments used to detect them, and how they have enabled the discovery of fundamental phenomena. It is an impressive tour of the most important advances in astrophysics, one that is enlivened by anecdotes and background material on the discoveries and discoverers. Harwit's coverage is largely fair, although it is possible to quibble with details.

In the second half of the book, Harwit estimates anew the number of astronomical phenomena likely to still be discovered. He argues that astronomers have identified 60 distinct phenomena up to now and, using statistical inference, calculates that his previous estimate of roughly 100 total phenomena was accurate. Harwit suggests that extrasolar asteroids traversing the solar system and as-yet-unknown properties of dark matter and dark energy are possible new messengers that could enable the discovery of some of the 40 phenomena still out there to be uncovered.

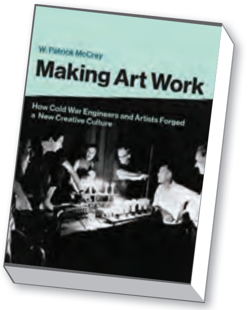
In the final chapter, Harwit reflects on how we should move forward in astronomical research given those limitations. Although he has few concrete recommendations, he nevertheless expounds on a wide range of issues: whether large research consortia should be favored over smaller efforts, whether expensive new instruments require international collaboration, how to keep young scientists in the field when they can use their analytic skills to earn far higher salaries elsewhere, and how to move humankind to barren planets.

Cosmic Messengers should be of interest to a wide audience of astronomers, other scientists, historians of science, government agency planners, and anyone who wants to see the fruits of curiosity-driven research. It will also be a valuable resource to students and others aiming to place their research into a much larger context.

James Moran
Harvard University
Cambridge, Massachusetts

The two cultures, revisited

Making Art Work
How Cold War Engineers and Artists Forged a New Creative Culture
W. Patrick McCray
MIT Press, 2020. $45.00



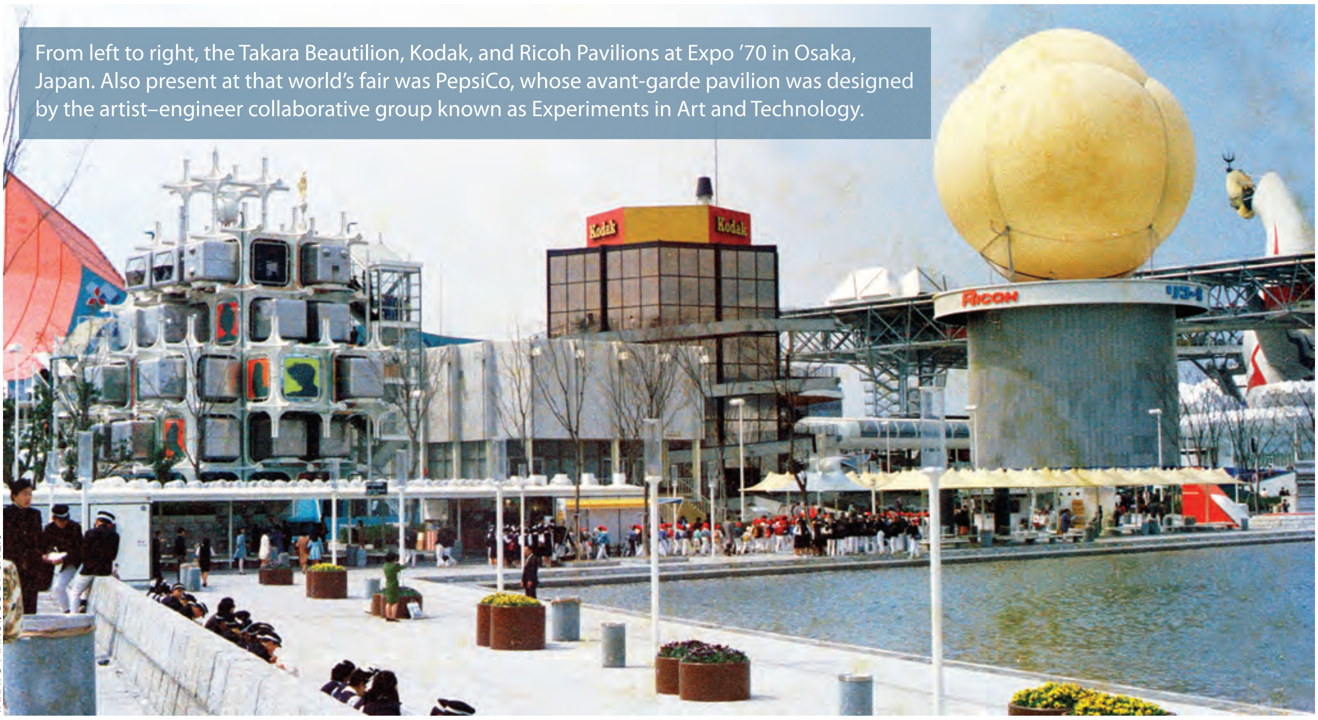
The 1959 pronouncement of C. P. Snow that the humanities and sciences make up two separate cultures has been a lightning rod for discussion ever since the novelist and physical chemist put forward his famous thesis. In the 1960s, for example, the pages of Physics Today were rife with responses to Snow (see Physics Today, September 1961, page 62; July 1966, page 160; and the article by Jerome Ashmore, November 1963, page 46). But was the divide as extreme as Snow believed it to be?

According to the historian of science W. Patrick McCray, it was not. In his new book, *Making Art Work: How Cold War Engineers and Artists Forged a New Creative Culture*, McCray delves into collaborations in the 1960s between engineers at companies like IBM and Bell Labs and modern artists in the postwar US such as Robert Rauschenberg, Claes Oldenburg, and Deborah Hay. He illustrates how artistic and scientific cultures were not irreconcilable but complementary: During that decade, some of the biggest names in the arts world relied on the technical skills of engineers to bring their artistic ideas to life.

