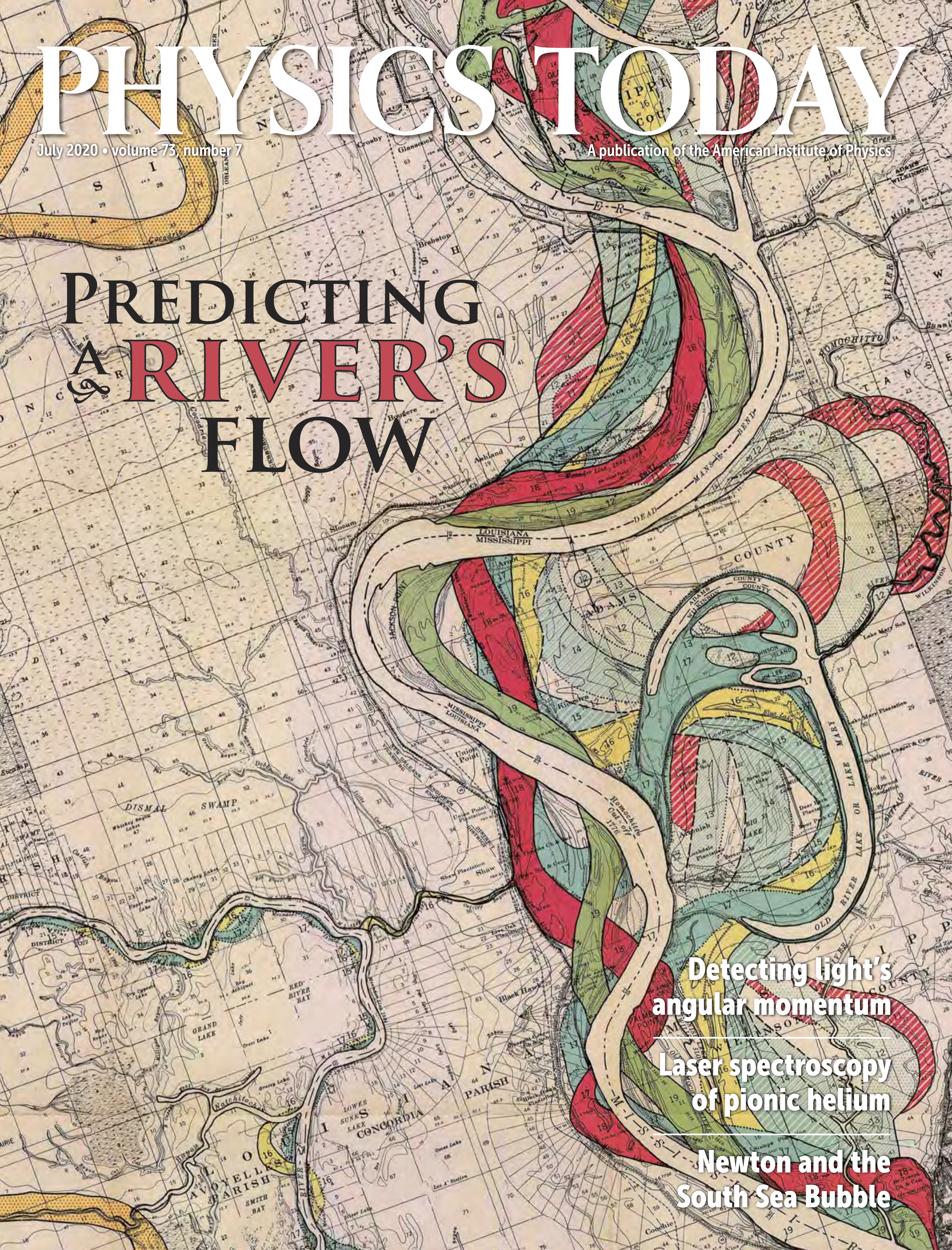
PHYSICS TODAY
July 2020 • volume 73, number 7
A publication of the American Institute of Physics

Predicting a River's Flow



Detecting light's angular momentum

Laser spectroscopy of pionic helium

Newton and the South Sea Bubble


Studying joint models with OCT images today.
Helping Arthritis patients tomorrow.
Accelerate possibility with affordable OCT.
ART AUTHENTICATION
BIOFILM ANALYSIS
BIOMEDICAL ENGINEERING
CARDIOLOGY
COATING MEASUREMENT
COUNTERFEIT DETECTION
CURRENCY AUTHENTICATION
DENTISTRY
DERMATOLOGY
DEVELOPMENTAL BIOLOGY
ELECTRICAL ENGINEERING
GASTROENTEROLOGY
GEM STONE IDENTIFICATION
GYNECOLOGY
SUMMER SPECIAL // OQ LabScope 2.0 /X
Generates 80,000 Lines Per Second
// Now $13,995.
Usually $14,995. Ends 8.30.2020
LOW-COST, HIGH-PERFORMANCE OCT SYSTEMS STARTING AT $9,995
LUMEDICA
LUMEDICASYSTEMS.COM

SR830 ***Lock-In Amplifier***

... the industry standard

SR830 ... $4950 (U.S. list)

- 1 mHz to 100 kHz range
- 100 dB dynamic reserve
- 5 ppm stability
- Auto gain, phase and reserve
- Harmonic detection (2F, 3F, ... nF)
- Time constants from 10 μs to 30 ks (6, 12, 18, 24 dB/oct rolloff)
- GPIB and RS-232 interfaces

The SR830 is the most widely used lock-in amplifier in the world. It has impressive analog specifications, many useful features, and it's reasonably priced. With over 100 dB of dynamic reserve and 0.01 degree phase resolution, the SR830 will handle the most demanding synchronous detection applications. It comes with convenient auto measurement features, a wide selection of time constants, and a built-in sine source.

Everything you need in a lock-in amplifier!

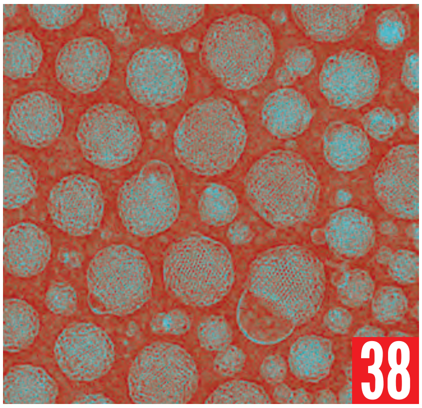
SRS Stanford Research Systems

www.thinksrs.com/products/lockin.html
Tel: (408)744-9040

PHYSICS TODAY

July 2020 | volume 73 number 7

FEATURES



ON THE COVER: Changes in rivers' flows can lead to destructive flooding or catastrophic famine. Those changes can happen over hundreds of years, like the meandering path of the Mississippi River, shown in this 1944 map, or they can take just a few hours, like in a flash flood. On **page 46**, Sean Fleming and Hoshin Gupta explain how researchers create hydrological models to understand river physics, manage water supplies, and inform public safety measures. (Harold Fisk/Army Corps of Engineers.)

Recently on PHYSICS TODAY ONLINE

www.physicstoday.org

#SHUTDOWNSTEM

▶ Combating racism
On 10 June scientists responded to the Strike for Black Lives and #ShutDownSTEM initiatives by spending the day learning about and discussing anti-Black racism and what to do about it. PHYSICS TODAY's coverage includes links to resources for sustaining the push for change.
physicstoday.org/forchange

ROYAL IRISH ACADEMY

▶ Elizabeth Fulhame
In 1794 Elizabeth Fulhame postulated the mechanism of catalysis decades before the term was coined. Yet the Scottish chemist is rarely mentioned in histories of the period. Claire Jarvis profiles the trailblazing chemist and her defiance of a society that thought little of female scientists.
physicstoday.org/Jul2020a

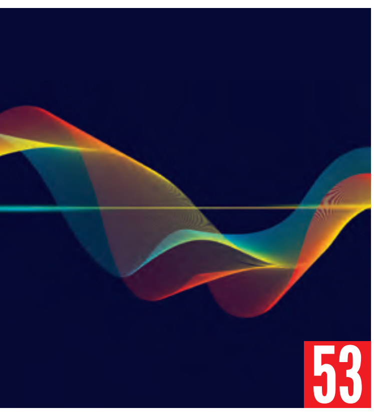
NASA-ISS

▶ BECs in space
An experiment aboard the International Space Station recently produced the first Bose–Einstein condensate in microgravity. Researchers report observation times of more than a second and unexpected BEC behaviors that could prove useful for applications such as high-precision spectroscopy.
physicstoday.org/Jul2020b

PHYSICS TODAY (ISSN 0031-9228, coden PHTOAD) volume 73, number 7. Published monthly by the American Institute of Physics, 1305 Walt Whitman Rd, Suite 300, 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 300, 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



FROM THE EDITOR

Freely and openly

Charles Day

In the autumn of 1990, I took a month off between the end of my postdoc at the Institute of Space and Astronautical Science in Sagamihara, Japan, and the start of my next job at NASA's Goddard Space Flight Center in Greenbelt, Maryland. During that sojourn I visited the Institute of Astronomy at Cambridge University, where I had done my PhD.

Among the new people I met there was Ian Stevens, who coincidentally had just finished a postdoc at Goddard. I no longer remember the circumstances of our encounter. Most likely it was during the 11:00am coffee break, when almost all of the institute's members gather in the library to converse over freshly brewed coffee. However, I can't forget the outcome. Stevens and I decided to work together. Our paper, "An x-ray excited wind in Centaurus X-3," appeared just over two years later.[1]

Collaboration pervades scientific research. Of the 150 or so news stories I've written for PHYSICS TODAY's Search and Discovery department, just one was about a paper by a single author, Rodney Baxter ("Order parameter of the chiral Potts model succumbs at last to exact solution," November 2005, page 19). Sometimes the work I reported on resulted from a collaboration among the members of a single group. Other times it resulted from a collaboration between groups.

One of my favorite examples of intergroup collaboration is the first observation of the quantum spin Hall effect (QSHE) in a heterostructure made of a layer of mercury telluride of just the right thickness sandwiched between layers of mercury cadmium telluride (see PHYSICS TODAY, January 2008, page 19).

The intricate setting has an interesting back story. Condensed-matter theorists began investigating the QSHE in the early 2000s. It arises when an insulator's band structure includes two spin-polarized edge states. Because of the states' topology, the corresponding edge currents flow without dissipation. The QSHE is not only interesting; it also has potential applications. The effect proved difficult to measure. It was too puny in the first prospective material, graphene. A proposal to realize it in a mechanically strained semiconductor was experimentally infeasible.

A hint of a way forward arrived when one of the theorists who elucidated the QSHE, Shoucheng Zhang of Stanford University, met one of the experimenters who eventually observed it, Laurens Molenkamp of the University of Würzburg, at a conference in South Korea. Molenkamp told Zhang about the narrow-bandgap semiconductor, HgTe. Spin-orbit coupling is so strong in HgTe that it splits the valence band and raises the higher of the split bands above what would otherwise be the conduction band. Could band inversion be a way to create the QSHE? the experimenter asked the theorist. When they met again a year later in Singapore, they agreed to collaborate.

A crucial step was taken from a general search for the QSHE in HgTe to a specific one in a heterostructure of HgCdTe/HgTe/HgCdTe when Molenkamp gave Zhang an electronic copy of the doctoral thesis of Alexander Pfeuffer-Jeschke. In 2000 Pfeuffer-Jeschke had calculated the threshold thickness below which the bandgap of the HgTe layer becomes zero and then becomes positive under the increasingly strong influence of the HgCdTe layers. He wrote his thesis in German, a language that Zhang had acquired as an undergraduate at the Free University of Berlin. Zhang read the thesis on the flight back from Singapore.

The story of the discovery of the QSHE is not only one of collaboration. It's also one of the free movement of people and ideas between countries. Molenkamp works in Germany and is from the Netherlands. The experiment that he and his collaborators carried out was sketched out in a *Science* paper[2] written by Zhang and two of his then students, B. Andrei Bernevig, who is from Romania, and Taylor Hughes, who is from the US. Zhang grew up in Shanghai.

I asked Molenkamp to fact-check the first draft of this editorial. His emailed reply closed with the observation that physics has been an international undertaking for a century. "There is just a rather limited number of professionals in the field," he wrote. "And you really only get things done when you exchange ideas. Freely and openly. And respecting scientific discourse."

References

1. C. S. R. Day, I. R. Stevens, *Astrophys. J.* **403**, 322 (1993).
2. B. A. Bernevig, T. L. Hughes, S.-C. Zhang, *Science* **314**, 1757 (2006).

PT

THE NORWEGIAN ACADEMY OF SCIENCE AND LETTERS

ANNOUNCES THE 2020 LAUREATES OF

THE KAVLI PRIZE

THE KAVLI PRIZE IN ASTROPHYSICS

Selection Committee: Viggo Hansteen (Chair), University of Oslo, Norway; Alessandra Buonanno, Max Planck Institute for Gravitational Physics, Germany; Andrea Ghez, University of California, Los Angeles, USA; Robert C. Kennicutt, Jr, University of Arizona, US; Irwin I. Shapiro, Harvard University, USA

for groundbreaking research in the field of observational X-ray astronomy, covering a wide range of topics from gas flows in clusters of galaxies to supermassive black holes at the heart of galaxies

ANDREW FABIAN, University of Cambridge, *UK*

THE KAVLI PRIZE IN NANOSCIENCE

Selection Committee: Bodil Holst (Chair), University of Bergen, Norway; Gabriel Aeppli, Paul Scherrer Institut, Switzerland; Susan Coppersmith, University of New South Wales, Australia; Shuit-Tong Lee, Soochow University, China; Joachim Spatz, Max Planck Institute for Medical Research, Germany

for sub-ångström resolution imaging and chemical analysis using electron beams

HARALD ROSE Universität Ulm, *Germany*
MAXIMILIAN HAIDER CEOS GmbH, *Germany*
KNUT URBAN Forschungszentrum Jülich, *Germany*
ONDREJ L. KRIVANEK Nion Co, *USA*

THE KAVLI PRIZE IN NEUROSCIENCE

Selection Committee: Kristine B. Walhovd (Chair), University of Oslo, Norway; Alexander Borst, Max Planck Institute of Neurobiology, Germany; Catherine Dulac, Harvard University, USA; Mary E. Hatten, The Rockefeller University, USA; Denis Le Bihan, NeuroSpin, CEA, France

for the transformative discovery of receptors for temperature and pressure

DAVID JULIUS UC San Francisco, *USA*
ARDEM PATAPOUTIAN Scripps Research, *USA*

kavliprize.org

The Kavli Prize is a partnership among The Norwegian Academy of Science and Letters, The Norwegian Ministry of Education and Research, and The Kavli Foundation (USA).

ITER and the prospects for commercial fusion

The article "The challenge and promise of studying burning plasmas" by Richard Hawryluk and Hartmut Zohm (PHYSICS TODAY, December 2019, page 34) contains a nice description of the physics involved in a burning plasma, which ITER, the international prototype fusion energy reactor, hopefully will produce. But ITER has had a troubled history. It was approved in 2005 for an estimated construction cost of approximately $5 billion, and deuterium–tritium experiments were expected to start in 2027. At present, the estimated cost has mushroomed to at least $25 billion in today's dollars, and the start date for D–T experiments has slipped to 2035, with their completion expected around 2040.

Realistically, though, ITER's development path is unlikely to produce commercially competitive electricity in this century. According to the ITER website (www.iter.org), the reactor is designed to produce a 10-fold or better return on energy; that is, it should produce 500 MW of fusion power from its 50 MW of input heating power.

Let's assume that ITER achieves that return in about 2040. What would that mean for power production? Electricity is generally produced with an efficiency of around one-third, so as a power plant, ITER would generate approximately 170 MW of electricity (MWe). Yet it requires 50 MW of beams or microwaves to power it. But beams and microwaves are themselves produced at around one-third efficiency, meaning that they would require 150 MW of input power. That would leave virtually nothing for the power grid. A typical commercial power plant, by comparison, will generate about 3 GW of heat or 1 GW of electricity.

For an ITER-like tokamak to be economically integrated into the grid, it would need its gain increased by at least a factor of three or four, its power increased by about a factor of six (to be on par with a typical commercial power plant), and both its size and cost reduced. Such a tokamak would deliver at least an order of magnitude more power to the wall and diverter plates. These requirements are not minor details! In all likelihood, reaching them would take decades and tens of billions of dollars, assuming they could be accomplished at all.

In addition to these obvious difficulties, tokamaks are limited in pressure and density, as Hawryluk and Zohm point out. They are also limited in current. The limits are not controversial; they have been well established theoretically and confirmed experimentally. Yet the constraints they place on fusion power, which I have called "conservative design rules,"[1,2] have been ignored by the tokamak community. Furthermore, conservative design rules have been in the literature for a decade. I have given many presentations on them at fusion labs and other places, and they have never been challenged, in print or in my seminars.

As long as tokamaks remain so constrained, they are unlikely to generate economic power. However, there is an alternative. As a breeder of nuclear fuel, an ITER-like tokamak would work well. Most likely it could economically breed uranium-233 from thorium. It would be a much more prolific fuel producer than a fission breeder of equal power and could become the basis of a worldwide, sustainable, carbon-free, nuclear infrastructure. Furthermore, it might well be able to do so soon after midcentury, assuming ITER is successful.[1,2]

References

1. W. Manheimer, *J. Fusion Energy* **33**, 199 (2014).
2. W. Manheimer, *IEEE Access* **6**, 64954 (2018).

Wallace Manheimer
(wallymanheimer@yahoo.com)
Allendale, New Jersey

The article by Richard Hawryluk and Hartmut Zohm addresses several interesting issues in ITER's march toward successful plasma burning. In particular, ITER's design, as presented in the article, relies on the high-confinement mode—the H-mode—that may be achieved with a thermal barrier believed to arise from turbulence-generated zonal flow.[1] As the discoverer of that type of flow, I highlight here a fundamental issue in the H-mode tokamak reactor. If operation depends on the formation of zonal flow possibly created by the self-organization of a type of Hasegawa–Mima turbulence, it is crucial that such turbulence is continuously generated. The scheme may be considered to be dynamic confinement, as compared with the classic static confinement scheme based on a magnetic bottle.

Dynamic confinement requires continuous injection of free energy via RF or neutral beams to sustain the turbulence and hence the thermal barrier—and, in effect, the pressure profile. For that process to take place on a steady-state basis, one must assume that the injected energy is lost continuously. Therefore, if the ITER design is based on dynamic confinement, ITER should be viewed as a power amplifier—that is, the fusion energy output should be regarded as amplified injected free energy. Since the injected power should be considered lost through an inverse cascade of turbulent energy, ignition criteria, such as the well-known Lawson criterion, that are based on energy confinement time become irrelevant: The energy is not confined. Here the crucial time scale is not of energy confinement but of maintaining the plasma pressure profile sustained by the zonal flow.

Reference

1. A. Hasegawa, M. Wakatani, *Phys. Rev. Lett.* **59**, 1581 (1987).

Akira Hasegawa
(a.hasegawa@solitoncomm.com)
Osaka University
Suita, Japan

After the ITER project was initiated, decarbonization of global energy production has emerged as an urgent objective for industrial society. Thermonuclear fusion is an inherently safe, ubiquitous, zero-carbon source of energy on a scale large enough to power a growing global industrial economy. It is therefore appropriate to examine fusion research by asking whether economically viable fusion power plants will come on line soon enough to affect climate change. The field has seen advances in high-temperature superconductors, inexpensive high-field permanent magnets, and other materials and in modeling and manufacturing. Those new technologies enable a range of heating and confinement approaches far beyond what was conceivable at ITER's inception.

The 2019 *Final Report of the Committee on a Strategic Plan for U.S. Burning Plasma Research* from the National Academies of Sciences, Engineering, and Medicine makes the following recommendations:

- First, the United States should remain an ITER partner as the most cost-effective way to gain experience with a burning plasma at the scale of a power plant.
- Second, the United States should start a national program of accompanying research and technology leading to the construction of a compact pilot plant that produces electricity from fusion at the lowest possible capital cost.[1]

Since the publication of that report, much has changed in the US fusion community. In the public sector, the American Physical Society's division of plasma physics led a community planning process that identified scientific and technological opportunities in plasma physics and fusion energy sciences and provided the Department of Energy's Fusion Energy Sciences Advisory Committee with strategic recommendations for addressing those opportunities.[2]

The private sector has invested hundreds of millions of dollars in pursuing novel approaches to accelerate the deployment of commercial fusion power plants (see, for example, PHYSICS TODAY, August 2018, page 25, and February 2019, page 28). The BETHE Project (Breakthroughs Enabling Thermonuclear-Fusion Energy) under the Advanced Research Projects Agency–Energy recently released a funding-opportunity announcement that targets transformational research "with the potential to be disruptive in the marketplace" and specifically encourages development of "timely, commercially viable fusion energy."

ITER will make important contributions to the mission of deploying commercially viable fusion. And public, private, and partnership groups are aggressively pursuing ways to accelerate commercial deployment. New designs may eliminate the physics challenges cited in the article by Richard Hawryluk and Hartmut Zohm.

ITER will not be on the critical path for commercial deployment, and the answer to when we will have commercial fusion is this: sooner than you think.

References

1. National Academies of Sciences, Engineering, and Medicine, *Final Report of the Committee on a Strategic Plan for U.S. Burning Plasma Research*, National Academies Press (2019), p. 1.
2. American Physical Society Division of Plasma Physics, *A Community Plan for Fusion Energy and Discovery Plasma Sciences*, APS (2019).