Making Art Work centers on three figures and examines the communities they worked in as both managers and makers. The first is Frank Malina, an aeronautical engineer-cum-artist who founded the arts and technology journal *Leonardo*. The second figure is the artist Gyorgy Kepes, who developed a visual-design program at MIT and founded the university's Center for Advanced Visual Studies. The bulk of *Making Art Work* focuses on the final individual, Billy Klüver, a Bell Labs electrical engineer with close ties to the New York avant-garde art scene who founded Experiments in Art and Technology (E.A.T.), an organization that fostered connections and collaborations between artists and engineers.

McCray adroitly moves between intellectual concepts from a range of disciplines. Historians of science and technology will be able to comfortably orient themselves in McCray's analyses of big science, paradigm shifts, and tacit knowledge. Likewise, historians of modern art will appreciate his careful description of artist–engineer collaborations like the performance series *9 Evenings: Theatre and Engineering* and the E.A.T.-designed Pepsi Pavilion at Expo '70 in Osaka, Japan.

Importantly, McCray observes that engineers often served as "invisible technicians," a category defined by the historian of science Steven Shapin in a 1989 article about skilled craftsmen whose contributions to 17th-century English laboratory experiments were omitted from later histories of science. Like those early-modern craftsmen, the engineers who



From left to right, the Takara Beautilion, Kodak, and Ricoh Pavilions at Expo '70 in Osaka, Japan. Also present at that world's fair was PepsiCo, whose avant-garde pavilion was designed by the artist–engineer collaborative group known as Experiments in Art and Technology.

collaborated with artists often failed to gain recognition: Although many artist–engineer collaborative works appeared in museums and art galleries, individual engineers regularly went unnamed in exhibition materials and press coverage. Along similar lines, McCray points out that the art-and-technology movement was largely dominated by white men. In the Los Angeles County Museum of Art's 1971 *Art and Technology* exhibition, for example, all 16 collaborative works featured were by white men.

Making Art Work differs from other studies of the art-and-technology movement in its focus on the history of engineering. Previous works have argued that the movement waned in the late 1960s because of increasing criticism of the artists for accepting money from and collaborating with corporations, which were often profiting from the US war effort in Vietnam. But McCray asserts that broader economic trends in the field of engineering were also a reason for the movement's decline: They disrupted a generation-long period of job stability for US engineers and made art-and-technology collaborations less feasible for them. That argument suggests that scholars should take a closer look at the history of engineering in the 1970s.

Making Art Work concludes by outlining what McCray terms a second wave of art-and-technology collaborations that began in the 1990s. The new wave was presaged by the development of early consumer technology like the Sony Portapak, a portable video camera that came to market in 1965, but the widespread adoption of personal computers in the 1980s and the rise of the internet in the 1990s truly heralded a new age.

The second wave has already weathered periods of pessimism—after both the dot-com bubble and the Great Recession. It has also seen periods of optimism that have manifested in initiatives like STEAM, which attempts to integrate the arts into traditional STEM (science, technology, engineering, and mathematics) curricula. As McCray shows, however, attempts to blend the arts and sciences have foundered in the past because of different—although not incompatible—ways of thinking, communication styles, and common knowledge, which indicates that the two cultures have not yet found a smooth connection point. I have argued elsewhere that there are similar difficulties in bridging the histories of science and art.

Although *Making Art Work* is primarily intended for historians, it should also appeal to a wider audience because it addresses core questions about how humans worked with and reacted to technology in the mid 20th century. As McCray shows, the seemingly impermeable barriers between science and art have in fact been highly porous. The two cultures do, it seems, have common ground.

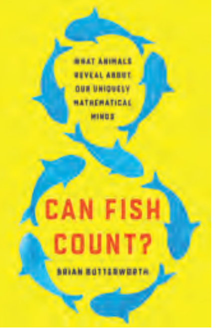
Leib Celnik
Washington, DC

NEW BOOKS & MEDIA

Can Fish Count?

What Animals Reveal About Our Uniquely Mathematical Minds

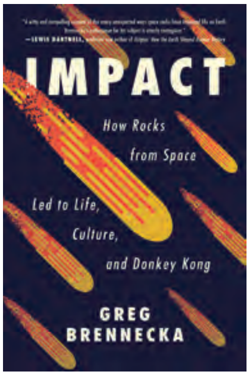
Brian Butterworth
Basic Books, 2022. $30.00



What are numbers? What does it mean to count? In *Can Fish Count?*, Brian Butterworth, a cognitive psychologist who specializes in the psychology of mathematical concepts, investigates how human—and animal—minds became mathematical. Remarkably, humans have had concepts of numbers for over 10 000 years: Butterworth cites a study demonstrating that counting words like "one," "two," and "three" are some of the slowest changing words in most languages across the world, which means that they likely were used to express concepts that human ancestors may have possessed 100 000 years ago. Counting seems to be a fundamental skill possessed by most animals: Although they don't have counting words, even tiny minnows have sufficient numeric ability to distinguish larger schools of fish from smaller ones. —RD

Impact

How Rocks from Space Led to Life, Culture, and Donkey Kong



Greg Brennecka
William Morrow, 2022. $28.99

Meteorites may seem mundane, but they might just be some of the most scientifically valuable objects out there, writes Greg Brennecka, a cosmochemist, in his new book. Using pop-culture references and a wry tone, *Impact* presents readers with a comprehensive overview of meteorite studies from prehistoric times to the present day. As Brennecka notes, the study of space rocks helps us understand how the solar system formed, whether other planetary systems are out there, why life developed, and how unique the existence of life is. Meteorites also influenced human society: Before the Iron Age, meteorites were the only known source of the highly prized metal. As a result, meteoritic iron is typically present in artifacts that denoted significant social status, like a regal ceremonial dagger found in King Tutankhamen's tomb. Although the book's title veers into hyperbole, Brennecka convincingly argues that meteorites comprise the universe's historical record. —RD

Lerner Maker Lab

Lerner, 2022. $266.65

Aimed at elementary school students, particularly those in second through fifth grades, this database features more than 500 art, cooking, and science projects designed to inspire creativity and learning. Monster magnets, miniature windmills, balloon rockets, black-bean veggie burgers, and holiday decorations are among the available projects. Each one includes a basic description, a list of materials needed, and detailed instructions with illustrations. Although the materials generally consist of common household items, the directions occasionally call for a few specialty components, such as a luxmeter or pager motor. The website strives to be accessible, interactive, and kid friendly, but some projects do recommend adult supervision. —CC

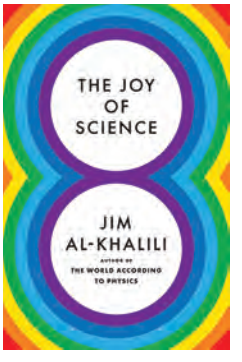
How to Bake a Universe

Alec Carvlin; ill. Brian Biggs
Norton Young Readers, 2022. $18.95

Aimed at readers ages 6–8, *How to Bake a Universe* uses the analogy of making a cake to explain the origin of the cosmos. The author, Alec Carvlin, starts his imaginative storytelling with a young chef protagonist putting an empty baking pan into an oven set to Absolute Hot. The chef then monitors the baking process as matter is created and eventually forms stars. The colorful drawings by Brian Biggs help bring the story to life. To give the book a bit more heft in the science department, a notes section, glossary, and timeline are provided, which will aid parents and other adult helpers in explaining some of the many difficult concepts that are presented. —CC

The Joy of Science

Jim Al-Khalili
Princeton U. Press, 2022. $16.95



Teaching nonscientists how to live more scientifically is the aim of Jim Al-Khalili in his new book, *The Joy of Science*. A theoretical physicist and host of the BBC radio program *The Life Scientific*, Al-Khalili focuses on science as a way of thinking and extols the wonders it can reveal. He bases his approach on the scientific method and, in eight short lessons, offers advice on such topics as recognizing objective truth, comprehending complex ideas, avoiding confirmation bias, and combatting disinformation. Using a light conversational tone and nontechnical language, Al-Khalili explains how science provides a reliable way of learning about the world, one that allows us to see beyond our limited senses and develop a more profound and personal appreciation of our surroundings. —CC PT

Focus on software, data acquisition, and instrumentation

The descriptions of the new products listed in this section are based on information supplied to us by the manufacturers. PHYSICS TODAY can assume no responsibility for their accuracy. For more information about a particular product, visit the website at the end of its description. Please send all new product submissions to ptpub@aip.org.

Andreas Mandelis

Triode high-voltage power supply

Spellman High Voltage Electronics has upgraded its triode high-voltage power supply, which is designed to drive thermionic scanning electron microscopes. The EBM-TEG series shares the same triode supply as the previous EBM but adds a detector drive with a photomultiplier tube, scintillator, and collector outputs. The integrated unit provides a highly regulated, low-noise, ultrastable accelerator supply programmable from 0 to −30 kV at 170 μA, and floating filament and bias supplies referenced to the accelerator to control the beam. High-voltage outputs drive the detector. Programming signals use differential analog inputs to minimize external noise and to offset voltages effects. A ground-referenced emission current monitor and a filament failure signal are provided, and the units are protected against overcurrent and overvoltage, arcs, and short circuits. ***Spellman High Voltage Electronics Corp****, 475 Wireless Blvd, Hauppauge, NY 11788, www.spellmanhv.com*

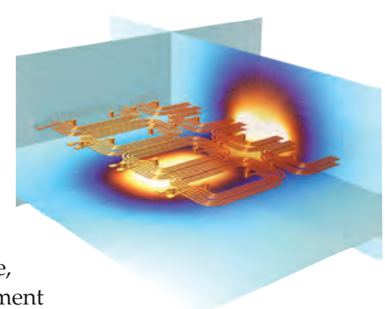
Downloadable calculator app

The latest version of the APE Optics Calculator app is available on Google Play. The app offers fast, easy access to commonly needed equations in ultrafast and nonlinear laser optics. Besides basic functions, the company has developed and integrated new features such as the ability for users to share calculation results with team members. Photon energy (measured in electron volts) has been added to most calculations, and explanatory text is offered for all calculations and more. ***APE GmbH****, Plauener Strasse 163–165, Haus N, 13053 Berlin, Germany, www.ape-berlin.de*

Compact hexapod for six-DOF motion

Aerotech's HEX150-140HL miniature hexapod positioning system can move with six degrees of freedom (DOF): precise translation in the x, y, and z directions and rotation around each of those axes (θ_x, θ_y, and θ_z). According to the company, the HEX150-140HL's minimum incremental motion and linear travel are better than those of other available products. The flexible system combines high precision and high load capacity, and its peak-to-peak specifications yield consistent, reliable six-DOF motion. It features minimum incremental motions of 20 nm in the x, y, and z directions and 0.04 arcsec in θ_x, θ_y, and θ_z; high payload-lift capabilities up to 7.5 kg; backdriving resistance to 100 N; and multiaxis-motion coordination, including servo, stepper, and piezo. Easy to control and program with virtual pivot-point adjustment, the HEX150-140HL can address space-constrained, multiple-DOF applications that require a fine positioning resolution. Those include photonic device manipulation, optics inspection and alignment, optical wafer probing, and synchrotron and beamline sample manipulation. ***Aerotech Inc****, 101 Zeta Dr, Pittsburgh, PA 15238-2811, www.aerotech.com*