Jesse Treu
(jtreu@stellarenergyfoundation.org)
Matthew Miller
Peter Burnim
Stellar Energy Foundation
Princeton, New Jersey

▶ **Hawryluk and Zohm reply:** Wallace Manheimer points out that remaining scientific and technological issues will need to be addressed to produce commercially competitive electricity from nuclear fusion on the completion of the ITER research program. True, ITER was not designed to produce net electricity, but power generation can be accelerated by doing more R&D in parallel with ITER's construction. The fusion community recognizes that; it was a central topic of the research initiatives highlighted in the 2019 National Academies of Sciences, Engineering, and Medicine report on burning plasma research.[1] Examples of such parallel activities that are underway include the proposed China Fusion Engineering Test Reactor and the new initiatives Jesse Treu and colleagues mention. Those efforts will profit from both the technological progress made during ITER construction and the physics understanding gained from its operation.

The "conservative design rules" Manheimer mentions are indeed recognized by the tokamak community. Research initiatives around the world are actively addressing them and seeking scientific and technological innovations to increase fusion power for a given facility size. Among them are increasing the magnetic field strength by using high-temperature superconducting magnets and increasing the power heat flux capabilities through the use of liquid-metal divertors.

Manheimer notes that fusion at an ITER-like tokamak can be used to breed nuclear fuel, such as uranium-233 from thorium-232. That may decrease the R&D effort in order to extend the parameters for fusion power plants, but it will entail other technological developments to breed ^{233}U from Th and incorporating them into a nuclear fission economy. Some ITER partners have considered that approach, but others find it less attractive because of its link to fission.

Treu and coauthors note the strong need for innovation, as identified in the National Academies report; the community consensus recommendations for a strategy to address those opportunities; and the increased role of private companies to accelerate development. We share with them the hope that new designs for fusion facilities combined with results from ITER can accelerate the development of commercial fusion.

Finally, the triggering mechanism for the transition to the high-confinement barrier (H-mode) in the edge region is an ongoing area of research that is building on early work by Akira Hasegawa, Masahiro Wakatani, and many others. Experimentally, the criterion for entering the H-mode is characterized in terms of an ion heat flux. For a fully developed H-mode state, the shear flow necessary to suppress turbulence in the edge region is provided by the zeroth-order diamagnetic flow—that is, the pressure gradient and not the turbulence itself. Hence, sustaining H-mode confinement after the transition has occurred does not rely on additional external energy or momentum input; it can be sustained by the heat flux arising from the α-particle heating. We now realize that this important aspect regarding sustaining the H-mode was not clear enough in our article.

Reference

1. National Academies of Sciences, Engineering, and Medicine, *Final Report of the Committee on a Strategic Plan for U.S. Burning Plasma Research*, National Academies Press (2019).

Richard Hawryluk
(rhawrylu@pppl.gov)
Princeton Plasma Physics Laboratory
Princeton, New Jersey
Hartmut Zohm
(zohm@ipp.mpg.de)
Max Planck Institute for Plasma Physics
Garching, Germany

A proposal for APS action on police brutality

The shocking death of George Floyd on 25 May 2020 is unfortunately not a singular event in the US. Eric Garner, Michael Brown, Freddie Gray, Breonna Taylor, and Amadou Diallo—who was struck by 19 of the 41 bullets fired by four New York City police officers—are just part of a long list of unarmed Black Americans who have died at the hands of the police. That pattern of violence is deeply rooted in the history of Black–white relations in the US and the failure of the leaders of this country to deal with systemic racism. It is worth asking what an organization with a diverse membership such as the American Physical Society can do to effect change.

Yearly, APS hosts large meetings in cities around the US. Perhaps the time has come for the organization to include the treatment of Black Americans by the local police as a key criterion in choosing cities to host the society's meetings. Cities in which Black people have died at the hands of the police—Baltimore (a frequent destination for APS meetings), New York, and St Louis, for example—might be removed from future consideration. The message APS would send would be powerful. Not only would it say that we do not condone racist policing, it would support the protection of the society's members and guests who might, for example, venture out onto the said city's streets while being Black.

The request here is simple: Stop doing business in cities in which the police kill Black citizens. The action would be similar to the divestment movement in South Africa in the 1980s, which was pivotal in ending apartheid. We hope other professional societies follow suit, but as with all causes, someone has to be the first. Physicists should be proud to take the lead here.

Philip Phillips
(dimer@illinois.edu)
Michael B. Weissman
University of Illinois at Urbana-Champaign

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.

The language of wildfires

The Quick Study "Fluid dynamics of wildfires" by Rod Linn (Physics Today, November 2019, page 70) applies physics principles to the increasingly urgent global issue of wildfires. It is both thought provoking and relevant.

Included in the piece are the phrases "rising air draws in air below it . . . in much the same way as air is drawn into a fireplace and rises up a chimney," "a rising plume can draw cool air," and "ambient wind drawn into the gaps." In those contexts, the dictionary meanings of the word "draw" seem to invite the common misapprehension that the heated air of a wildfire is somehow pulled or drawn upward from above.

Heated gases from wildfires are not drawn upward, nor do they rise spontaneously. Wildfire gases are better characterized as being pushed, pressed, or buoyed upward from beneath, ultimately by ambient pressure. The heated gases of a wildfire, much like air bubbles rising in water, are driven upward by a hydrostatic pressure gradient. The greater pressure is at the lower elevation, as established by gravity. Absent gravity, the gradient would disappear and the heated gases would not rise but would instead smother the wildfire.

Precise language is important when talking about science, particularly to accurately represent wildfires as being pushed from behind or below rather than to suggest, as is common in popular parlance, that they are drawn or pulled upward from above.

Evan Jones
(revwin@yahoo.com)
Sierra College
Rocklin, California

Laser spectroscopy: A new way to study pions

By creating exotic atoms containing an ever-broadening range of particles, researchers can extend the tools of atomic physics to new realms.

The best atomic clocks, which mark time by measuring the frequency of an atomic resonance, operate with an uncertainty of around 10^{-18}. If left to run for the age of the universe, they'd lose less than a single second. That astonishing precision is made possible by laser spectroscopy.

The quantum mechanical laws that give rise to atomic structure aren't limited to protons, neutrons, and electrons. One can, in principle, construct a hydrogen-like atom out of any two particles of opposite charge, or replace any electron in a larger atom with any other negatively charged particle. Laser spectroscopy on those exotic atoms can probe the properties of both the unusual particles and the ordinary ones.

In particular, by studying the spectrum of antihydrogen, researchers can search for differences between matter and antimatter (see PHYSICS TODAY, February 2017, page 16). And muonic hydrogen, in which a muon replaces the electron, has been used to measure the charge radius of the proton: The muon, 200 times more massive than the electron, occupies a smaller atomic orbital that's more sensitive to the proton's nonzero size.

Now Masaki Hori of the Max Planck Institute of Quantum Optics and his colleagues have made the first laser spectroscopic measurement on an atom containing a meson.[1] Specifically, they've studied pionic helium, in which one of the electrons in an ordinary helium atom is replaced by a negatively charged pion. With a lifetime of just 26 ns, the charged pion is the shortest-lived particle ever to be probed by exotic-atom laser spectroscopy.

Because an atom's spectroscopic resonances depend on the masses of its constituent particles, Hori and colleagues' technique could eventually lead to an improved measurement of the charged pion's mass, currently known to a fractional precision of 10^{-6}. In turn, through kinematic analysis of the charged pion's main decay channel—which yields a muon and a muon antineutrino—the pion mass measurement could provide a much-needed additional constraint on the neutrino mass.

MASAKI HORI

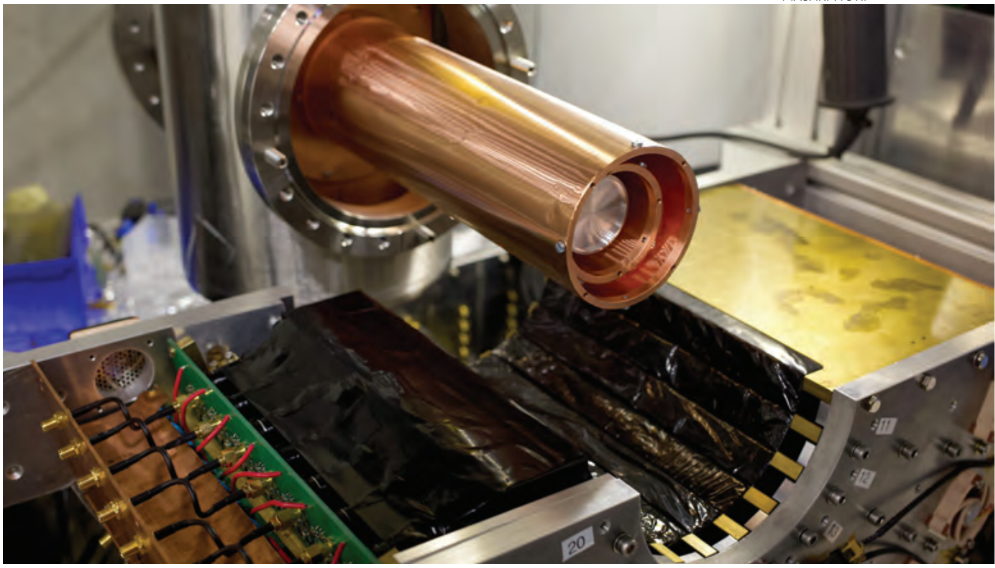


FIGURE 1. PION PHYSICS IN A TUBE. In this disassembled experiment, a 4.2-cm-diameter aluminum cylinder, just visible inside the copper radiation shields, is a vessel for holding superfluid helium. When a pion beam streams through the cylinder, some of the pions displace atomic electrons to create pionic helium atoms. A suitably tuned laser beam passing through the cylinder in the opposite direction excites some of the pionic atoms into short-lived states in which the pion and nucleus promptly and destructively collide. The fission products of the annihilated atoms, detected with an array of plastic scintillators surrounding the cylinder (half of them shown here), indicate that the laser was on resonance with an atomic transition.

Exotic atoms

The pion's short lifetime poses an experimental challenge. Positrons and antiprotons, the components of antihydrogen and numerous other exotic atoms, live effectively forever—as long as they're kept away from their antiparticles—so there's no limit on how long they can be cooled, trapped, or shuttled around with electric and magnetic fields. Even the muon, with a lifetime of 2 μs, survives long enough to be decelerated and inserted into an atom.

The pion allows no such leeway. There's no time to do anything more sophisticated than take the pions straight off a beamline, smash them into a target element of interest, and hope that some of them displace atomic electrons to form pionic atoms.

Unfortunately, the resulting atoms have even shorter lifetimes than the pions they contain. Isolated charged pions decay through the weak interaction. But if a pion comes across an atomic nucleus, it reacts with one of the nucleons through the much faster strong interaction, destroying both the nucleus and the pion.

Typically, a pion enters an atom in an orbital of high principal quantum number n, then immediately tumbles into a lower-energy orbital, emitting photons and casting electrons out of the atom as it goes. It makes its way to an orbital that has large overlap with the nucleus, then blows it to bits. The whole process, which takes a few picoseconds or less, has been well studied. Indeed, it's from the energies of the photons emitted from pionic nitrogen that we know the pion mass as well as we do.[2] But the short-lived atoms can't be studied with the accuracy of laser spectroscopy.

Helium promised an exception. If a

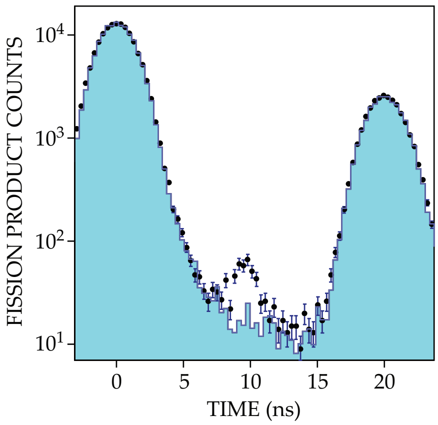


FIGURE 2. MOST PIONIC HELIUM ATOMS self-destruct immediately, even without laser excitation. So when pions arrive in bunches spaced by 20 ns, their resulting fission products are likewise detected in bunches, as shown by the blue histogram of data taken with the laser switched off. The only way to see a spectroscopic signal from the few longer-lived pionic atoms is to time the laser pulses to arrive between the pion bunches. The black data points show a successful spectroscopic detection with the laser tuned to 1631.4 nm. (Adapted from ref. 1.)

pion enters a helium atom in an orbital of high angular-momentum quantum number l (say, $n = 16$, $l = 15$), it's left with nowhere to go. It can't move to another orbital without a large change in angular momentum, and the remaining atomic election in a tightly bound, zero-l orbital can't do much to compensate. Because the high-l orbital has little overlap with the nucleus, the pionic helium atom could live for almost as long as the pion itself. With the right tools and techniques, that's long enough to probe with a laser.

Finding the resonance

Hori and colleagues booked time at the most powerful pion beamline in the world, at the Paul Scherrer Institute (PSI) in Switzerland. Protons from the PSI ring accelerator collide with a chunk of graphite to create a stream of charged pions for user experiments. (The J-PARC accelerator in Japan creates pions the same way; see PHYSICS TODAY, June 2020, page 14.)

The experimental apparatus for making and detecting pionic helium is shown in figure 1. To hold the liquid helium for their target, the researchers needed a vessel with walls sturdy enough to contain the cryogenic fluid but thin enough for the pions to pass right through. They chose a tube-shaped container made of aluminum with walls 0.5 mm thick. "It took a year of failed prototypes, as the thin walls kept buckling," says Hori, "but we finally got a chamber that didn't leak or deform at cryogenic temperature."

None of the pionic helium atoms ever leave the liquid helium environment. Almost all of them are destroyed through the energetic collision of the pion and the nucleus that sends nucleons flying in random directions. The nucleons, like the pions, pass right through the thin aluminum walls, and they're detected by an array of plastic scintillators surrounding the helium target.

Some 98% of pionic helium atoms form in low-l states that decay almost instantaneously; the remaining 2% occupy high-l states that meet a more leisurely demise over the ensuing tens of nanoseconds. If those rare remaining atoms could be excited by a laser pulse into a lower-l state, they'd be prompted to undergo fission all at once. Their decays would show up as a blip in the scintillator trace that's present when the laser is on and absent when it's off—a telltale sign that the laser is tuned in to an atomic resonance.

The timing needed to be precise. The PSI accelerator generates protons, and thus pions, at 50 MHz, or in bunches 20 ns apart. "Essentially, the experiment is temporarily blinded every 20 ns by the arrival of new particles," explains Hori. "That time window seemed too short for spectroscopy. I asked if they could increase the spacing, but it would have been too much of a disturbance for the other experiments at the facility."

Ultimately, the experimenters took the risk and proceeded anyway. To know where to focus their spectroscopic search, they enlisted the help of theorist Vladimir Korobov of the Joint Institute for Nuclear Research in Dubna, Russia, to calculate the expected transition energies of pionic helium. At first, they sought to excite transitions from the $n = 16$, $l = 15$ state, the high-l state expected to be most populated. But despite months of work, they found nothing.

Running out of beam time and grant money, they tried exciting a transition from the $n = 17$, $l = 16$ state. Within days, they saw a signal, shown in figure 2 as the difference between the black data points (collected with the laser on) and the blue histogram (laser off). The signal peak nestled between two much larger peaks, created by the short-lived states, so it was visible only when the data were plotted logarithmically.

But there was a problem: The resonance wasn't quite where it was supposed to be. Korobov had calculated that it should be centered on a frequency of 183 681 GHz (equivalent to a wavelength of 1632.1 nm), and the laser frequency that generated the signal was 0.04% higher than that. The difference is far too great to be explained by uncertainty in the pion mass. Instead, the researchers attribute it to atomic collisions in the liquid helium that distort the pionic atoms' orbitals and shift their energies. The same mechanism may explain why they didn't see any signal from the $n = 16$, $l = 15$ state—collisions may distort some atomic states so much as to render them effectively unbound. Says Hori, "This is a far more fragile system than any other exotic atom, or even any normal atom, that's been studied with laser spectroscopy before."

To get a handle on the collision effect, Hori and colleagues plan to swap out the liquid helium target for a gaseous one whose density they can adjust. A more dilute gas means fewer atomic collisions. By measuring the resonance in gases of different density and extrapolating to zero, they should be left with the resonant frequency of an isolated pionic helium atom. From there, they'll be able to calculate the pion mass. If they encounter no measurement limitations other than the resonance's natural linewidth, Hori estimates that they could reach a fractional uncertainty as low as 10^{-8}.

Neutrinos and more

An improved measurement of the pion's mass will make possible a more precise analysis of its decay into a muon and muon antineutrino. Solving the kinematic equations for the neutrino's mass gives an expression that depends on the square of the charged pion mass. (It also depends on the square of the muon mass, but that's much better known.) So although the pion mass is known to within 240 eV, the best that can be said of the muon neutrino from pion decay is that its mass is less than 190 keV. By pinning down the pion mass by an additional two orders of magnitude, laser spectroscopy could help to refine that limit.

With the help of theory, one can arrive at a far tighter constraint. Current understanding of neutrino physics holds that all three known flavors of neutrino—electron, muon, and tau—are mixtures of the same three mass states. The mixing allows neutrinos to transform from one flavor into another, and studies of that flavor oscillation, which measure the differences between the squares of the masses, show that all three masses are within tens of meV of one another. Furthermore, the KATRIN (Karlsruhe Tritium Neutrino) experiment, a study of nuclear beta decay, found[3] that the electron antineutrinos emitted in that process can't be more massive than 1.1 eV.

That line of reasoning implies that the muon neutrino mass is also in the neighborhood of 1 eV or less, and certainly not as large as 190 keV. But neutrino mass is a mysterious thing. It's not part of the standard model of particle physics, and there may yet be theoretical surprises in store. Experiments that directly probe each neutrino flavor's mass may still be valuable.

Now that pions can be inserted into atoms and manipulated with lasers, Hori hopes that laser spectroscopy will prove to be a more general technique for studying other mesons. In addition to the negatively charged pion—a bound state of a down quark and an up antiquark—another appealing target is the negatively charged kaon, which replaces the down quark with a strange quark. The kaon's lifetime, at 12 ns, is slightly less than half as long as the pion's. And its mass, with a fractional uncertainty of more than 10^{-5}, is in need of a more precise measurement.

Johanna Miller

References

1. M. Hori et al., *Nature* **581**, 37 (2020).
2. M. Trassinelli et al., *Phys. Lett. B* **759**, 583 (2016).
3. M. Aker et al. (KATRIN collaboration), *Phys. Rev. Lett.* **123**, 221802 (2019).

Cold, supersaturated urban air could be accelerating pollutant particle growth

A new experiment suggests that ammonium nitrate particles nucleate and quickly grow in winter conditions.

In early April, city dwellers of India's northern provinces experienced once-in-a-lifetime views of the snow-capped Himalayas, thanks to a rare absence of smog. The foggy mixture of particles and gases will certainly reappear in the region once coronavirus restrictions are lifted and the air pollution returns. Particulate matter in the atmosphere is a leading cause of lung disease and may contribute to neurological diseases such as Alzheimer's. A better understanding of the pollutants that lead to smog formation is necessary for any mitigation effort to succeed.

Some particles in the atmosphere, called primary particles, form directly from combustion or mechanical generation. Other particles, called secondary particles, form from trace gases that condense and stick to the surfaces of existing particles, and sometimes undergo complex chemical evolution. Winter smog, shown in figure 1, is a mixture of both types of particles. Rapid growth is essential to the survival of secondary particles, but how they can grow so quickly in polluted urban air is a puzzle. Observations suggest that in urban air, vapors and particles are almost in equi-

FIGURE 1. SMOG LOOMS OVER PRAGUE on a February day. The pollutant, which forms when particulate matter reacts with UV radiation in the atmosphere, is a major contributor to respiratory disease. Still, it's unclear how the large particles found in smog can grow in urban air. (Courtesy of Vojife/Wikimedia Commons, CC BY 3.0.)

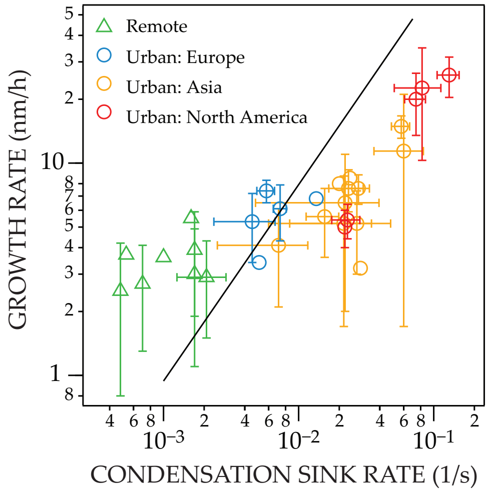


FIGURE 2. ATMOSPHERIC PARTICLE GROWTH RATES, shown here on the vertical axis, are only modestly faster in many urban environments (red, orange, and blue dots) than in remote ones (green triangles). The condensation sink rate (horizontal axis) describes how quickly newly formed clusters are scavenged by existing particles and is higher in urban environments than in remote ones. The persistence of those clusters depends on the ratio of the condensation sink and growth rates (the slope of the black line corresponds to a constant ratio), with small urban particles likely to be lost by scavenging unless they grow quickly. (Adapted from ref. 1.)

librium, so particles only grow slowly.

Now Neil Donahue of Carnegie Mellon University and colleagues have shown in a laboratory experiment that common nitrogen-containing vehicle emissions may accelerate the particle-formation process.[1] Those compounds have previously been thought to exist in concentrations near their equilibrium values in the atmosphere and therefore not to play a role in particle formation or early growth. The researchers identified conditions under which the vapor pressures of precursor pollutants in cold temperatures, typical in some winter months and in the upper atmosphere, allow nitric acid and ammonia to condense and quickly grow into large secondary particles of ammonium nitrate. Those particles can add significantly to the total number of particles in the atmosphere and contribute to the associated health effects.

Growing pains

Internal combustion engines, incinerators, cooking fires, and other sites of combustion in cities lead to high atmospheric concentrations of gas-phase pollutants, many of which readily condense into tiny clusters of molecules a few nanometers in size. But it's not easy for a small cluster of molecules to grow enough to form a larger stable particle, typically characterized as a cluster 10 nm or larger.