Modeling and simulation software



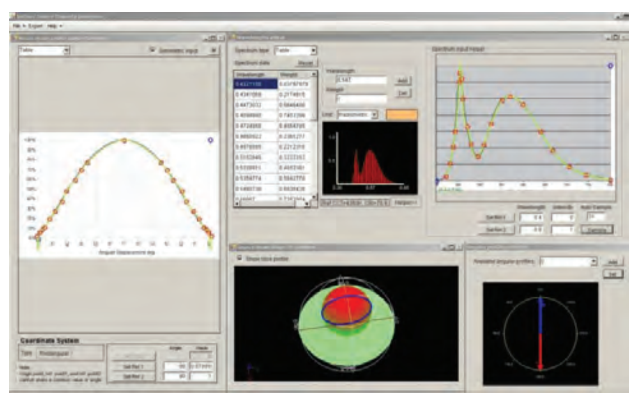
Comsol has updated its Multiphysics software for creating physics-based models and simulation applications. Version 6.0 introduces a feature for model management, a module for uncertainty quantification analysis, and updates and performance enhancements. The Model Manager enables efficient simulation-data management through version control; tracking changes; and advanced search functionality within models, computer-aided-design data, and other related external files. It provides a structured workspace where colleagues can collaborate within organizations and with external parties. Based on probabilistic design methods, the Uncertainty Quantification Module facilitates the production of more complete, accurate, and useful multiphysics models. Other updates in Version 6.0 include the improvement by a factor of 10 in speed and memory consumption performance for certain engineering applications, more efficient electromagnetic simulation of printed-circuit-board designs, and a new flow-induced noise realm for acoustics modeling. ***Comsol Inc****, 100 District Ave, Burlington, MA 01803, www.comsol.com*

App for temperature-data logging

CAS DataLoggers reports that TandD has enhanced its T&D Thermo app for both iOS and Android users. With the mobile application, users of some TandD data loggers can carry out operations such as making settings, downloading recorded data, and viewing graphs via smartphones and tablets. Version 2.2 adds support for TandD's recently released TR-7A series of loggers. Background push notifications have been enabled to send alarm warnings from the WebStorage Service to mobile devices, provided users have confirmed that both their phones and the T&D Thermo App have "Background Refresh" enabled. (That separate utility does not replace the emailed warning notifications from the WebStorage Service.) The update also adds a function to output log and settings from the Settings Table Screen and a function to create PDF reports of data downloaded using Bluetooth or direct communications over a wireless local area network. ***CAS DataLoggers****, 8437 Mayfield Rd, Unit 104, Chesterland, OH 44026, www.dataloggerinc.com*

Antenna system for direction finding and monitoring

The R&S ADD557SR antenna system from Rohde & Schwarz features both vertical and horizontal polarization and covers the frequency range from 20 MHz to 6 GHz for direction finding, and from 20 MHz to 8.5 GHz for monitoring. It is suitable for use with R&S DDF550 wideband direction finders and DDF5GTS high-speed scanning direction finders in fixed, mobile, and transportable applications. The versatile system contains a separate omnidirectional-monitoring antenna output and can be switched between the two polarizations. To avoid saturating the active antenna elements and enable direction finding even in noisy situations, the system can be switched into passive mode in the very high and ultrahigh frequency ranges. The R&S ADD557SR is capable of superresolution: The advanced direction-finding technology can simultaneously determine the bearings of multiple emissions on the same frequency and spot signals in the spectrum that other emissions have concealed and that are undetectable to other direction finders, according to the company. ***Rohde & Schwarz GmbH & Co KG****, Muehldorfstrasse 15, 81671 Munich, Germany, www.rohde-schwarz.com*



Optical-design software

Lambda Research has unveiled TracePro 2022, the latest version of its software for the design and analysis of illumination and optical systems. The main improvements in the new release are the incorporation of a geometric modeler, KOSMOS/KCM, and an updated graphical user interface. The new modeler gives TracePro the capability to model asymmetric and free-form optical surfaces with the accuracy required for optical ray tracing. As a result, the company was able to expand the lens element capability to include a total of 20 different surface shapes, and it anticipates adding more shapes. Computer-aided-design translators such as STEP, IGES, and SAT are now bundled with TracePro at no additional charge. ***Lambda Research Corporation****, 25 Porter Rd, Littleton, MA 01460, www.lambdares.com*

Software-defined test instrumentation

Liquid Instruments has released two updates to its new hardware platform, Moku:Pro, that increase the system's versatility and further enable it to replace a suite of common test instruments with one streamlined rack-mounted device. Multi-instrument Mode allows multiple instruments to be dynamically combined on a single field-programmable gate array (FPGA) chip. For the first time, according to the company, users can configure multiple test instruments to run simultaneously on a standalone piece of hardware and implement custom code directly onto the FPGA. With the Moku Cloud Compile subscription service, users can design and deploy custom signal-processing algorithms to Moku:Pro's FPGA. The cloud-based tool is fully compatible with the Multi-instrument Mode and delivers a complete solution to optimize custom-built test suites for specific applications and requirements. The new modes are designed to serve the quantum computing, aerospace, semiconductor, lidar, and automotive industries. ***Liquid Instruments***, *740 Lomas Santa Fe Dr, Ste 102, Solana Beach, CA 92075, www.liquidinstruments.com*

High-frequency mixed-signal oscilloscope

Tektronix has launched the newest version of its 5 Series mixed-signal oscilloscope (MSO). The 5 Series B MSO is more versatile and provides added support for power-integrity testing and enhancements for debugging and validation. Updates include a new auxiliary trigger input that allows the oscilloscope to be synchronized to an external signal without using any of the four, six, or eight input channels. The maximum frequency output for the optional built-in arbitrary function generator has increased from 50 MHz to 100 MHz, which enables higher-frequency stimulus for Bode plots and impedance measurements. A faster processor makes controls more responsive. For users working and collaborating outside the laboratory, the 5 Series B MSO works with new tools for off-line analysis and cloud data storage. ***Tektronix Inc***, *14150 SW Karl Braun Dr, PO Box 500, Beaverton, OR 97077, www.tek.com* PT

OBITUARIES

Kyozi Kawasaki

Kyozi Kawasaki, one of the major contributors to statistical physics over the past half century, passed away on 12 November 2021 in Fukuoka, Japan.

Born on 4 August 1930 in Shiga prefecture, Japan, Kyozi was a teenager during World War II. The hardships he incurred during and after the war greatly influenced his choice of career as an adult. Kyozi received both his BS and MS in physics from Kyushu University, followed by a PhD in physics in 1959 from Duke University, where he was greatly inspired by Michael Buckingham and William Fairbank.

After graduation, Kyozi held several academic positions in the US (at MIT and Temple University) and in Japan (at Nagoya, Kyushu, and Kyoto Universities). As a young researcher, he worked at MIT with Irwin Oppenheim, and at Kyoto University he collaborated with Hazime Mori. In 1976 Kyozi accepted a professorship at Kyushu University, and he remained there until his retirement in 1994. But he continued to be active in science and taught physics at Chubu University and the Fukuoka Institute of Technology.

Kyozi worked on a broad range of statistical-physics systems, primarily exploring their dynamics. His papers contain seminal results that led to new predictions and influenced a great number of theoreticians and experimentalists alike. In 1966 Kyozi presented a key model of spin-exchange dynamics in which the order parameter is conserved. Now known as the Kawasaki spin dynamics, it was used later in a vast number of simulations and theoretical studies of ordering dynamics.

In the late 1960s, Kyozi constructed a general mode-coupling theory (MCT) of systems undergoing a phase transition close to a critical point. More specifically, he showed that critical dynamics in near-critical fluids is governed by hydrodynamic interactions. His prediction of the decay rate of critical fluctuations agreed remarkably well with experiments on light scattering and sound attenuation in fluids. His MCT was later applied to magnetic systems and superfluids. In the mid 1970s, Kyozi expanded his tour-de-force MCT to general spatial dimensionalities and clarified the relationship between his theory and the dynamic renormalization group theory. In the late 1970s, Kyozi started to study phase ordering associated with phase transitions. In particular, Kyozi laid the foundation for future work through his predictions of how hydrodynamic interactions modify spinodal decomposition and fluids under shear flow.

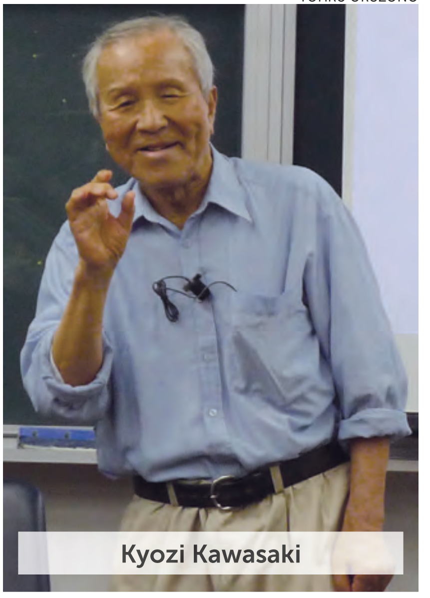
It is worth emphasizing that Kyozi was interested in transport properties as a fundamental problem in statistical physics. Together with Oppenheim, he carried out in 1965 a density expansion of transport coefficients in a dilute gas that resulted in logarithmic correction terms. His MCT made it possible to evaluate the anomalous part of the transport coefficients. In 1973 Kyozi and James Gunton proposed a theory for nonlinear transport coefficients.

Throughout his career, Kyozi emphasized the pathway from equilibrium to nonequilibrium statistical physics, and he stressed the importance of nonlinear, nonequilibrium statistical physics. His other contributions include theories for random-interface growth, defect and vortex dynamics, multicomponent membranes and vesicles, liquid crystalline elastomers, and mesoscopic phases of polymeric systems. His more recent work dealt with a mode-coupling approach to the glass transition and its precursors.

Kyozi always paid special attention to educating and stimulating his younger peers. As a consequence, he was successful in choosing excellent research associates and graduate students, many of whom hold academic positions throughout Japan.

Kyozi was a corecipient of the 2001 Boltzmann Medal, the highest distinction in statistical physics, which is given once every three years by the International Union of Pure and Applied Physics. His contributions to science were also recognized by the 1972 Nishina Memorial Prize, a 1992 Humboldt Research Award, and a 2000 Toray Science and Technology Prize, among other distinctions.

TOHRU OKUZONO



Kyozi Kawasaki

Kyozi Kawasaki will be remembered not only for his great scientific achievements but also for his support, gentle demeanor, and love of science, which made him an exemplary role model for generations of younger scientists, in Japan and worldwide. The physics community has lost one of its greatest and most creative members. His legacy, though, will continue to influence statistical physics for many years to come.

David Andelman
Tel Aviv University
Tel Aviv, Israel
Helmut Brand
University of Bayreuth
Bayreuth, Germany
Takao Ohta
Akira Onuki
Kyoto University
Kyoto, Japan PT

RECENTLY POSTED NOTICES AT
www.physicstoday.org/obituaries

Dennis Holliday
22 November 1935 – 18 March 2021

William (Bill) Carl Schieve
28 April 1929 – 19 September 2020

David G. Seely
14 September 1958 – 3 September 2020

Herschel S. Pilloff
15 August 1940 – 27 February 2020

Bruce Jakosky (bruce.jakosky@lasp.colorado.edu) is a professor of geological sciences and an associate director of the Laboratory for Atmospheric and Space Physics at the University of Colorado Boulder. For 18 years he was the principal investigator of NASA's *Mars Atmosphere and Volatile Evolution* mission.