Many nascent clusters simply evaporate due to an effect identified in 1871 by Lord Kelvin: Smaller clusters have greater curvatures than larger ones, and therefore tend to have higher surface vapor pressures. As a result, smaller clusters evaporate more readily. For molecular clusters with diameters less than about 10 nm, molecules are more likely to escape than remain stuck together.

Other newly formed clusters are scavenged by existing particles. As tiny clusters grow, they reach a size where they can get caught in the turbulent wake of a larger, existing particle and stick to its surface. As a result, the total number of particles does not increase. The scavenging of new clusters by existing particles is known as the condensation sink rate.

The ability of small, newly formed clusters to grow into larger, stable particles depends on the ratio of the condensation sink and the cluster growth rate. Observations from Europe, North America, and Asia, summarized in figure 2, show that particle growth rates in urban areas are only a few times greater than growth rates in remote areas, whereas the condensation sink rate in urban areas is up to two orders of magnitude larger. The relatively efficient particle removal rate in urban areas, combined with the relatively slow growth rate of particles (10–20 nm/hr), has led atmospheric researchers to seek a mechanism by which particles can reach a diameter more than 10 nm within seconds to minutes within the rich chemical mixture of urban atmospheres.

Cows, cars, and CLOUD

Ammonia is the most significant base gas in the atmosphere, where it contributes to the formation of ammonium nitrate in the presence of nitric acid. In rural areas, ammonia is closely associated with livestock dung; in urban areas, vehicle catalytic converters may generate a level of ammonia similar to that produced in some agricultural regions.[2] The

FIGURE 3. THE CLOUD (COSMICS LEAVING OUTDOOR DROPLETS) CHAMBER at CERN is a 26-cubic-meter tank in which temperature and gas concentrations are carefully controlled to simulate natural atmospheric conditions between the ground level and the stratosphere. The proton synchrotron serves as an artificial source of cosmic rays, which simulates natural atmospheric conditions and may stabilize particles that form from the ionized gases in the chamber. External instrumentation allows for constant monitoring of the chamber's contents. (Courtesy of CERN/Neil Donahue.)

most common pathway for the formation of ammonium nitrate in a relatively clean atmosphere keeps the compound near chemical equilibrium with gaseous ammonia and nitric acid. As a result, that mechanism is unlikely to increase the number density of particles unless the atmosphere becomes supersaturated with nitric acid and ammonia.

To investigate what conditions may drive supersaturation and new particle formation in an urban setting, Donahue and his colleagues studied mixtures of common pollutants in the CLOUD (Cosmics Leaving Outdoor Droplets) chamber at CERN (figure 3). The facility—modeled after cloud chambers used in particle physics in the 1950s—provides an environment in which to simulate vapor and particle interactions under conditions representative of the atmosphere from the lower troposphere up to the stratosphere. UV light generated by a built-in fiber-optic system mimics the energy provided by the Sun's cosmic rays. That radiation ultimately controls how molecules ionize at different altitudes.

In a series of experiments designed to replicate winter conditions in polluted cities, the researchers introduced various combinations of nitric acid, ammonia, sulfuric acid, and their precursor molecules into the CLOUD chamber. They exposed the gases to a range of temperatures typical of the lower atmosphere and varied the radiation input to control ion distribution. Mass spectrometers and particle counters that encircle the chamber determined the resulting concentration of gases and the properties of newly formed particles.

The team found that at temperatures warmer than –15 °C, ammonium nitrate and sulfuric acid followed a known pathway to form small clusters of molecules.[3] Once clusters of ammonium sulfate reached a threshold size, ammonium nitrate began to condense onto the clusters, resulting in rapid particle growth—above 100 nm/hr—as long as the temperature remained below 5 °C. At temperatures colder than –15 °C, which can occur in humid air outflows above some clouds, the nitric acid and ammonia gases nucleated through an acid–base stabilization mechanism to form particles of ammonium nitrate, which then grew via additional condensation of ammonium nitrate.

Calculations showed that the critical size at which ammonium nitrate begins rapid growth depends on the maximum concentrations of the ammonia and nitric acid.[4] Once that size is reached, rapid growth continues until the precursor vapors return to the chemical equilibrium commonly found between ammonia, nitric acid, and ammonium nitrate.

Large-scale atmospheric transport models do not take into account the newly proposed path to particle formation. The CLOUD experiments suggest that it may be important instead to consider those vapors using dynamic calculations. The researchers propose that variations in temperature and emission sources in cold urban settings provide the conditions necessary for localized supersaturation of nitric acid and ammonia. By imposing regulations to control nitrogen emissions, urban planners could reduce the concentration and growth of particles and improve air quality.

Rachel Berkowitz

References

1. M. Wang et al., *Nature* **581**, 184 (2020).
2. B. J. Finlayson-Pitts, J. N. Pitts Jr, *Chemistry of the Upper and Lower Atmosphere*, Academic Press (2000).
3. J. Kirkby et al., *Nature* **476**, 429 (2011).
4. J. Kontkanen et al., *Atmos. Chem. Phys.* **18**, 13733 (2018).

Chip-scale sensor detects light's orbital angular momentum

The photocurrent in an unusual material could facilitate an increase in the information density of optical communications.

Light often has spin angular momentum, more commonly referred to as left or right circular polarization. Only in the past 30 years have researchers been able to impart orbital angular momentum (OAM; see the article by Miles Padgett, Johannes Courtial, and Les Allen, PHYSICS TODAY, May 2004, page 35). Instead of the usual flat wavefront, light with nonzero OAM has a helical wavefront that turns like a corkscrew in the direction of propagation. The quantum number m describes how tight the light's corkscrew motion is. For example, $m=1$ means a full clockwise rotation in a wavelength, and $m=-2$ means two full counterclockwise rotations in a wavelength.

OAM provides additional degrees of freedom for encoding information in optical communications. Modern optical networks rely on wavelength and intensity modulations to send data; the addition of polarization and OAM could pack in more information without boosting traffic. What's more, unlike polarization, which has two modes, m can take any integer value. Modes of different m are also orthogonal and wouldn't interfere with one another during transmission.

Ordinarily, measuring OAM requires a roomful of bulky optics. But Ritesh Agarwal at the University of Pennsylvania and his colleagues have found a compact way to detect light's OAM through the induced photocurrent in a device that's just tens of micrometers across.[1] Their technique could make it easier to read out information encoded in the OAM.

Gaining momentum

To measure m, researchers typically interfere light with a reference beam or send it through a hologram or specially designed aperture. The nature of the interference or diffraction pattern manifests the OAM. But those measurements not only require table-scale setups; they are also limited in the number of modes they can measure. For example, most holograms can detect only one particular value of m, so a new one must be swapped in for each mode.

In CCDs and other solid-state sensors, light is detected through its interactions with matter. A material's response depends primarily on the local incident number of photons—that is, the intensity. Phase information is usually lost when intensity is converted to photocurrent. Because a clockwise helical wavefront and a counterclockwise one have the same donut-shaped intensity profile, a phase-sensitive measurement is essential for characterizing light with OAM without introducing holograms and apertures.

Agarwal has been contemplating such a measurement for five years. Back in 2015 he and his team found a way to detect a light beam's circular polarization—its spin angular momentum—with a silicon nanowire.[2] They were interested in extending the capability to OAM but weren't sure how to do it. A couple years later, a breakthrough came when the team was trying to understand the properties of a new group of materials, Weyl semimetals.

Mean-Weyl . . .

Identified for the first time in 2015, Weyl semimetals have a broken symmetry, generally inversion symmetry, that allows them to host quasiparticles that behave as Weyl fermions—massless, charged, and chiral. (For more on Weyl semimetals, see Physics Today, December 2019, page 24.) Agarwal and his colleagues found that the photocurrent in Weyl semimetals was sensitive not to the light intensity but to the intensity gradients.[3]

Photocurrents in most materials depend only on the light's local intensity. But when excited at a normal angle, a crystal with mirror symmetry reacts not to the local intensity but to how the local intensity compares with the surrounding intensities, or the intensity gradient. In that case, the photocurrent is nonlocal and results from an imbalance, in both real and momentum space, in the density of excited electrons. Mathematically, it is described by the terms beyond the dipole approximation in the multipole expansion of the light–matter interactions. The symmetry cancels out the dipole response, which would otherwise obscure the other terms. Although Weyl semimetals aren't the only materials with the requisite symmetry, they have the added advantage of a nonlinear optical response that's among the largest identified to date.

In their intensity-gradient experiment, the researchers measured a circular photocurrent that could be explained only by the nonuniform distribution of the light. But in addition to a nonuniform intensity, light with nonzero OAM has a circling phase gradient that depends on *m*. Could a photocurrent sense a phase gradient as well? The group's calculations indicated that the measurement should be possible.

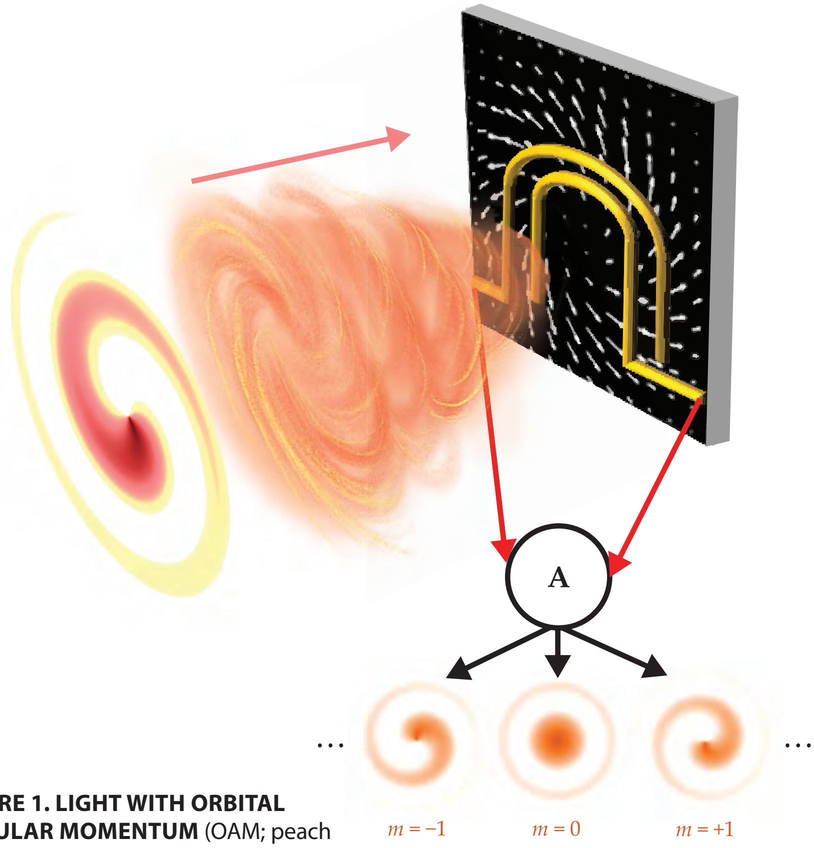


FIGURE 1. LIGHT WITH ORBITAL ANGULAR MOMENTUM (OAM; peach swirl on the left) excites photocurrents in a material (black rectangle). Usually, the photocurrent depends only on the light intensity, which doesn't vary with OAM. But in materials with the right symmetry, it arises in part from light's phase gradient (white arrows)—that is, the transfer of OAM from photon to electron. Using specially designed U-shaped electrodes (yellow), researchers measure the photocurrent with an ammeter (A) to determine the OAM quantum number *m*. (Adapted from ref. 1.)

A new phase

Agarwal and his graduate student Zhurun Ji prepared a 50- to 200-nm-thick layer of the Weyl semimetal tungsten ditelluride and designed U-shaped electrodes, shown in figure 1, to probe the photocurrent generated by light with OAM. The electrode shape needs to distinguish the current due to the OAM phase gradient from the current due to intensity gradients and other effects. Although a few other electrode shapes work, the U shape is the best at measuring a radial current and unambiguously canceling out the other effects.

Before reaching the WTe_2 detector, a beam with OAM passes through a rotating quarter-wave plate that cycles the polarization from 0° linear to left circular to 90° linear to right circular and back to 0° linear. The measured radial photocurrent, shown in figure 2a for beams with $m = +1$ and -1, has three components: one that depends on the linear polarization state; one, J_c, that depends on the circular polarization state; and one that is independent of polarization and includes thermal currents. To distinguish between the effect of the intensity gradient and that of the phase gradient, the researchers exploit the components' differing dependencies by comparing beams with OAM $+m$ and $-m$, which have opposite helicities but the same intensity profiles.

The photocurrent is not the same for left and right circular polarization states (the left and right peaks in figure 2a). And the difference, arising from J_c, is around the same magnitude but swapped for

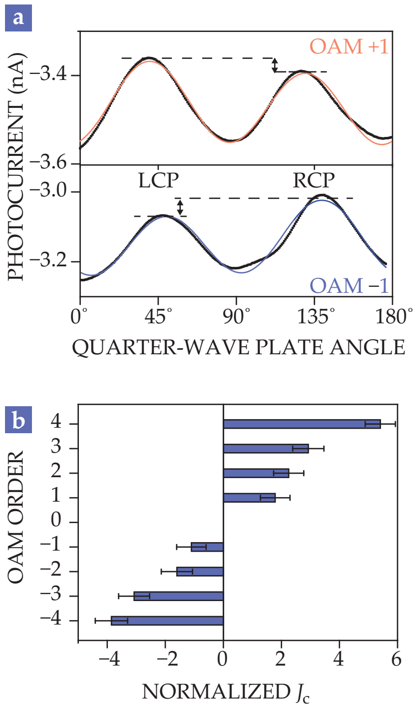


FIGURE 2. A WEYL SEMIMETAL'S PHOTOCURRENT is sensitive to light's orbital angular momentum (OAM). **(a)** Before reaching a thin layer of tungsten ditelluride, light passes through a rotating quarter-wave plate, which cycles the polarization from linear to left circularly polarized (LCP) to orthogonal linear to right circularly polarized (RCP) and back to the original polarization. Part of the current depends on circular polarization and leads to the difference in the left and right photocurrent peaks, which correspond to LCP and RCP. That difference is around the same magnitude but swapped for OAM +1 (upper curve) and −1 (lower curve). **(b)** The photocurrent component J_c that varies with circular polarization is proportional to the magnitude and helicity of light's OAM. (Adapted from ref. 1.)

OAM +1 and −1. The researchers found the same behavior for all OAM modes that they measured. It turns out that J_c corresponds to the photocurrent response from the phase gradient and grows proportionally with $|m|$, as shown in figure 2b for OAM up to ±4. It results from light transferring its OAM and energy simultaneously to the electrons in the material. Because the light's phase isn't uniform, there is once again a spatial imbalance of excited electrons and a net current either along or perpendicular to the gradient.

To complement the chip-scale detection of OAM, researchers have reduced the production of light with OAM to chip-scale too. Published back to back with Agarwal's paper, the University of Pennsylvania's Liang Feng and his colleagues report producing light with five different OAM values with the same microlaser.[4] In their scheme, light circling in a microscale semiconducting ring was scattered by ridges on the inner surface of the ring such that the light's circling became a helical propagation. Control over the OAM chirality came from two waveguide arms, one pumped in clockwise light and the other, counterclockwise.

Heather M. Hill

References

1. Z. Ji et al., *Science* **368**, 763 (2020).
2. S. Dhara, E. J. Mele, R. Agarwal, *Science* **349**, 726 (2015).
3. Z. Ji et al., *Nat. Mater.* **18**, 955 (2019).
4. Z. Zhang et al., *Science* **368**, 760 (2020). PT

Stuck at home, physicists pivot to combat COVID-19

Consultations with biologists and medical experts bear fruit.

In late February Mahesh Bandi was in Trieste to give lectures at the Abdus Salam International Centre for Theoretical Physics. Northern Italy was becoming a hot spot for COVID-19, and travel to other parts of Europe was restricted; instead of continuing to Paris as he had planned, Bandi returned to his job at the Okinawa Institute of Science and Technology. Before leaving Italy, he gave away the face masks he had brought from Japan, figuring he could get more back home. He was wrong.

Since returning to Japan, Bandi has set aside his usual research—on granular materials, statistical physics of sustainable energy, the stiffness of the human foot, and more—and instead has focused on creating reusable face masks and other personal protective equipment to help medical workers stay safe from infection. He is also assisting an international agency in real-time analysis and modeling of the pandemic's spread dynamics. "It's the first time in my life that I feel all my learning is really helping people," he says.

Scientists around the world have turned their attention to COVID-19. Among the projects physicists are working on are apps for proximity tracing, mobile care deployment, and supply-chain simplification; ventilators from easily sourced parts; numerical simulations of ways to reopen civic life and reboot the economy; and molecular simulations to search for therapeutics and vaccines.

Besides creativity and handiness with electronics and computers, physicists have expertise in managing big data and large teams. And approximating is part of every physicist's training, notes John Langford, a researcher at Microsoft who helped design protocols for contact-tracing apps. "We are in a situation where the fatality rate is unclear, where many things are unclear. If you are comfortable dealing with approximate information, you can start to act now."

PRINCETON OPEN VENTILATION MONITOR COLLABORATION

THE MONITOR ON THIS PRESSURIZED HELMET consists of a sensor (orange box) hooked to the expiratory hose that feeds data about the pressure inside the helmet and air flowing out to an interface box fitted with a customized Raspberry Pi computer. From those data, the pressure exerted by the lungs, the volume of air per breath, and breaths per minute can be determined. The monitor was built by Princeton University scientists in consultation with medical workers to treat COVID-19 patients.

Says Gordon Watts, a particle physicist at the University of Washington, "Everyone is eager to contribute, but we physicists have almost zero training in epidemiology, medical devices, or anything like that. It's important to engage without injecting noise."

Researchers' efforts to help combat COVID-19 have sprung up largely spontaneously. At the same time, physicists have created networks to facilitate interdisciplinary collaborations, spread successful ideas, reduce duplication, and maximize impact. "Right away, it was clear the problem will be with us for a while, it will take time to develop a vaccine, and there is a real risk of a second wave," says Princeton University particle physicist Peter Elmer, who works on the CMS experiment at CERN. "Universities and national labs have international capabilities. If we cannot respond and do

MAXIMILIEN BRICE/CERN

A FACE SHIELD PRODUCTION CENTER AT CERN gets a visit from director general Fabiola Gianotti (left). Francisco Perez Gomez (center) was one of the technicians who initiated CERN's manufacturing of face shields. His brother, Juan Carlos Perez (right), an engineer in the lab's magnet group, started a second production line at a different site. Each line made around 500 face shields a day.

something useful in this environment, who can?"

In March, Elmer and a handful of like-minded colleagues started Science Responds, an informal network to help nonmedical researchers contribute to combating the global pandemic and its fallout. CERN, meanwhile, set up a task force to coordinate ideas and support efforts related to COVID-19. "The task force is acting as a matchmaker so that ideas can be pursued immediately," says task force chair Beniamino Di Girolamo.

From visors to ventilators

Technicians at CERN churned out 15 000 face shields by hand over seven weeks. Most were distributed regionally to health workers, police officers, customs agents, and others who have significant exposure to the public. "That has been the most successful initiative in terms of immediate impact," Di Girolamo says. Lab technicians also produced 10 metric tons of hand sanitizer from a recipe provided by the World Health Organization and used 3D printing to make washable face masks. Commercial manufacturers are putting CERN's mask design into mass production, he says.

Additionally, demand for CERN's emergency vehicles soared. Typically they get called perhaps 10 times a year to take patients to the hospital for incidents unrelated to the lab. For a brief period this spring, during the peak of the COVID-19 outbreak in Geneva, they were used for the local community about 20 times per week.

CERN has also dedicated a large portion of its computational and data resources to pandemic-related projects. For example, the distributed project Folding@home, which simulates protein dynamics to help develop therapeutics, runs on computers at CERN and other labs. And scientists from the UK's Diamond Light Source and elsewhere store data from COVID-19 studies at CERN. (For more about COVID-19 work at major facilities, see PHYSICS TODAY, May 2020, page 22; for a story about physicists modeling the pandemic's epidemiology, see June 2020, page 25.)

CERN scientists designed a ventilator optimized for use in poor countries. It aims to be cheap and easy to manufacture, run on batteries, function at a wide range of temperatures, and directly connect to mobile pressure tanks, since it may be needed in areas that lack medical gas outlets. CERN's is one of many ventilator projects around the world. "They differ in details," says Di Girolamo, "and there is enough need for all of them."

Physicists are also behind the Mechanical Ventilator Milano, which is based on a simple design that uses only a few parts to enable rapid, inexpensive production. Features include push-button measurements of alveolar pressure and the ability to boost oxygenation immediately following intubation. It was developed with the help of anesthesiologists "working in the COVID-19 wards in Italy and abroad," says project spokesperson Cristiano Galbiati, who splits his time between Princeton and Italy's Gran Sasso Science Institute.

While in Milan in March, Galbiati heard through friends about the paucity of ventilators. He thought, "We scientists have the know-how to work with gases and should be able to find a way to put oxygen in the lungs of patients." Hundreds of his colleagues from the DarkSide dark-matter experiment joined the effort. "The project never stopped," says Galbiati. "It was going on in nine different time zones." On 1 May, barely six weeks after its conception, the ventilator received emergency use authorization from the US Food and Drug Administration.

Some other Princeton scientists asked clinicians at the University of Pennsylvania hospital how they could help fight COVID-19. A group that grew to more than 20 then took on the task of creating an inexpensive monitoring and alarm system for a noninvasive ventilator. The ventilator, which consists of a helmet that delivers pressurized air, reduces the amount of time patients need to be intubated on an invasive ventilator. For some patients, the helmet ventilator eliminates the need for intubation entirely. "It's basically a bag over the head," says team coordinator and Princeton physicist Andrew Leifer, "so if something happens, patients can asphyxiate."

Commercial monitoring solutions exist, says Leifer, "but the supply chain is crushed. A critical aspect of our design is that it's super robust to the medical supply chain." Within three weeks of getting started on 1 April, and working largely in their homes, the team members had built a prototype monitor. In six weeks they had produced 40 units. In May the design was tested on human volunteers, and the team is applying for emergency use authorization so it can be deployed in hospitals. The project has also piqued the scientists' interest about lung function. Says Elmer, "We could use the monitors to get data sets to see how patients' lungs evolve, to get a full picture of lungs in COVID-19 patients."

MAHESH BANDI/OKINAWA INSITUTE OF SCIENCE AND TECHNOLOGY

THIS COTTON CANDY MACHINE (diameter 70 cm) was retooled to make a high-quality filter material for face masks. The technique could enable inexpensive, local fabrication, especially in less developed regions of the world.

"Some good in this mess"

Savannah Thais, a postdoc at Princeton, uses the Science Responds network to enlist physicists and data scientists as contributors to projects related to COVID-19. So far, the network has a core of about 20 scientists and more than 200 participants total. "There is tons going on," she says. "It can be hard to parse. We want to be a central information hub and provide opportunities for conversation." An emerging goal, she adds, is to use the experience from particle physics to help other researchers with remote interactions.

When Washington University in St Louis biochemist Greg Bowman of

Folding@home reached out to Science Responds for advice on running a large distributed project, Rice University's Christopher Tunnell was happy to help. "Working in teams of hundreds, even thousands, has become second nature for us, but a lot goes into making it work well," he says, alluding to his research in computational astroparticle physics. Bowman's question "was not a physics question, it was an organizational question," Tunnell says. He shared a document that outlines how XENON, the dark-matter experiment he works on, handles organizational issues, publications, partnerships, and so on.

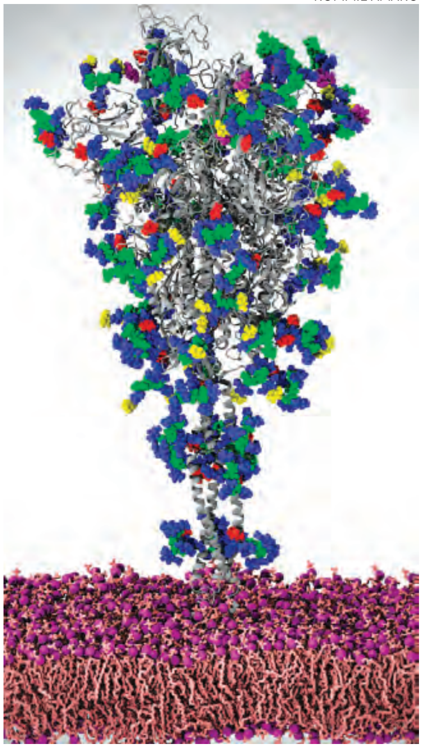
When Tunnell discovered that the pandemic was interfering with students getting summer internships this year, he hired eight of them. "These are top computer science and engineering students, and they would usually go to prestigious tech companies," he says. Some of the international students would have had to leave the country. The hires are small scale, he says, "but at least it's some good in this mess." (See also the story "Pandemic creates hurdles for academic job seekers," PHYSICS TODAY online, 23 April 2020.)

Thais, for her part, is applying machine learning and other techniques to address socioeconomic impacts of COVID-19. On 17 May she and a team she assembled largely through pandemic-related hackathons launched a dashboard to track economic distress, health impacts, and other factors in vulnerable populations. It aims to inform policymakers, nonprofit organizations, and health care providers. (One of Tunnell's computer science interns is collaborating with Thais on that project.)

The novel coronavirus was an easy pivot for Rommie Amaro, a biophysicist at the University of California (UC), San Diego, who uses computational and theoretical methods to understand and design proteins and molecular machines. The virus "is one of the most complicated structures we have come across," she says. She and her colleagues are simulating the spike glycoprotein, which protrudes from the virus's surface.

Computational methods are essential because the protein is hard to crystallize. "We will get the complete structure, which you cannot get experimentally," she says. "Sugars act like a shield over the surface, and we are looking for chinks in that armor." The work, she says, could

ROMMIE AMARO



A SNAPSHOT FROM A SIMULATION OF THE SPIKE PROTEIN on the surface of the novel coronavirus. Rommie Amaro of the University of California, San Diego, and colleagues are using simulations to look for vulnerable points in the protein.

contribute to drug and vaccine design. COVID-19 will ultimately be stopped by a vaccine, says Amaro's UC Berkeley colleague Teresa Head-Gordon, a codirector of the Molecular Sciences Software Institute. For now, she says, "we are building capability to be ready for the inevitable next time a new virus is introduced and to more rapidly explore a greater range of therapeutics."

At the University of Newcastle in Australia, Paul Dastoor, a physicist who heads the Centre for Organic Electronics, is exploiting his work on horse fertility to develop a rapid fluorescence test based on the viral load in saliva. Scientists at the center also jumped into action to make face shields from a flexible plastic substrate they had used for solar panels. "We thought of the idea on a Friday," he says. "On Monday people came into the lab at 6:00am, and two hours later the first prototypes appeared." After getting feedback from hospital workers, the lab built a machine that can make up to a thousand shields per day.

Bandi of the Okinawa Institute of

Science and Technology and Manu Prakash of Stanford University independently dreamt up the idea of using cotton candy machines to make face masks. Polystyrene or polypropylene, the same materials used for N95 masks, is fed into the machine, melts, and flies out through holes. It solidifies into an unwoven fabric that can form the basis for a high-quality mask. Both researchers have other COVID-related projects, such as washable masks, full-face coverings made from off-the-shelf snorkel masks for health workers, and more. They both aim to create protective gear that can be manufactured cheaply in developing countries. "In places where there is no supply chain, it changes the dynamics of what can be made locally," says Prakash.

Trace, test, treat

Some scientists are working on contact-tracing apps to help identify and notify people who may have become infected from being near someone who tests positive for COVID-19. The new apps tend to be variations on the same theme: If two smartphones that come within Bluetooth range remain in proximity long enough—say, 20 minutes—they retain a record of that contact for a limited time, perhaps one month, in the form of an anonymized code. If someone subsequently tests positive for COVID-19 and voluntarily uploads that information, the contact codes retained on their device are used to inform others who had phones in proximity.

Many of the apps under development use Bluetooth only. University of Toronto computational chemist Alán Aspuru-Guzik is working on an app that also uses GPS. The advantage, he says, is that it helps identify infection hot spots. He adds that proximity-tracking apps should be used together with traditional manual contact tracing. The manual intervention is key for areas that don't have good coverage, and for people who don't opt in to contact tracing by app or don't have smartphones.

For app designers, challenges include making sure that proximity with a barrier in between—say for people in adjacent offices—does not generate a contact record, certifying that a reported positive is valid, and attracting sufficient participation for the proximity tracing to be useful. Perhaps most challenging of all, however, is protecting privacy and assuring users that the information will be used only for the stated public health purpose.

One approach to gaining the public's trust is to place the data in nongovernment hands. "People may trust CERN more than their own governments," says CERN computer scientist Mario Lassnig. He and colleagues are considering a proximity tracer just for the world's high-energy-physics labs. For now, CERN is using a system that was repurposed from its original function of keeping track of radioactive equipment. The development of apps is longer term, Lassnig says, "so we have them in place in a second or third wave of the pandemic, and we are not sitting ducks like we are now."

Microsoft's Langford notes that testing, tracing, and isolation are all needed to suppress the disease. "You can trade off," he says, "but it's not fun. You have to isolate more people if you minimize testing."

Toni Feder

Physics-related companies report mostly mild economic symptoms from COVID-19

Most makers of research tools for physical scientists report stable business. Some of their products are being used in the fight against the novel coronavirus.

OCEAN INSIGHT

SINCE THE OUTBREAK OF COVID-19, Ocean Insight has seen increased interest in its spectral systems for the analysis of biological samples and UV disinfection sources. Here the company's spectrometers and accessories are used to measure fluorescence of proteins.

Sitting in his UK Office, Henry Langston began receiving reports as the COVID-19 pandemic progressed from one part of the globe to another. Langston, chief commercial officer of Ocean Insight, says it began with employees being unable to return to the company's Shanghai offices from their New Year's travels. "Then we were sending masks to our workforce in China, and three weeks later, they were shipping masks to Europe and the US," he says. "We've seen different geographies and countries go through all sorts of different levels of lockdowns."

A 250-person company that makes UV, visible, and IR spectrometers, Ocean Insight serves mostly the biomedical diagnostics market. One of its largest customers makes blood analyzers that are used in intensive care units. "A big aha moment for us was understanding that anybody on a ventilator must have their blood checked every one to two hours when they are anesthetized," says Langston. "That made us realize how directly we were on the front lines and how important it was for us to be resilient as a business."

The company has received inquiries about utilizing spectroscopy to analyze UV light used in disinfecting personal protective equipment, Langston says. It's also looking at novel diagnostic approaches such as colorimetric assays and at fluorescent tagging for antibody identification.

Physics Today reached out to a dozen suppliers of research tools used by physical scientists. Most reported weathering the pandemic with, at worst, only a slight downturn in sales. For many, the diversity of their product lines, the multiple markets they serve, and the high-tech nature of their customers have insulated them from the worst effects of the virus. "We are a lucky crowd because the business we are in is inside the physics community," says Jan Benhelm, chief marketing officer of Zurich Instruments. "We are physicists and engineers dealing with other physicists and engineers, and we are all highly educated and motivated and passionate people. It's a very good environment."

"We're lucky with photonics and the world that we're in," echoes Langston. "With a lot of different end-user markets and applications, technology is even more important in times like this."

Applied Physics Technologies, which makes components for electron microscopes, was the only surveyed company to report that its operations were seriously affected by the pandemic. Although it remained open as an essential business, the Oregon company had to lay off or furlough two-fifths of its 50-member workforce, says CEO Marcus Straw. Revenues are down 20% so far this year compared to 2019, when sales totaled \$6.1 million for the year. But Straw says the coronavirus is only partly responsible for the downturn; half is due to a slowing in the semiconductor industry that was underway prior to the pandemic. The deep pockets of Hitachi, which acquired the company last year, have helped to cushion the effects, Straw says. "We're not going to pull back on our growth plans. We think we will start seeing a slow ramping up in our business around September."

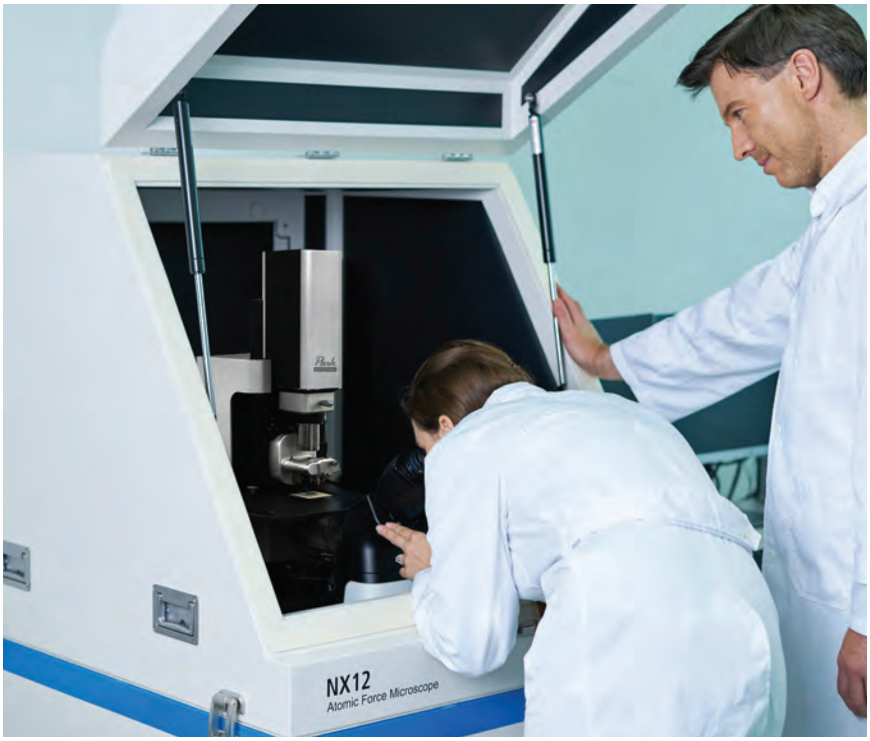
RBD Instruments, a 10-employee enterprise based in Oregon with four product lines—compact Auger electron analyzers, sputter ion guns, picoammeters, and water-vapor desorption systems—has seen declines in one offset by increases in others. "They're kind of averaging out. We're making a sale here and a sale there on everything," says co-owner Randy Dellwo. The company applied for and received a Paycheck Protection Program (PPP) loan from the Small Business Administration. Its products are used mainly by the semiconductor industry and in metal and surface R&D.

Zurich Instruments, a manufacturer of lock-in analyzers, impedance analyzers, and commercial quantum computing control systems, just hired a physicist for its Boston office and is looking to hire more, says Benhelm. About one-quarter of the company's 90 employees worldwide hold PhDs in physics, and most of those work in sales and marketing. The overall business situation is stable, the supply chain hasn't been disrupted, and almost all its customers—universities, national labs, and multinational companies—continue to place orders and take deliveries, he says.

Euclid Techlabs' 20 employees include 18 with PhDs in physics or engineering, says Ilya Ponomarev, chief business development officer. The Maryland company collaborates with SLAC, Brookhaven, and other US national laboratories in developing accelerator components for industrial and medical applications. It's applied for a grant from the Defense Department to develop a process for using accelerators to rapidly disinfect personal protection equipment. Negotiations with several customers, including two in China, have had to be put on hold, as has work on several Small Business Innovation Research awards that required experimental work at national labs. The company has secured a loan through the PPP, Ponomarev says.

Park Systems, headquartered in Suwon, South Korea, has seen sales of its atomic force microscopes (AFMs) continue to grow despite the pandemic, albeit at a slower pace than the 25–30% compounded growth it has seen in recent years, says Keibock Lee, president of Park America. The company's first quarter, ending 31 March, was its best ever. Park has about 300 employees and annual sales of $70 million. AFMs are used in R&D at academic and national labs and for quality assurance in industry. Founder and namesake Sang-il Park is a PhD physicist, and, says Lee, the company's North American sales managers all have PhDs in physics, materials science, or chemistry.

PARK SYSTEMS



ATOMIC FORCE MICROSCOPES such as this NX12 model from Park Systems have been used for research on previous coronaviruses, including the virus that caused the severe acute respiratory syndrome pandemic in 2002.

Lee adds that it's very likely the company's instruments are being used in research to counter severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), because AFMs were vital in probing earlier coronaviruses.

Lumina Power, a 50-employee Massachusetts company, builds power supplies that drive laser diodes and charge capacitors. The firm has seen a 10-fold increase in demand for one of its products, a power supply that drives the UV flashlamps of a hospital room sterilizer made by Xenex. Lumina president Tung Huynh says the company is on track to record $3 million–$4 million from that one product, which would represent 20–25% of its typical annual revenues. That's helped to cushion a big coronavirus-induced falloff in demand from its primary customers, manufacturers of dermatological laser instruments.

Teledyne Hastings makes vacuum gauges, thermal mass flow meters, and mass flow controllers. Early on, the Virginia company was designated by its customers as an essential business, based on its products' end uses in liquid-oxygen cryogenics, flow meters for personal protective equipment, and leak testing. Its flow controllers also are used in research, says director of sales and business development Douglas Baker. Business has held up well in the downturn largely because of large pre-pandemic orders placed by semiconductor manufacturers, he says. He isn't sure whether that demand will last, however. "I'm proud to say we're doing everything we can to help out" in the coronavirus effort, Baker says.

Wavelength Electronics, which counts physicists, chemists, and engineers among its 40 employees, makes electronic controllers for diode and quantum cascade lasers and controllers for thermoelectric and resistive heaters. Its markets range from agriculture to medicine to railroads. A slight decline in overall sales has been accompanied by an increase in orders for a component of immunoassay machines that are often used in vaccine research, says Lisa Mueller, a Wavelength marketing coordinator.

Based in Bozeman, Montana, Wavelength never closed its doors. "We were getting letters before the stay-in-place order that were saying 'we are essential companies and you are essential suppliers so we need you to stay operating,'" Mueller says.

Breathing easier

Sales at OnScale haven't been seriously affected by the coronavirus, says Ian Campbell, CEO of the 35-employee startup. The Silicon Valley company builds engineering simulation platforms based on massively scalable multiphysics solvers—mathematical models that embody physics to tackle problems in engineering. The simulations can run either on local workstations or in the cloud to design items such as sensors, biomedical devices, and 5G components, he says. Since the COVID-19 outbreak, engineers at Intel, Google, Siemens, Philips, and other customers have been running solvers in the cloud from laptops at home, instead of on powerful onsite workstations.

In a collaboration known as Project BreathEasy, OnScale has teamed with Lexma to produce "digital twins" of actual COVID patients' lungs, using computational fluid dynamics and structural analysis of the patients' CT scans and x rays. The goal of the project, now in its test phase, is to maximize ventilator resources by helping doctors decide how patients should be ventilated, the proper ventilator settings, and duration of intubation. OnScale's other coronavirus-related development projects include a partnership with the Institute for Transformative Technologies on a nonventilator breathing support apparatus and microfluidics lab-on-a-chip applications for medical testing and treatment with a separate partner.

Google Gradient Ventures and Intel Capital have been strong backers, but Campbell says that despite being a "COVID-proof" business, OnScale is finding it tough during the pandemic to attract additional venture capital to support its operations.

Toptica Photonics is a Germany-based developer and manufacturer of laser systems for scientific and industrial applications. Products include diode lasers, ultrafast fiber lasers, terahertz systems, and frequency combs. More than one-quarter of its 350 employees have PhDs, the majority in physics.

Mark Tolbert, president of Toptica's US subsidiary, says sales are still growing. "Where I predict the biggest impact for our domestic business is at academic labs," he says. "Universities are having such a hard time even reconsidering reopening." Yet many are placing orders for equipment needed to perform their funded research.

Evolving sales practices

Adapting to a virtual environment has been challenging for Toptica and others. "I have the philosophy that people need people, so I'm expecting a significant downward trend in sales from the loss of in-person interactions at conferences and sales calls," Tolbert says. Before the pandemic, "we would have hundreds of new opportunities coming in from a conference like CLEO that aren't there."

In lieu of face-to-face interactions, Toptica has turned to virtual meetings. The company has experienced a "shockingly high" acceptance rate for its meeting invitations, says Tolbert. Direct virtual interactions, webinars, and remote demos have been successful, but virtual exhibitions have not had similar results. "We go through the dialog, and obviously we can't move forward because nothing's open. But we can get the process started," he says. "We're going to learn one of two things from the pandemic: Either we'll be hugely impacted because we're not getting new business, or that we were overspending on our marketing accounts," he adds.

Before this year, Zurich Instruments typically sent representatives to 80 conferences, trade shows, and user meetings annually. After the American Physical Society's March meeting was canceled on short notice, says Benhelm, "we basically flipped a switch from one day to the next" and rolled out webinars and virtual user meetings. One webinar on transport measurements in nanostructures attracted 300 attendees, "certainly a large crowd by any of our measures," he notes. "This has been a great success and a blast for the team."

Some coronavirus-forced changes to sales and marketing may endure when the pandemic ultimately ends. Campbell, for instance, says it's very likely that some people will continue to work from home indefinitely, a development that would benefit OnScale.

David Kramer PT

Andrew Odlyzko is a professor of mathematics at the University of Minnesota in Minneapolis. He has published in mathematics, computing, and communications and is now concentrating on the interaction of finance and technological innovation.

Isaac Newton and the perils of the *financial South Sea*

Andrew Odlyzko

Despite Newton's general brilliance and his expertise in finance, groupthink led him to plunge into the South Sea Bubble and lose much of his fortune.

Brilliant scientists have been known to do foolish things, but Isaac Newton's financially disastrous moves during the South Sea Bubble of 1720 are a particularly remarkable blunder. When it was founded in 1711, the South Sea Company was primarily a scheme for managing British government debt. Newton was an early investor and profited nicely as the price of South Sea stock rose over the course of the 1710s. However, in 1720 the company's stock experienced one of the most legendary rises and falls in financial history. Newton decided in the early stages of that mania that it was going to end badly and liquidated his stake at a large profit. But the bubble kept inflating, and Newton jumped back in almost at the peak. His experience provides an instructive example of how even brilliant thinkers can go astray in an environment that lends itself to collective delusions as a result of the proliferation of misinformation and disinformation (see figure 1).

The story of Newton's losses in the South Sea Bubble has become one of the most famous in popular finance literature; surveying his losses, Newton allegedly said that he could "calculate the motions of the heavenly bodies, but not the madness of people." But for a long time, only a few pieces of reliable information were available about Newton's investments. The recent discovery of extensive additional documents, many of them in the archives of the Bank of England, provides considerably more detail about Newton's financial travails.[1] Unlike many other anecdotes about famous figures, the colorful story of Newton and the South Sea Bubble was largely correct. In some ways it even understated the extent of his mistakes.

South Sea Bubble basics

The literature on the South Sea Bubble is voluminous,[2] but many of the existing accounts are faulty. For example, some inaccurately claim that the South Sea Company was a fraudulent enterprise from the start, or that it collapsed after the crash of the bubble in the fall of 1720. The truth, as usual, is far more complicated.

JANUSZ PIEŃKOWSKI/ALAMY STOCK PHOTO

Isaac Newton on the British £1 note, issued from 1978 to 1988.

The South Sea Company was established in 1711 to deal with a pressing financial problem. The British government had a large backlog of unpaid bills, largely from contractors supplying the British military during the War of the Spanish Succession. The government offered its creditors South Sea stock, a product similar to shares in a modern corporation. The stock did not promise full repayment of the money creditors were owed, but it did promise them regular payment of interest.