How did Mars lose its atmosphere and water?

Bruce M. Jakosky

They were mostly lost to space early in Mars's history, in processes driven by the Sun's UV photons and solar wind after Mars lost its magnetic field.

Mars today is a cold, dry planet. Its temperature averages 50 K below the freezing point of water. And its atmosphere is too thin for water to persist as a liquid. Geological evidence, however, shows that liquid water was abundant on the surface of ancient Mars (see my article with Michael Mellon, Physics Today, April 2004, page 71). The planet has features that imply the existence of rivers, streams, and shorelines early in Mars's history (see figure 1).

The Sun was 30% dimmer then, so a thick greenhouse atmosphere must have been warming the planet. Carbon dioxide makes up 96% of today's atmosphere and was likely the largest contributor to that greenhouse effect, though it may not have been the only greenhouse gas on Mars. Where did the CO_2 from that earlier, thicker atmosphere go? Where did the water that carved the channels and eroded the surface go? Can water and CO_2 be put back into the atmosphere and make Mars warm again?

It takes no more than a few bars of pressure from atmospheric CO_2 to raise the temperature to the melting point of ice, and Mars initially may have had as much as 20 bars. By comparison, the total atmospheric pressure at Earth's surface is about 1 bar. The abundance of water on Mars can be expressed as a global equivalent layer (GEL), which is the depth of the water if all of it existed at the surface as a liquid and was spread uniformly over the planet. Mars's observed morphological features would have required a GEL of at least 50 m. Water may also reside as groundwater or as ice in the crust, in an amount that could raise the GEL by nearly an order of magnitude. Altogether, those reservoirs yield about 500 m GEL of water on the planet. By comparison, Earth's oceans, if spread over the entire planet, would form a layer about 2 km thick.

MAVEN

Planets can lose water and CO_2 above their surfaces in two ways. The Sun and its solar wind can strip water vapor and gaseous CO_2 from the top of the atmosphere into space. The two compounds can also diffuse into the subsurface. There, CO_2 and water can react with crustal materials to form CO_2- and H_2O-rich minerals.

The *Mars Atmosphere and Volatile Evolution*, or *MAVEN*, spacecraft has been tracking the stripping of the Martian atmosphere since 2014. Although the atmosphere is losing gas today at a rate of only about 2–3 kg/s, rates would have been much higher early in Mars's history, when the Sun's extreme UV rays and the solar wind were more intense. But by observ-

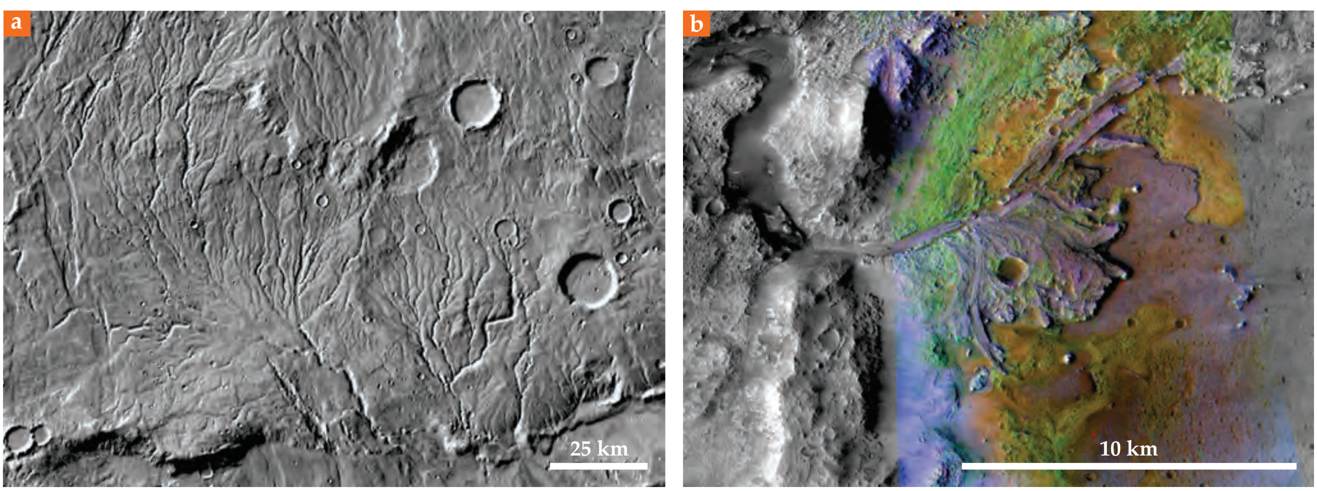


FIGURE 1. GEOLOGICAL EVIDENCE for the presence of liquid water on ancient Mars. **(a)** Valley networks on the oldest surfaces are thought to have been carved gradually by stable liquid water. **(b)** A river delta formed from debris deposited where water flowed into Jezero Crater. Colors represent compositional information derived from orbital spectroscopy that relates to water-bearing minerals. (Courtesy of NASA.)