The South Sea Company received the funds from the government to pay that interest to the creditors. The company also held a monopoly on British trade with the west coast of the Americas and part of the east coast of South America—hence the name "South Sea Company." The company profited from the sale of some British goods and, more significantly and more grimly, enslaved Africans. The enterprise enticed investors with the promise that profits from trade in the South Seas would add generously to their interest payments.

During the late 1710s, the South Sea Company was a rather dull operation that simply passed the government's payments to its investors. Scholars don't have solid information about the profitability of its trading activities during that period; the evidence strongly suggests the company's commercial operations lost money in those early years. However, the trade monopoly helped inspire dreams of future riches among the public. Newton, who also owned government bonds and some investments in the Bank of England, was an early investor and added to his stake as time went on. His South Sea investment was initially quite profitable; the prices of financial securities increased as the long period of draining wars ended and peacetime economic activities began to grow.

The economic recovery of the late 1710s inspired a new vision for the South Sea Company. The British government announced that the company would take over most of the British national debt in 1720. The result was the South Sea Bubble. The price of South Sea stock soared through the summer of 1720, and then it underwent a precipitous collapse in September, most likely because investors began to realize their profit expectations were unrealistic. By October, the stock was worth less than a quarter of its peak price (see the box on page 33).

But the collapse was not the end of the South Sea Company. After some financial restructuring, it continued to exist until the middle of the 19th century, almost exclusively as a private agent handling the paying out of government funds to holders of the national debt. Those who bought South Sea stock before mid 1719 and simply held onto it for a few years prospered, as they were rewarded with a roughly 50% capital gain on top of generous dividends. Those returns were enabled in part by the company's financial moves in 1720, which amounted to a Ponzi scheme that rewarded early investors with money from new investors.

Newton's finances

In 1720 Newton was almost 80. His significant scientific achievements were decades in the past, and he was not doing any original research. Even so, he was active. He had left his academic position at Cambridge University to become the warden of the Royal Mint (figure 2) in 1696, where he played an important role in carrying out the great recoinage of the 1690s. In 1699 he became master of the Mint, and he would remain in that post until his death in 1727. He continued to lead the Royal

FIGURE 1. A SATIRICAL 1720 PRINT about the chaos of financial bubbles, including the South Sea Bubble. (Image from the Wellcome Collection, CC BY 4.0.)

Society as its president and was a celebrity that foreign dignitaries visiting London were eager to meet. Both private individuals and the British government called on him for technical advice on various problems, such as finding longitude at sea. His physical and mental decline apparently started after 1720, as Richard Westfall details in his authoritative biography.[3]

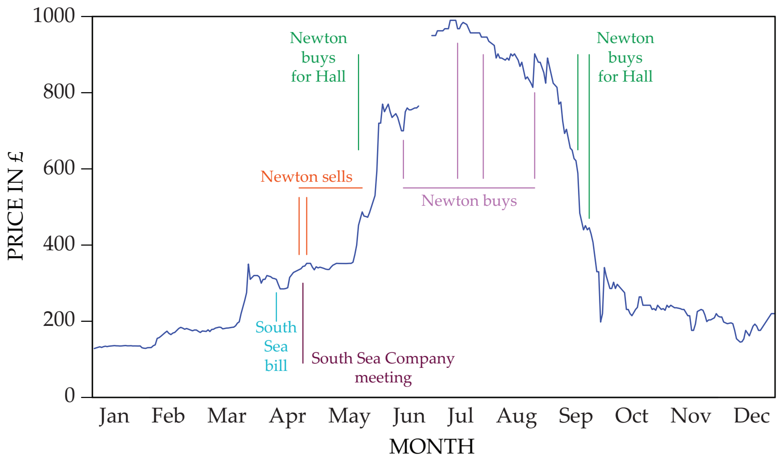
Newton was also wealthy by the standards of the day. Master of the Mint was a lucrative post. His position at Cambridge had paid about £100 per year; by contrast, his earnings at the Mint, including both a salary and a fraction of the Mint's production of coins, averaged close to £2000 per year. By 1720 he was also earning well over £1000 per year from dividends on his investments. His total annual income of more than £3000 put him in the top 1% of the population, and not far from the top 0.1%. He was recognized as a member of the British elite and lived well, with a horse-drawn coach and a retinue of servants. But he did not overspend, was charitable, and saved a substantial fraction of his earnings.

Land ownership was the traditional marker of British wealth and social status. Newton, however, never acquired any significant real estate. He was among the first to put his money mainly into financial instruments, a relatively new possibility at the time. His investments were primarily in government bonds and in securities of large joint stock companies such as the Bank of England and the South Sea Company.

Newton's net worth shortly before the South Sea Bubble started was just over £30 000. That is also the approximate value

SOUTH SEA STOCK PRICES AND INVESTMENTS BY ISAAC NEWTON DURING THE BUBBLE OF 1720

The graph here tracks the price of South Sea stock, with prices adjusted for stock dividends. The horizontal lines denote approximate date ranges for purchases and sales, and vertical lines mark documented dates of actual transactions or instructions for transactions. The labels "Newton buys for Hall" represent some of the purchases of South Sea securities by the Hall estate, of which Newton was an executor. The "South Sea bill" line marks 7 April, the date the main legislation authorizing the South Sea scheme was passed, and the "South Sea Company meeting" line denotes a 21 April gathering of South Sea investors. The gap in prices at the end of June was caused by a change from transactions for immediate cash settlement to what are today called "futures."



of his estate at his death in 1727, so the bubble hardly ruined him financially. However, he did lose a substantial amount. After making an early profit of about £20 000 in the early stages of the bubble, Newton put nearly all his money into the doomed venture. By mid 1721, his net worth was down to about £20 000; he had lost all his early profits and a good bit more besides.

Newton's financial moves during the bubble

Our knowledge of Newton's finances appears to be fairly complete for the period after the crash of the South Sea Bubble. Deducing what Newton did during the crucial period in 1720, however, involves making inferences from incomplete data. The box above summarizes what we know about Newton's investments in the South Sea Company over the course of 1720. At the start of that "fatal year," as it was almost universally called for some time after the crash, slightly less than half of Newton's assets were in South Sea stock; the rest were in government bonds.

The two vertical lines marked "Newton sells" represent two instructions he issued in April 1720 for the sale of his South Sea stock; the horizontal line represents a period during which Newton was acquiring a large holding of British government bonds. Based on his overall wealth, we can assume that the money for that came from the liquidation of his South Sea stock.

It appears that in June of that year, Newton liquidated almost all of his newly acquired stake in government bonds and used the proceeds to buy South Sea stock. That move is shown in the first of the four vertical lines marked "Newton buys"; the other three vertical lines denote documented instances when Newton either bought South Sea stock directly or converted some of his other government bonds into it. We know he continued to pour money into the South Sea stock even as its price was beginning to slide, before the precipitous collapse in September 1720. By that time, essentially his entire fortune was invested in South Sea.

Unfortunately, we have no direct evidence of what motivated Newton's financial decisions. However, we can obtain some additional insights by examining Newton's involvement in the estate of his friend Thomas Hall, another wealthy civil servant who named Newton as one of the executors of his holdings. We have a complete record of the financial transactions of the estate, since the executors had to be able to account for their actions.

Newton did not make decisions for the Hall estate alone; he had to work with three other executors and consult Francis Hall, Thomas's son and the principal beneficiary. Although the estate's decisions do not represent Newton's views alone, he must have been in broad agreement with the group's resolutions, and his own thinking was surely influenced by the discussions that took place. As seen in the box, the Hall estate also invested in South Sea stock. An early purchase, apparently made at the request of Francis Hall, took place just as Newton's purchases of government bonds were ending. The last two purchases occurred as the stock was more than halfway down its slide and show real faith in the South Sea project despite the price collapse.

Newton's early skepticism and later conversion

Let us now consider in more detail the investment scene in April 1720, when Newton made an early decision to sell his stock. The project for the South Sea Company to take over almost all of the British national debt was presented to Parliament at the end of January 1720 and, after heated debate, accepted at the beginning of February. However, the South Sea stock price did not vary much from late March until the end of May. What could have induced Newton to sell out in the middle of that period?

My analysis of the available financial documents strongly suggests that Newton decided to liquidate his entire stake in the South Sea venture in the space of less than one week, on Tuesday and Saturday, 19 and 23 April. Those two days neatly bracket a meeting of South Sea investors on 21 April. We do not know much about what transpired at that meeting, but according to news reports, many people were in attendance for the long discussions and presentations. It could be that the meeting helped strengthen Newton's skepticism and led him to decide to get out. Significantly, Thomas Guy, an eccentric

FIGURE 2. THE TOWER OF LONDON, the location of the Royal Mint during Newton's time. (Photo by N. H. Fischer/Wikimedia Commons/CC BY-SA 4.0.)

investor who became famous for making a massive fortune in the bubble, started liquidating his stake the day after the stockholders meeting.[1] Unlike Newton, Guy did not get seduced into buying back in. He kept his profits and used them to establish Guy's Hospital, famous for its contributions to medicine in the past three centuries.

April 1720 also saw the publication of an unusually large number of pamphlets and newspaper articles about the economic fundamentals of the South Sea project, some of which Newton surely would have seen or discussed with contemporaries. On 14 April, a week after the passing of the South Sea Act, the South Sea Company offered some of its new shares to the public in the first of four sales. To stir up enthusiasm for the first sale, the company's managers apparently arranged for the publication of an anonymous article in the newspaper *Flying-Post* on 9 April. The piece presented a vision for nearly infinite investor returns: the higher the price the South Sea Company could obtain for its new shares, the better its investors would fare.

The *Flying-Post* article reached its astounding conclusion by a certain amount of what today might be called "mathiness," quantitative reasoning that confuses the issue instead of clarifying it. If you own all 500 shares of a company with assets of $100 000, each share is worth $200. If you then sell 2000 new shares at $400 each, your company now has assets of $900 000 and 2500 shares, so each share is worth $360. Modern startups often attract early investors with similar promises. But a legitimate startup will offer some justification for the increased price that new investors pay, such as a new technological breakthrough achieved with funds from early investors. The *Flying-Post* article offered no evidence of any new ingredient that would produce increased profits. In reality, what it was describing was an 18th-century Ponzi scheme.

Three convincing rebuttals of the *Flying-Post* article have survived from that period. One was authored by Archibald Hutcheson, the most famous debunker of the South Sea project. His pamphlet appeared on 21 April, the date of the South Sea Company stockholders meeting. It is possible that Newton was influenced by those refutations, or he may have reached the same conclusion himself.

But the bulk of British investors did not divest. How could people disregard the refutations of the *Flying-Post* puff? A letter from a prominent private banker in London, written in June 1720, nicely illustrates the fear of missing out that investors often experience during a mania: Despite his own long-standing skepticism about the South Sea project, the banker wrote that "when the rest of the world is mad we must imitate them in some measure."[4] Further, even though the math behind the *Flying-Post* claims was risibly fallacious, the South Sea Company's trading monopolies remained potentially profitable.

What is most remarkable is that aside from the *Flying-Post* article, the South Sea managers never presented the public with a business plan explaining how they would generate substantial returns for their shareholders. In fact, I can say with confidence that South Sea promoters lacked an even halfway plausible business plan. The one they were forced to present to Parliament after the bubble burst indicated that they were planning to pay the high dividends they promised mainly out of the funds they expected to obtain from sales of shares—in other words, they had no plan beyond the misleading and overly optimistic scheme their anonymous friend or agent outlined in the *Flying-Post*.[5] But that was not known to the public before the collapse. The mania reached a new level of intensity in June, and Newton was swept up in the groupthink that ruled the British investing public.

Newton as investor

It is noteworthy that once Newton decided to go back into South Sea stock, he moved essentially all his financial assets into it, and he did so rapidly with a series of investments in mid-June. It seems he had become convinced the purchase was a promising venture. In contrast, the Hall estate left much of its assets in Bank of England stock. Newton's banker and deputy at the Mint, John Francis Fauquier, also went down a different path; he did not commit all his funds to the South Sea Company, and he invested on a schedule different from Newton's. Newton was clearly making his own investment decisions and not simply following others' advice.

In general, Newton was intimately familiar with commodities and finance. He had to be, since his job as master of the Mint required him to make many decisions that depended on market prices and conditions. Did he also visit Exchange Alley (see figure 3), where financial dealings were concentrated in the early 18th century? The alley was not regarded as reputable, and reports of visits to it by the aristocracy, especially by aristocratic ladies, scandalized polite society. A claim exists in the literature that Newton was seen in Exchange Alley, based on a letter by the famous London publisher Jacob Tonson that

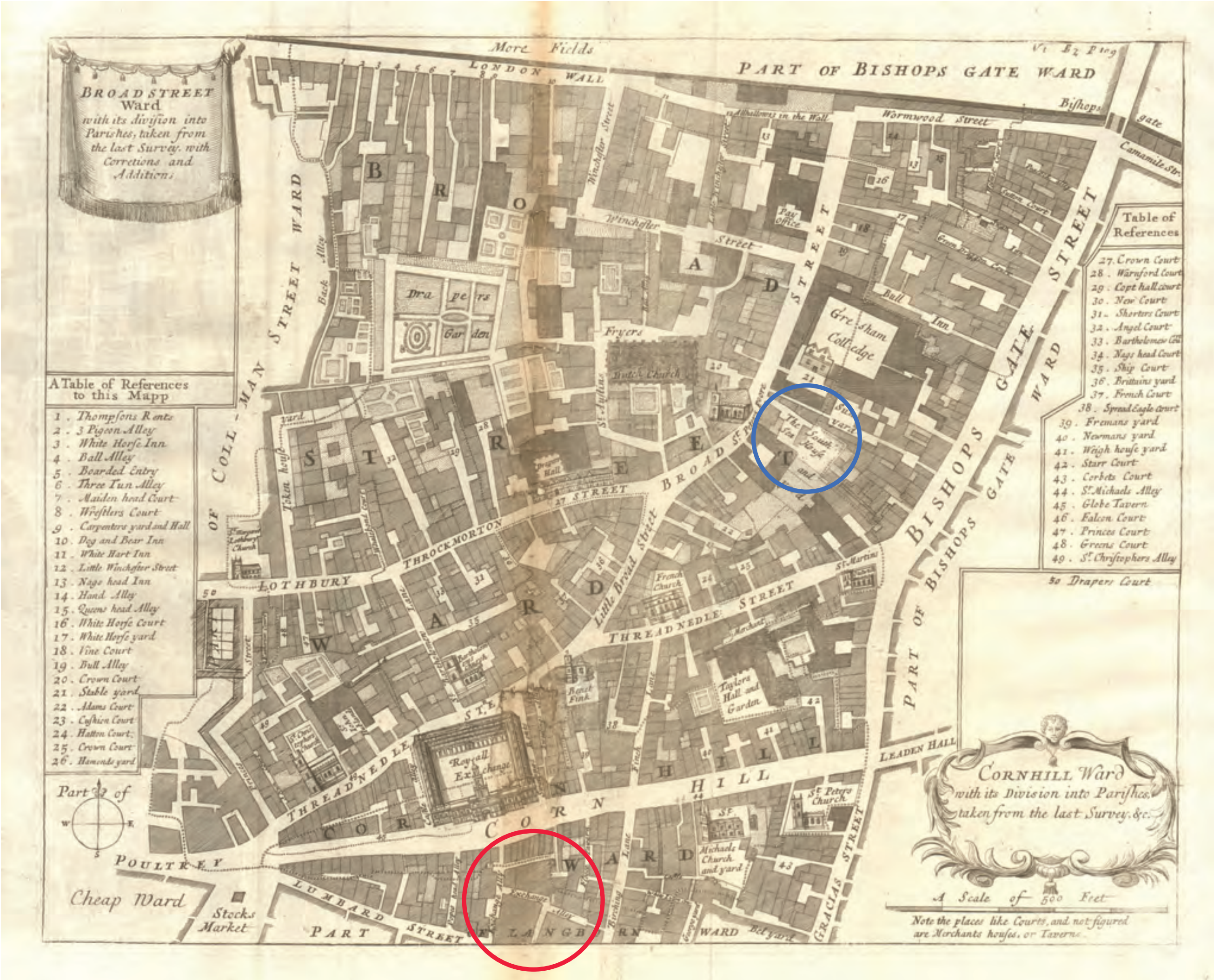


FIGURE 3. A 1720 MAP OF LONDON'S BROAD STREET WARD. Exchange Alley (in the red circle) was the share-trading center of London in 1720. The South Sea House (in the blue circle) was the headquarters of the South Sea Company. (Image from Antiqua Print Gallery/Alamy Stock Photo.)

mentions the possibility of meeting a "Mr. Newton" there.[6] However, it seems likely that Tonson would have referred to such a famous figure by his full appellation, "Sir Isaac Newton." So Tonson probably had another Newton in mind in his letter. (Several affluent Newtons lived in Britain at that time.)

Whether he ever visited Exchange Alley or not, we have evidence of Newton paying attention to the operations of individual companies. In a letter to his friend the mathematician Nicolas Fatio de Duillier, Newton declined to invest in a company that Fatio was promoting. He not only noted the low price of that company's stock, he also wrote that its fundamentals were not good, since he had learned its "rents in Scotland are ill paid & difficultly collected."[7] The letter to Fatio also stated that "I lost very much by the South Sea company which makes my pockets empty, & my mind averse from dealing in these matters." (Newton's statement is an exaggeration, as his pockets were not empty; he was still quite rich. But his recent losses were a handy excuse.)

Thus Newton was clearly financially sophisticated. However, unlike Guy and a small number of other investors, he was not astute enough to avoid the disaster of the South Sea Bubble. In that he was part of an immense crowd—an estimated 80–90% of all British investors—who were drawn into that economic spasm. Unlike astronomy, mathematics, and physics, finance was not an area where Newton towered over his contemporaries.

Newton, finance, and crowd psychology

The number of investors who were taken in by the South Sea Bubble reflected to some extent the relatively low level of financial sophistication in the early modern society. Investors—even experienced ones like Newton—had only recently been introduced to new financial institutions and products. On the other hand, contemporary examples abound of the investing public being bamboozled by specious visions of infinite investment returns. There is little sign that has changed over the ages.

And we should not overestimate the financial naivete of the public in the early 18th century. In Newton's time, professionals were available to assess the value of conventional financial products. Among them was Newton's friend Abraham de Moivre, an important mathematician who was at the forefront of the development of probability theory and actuarial science. De Moivre earned part of his income by providing advice on gambling wagers and on valuations of leases and annuities.[8] In addition, frequent government lotteries spawned a subindustry of intermediaries that enabled people to vary their levels of risk by purchasing a small part of several lottery tickets.

The South Sea Bubble posed a more challenging problem. No well-defined money flows, or even probability distributions, existed. Aside from the Ponzi arguments, which deceived many, all sorts of vague rumors of flourishing trade, strategic alliances, and the like sprang and spread. Some observers, such as Hutcheson, saw they were unrealistic, but most of the public, including Newton, fell for the humbug.

Newton's failure to discern the South Sea Bubble's unsustainability illustrates a fact that is sometimes forgotten when we discuss him: He was brilliant, but he was not a universal genius. He did much to improve the efficiency of the Royal Mint by reducing the variation of gold coins in weight and fineness, for which purpose he formulated the cooling law that bears his name.[9] He aided in setting up the Longitude Prize and he helped dispose of many crackpot submissions to the body in charge of awarding that prize, but he was wrong in his skepticism about the feasibility of using clocks in that setting.[3] And, of course, he spent years on alchemy and theology, with little to show for his efforts.

That Newton was able to achieve so much in astronomy, mathematics, and physics was due not just to his own genius but to his starting that work at just the right moment, when solid foundations had been laid by the likes of Tycho Brahe, René Descartes, Galileo Galilei, Christiaan Huygens, Johannes Kepler, and Blaise Pascal. In those areas he really was able to "stand on the shoulders of giants." In the finance of the South Sea Bubble, he was standing in a swamp, and so even his brilliance did not save him from losses.

This article is based on my paper "Newton's financial misadventures in the South Sea Bubble," Notes and Records: The Royal Society Journal of the History of Science, *volume 73, number 1 (March 2019), page 25.*

REFERENCES

1. A. Odlyzko, *Notes Rec.* **73**, 25 (2019).
2. A. Anderson, *An Historical and Chronological Deduction of the Origin of Commerce, from the Earliest Accounts to the Present Time* […], 2 vols., A. Millar (1764); J. Carswell, *The South Sea Bubble*, rev. ed., Alan Sutton (1993); E. Chancellor, *Devil Take the Hindmost: A History of Financial Speculation*, Farrar, Straus and Giroux (1999); R. Dale, *The First Crash: Lessons from the South Sea Bubble*, Princeton U. Press (2004); Viscount Erleigh (G. R. I. Reading), *The South Sea Bubble*, G. P. Putnam's Sons (1933); W. R. Scott, *The Constitution and Finance of English, Scottish and Irish Joint-Stock Companies to 1720*, 3 vols., Cambridge U. Press (1910–12).
3. R. S. Westfall, *Never at Rest: A Biography of Isaac Newton*, Cambridge U. Press (1980).
4. R. Martin-Holland, *J. Inst. Bankers* **33**, 156 (1912), p. 158.
5. W. Deringer, *Calculated Values: Finance, Politics, and the Quantitative Age*, Harvard U. Press (2018), pp. 215–22; P. Harsin, ed., *John Law: Oeuvres complètes*, 3 vols., Librairie du Recueil Sirey (1934), vol. 3, pp. 231-35.
6. A. E. Murphy, *Richard Cantillon: Entrepreneur and Economist*, Oxford U. Press (1986), p. 191.
7. I. Newton to N. Fatio de Duillier (14 September 1724), auctioned at Christie's on 8 June 2005, www.christies.com/lotfinder/Lot/newton-sir-isaac-1642-1727-autograph-letter-signed-4515605-details.aspx.
8. D. R. Bellhouse, *Abraham de Moivre: Setting the Stage for Classical Probability and Its Applications*, CRC Press/Taylor & Francis (2011).
9. A. Belenkiy, *J. R. Stat. Soc. A* **176**, 481 (2013). PT

ICEoxford receives Queen's Award for Enterprise in Innovation, crowns its best year yet

Chris Busby
Managing Director, ICEoxford

ICEoxford was founded in 2004 to shake up the dry cryogenics market. We have worked closely with scientists at the cutting edge of their discipline to understand their exact experimental requirements, and that approach has paid off.