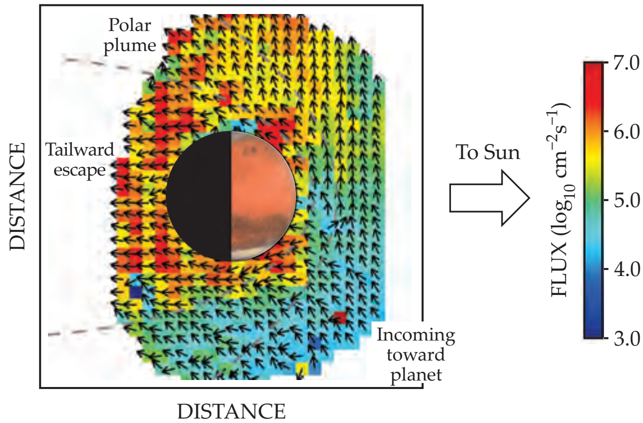


FIGURE 2. A YEAR'S WORTH of data on oxygen escaping from Mars. Colors represent the flux of oxygen atoms, and arrows show the average direction of their movement. (Adapted from an image courtesy of the American Geophysical Union.)

ing the processes today and knowing some history of the Sun, planetary scientists can extrapolate the loss rate into the past and estimate the total loss through time.

They use the observed enrichment of the heavier of stable isotope pairs, such as D/H, $^{13}C/^{12}C$, $^{15}N/^{14}N$, and $^{38}Ar/^{36}Ar$, to infer the fraction of each gas that has been lost (see Physics Today, May 2015, page 12). In each case, the lighter isotope is removed to space more easily and leaves behind an increased concentration of the heavier isotope.

Loss of hydrogen, oxygen, and carbon are of interest to scientists because of their connections to the climate-related gases H_2O and CO_2. Using the measurements of *MAVEN* and other spacecraft, for example, researchers can track H atoms from their occurrence in atmospheric water vapor in the lower atmosphere to an extended corona that reaches from the upper atmosphere to at least as far as 10 Mars radii.

The inferred H loss rate varies by as much as an order of magnitude through the Martian seasons. That variation results from increased atmospheric dust near perihelion that raises temperatures and allows water to be carried to higher altitudes where it can escape more readily. Extrapolating the loss through time, scientists have found that more than 25 m GEL of H_2O may have been lost to space.

Scientists can also track the historical loss of O to space by recognizing that it can come from either CO_2 or H_2O. Oxygen is removed either as ions from the upper atmosphere or by photochemical processes that break up CO_2 or H_2O and eject the O atom (see figure 2). Carbon is harder to track, but researchers can estimate the total amount lost to space from the $^{13}C/^{12}C$ isotope ratio, from the current estimated loss rate for O, and from the rate of photodissociation and loss of atmospheric carbon monoxide. Those amounts yield an integrated loss through time of at least 1–2 bars of CO_2.

The dynamo, reactions, and the poles

Planetary scientists think that the loss of H was much larger early in history when the Sun's UV radiation intensity was much higher. The stripping of C and O by the solar wind began about 4.1 billion years ago, when the Martian magnetic field shut off with the death of the planet's dynamo (see Physics Today, October 2021, page 17). At that point in time, no global magnetic field existed to protect the atmosphere from the onslaught of the solar wind.

Carbon dioxide can react with minerals to form carbon-bearing minerals in the crust. Orbital surveys have identified it in surface materials, in buried materials that have been exposed at the surface, and in meteorites found on Earth that have come from Mars. Carbonates at or near the surface hold the equivalent of no more than a couple of tens of millibars of CO_2—counted as if it all were released into the atmosphere. Deeply buried carbonates could hold up to a bar of CO_2, but their abundance is difficult to quantify, as they are exposed and visible in only a few locations.

Today, water is locked up in the polar caps and in mid- and high-latitude ground ice. Those sinks hold about 20–30 m GEL of H_2O. A much bigger sink may be water bound in hydrated minerals at the surface that spacecraft have identified spectroscopically from orbit.

Scientists can extrapolate that data to estimate the water locked up in the minerals hidden beneath the surface, but with large uncertainties. The most likely amount of stored water is between 130–260 m GEL, although a reservoir more than 500 m deep is possible. The large uncertainty arises from the difficulty of predicting a global abundance from a small number of exposures of minerals.

Clearly, the removal of CO_2 from the atmosphere to space and the crust can account for the loss of a thick early atmosphere. But the water history is more complicated. In addition to being lost to space and the crust, water has been released to the surface in catastrophic floods. That water would have eventually evaporated over time or percolated into the crust, and volcanic materials would have outgassed. Today H_2O from the polar ice caps moves through the atmosphere and exchanges with midlatitude ground ice or condenses at the opposite pole. The amounts of water in each case are difficult to quantify. Despite the uncertainties, however, scientists are getting a consistent story of what drove the evolving climate and water on Mars.

Can the CO_2 that is locked in the crust be mobilized and restored to Mars's atmosphere today? If that were possible, it could support greenhouse warming that would raise the planet's temperature and allow liquid water to become widespread again. Unfortunately, the bulk of the CO_2 that remains on Mars is distributed globally and would require strip-mining and processing of the entire planet to release.

Don't give up hope, though. One can imagine manufacturing boutique, high-efficiency greenhouse gases on Mars that could—someday, at least—provide a substantial greenhouse effect.

Additional resources

▶ M. S. Chaffin et al., "Martian water loss to space enhanced by regional dust storms," *Nat. Astron.* **5**, 1036 (2021).

▶ J. P. Grotzinger et al., "A habitable fluvio-lacustrine environment at Yellowknife Bay, Gale Crater, Mars," *Science* **343**, 1242777 (2014).

▶ B. M. Jakosky, "The CO_2 inventory on Mars," *Planet. Space Sci.* **175**, 52 (2019).

▶ B. M. Jakosky, "Atmospheric loss to space and the history of water on Mars," *Annu. Rev. Earth Planet. Sci.* **49**, 71 (2021).