Our growing team has delivered hundreds of dry and wet systems worldwide to universities, research facilities, and government institutions. We work with all our customers to meet or exceed their specifications within their budgets. Continuous, collaborative innovation has given us an edge to compete globally with bigger, more established cryogenics manufacturers.

Traditionally, R&D companies have rarely emphasized customer service. But by building relationships, we have been able to truly collaborate with our customers and become better at targeting areas for improvement. From our site in West Oxfordshire, UK, we can remotely access and control systems anywhere in the world to provide the highest levels of aftercare.

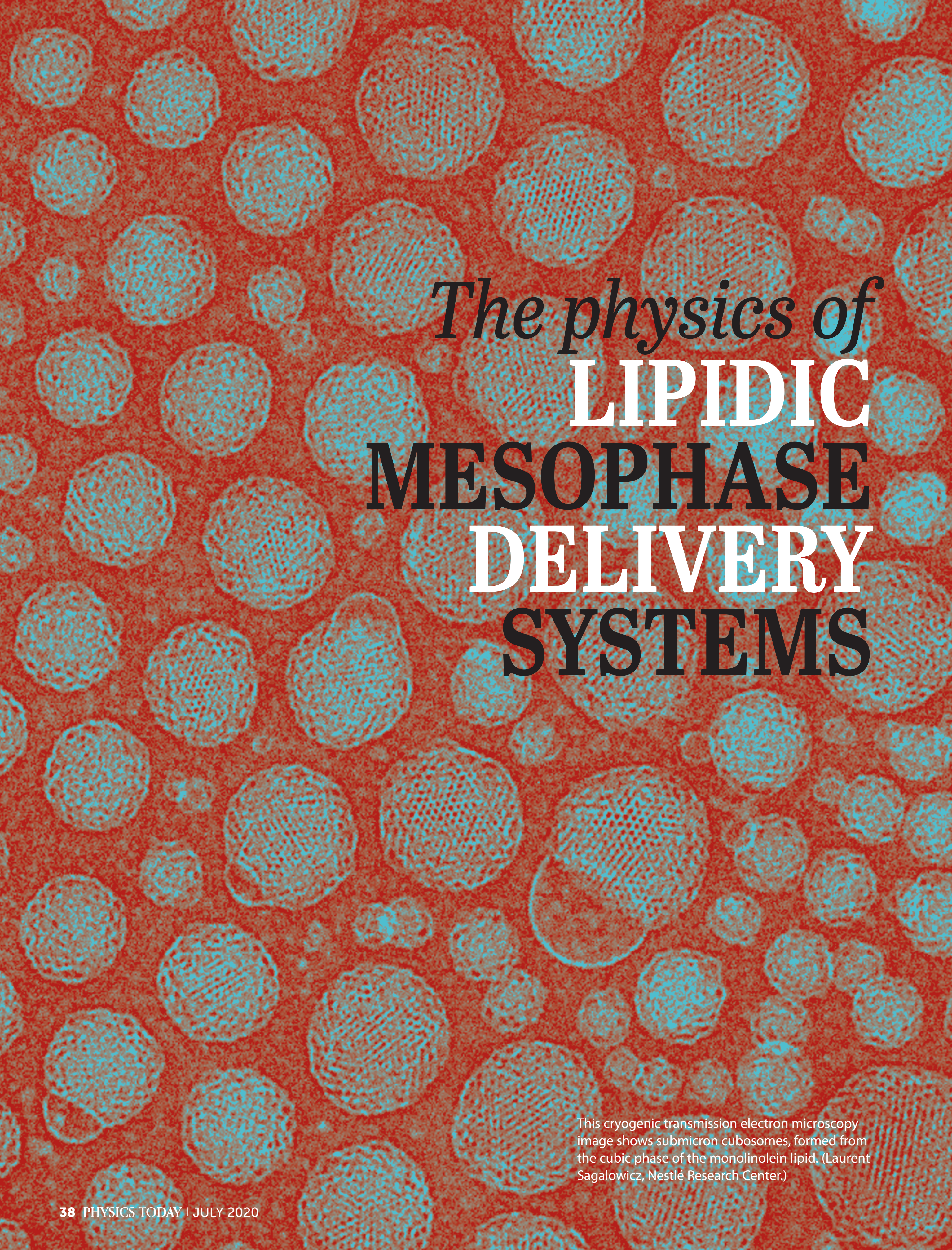
In just 16 years we have developed what is now one of the world's widest ranges of cryogenic products. Our cutting-edge offerings operate at 4 K, 1 K, 300 mK, and 10 mK for both wet and dry systems. Our popular DRY ICE 1.0K Series spans large sample space systems through to compact benchtop designs. The series' newest addition is the DRY ICE 0.8K Benchtop. This cryostat incorporates a large sample-space designed to accommodate a large experimental heat load in a modular benchtop system, perfect for quantum computing and other applications down to temperatures of <0.8 K.

The DRY ICE 1.0K system was initially developed for scientists at MIT to optimize the performance of single photon detectors and manage a large experimental heat load. The platform can reach 0.8 K while achieving greater than 360 mW of power in the sample area at 1.0 K, five times as high as previous technologies. With the addition of an AV frame, the DRY ICE 1.0K reduces vibration to less than ±1 µm at the sample. The system can be further adapted to achieve optical access to the sample within a magnetic field.

Since installing the first DRY ICE 1.0K, we have redesigned the unit to suit the requirements of a US industry customer, a Silicon Valley quantum computing startup venture, and the University of Bristol.

Earlier this year the UK government announced that we had received a Queen's Award for Enterprise in Innovation for DRY ICE 1.0K. The award is the most prestigious for a UK business of any size, throughout every sector. Previous winners have reported a corresponding increase in company growth, and we are aiming to do the same. We significantly exceeded our sales target for the last financial year while growing margins and profitability, and we were thus able to continue to expand our team.

Photonic quantum computing is one of our most rapidly developing markets. We are confident our products will help bring quantum technologies to the marketplace and deliver benefits to business, medicine, health, public administration, and the home. ICE is a dynamic company proud to be at the forefront of this emerging technology.



The physics of LIPIDIC MESOPHASE DELIVERY SYSTEMS

This cryogenic transmission electron microscopy image shows submicron cubosomes, formed from the cubic phase of the monolinolein lipid. (Laurent Sagalowicz, Nestlé Research Center.)

Simone Aleandri is a senior pharmaceutical research scientist at the University of Bern, and **Raffaele Mezzenga** is a professor of food and soft materials at ETH Zürich, both in Switzerland.

Simone Aleandri and Raffaele Mezzenga

Effectively releasing drugs in the body depends on the physical and chemical traits of a special class of liquid crystals.

An ideal drug delivery system should be biodegradable and biocompatible and should incorporate the active pharmaceutical ingredient without losing or altering its activity. To optimize the drug's therapeutic time window, the system should provide an efficient and controlled release mechanism for a drug at a specific location *in vivo*. Lipidic mesophases, a special class of materials belonging to the lyotropic-liquid-crystal family, meet those criteria. They thus share characteristics of both crystals and liquid materials: Like crystals, their molecules have short-range positional order, and like liquids, they lack long-range positional order. Lipidic mesophases exhibit distinct structures based on their composition, and they constitute an attractive alternative drug delivery system, capable of incorporating molecules of different polarity and size and protecting them from chemical and physical degradation.

As figure 1 depicts, when lipid molecules are hydrated, they can self-assemble in different arrangements that correspond to the lamellar, hexagonal, cubic, and micellar cubic phases, among others.[1] The lamellar phase consists of a two-dimensional stack of amphiphilic bilayers separated by aqueous layers. Each bilayer consists of two monolayers packed tail-to-tail to minimize contact between the hydrocarbon chains and water. In that arrangement, water partitions nearly exclusively to the lipid polar heads and forms slabs of water–lipid heads. The lamellar phase is denoted L_α if the lipid tails are disordered in the 2D layers, L_β if they are organized in a lattice. The hexagonal phase consists of an ensemble of cylindrical micelles packed on a hexagonal lattice. The water–lipid heads form the cylinders, whereas the lipid tails fill the continuous matrix.

The bicontinuous cubic phases are composed of two sets of water channels separated by curved bilayers in the 3D space such that every point on the bilayer midplane surface is a saddle point with a zero-mean curvature.[2] Those bilayers encompass a system of aqueous channels and form structured yet flexible networks that are nonbirefringent and optically transparent. Specifically, lipid molecules in bicontinuous cubic phases form a highly curved continuous bilayer that separates the two interpenetrating but nonintersecting aqueous channel networks. The structures can display a double gyroid (Ia3d, with threefold connectivity of aqueous channels), double diamond (Pn3m, with fourfold connectivity), and primitive (Im3m, with sixfold connectivity) symmetry.[3] Finally, the inverse micellar phase comprises spheroidal micelles arranged in Fd3m or Fm3m symmetry, or in a disordered state L_2. Most lipids do not show those phases in pure water and need oil-based components, such as oleic acid or limonene, to form the micellar cubic phase.

The critical packing parameter CPP—defined as the ratio between the hydrophobic chain volume and the hydrophobic chain length in the molten state times the cross-sectional molecular area—can predict qualitatively the structural phase given by each lipid in the presence of water. However, because of the complexity of bicontinuous cubic phase structures, CPP theory cannot fully describe the system.[4] Lipids, such as phytantriol (PHT), monoolein (MO), and monolinolein (MLO), and phospholipids are the most well-known macromolecules capable of forming lyotropic liquid crystals in water via self-assembly. MO and MLO, which are generally recognized as safe for human and animal use by the US Food and Drug Administration, are the most widely used materials for encapsulating a broad range of drugs with various sizes and polarities. Their phase diagrams, shown in figure 2, describe their behavior at thermodynamic equilibrium, where the system is at its lowest state of free energy. The lipid–water mixtures show lyotropic-liquid-crystalline behavior: Changing the water content or the temperature results in a different structure. By increasing the water content, the lamellar phase of MO, for example, transforms first to an Ia3d and then to a Pn3m phase, which in turn will swell until reaching its maximum hydration beyond which Pn3m and water coexist.

On the other hand, increasing the system's temperature induces a transition from the cubic phase to an inverse hexagonal phase.

Once the maximum hydration level has been reached, the Pn3m phase cannot swell anymore. The excess water given to the system remains confined outside the cubic structure, and the Pn3m bicontinuous cubic phase coexists with excess water at thermodynamic equilibrium. The Im3m bicontinuous cubic, hexagonal, and the micellar Fm3m phases show a similar behavior: Excess water coexists thermodynamically with lipidic mesophases.That behavior is a crucial and distinctive feature of lipid mesophases and makes them valuable as drug delivery systems. The thermodynamic equilibrium allows researchers to design nanostructures that remain stable for the entire drug-delivery process, which makes lipid mesophases more advantageous compared to other nanostructured fluids, such as thermodynamically unstable nanoemulsions.[5]

When mesophases coexisting with excess water have reached their maximum hydration level, a dispersion of bulk phases can be formed by adding energy to the system. For sonication, the most commonly used approach to form dispersions, a lipid film is hydrated in the desired buffer containing a stabilizing polymer, and sound waves agitate the film until the sample becomes a homogenous dispersion.[6] The water-dispersed forms of the parent cubic- and hexagonal-phase bulk counterparts are cubosomes and hexosomes. They possess the same nanostructure as the parent bulk phase but exhibit a much lower viscosity, improved fluidity, and larger surface area. Such dispersions are commonly prepared by mixing lipids with a stabilizing polymer that acts as a surfactant and improves the nanostructure's shelf life. With those dispersions, researchers can deliver poorly water-soluble drugs, selectively target malignant cells, and use molecular detection to identify and deliver radioactive drugs for radiation therapies.[7]

Phase characterization techniques

Researchers routinely use small-angle x-ray scattering (SAXS) to determine the lipid phase, and thus construct the phase diagram, for different lipid–water systems. That method relies on constructive interferences in the reciprocal space from many ordered scattering planes that belong to the mesophase. An x-ray beam is directed at the lipid sample, and the resulting scattering pattern gives a characteristic set of rings, or maxima, that correspond to Bragg reflections. Their positions in the reciprocal space depend on the Miller indices of the mesophase scattering planes, and the sequence of Bragg reflections consequently identifies the symmetry of the mesophase studied. SAXS allows for the lattice parameter—the size of the repeat unit cell—to be determined. When the parameter is translated, researchers can reconstruct the entire mesophase in 3D. The lattice parameter and the phase composition enable scientists to determine structural features of the mesophase, such as the width of a lipid bilayer and the diameter of water channels, micelles, and slabs.

Besides SAXS, cryogenic transmission electron microscopy (cryo-TEM) also characterizes lipidic mesophases. Rather than probing reciprocal space, it uses direct measurements of real-space structures. Using cryo-TEM, scientists have obtained high-resolution images of bicontinuous cubic mesophases, which enables direct visualization of their internal channels, size, and other important structural information. Since SAXS probes structure in the reciprocal space and cryo-TEM in real space, taking the Fourier transform of the lipid structures in the cryo-TEM images results in a phase assignment equivalent to that obtained via SAXS.[8]

Rheology of lipidic mesophases

One of the first features to consider when designing drug delivery systems based on lipidic mesophases is the rheological behavior of the mesophases, which varies tremendously. Measurements of the storage moduli G' and the loss moduli G'' as a function of the shear frequency ω reveal important information about the rheological behavior of the mesophases. Different viscoelastic regimes emerge because of the diverse topologies of the mesophases, and each regime's water–lipid interface has its own specific relaxation time. The viscosity of mesophases increases progressively from the less viscous lamellar phase to the viscoelastic hexagonal and the highly elastic bicontinuous cubic phases.[9] Specifically, the lamellar phase behaves as a plastic fluid that has extensive energy dissipation mechanisms due to the parallel slip of the lamellae. The hexagonal phase behaves as a viscoelastic material with a relaxation time of about 0.5–1 s. In contrast, the bicontinuous cubic phases show complex rheological behavior simulated by a multiple Maxwell model. Those phases have characteristic times of up to several tens of seconds that are associated with the relaxation of the

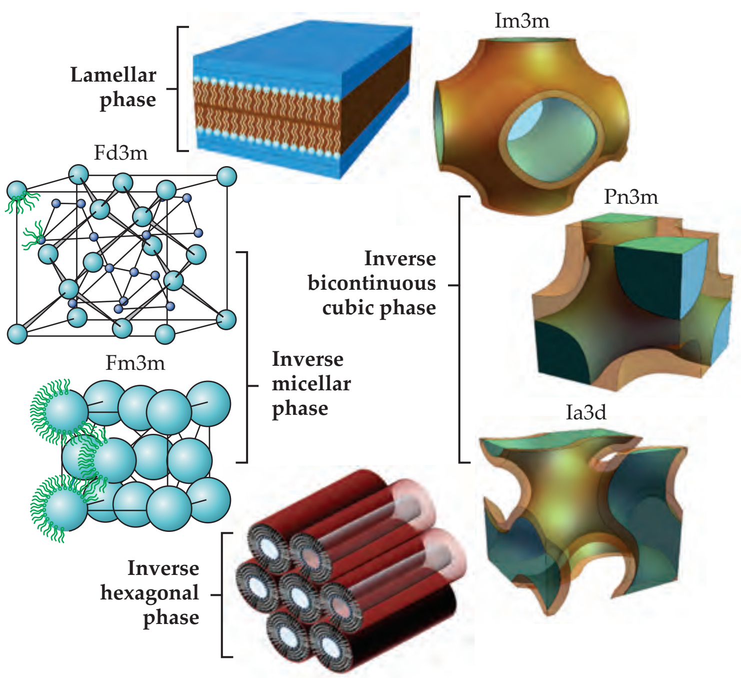


FIGURE 1. THE LIPID MESOPHASE STRUCTURES shown here offer a means to protect and transport drugs through the human body. The inverse bicontinuous cubic phases can display three different symmetries: the double gyroid (Ia3d) has threefold connectivity of the aqueous channels; the double diamond (Pn3m) displays fourfold connectivity; and the primitive (Im3m), sixfold connectivity. The inverse micellar cubic phase comprises spheroidal micelles arranged in Fd3m or Fm3m symmetry. (Adapted from ref. 1.)

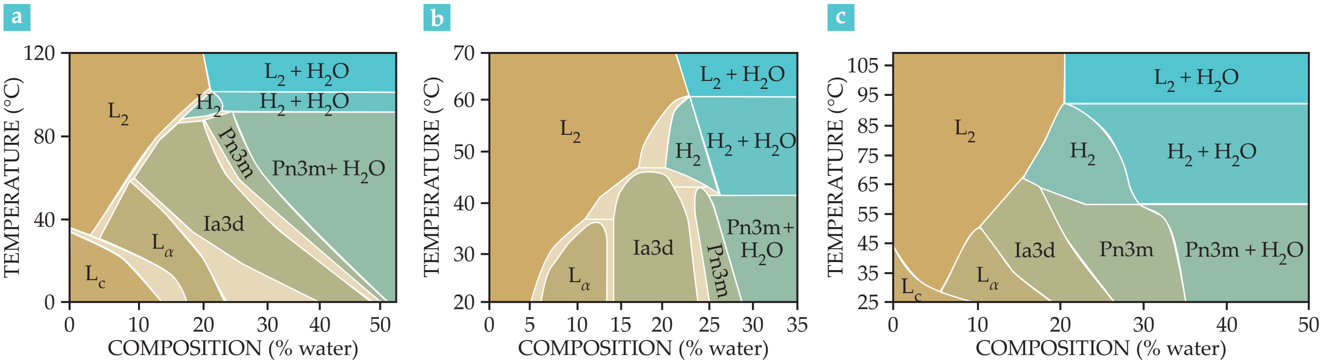


FIGURE 2. THE PHASE DIAGRAMS of the lipids **(a)** monoolein, **(b)** phytantriol, and **(c)** monolinolein show the different structures each can form at thermodynamic equilibrium for different water compositions. Two symmetries are possible for the lamellar phase: the crystalline lipid tail form L_c, also known as L_β, and the amorphous lipid tail form L_α; three for the inverse bicontinous cubic phase (Ia3d, Pn3m, Im3m); and three for the micellar cubic phase—Fd3m and Fm3m if the micelles are organized in a lattice and L_2 if in a disordered state. (Adapted from ref. 4.)

lipid head group–water interface. The variation of the relaxation time as a function of the temperature or composition captures subtle cubic-to-cubic phase transitions and the more trivial cubic and excess-water transition.

Although the viscous cubic phase has an optimal rheology to act as a drug depot, its high viscosity could make the subcutaneous administration of drugs challenging. Water can overcome that challenge by triggering the *in situ* formation of gelled lipidic structures, starting from a liquid or less viscous material.[10] For example, a technology developed by the Swedish company Camurus dissolves the drug in a lipid solution that transforms to gel upon contact with water. After subcutaneous injection, the precursor lipid solution swells from the water naturally present in the tissue and forms the lipid depot *in situ*.

Connecting mesophase structure and drug release

Mesophases can slowly release hydrophobic or hydrophilic biomacromolecules such as proteins, antibodies, RNA, and DNA. But some small hydrophilic drugs encapsulated in the water channels of a mesophase burst out of it. So far, the most commonly explored approaches to control a hydrophilic drug's release from lipidic mesophases include ionic interactions between the drug and the lipid polar group, variation of the water-channel dimensions, and the formation of a lipid prodrug, a compound that activates in the body to produce a working drug.

The release processes of hydrophilic molecules depend on the structure of the mesophase. The release phenomenon—a Fickian diffusion process—follows a first-order kinetic profile regulated by the size of the mesophase's aqueous channels, the size of the drug, and most importantly, the symmetry of the mesophase. To be released, hydrophilic drug molecules must diffuse through the water domains and cross the lipid domains. The ability of the lipids to keep a drug confined in the water channels controls its release. Drug diffusion depends on the dimensionality d of the water domains, defined as the number of orthogonal directions a molecule may explore without leaving the water channel.

In the micellar cubic mesophase, the drug is confined in the micelles with water domains of about 2 nm in diameter ($d=0$). In a hexagonal phase, the molecules move through cylindrical water channels ($d=1$), whereas in a lamellar phase they can move along the planar bilayers ($d=2$). In bicontinuous cubic phases, the diffusion is a 3D process ($d=3$) in which the mobility of embedded hydrophilic molecules is hindered only by the channel size—about 5 nm in diameter—and its tortuosity. In general, the diffusion coefficient of a hydrophilic drug decreases with the dimensionality of the phase. Therefore, the Pn3m phases ($d=3$) show the highest diffusion coefficients, whereas the micellar cubic phases ($d=0$) show the lowest, as illustrated in figures 3a and 3b. That the release rate of drugs from micellar cubic phases is low but nonzero highlights that partitioning of the drug within the lipid bilayers contributes to the release kinetics.

Researchers in 2015 introduced the structural control efficiency index (SCEI).[11] It provides an estimate of the kinetics of drug release for various phases and hydrophobic domains, which have different symmetries and permeability to drugs. The $\mathrm{SCEI} = \log(D_{max}/D_{min})/DS$, where D_{max} and D_{min} are the highest and lowest diffusion coefficients observed for a given system, and DS is the structural dimension span explored; $DS=3$ for MLO-based systems, whereas $DS=2$ for phosphatidylcholine-based systems. Figure 3c graphs the slope in the logarithmic plot of the diffusion coefficients versus the structural dimension and visualizes the SCEI. The hydrophobic lipid domain shows high permeability to a hydrophilic drug such as caffeine, which easily penetrates lipidic barriers, as indicated by the partition coefficient. On the other hand, a drug with high hydrophilicity and large size has a highly negative partitioning coefficient, the lowest diffusion coefficient, and optimal structural control.

For a given drug–lipid pair with the same dimensionality, topology controls the diffusion process. Diffusing molecules, for example, interact with the geometrical attributes of the three main symmetries of inverse bicontinuous cubic phases—Pn3m, Ia3d, and Im3m—which make a pivotal contribution in determining the diffusion mechanism. For small hydrophilic molecules such as glucose, at identical water content the diffusion coefficient decreases from a maximum to a minimum when going from Ia3d to Pn3m to Im3m. As a general rule, the less porous Im3m symmetry shows a slower transport efficiency at water content parity, compared with the more porous Pn3m geometry. Moreover, mesophase symmetry, lipid characteristics, and particle features must all be considered when developing

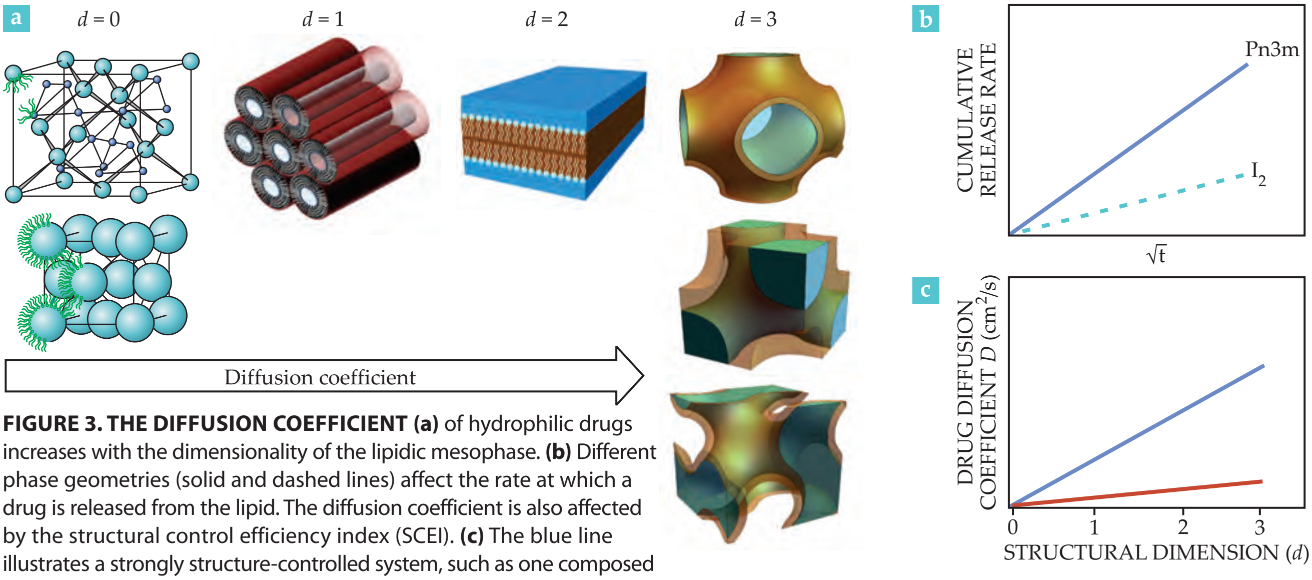


FIGURE 3. THE DIFFUSION COEFFICIENT (a) of hydrophilic drugs increases with the dimensionality of the lipidic mesophase. **(b)** Different phase geometries (solid and dashed lines) affect the rate at which a drug is released from the lipid. The diffusion coefficient is also affected by the structural control efficiency index (SCEI). **(c)** The blue line illustrates a strongly structure-controlled system, such as one composed of monolinolein (MLO), hexadecane, and glucose or a system of MLO, tocopherol, and glucose. The SCEI is about 1, which means on average the diffusion coefficient increases by an order of magnitude per additional dimension. On the other hand, for weakly structure-controlled systems, such as phosphocholine-limonene-caffeine, illustrated by the red line, the SCEI is close to zero. In that case, the hydrophobic lipid domain shows high permeability to the hydrophilic drug. (Figure by S. Aleandri and R. Mezzenga.)

systems with optimized release behavior.[12] In contrast and somewhat surprisingly, the water-channel connectivity does not influence diffusion, which is a result of the properties of the random walk in a lattice.[12]

Tuning with additives

Additives can modify the structure of the lipid mesophases. For example, adding an increasing amount of hexadecane or vitamin A can tune the phase of the MO–water system. The release profile, and therefore the drug diffusion coefficient, changes with the phase. In particular, the release from the cubic phase is significantly faster than the release from the micellar, lamellar, and hexagonal phases. The geometry of self-assembled mesophases is important in determining the release rate and confirms the hypothesis that the open or closed state of the aqueous channels influences the rate of drug release.

Typical lipidic mesophases with a Pn3m, Im3m, or Ia3d symmetry are characterized by water channels with a diameter of 3–5 nm. That geometric constraint prevents large hydrophilic molecules, such as hydrophilic proteins, hormones, and antibodies, from being included in the mesophase. But that major structural limitation can be overcome by additives that increase the water-channel dimensions, including hydration-modulating agents such as sucrose stearate, phospholipids, and cholesterol. Electrostatic swelling also increases the size of water channels, either by doping lipids with charged lipids capable of swelling the mesophases or by decreasing the chemical potential of water in the nanochannels so that the total free energy of the system requires more water to equilibrate. Electrostatic swelling has been demonstrated by several research groups,[13–15] and in those studies, the cubic phases have water channels more than five times as large as traditional lipidic mesophases.

Interestingly, researchers use a combination of neutral lipids such as monopalmitolein (MP) and charged ones such as one called DSPG to design ultraswollen cubic phases of double gyroid (Ia3d), double diamond (Pn3m), and double primitive (Im3m) structures. As shown in figure 4, the channels are sufficiently large to allow in-meso crystallization of membrane proteins with gigantic extracellular domains.[15] In such a case, the increasingly hydrated mesophases reenter the phase diagrams of the type Ia3d → Pn3m → Ia3d and Pn3m → Im3m → Pn3m. The reentrances depend on the amount of phospholipid used, which is unique in lipid self-assembly. An Ia3d phase normally converts to a Pn3m phase upon hydration and will swell until reaching its maximum hydration; similarly, Pn3m swells upon hydration and converts into an Im3m symmetry, which can swell further before reaching its maximum hydration.

The phase reentrances indicate the competing effects at play when electrostatic swelling and increasing hydration occur together. One possible scenario is that increasing the length scales upon swelling by raising the water content decreases the effectiveness of the electrostatic interactions, which moderates the swelling effect at extreme swelling ratios. That hypothesis requires validation, and the reentrances of the phase diagram for electrostatically swollen lipidic mesophases highlight how multicomponent lipidic mesophases continue to reveal fascinating and unexpected surprises. Electrostatic swelling of the mesophase via the addition of charged lipids is a promising tool to generate thermodynamically stable swollen cubic phases. With that approach, one could load large hydrophilic biomacromolecules: The resulting lipidic surface would be decorated by a negatively charged lipid and increase the gel's adhesion with positively charged inflamed tissues.

Adding even low concentrations of electrolytes to the aqueous phase of water–amphiphile systems can significantly modify the corresponding phase diagrams. MO–water mixtures, depending on the type and amount of electrolyte, swell and shrink the water channels. Sodium chloride or sodium sulfate dehydrates the monoolein head groups at the interface, which reduces their occupied effective area; the result is a shrinking

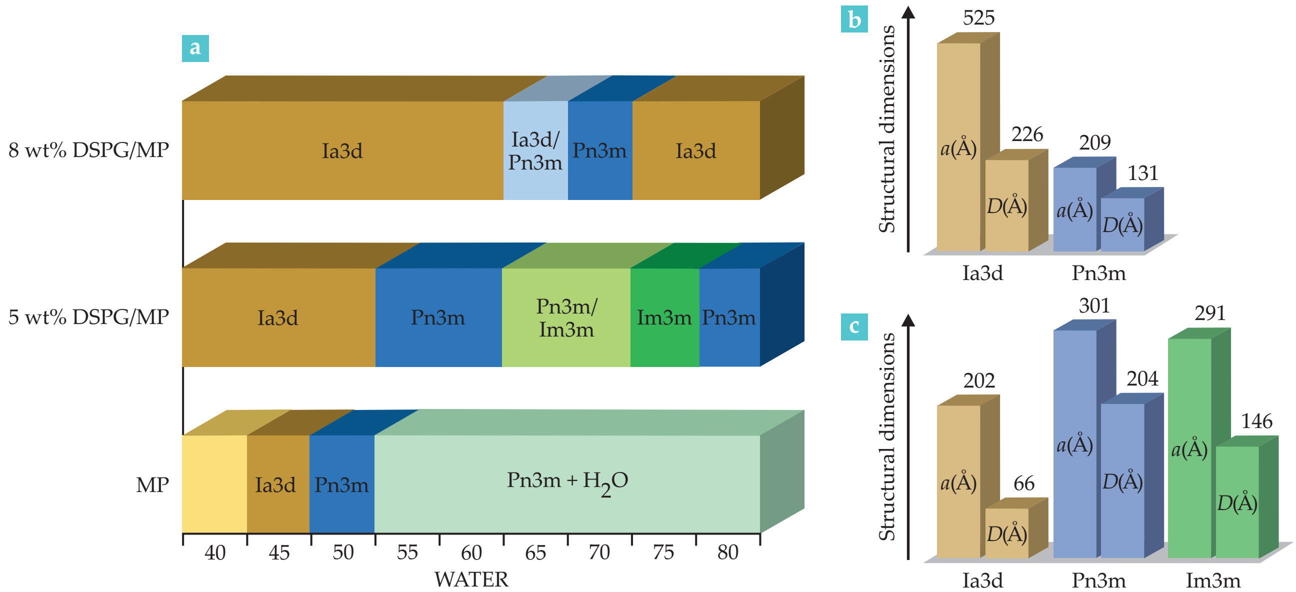


FIGURE 4. SWOLLEN LIPIDIC MESOPHASES transition from one structural phase to another with an increasing weight percent of water. **(a)** The bar graph shows how different combinations of a neutral lipid—monopalmitolein (MP)—and a charged lipid known as DSPG produce a unique range of mesophase structures. **(b, c)** The graphs on the right show how different percentages of DSPG and water swelling affect the structural dimension. The letters *a* and *D* stand for lattice parameter and diameter of aqueous channel, respectively. (Adapted from ref. 15.)

of the water channels. On the other hand, sodium iodide leads to an increase in the hydration of the lipid head groups, expands the interfacial area occupied by a monoolein head group, and swells the water channels of the cubic phase. Low concentrations of Ca^{2+} can shrink the lipidic mesophase water channels, and if the Ca^{2+} concentration increases up to 30 millimolar, it can transform the multilamellar vesicles—formed by MO and a negatively charged lipid such as DOPG—into a cubic phase.[10]

As depicted in figure 3, a drug's release can also be controlled by including additives that render the delivery system sensitive to internal stimuli—including pH and temperature—and external stimuli, such as light and magnetic fields. With appropriate molecular design, the delivery of therapeutic agents can be spatially and temporally controlled by triggered drug release. Such on–off switching is advantageous when continuous drug release may be toxic, and it should improve the therapy's effectiveness. Mesophases can host stimuli-responsive lipids without impairing structure and stability and can endow them with specific functionalities.

For example, figure 5a indicates that adding 3% by weight of vitamin A to the MO–water system leads to an efficient temperature-induced release of hydrophilic molecules. The system composed of MO, vitamin A, and water at 30 °C shows a Pn3m phase and a faster drug release[16] compared with the hexagonal phase obtained at 40 °C. The mesophases composed of monolinolein–water doped with pyridinylmethyl–linoleate show a hexagonal-to-cubic transition induced by the slightly acidic pH typical of cancerous tissues. Figure 5b indicates that for a pH of 5.5 versus 7.4, the phase transition shows a faster release, which means that the hydrophilic anticancer drug doxorubicin, loaded into the engineered mesophase, has an increased efficiency for killing cancer cells at low pH.[16]

With respect to external stimuli, the application of light offers high spatial and temporal precision. Modern lasers modulate drug release profiles and provide a high level of pharmacological control by regulating the light's wavelength, intensity, duration, and spot size. Light, therefore, can directly affect the nanocarrier's assembly and thereby lead to drug release. As figure 5c illustrates, a light-responsive mesophase composed of a mixture of monoolein and a small amount of a light-sensitive lipid produces a sequential light-triggered release when irradiated with UV light and retains an embedded hydrophilic dye when exposed to visible light.[17]

The outlook for lipidic mesophases

To date, inverse bicontinuous cubic and hexagonal phases have both been shown to hold a wide range of active pharmaceutical ingredients. Because researchers can develop and design new lipid molecules capable of self-assembling into lipidic mesophases, the use of mesophases will grow quickly in the near future. Although lipidic mesophases are an interesting drug delivery system, they haven't been as extensively commercialized as analogous products, such as liposomal delivery systems, emulsions, and nanosuspensions. To bring lipidic mesophase products to market will require a clinical translation process, and research has been more focused on gaining information about the physics of these systems. The pharmacokinetics and pharmacodynamics in the biological environment remain less advanced, and researchers are still uncertain how the kinetics of lipidic mesophase formation in the body affects tissue tolerability.

The Camurus company is developing a new technology using a lipid solution, containing the dissolved drug, that transforms into lipidic mesophases upon contact with water. The Camurus pipeline contains different liquid crystalline systems at various stages of clinical trials. For example, their FluidCrystal technology includes injection depot, topical bioadhesive, and nanoparticle formulations based on lipid self-assemblies. Once duly engineered, lipidic mesophases could outperform the more commonly used micelles, emulsions, and liposome

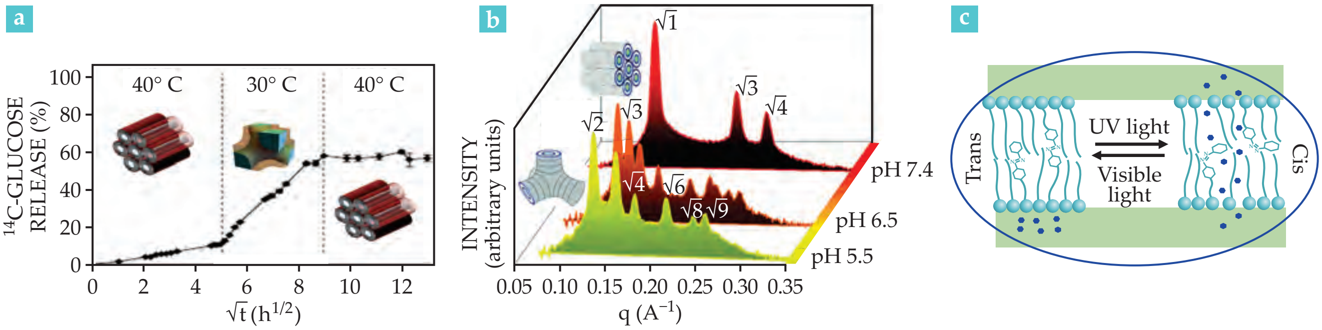


FIGURE 5. ENVIRONMENTAL STIMULI can trigger a lipidic mesophase to release the drug cargo it's carrying. **(a)** Compared with a lipid mesophase system at 40 °C, the system at 30 °C that includes vitamin A releases glucose at a faster rate. (Adapted from ref. 10.) **(b)** Small-angle x-ray scattering measurements show a phase transition with the change to a lower pH. (Adapted from ref. 16.) **(c)** UV light causes a small molecule embedded in a lipid bilayer to switch from a *trans* to a *cis* configuration, which makes the lipidic mesophase less packed. That increases the release of its cargo (solid blue circles) whereas visible light induces the opposite effect, and the molecules diffuse at a slower rate. (Adapted from ref. 17.)

systems in terms of local and systemic tolerability, encapsulation efficiency, physical instability, manufacturing and excipient costs, and more importantly, sustained-release control.

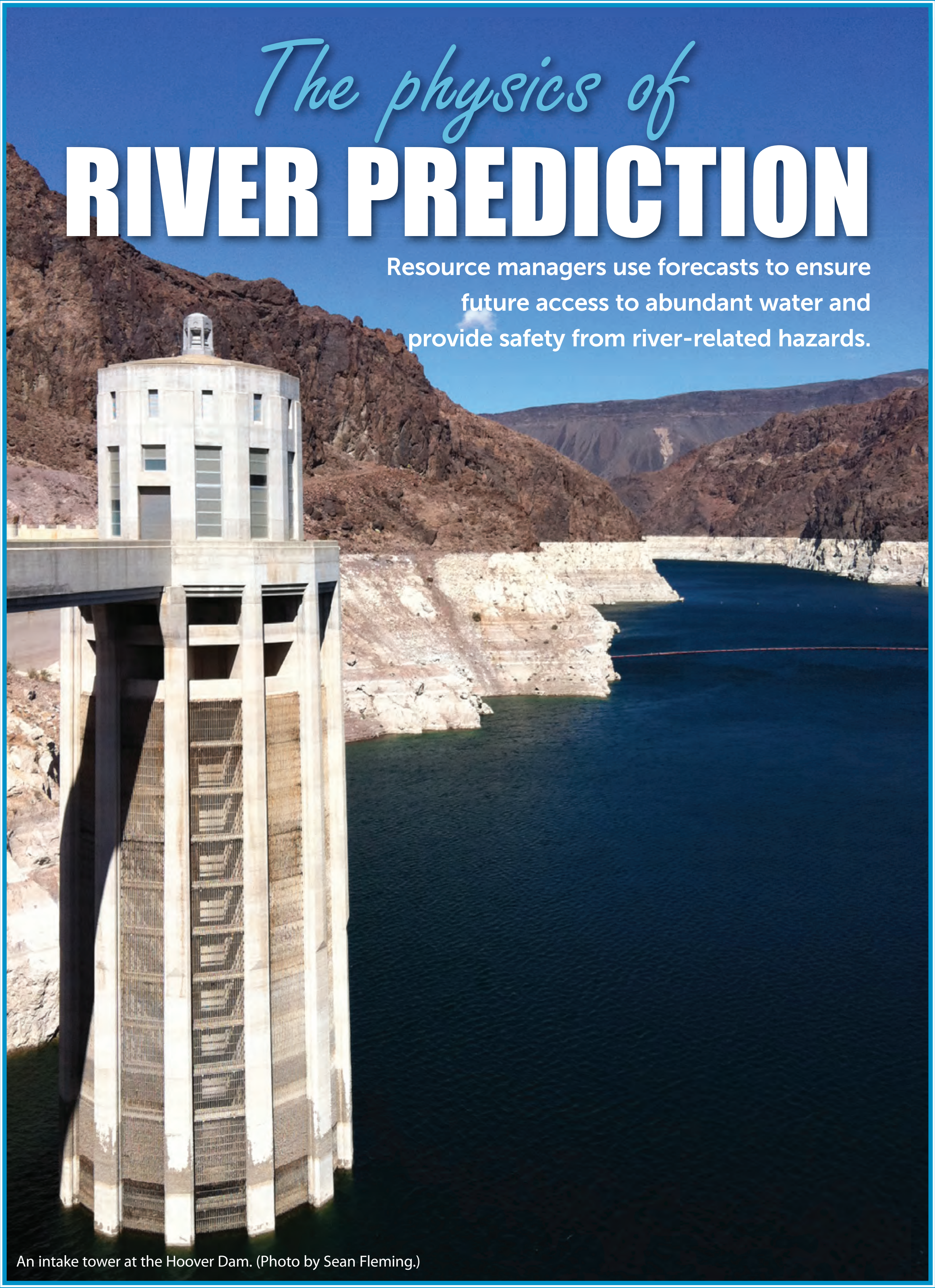
REFERENCES

1. R. Mezzenga et al., *Adv. Mater.* **31**, 1900818 (2019).
2. H. Qiu, M. Caffrey, *Biomaterials* **21**, 223 (2000).
3. V. Luzzati, F. Husson, *J. Cell Biol.* **12**, 207 (1962).
4. L. Sagalowicz et al., *Trends Food Sci. Technol.* **17**, 204 (2006).
5. C. Fong, T. Le, C. J. Drummond, *Chem. Soc. Rev.* **41**, 1297 (2012).
6. H. M. G. Barriga, M. N. Holme, M. M. Stevens, *Angew. Chem. Int. Ed. Engl.* **58**, 2958 (2019).
7. S. Aleandri et al., *Langmuir* **31**, 12770 (2015).
8. D. Demurtas et al., *Nat. Commun.* **6**, 8915 (2015).
9. R. Mezzenga et al., *Langmuir* **21**, 3322 (2005).
10. W.-K. Fong, T. Hanley, B. J. Boyd, *J. Control. Release* **135**, 218 (2009).
11. I. Martiel et al., *J. Control. Release* **204**, 78 (2015).
12. S. Assenza, R. Mezzenga, *J. Chem. Phys.* **148**, 054902 (2018).
13. A. I. I. Tyler et al., *Soft Matter* **11**, 3279 (2015).
14. H. Kim, Z. Song, C. Leal, *Proc. Natl. Acad. Sci. USA* **114**, 10834 (2017).
15. A. Zabara et al., *Nat. Commun.* **9**, 544 (2018).
16. R. Negrini et al., *Chem. Commun.* **51**, 6671 (2015).
17. S. Aleandri et al., *Langmuir* **31**, 6981 (2015).

PT

The physics of RIVER PREDICTION

Resource managers use forecasts to ensure future access to abundant water and provide safety from river-related hazards.



An intake tower at the Hoover Dam. (Photo by Sean Fleming.)

Sean Fleming is an applied R&D technical lead at the US Department of Agriculture's Natural Resources Conservation Service in Portland, Oregon, and an affiliate faculty member at Oregon State University in Corvallis and the University of British Columbia in Vancouver. **Hoshin Gupta** is a Regents' professor in the department of hydrology and atmospheric sciences at the University of Arizona in Tucson.