▶ R. D. Wordsworth, "The climate of early Mars," *Annu. Rev. Earth Planet. Sci.* **44**, 381 (2016). PT

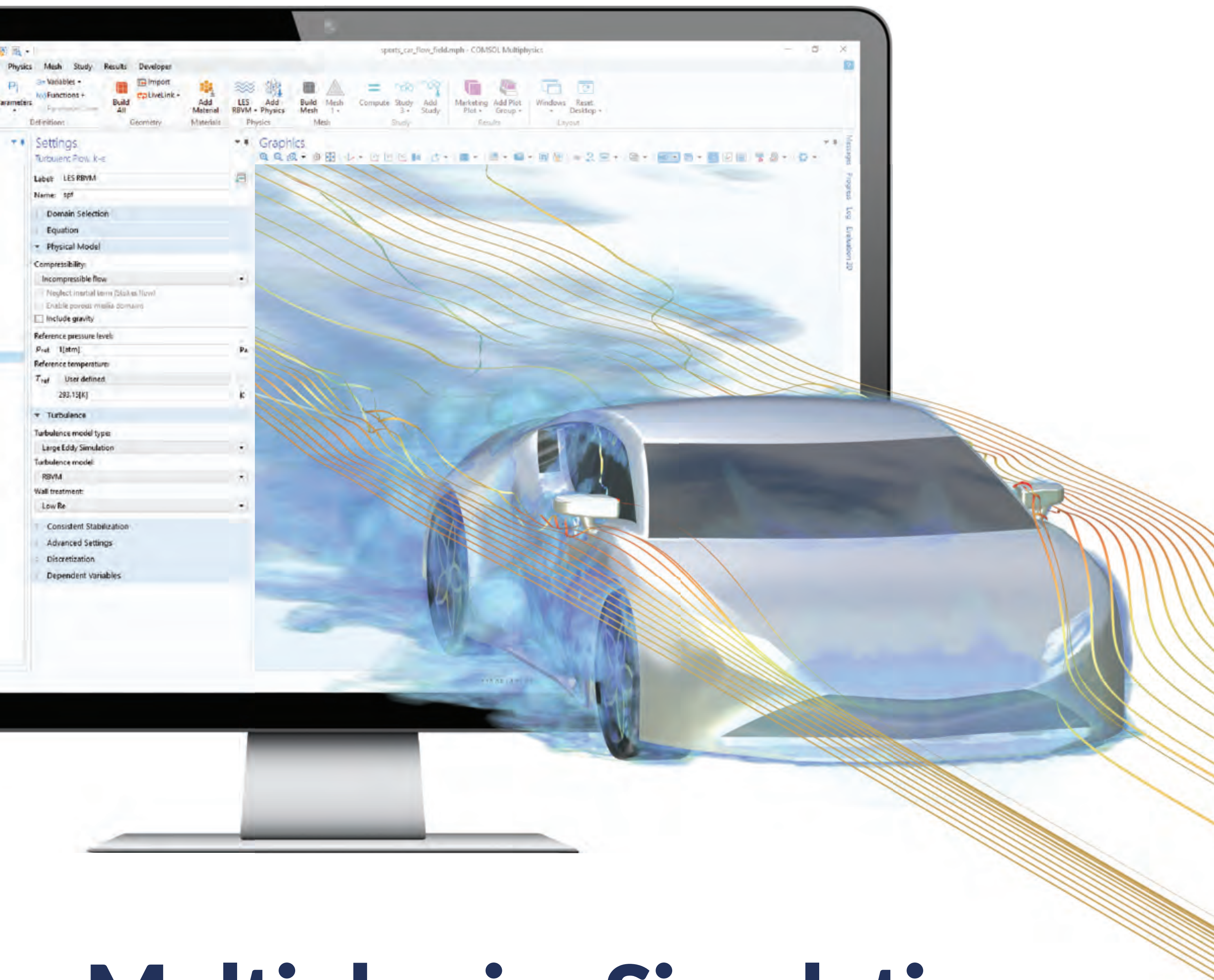
A wandering vortex

Vortices are found across a wide range of spatial scales and environments, including the water spiraling down a bathtub drain and the whirlpools of cooling plasma that sink into the Sun's interior. To create a vortex in the lab, a tank of water is typically mounted on a rotating table. The water drain is aligned with the rotation axis, and a vortex appears and remains above the drain. Natural vortices, however, typically follow a more complicated motion.

To model that more realistic situation, Rick Munro (University of Nottingham, UK) and Mike Foster (the Ohio State University) set up a closed cylindrical tank with the drain located at its midradius, filled it with water, and rotated it at an angular velocity of 0.6 rad/s. A vortex initially forms over the drain but quickly moves off by self-induction and follows a more-or-less circular path around the axis until it finally comes to rest. From the tank's outer rim, fluorescently dyed water spirals inward via the bottom boundary layer before moving upward as it swirls around the vortex core. That motion is seen in this photo, which was taken after 4.2 rotations of the tank. At that time, the vortex has moved a few millimeters off the drain and the dyed water has reached a height of 5 cm. After 25 rotations, it broadens and comes to rest as a hollow-core vortex. (Image courtesy of Cambridge University Press; R. J. Munro, M. R. Foster, *J. Fluid Mech.* **933**, R2, 2022.)

—AL

TO SUBMIT CANDIDATE IMAGES FOR BACK SCATTER VISIT https://contact.physicstoday.org.


Settings
Turbulent Flow, k-ε
Label: LES RBVM
Name: spf
Domain Selection
Equation
Physical Model
Compressibility:
Incompressible flow
Include gravity
Reference pressure level:
Reference temperature:
User defined
293.15[K]
Turbulence
Turbulence model type:
Large Eddy Simulation
Turbulence model:
RBVM
Wall treatment:
Low Re
Consistent Stabilization
Advanced Settings
Discretization
Dependent variables
Graphics

MathWorks®

MATLAB SPEAKS DEEP LEARNING

With MATLAB® you can build deep learning models using classification and regression on signal, image, and text data. Interactively label data, design and train models, manage your experiments, and share your results.

mathworks.com/deeplearning

Semantic segmentation for wildlife conservation.