Sean W. Fleming and Hoshin V. Gupta

Rivers support life and fuel civilization.[1,2] They provide water for drinking, irrigate food crops, and help build everything from cars to computers. Their waters drive hydroelectric turbines that generate clean energy. Rivers have even supported nuclear physics developments that changed the course of a war: The hydroelectric complexes of the Columbia Basin Project and the Tennessee Valley Authority enabled energy-intensive uranium and plutonium refinement for the Manhattan Project.

Rivers have always been crucial transportation pathways. The exploration, settlement, and economic development of the Americas depended acutely on river navigation. The Danube serves as a trade route in Europe, much as it did for the Romans 2000 years ago, and today it carries commercial freight across the continent.

Rivers also provide homes for precious ecosystems. They house fisheries, facilitate recreation, and bring in tourism dollars. That's not to mention their tremendous cultural value: American literature wouldn't be the same without Mark Twain's recollections of his experiences as a Mississippi riverboat pilot.

Rivers can also kill. Floods are the most devastating natural force in the US, and at times they have been architects of history. In 1948, the Columbia River flooded and wiped out the progressive Portland suburb of Vanport, Oregon. Fifteen people died, and the destruction permanently changed the area's racial dynamics. The flood also motivated the US and Canada to negotiate the Columbia River Treaty, intended in part to support flood control efforts. Droughts are subtler than floods but are considered more damaging globally.

Although the water wars predicted in 1995 by then World Bank vice president Ismail Serageldin are unlikely to materialize,[3] competition over scarce water resources has occasionally led to violence. For example, Israeli and Syrian forces skirmished in the mid 1960s over water resources in the Jordan River basin.

Predicting variability and long-term changes in river flows, like those shown in figure 1, is crucial for optimally managing water resources and sidestepping danger. Such hydrologic forecasts are made on a range of time scales and for many purposes. Flood prediction is done hours to days ahead of an event and is used to make emergency management decisions. Seasonal water supply forecasts are needed months in advance to inform the vast water and power management infrastructure in the mostly dry western US. For long-term assessment of the effects of land use and climate change, planning happens over decades or even generations. Successful prognostication depends on a cross-disciplinary set of quantitative modeling approaches.

How do watershed hydrology models work, and what physical principles do they embody? Each model can include contributions from numerous disciplines: civil engineering, geophysics, agricultural engineering, meteorology, climate science, glaciology, and others. Given that complexity, how do hydrologists pick which systems and processes to include in their models? How do they choose appropriate representations and implement them effectively? What, in short, is the physics behind river-prediction models?

What's under the hood?

River prediction models are usually implemented at the watershed scale. A watershed, also known as a catchment or river basin, is the entire upstream land area that drains to a certain point on a river. It's typically determined by topography, such as a mountain ridge separating one watershed from the next; a large example of such a ridge is the continental divide between the Columbia and Mississippi River basins.

Many geophysical and biophysical processes, including snow accumulation and melt, rainfall infiltration, groundwater flow, evaporation, and transpiration by plants make up a watershed's hydrology. By accounting for those processes, a model reproduces and predicts water flux dynamics throughout the basin and, ultimately, at a point of interest on the river. Often that point is chosen to coincide with a long-term river-flow measurement location, called a streamgage or hydrometric station, that collects observational data needed to build and test the model. The point of interest can also be chosen to help answer a practical question, such as whether a site would be appropriate for a power plant that requires water for cooling.

A model's output is normally a time series of the river's average flow rate, often in cubic meters per second, at one or more

FIGURE 1. RIVERS AND OTHER WATER RESOURCES evolve in response to natural and anthropogenic factors. **(a)** The advanced society that lived in Ancestral Puebloan (formerly known as Anasazi) ruins near Los Alamos, New Mexico, moved on around 1200 AD, in part because of the drying climate. **(b)** Nevada's Walker Lake is a remnant of ice-age Lake Lahontan; its water level has dropped 55 m since the 19th century, mainly from diversions for agriculture. **(c)** The McKenzie River springs from a volcanic aquifer at Tamolitch Falls in the Oregon Cascades in a dramatic demonstration of river–aquifer interactions. Accurately capturing the water-storage effects can be crucial for river-flow forecasting. **(d)** The Los Angeles River is a favorite filming location for car chases in Hollywood movies. Such highly urbanized rivers no longer have the natural water storage and release mechanisms seen in the McKenzie River. They are therefore more variable and flood-prone, which makes accurate forecasting even more crucial—and more challenging. (Photos by Sean Fleming.)

points of interest. The most common time-averaging frequency is daily, but it can range from subhourly to yearly. A model may also generate additional data, such as estimates of soil moisture or snowpack.

Virtual watersheds

Several approaches to hydrologic modeling are outlined in figure 2. The most explicitly physics-based option is process simulation, which relies on direct mathematical representations of water movement over and through Earth. That movement can be described in varying levels of physical detail.

The Navier–Stokes and continuity equations fundamentally govern fluid transport. To our knowledge, the complete nonlinear equations have never been directly implemented for watershed hydrologic modeling. And for good reason: They're notoriously difficult to solve numerically, which makes them impractical to implement at the spatial and temporal resolutions and scales typically required. However, with exascale computing on the horizon, applying the full equations may be an idea worth exploring.[4]

In 1969 Allan Freeze and Richard Harlan put forward what is widely considered to be the gold standard in hydrologic modeling.[5] Derived from the full Navier–Stokes equations under simplifying assumptions, their physics-oriented approach depicts and predicts nature through a system of coupled partial differential equations that represent water fluxes through landscape elements. (See box 1 for a sketch of their approach.) The approach has only occasionally been used for practical applications. It is receiving renewed interest, however, particularly from US Department of Energy laboratories that are developing multiphysics environmental models to run on high-performance computers.[6,7]

A less detailed yet widely used hydrologic model approximates water flow using the conceptual linear reservoir assumption. Under that assumption, a watershed's natural water-storage mechanisms—lakes, wetlands, soil moisture, aquifers, and so forth—act as a *de facto* reservoir whose output rate depends on how full it is. Combining that with continuity

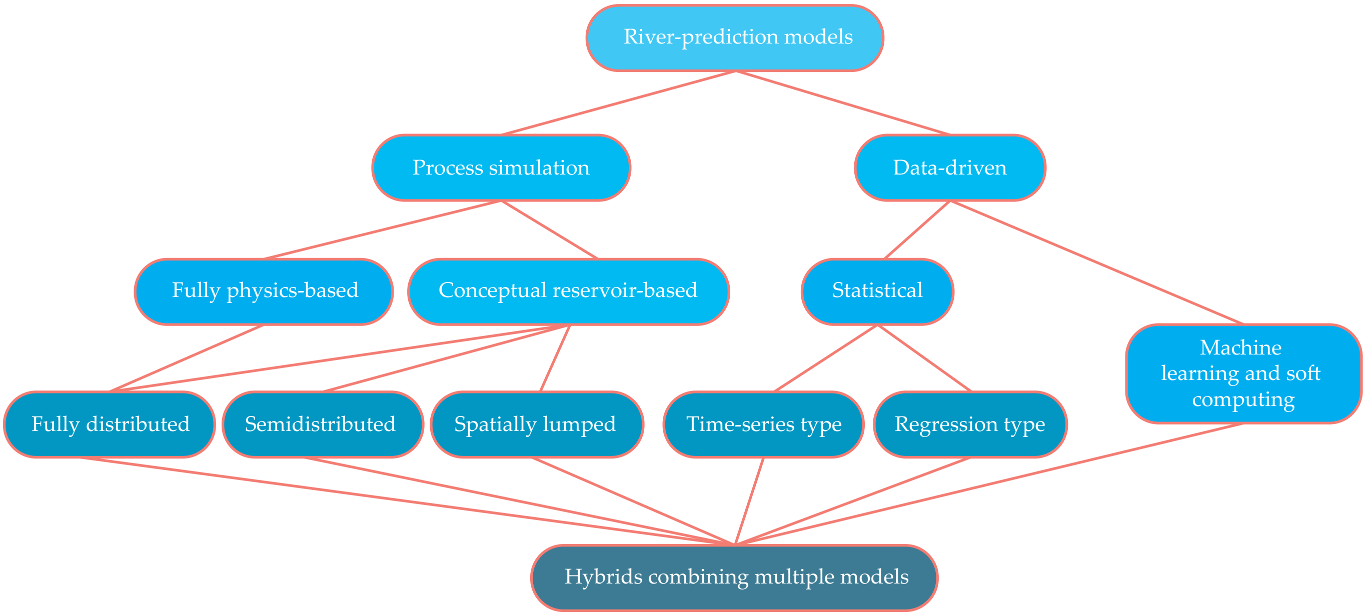


FIGURE 2. RIVER-PREDICTION MODELS fall broadly into two categories. Process-simulation models aim to be explicitly physics-based and are further classified by their rigor, their level of spatial resolution, and the geophysical and biophysical processes they include. Data-driven models use pattern-detection algorithms to implicitly capture the physics of river runoff generation. They relate predictors to outputs using some form of input–output mapping and are further categorized by their use of classical statistics or artificial intelligence. There is much overlap among the modeling types, and there are many opportunities to combine them.

gives the linear reservoir model for river prediction:

$$I(t) = Q(t) + k\,\frac{\partial Q(t)}{\partial t}$$

where Q is the predicted streamflow rate; I represents the conceptual reservoir's net rate of water input, which consists mainly of rainfall or snowmelt and may be adjusted for evapotranspiration losses; and k is a constant that accounts for how quickly streamflow responds to weather. Readers may recognize the model as a first-order linear time-invariant systems approach. It may seem simple, but it can become complicated when it's configured to capture more complex geophysical dynamics such as multiple reservoirs in series or parallel, spatially distributed reservoirs, and nonlinear reservoirs.

A similar but more straightforward treatment considers only continuity without explicitly including the linear reservoir assumption. In such a model, changes in a basin's water storage are equal to the difference between inputs, like rainfall and snowmelt, and outputs, like evapotranspiration and streamflow. Such water-balance models can be implemented using simple spreadsheet calculations and are sometimes used for practical water-planning tasks. They can also serve as frameworks for building intricate suites of interlinked submodels that represent various additional processes.

The concepts discussed so far represent only two things: the dynamics of water movement across and through the ground and the river basin's overall water balance. Those elements form the core of any mechanistic river-prediction model. But a river and its watershed can have many different components, such as trees, buildings, swamps, and ice fields. In practice, therefore, most process-simulation models are modular; in addition to their core, they usually integrate several submodels representing environmental factors that can affect streamflow.

Like the cores, submodels vary in their approaches and level of detail. They might represent transpiration by forests and crops, evaporation from lakes and the soil surface, snowpack accumulation and melt, or ice melt from mountain glaciers, which is distinct from seasonal snowmelt. They can also account for near-surface hydrometeorological dynamics, like the dependence of precipitation's phase on temperature and elevation, which can vary dramatically in rugged mountain watersheds. An estimate of $I(t)$ in a linear reservoir model might, for example, come from adding local rain flux from an atmospheric submodel to snowmelt flux from a snowpack one.

Models can also account for modifications of natural processes. Capturing land-use changes, such as certain forestry practices that reduce the tree canopy or urbanization of agricultural lands that increases impermeable area, can help predict corresponding shifts in evapotranspiration, snow dynamics, and infiltration. Those shifts can influence river flows by, for example, increasing flood frequency and severity.

An important attribute of any process simulation model is its degree of spatial distribution. A fully distributed model divides the watershed into a grid pattern to accommodate heterogeneity in watershed processes; it can account for details like a thunderstorm only producing rainfall in one part of the watershed. At the other end of the spectrum, a spatially lumped model is a parsimonious approach that treats the watershed upstream of the point of interest as a single homogeneous unit. The intermediate option is a semidistributed model that divides the watershed broadly by spatial proximity or elevation. Some version of either the conceptual linear reservoir or the water-balance method is typically used regardless of a model's degree of spatial distribution. However, fully distributed models occasionally implement the gold-standard approach, which typically requires a finite-difference solution that leads naturally to a grid-based division of the watershed.

Cyborg hydrologists

Unlike process-simulation models, data-analytics approaches to

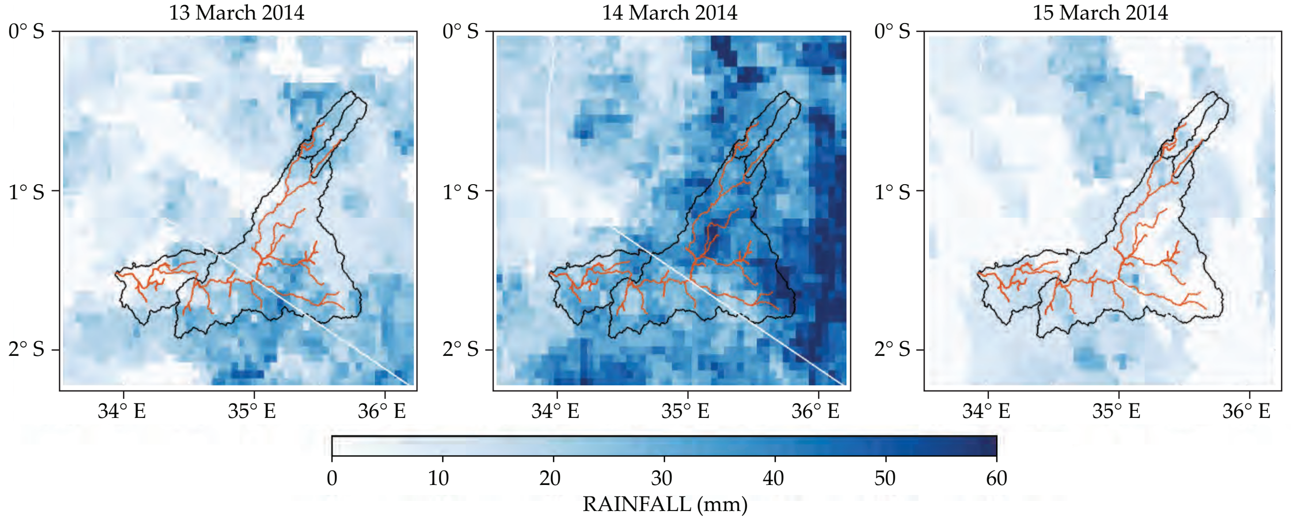


FIGURE 3. RAINFALL IN THE MARA RIVER BASIN is tracked by satellite and ground-based measurements and incorporated into the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) dataset for the region. These maps show daily CHIRPS precipitation; black, orange, and white lines denote, respectively, the watershed boundary, stream network, and Kenya-Tanzania-Uganda borders. Tirthankar Roy and colleagues at the University of Arizona (including one of us, Gupta) combined several satellite-based precipitation products, including CHIRPS, with an ensemble of hydrologic models to provide real-time probabilistic streamflow monitoring and forecasting for the Mara River.[17] Their effort is part of SERVIR, a joint program run by NASA and the US Agency for International Development and designed to help developing nations make environmental decisions. (Courtesy of Tirthankar Roy.)

river prediction view each watershed as a dynamical filter with input and output signals such as rainfall and streamflow. The model's job is to implicitly capture watershed processes in a transfer function that empirically maps the inputs to the outputs. Such top-down data-driven prediction methods use both statistical and machine-learning techniques, and they serve as a powerful and flexible complement to bottom-up mechanistic models.

Linear Gaussian statistical models have long been used for river prediction. For instance, the 1960s-era Thomas–Fiering model for short-term river forecasting applied standard linear time-series procedures, which are widely used across the natural and social sciences to make predictions from memory-rich datasets. (In fact, Harold Edwin Hurst and Benoit Mandelbrot's discovery of long-term memory in time series originated from their studies of Nile River flows. Since then, fractal dynamics and $1/f$ noise in hydrologic data have continued to attract physicists' attention.[8]) A more modern example of applying linear Gaussian statistics to river prediction is a probabilistic extension of principal component regression. Originally adapted to water-supply forecasting (WSF) by the US Department of Agriculture, it is commonly used by government agencies and hydroelectric utilities across the western US and Canada to predict seasonal snowmelt volume.[9]

Machine learning is a branch of artificial intelligence (AI) that uses algorithms to detect patterns in data and then uses those patterns to make predictions. One of us (Gupta) helped lead the charge 25 years ago to apply machine learning to hydrology;[10] that approach is now coming back in a big way as AI permeates the everyday world. (For more on applying AI to river prediction, see box 2). Even Google is getting in on the action by using AI for experimental large-scale flood forecasting in India.

Current R&D on machine learning for river prediction is bridging the gap between academic research and live operational forecasting systems. Hybrid solutions blend AI with specific technical and institutional requirements around river prediction, including ease of use and alignment with existing knowledge of the physical processes governing river flow.[4] For instance, one of us (Fleming) is currently retrofitting the US Department of Agriculture's proven WSF model with a physics-constrained AI metasystem that integrates automated machine learning.[11]

Data-driven models can easily test the effects of integrating new, potentially helpful information with established predictors. If an update is useful, it can quickly be deployed. For example, El Niño events can cause drier, warmer winters, lower snowpack, and reduced river-flow volumes in the Pacific Northwest; they tend to have the opposite effect in the US Southwest. Climatologists routinely summarize ocean temperature data indicative of such events in compact metrics like the Niño 3.4 index. Hydrologists can easily combine those indices with other predictor variables in a regression- or AI-based WSF model to improve its accuracy. Such practices are common in operational forecasting. However, climate science evolves rapidly, and data-driven models also make it simple to test the river-prediction value of emerging information. They are therefore crucial tools for ongoing hydroclimatic research.[12]

Modeling chains

Watershed hydrology is one component of a larger environmental framework, so multiple models are sometimes linked for a more comprehensive view. Gleaning inputs for river prediction models from the outputs of numerical weather-prediction models is a common practice. That chain forms the basis for operational flood forecasting, which provides crucial information for emergency management and dam safety; it facilitates decision-making around whether to issue evacuation notices or preemptively spill water from a reservoir. Government agencies and dam operators generate and use such information daily.

Outputs from watershed hydrology models can also be inputs to river-hydraulics models for mapping flood inundation and propagation. Those models predict where floodwaters will go and how far they'll reach. Their results are used for emergency planning, setting home insurance rates, and predicting floods in large rivers like the Mississippi, where flood waves from storms far upstream can take days to propagate downstream. Hydraulics models also inform physical habitat assessments; fish like certain kinds of flow patterns, so information about water flow can help protect and restore their habitat.

Basin-scale river-hydrology models coupled with numerical groundwater models can predict the details of river–aquifer interactions. Understanding those relationships has been important for addressing interstate water conflicts like a recent US Supreme Court case between Texas and New Mexico, which was based in part on the effects of groundwater extraction and surface water–groundwater interactions in the Rio Grande basin.

Model chains are also used to assess possible climate change implications for rivers. Outputs from global climate models can be used to drive river predictions; however, the outputs must first be downscaled and bias-corrected to adjust for systematic errors and to provide information about meteorological forcing at appropriate spatial and temporal scales. Climate change may also induce other environmental shifts, such as vegetation changes or glacier recession, that can have hydrologic implications. In those cases, a modeling chain would require intermediate steps to quantify such land-cover changes so they can be represented in the watershed model. Normally, there is no dynamic coupling within a chain; it is a one-way pipeline of offline models run sequentially by different research groups.

Complexity, selection, and ensembles

Many kinds of data are potentially relevant for river prediction. They include land surface characterization, such as maps of vegetation cover and impervious surface area; weather metrics, such as temperature, precipitation, wind speed, and solar radiation; digital elevation models and river network representations; and maps of hydrogeologic characteristics, such as soil types. Sources for those data include long-term environmental monitoring stations, airborne and satellite remote sensing (see figure 3), and outputs from other models. The choice of a river prediction model should reflect in part what data are available: Using a simple hydrologic prediction model when many types and large volumes of data are available may result in an unnecessary loss of information-generating capacity, whereas using a complex model with insufficient data may create a false sense of sophistication and potentially incorrect results.

Model choice is less about pros and cons and more about picking the right tool for the job. But that can be challenging. One overarching theme to river prediction models is that there are many, and they're diverse. Broadly speaking, physics-oriented models are great for testing our hydrologic knowledge because they use explicit representations of specific processes. Their virtual watersheds can also directly simulate predictive scenarios around climate change, urbanization, wildfire, and other long-term environmental shifts. Data-driven models, on the other hand, cost far less to build and run. They also tend to give more accurate short-term operational forecasts of flooding and water supply with more reliable quantitative estimates of predictive uncertainty.

Other selection criteria for a model are whether it represents all pertinent processes; whether it captures the problem's time and space scales; and whether uncertainty assessment is required. Pragmatic issues also arise, such as reliability, run time, implementation and operating costs, and stakeholder buy-in. New modular, customizable frameworks allow users to choose components. For hydrology, as in many other fields, the best predictive model is often an ensemble.

The value of predictability

River prediction is a high-stakes game, and the stakes are only getting higher. Even modest,

BOX 1. THE GOLD STANDARD

The classical physics-based approach to representing water flows in and across landscapes was first outlined in 1969 by Allan Freeze and Richard Harlan.[5] Their model is built primarily around four coupled partial differential equations. Below are the equations in Freeze and Harlan's original notation. Today's models often use more modern formulations, driven in part by advances in computational technology.

The first key element is the Richards equation, which governs water movement in the vadose zone, an unsaturated area above the groundwater table. In their paper, Freeze and Harlan reported that numerical solutions to the three-dimensional Richards equation were not yet possible. They therefore focused on a one-dimensional simplification that only predicted the vertical infiltration of rainfall or snowmelt into the soil:

$$\frac{\partial}{\partial z}\left[K(\psi)\left(\frac{\partial \psi}{\partial z}+1\right)\right]=C(\psi)\frac{\partial \psi}{\partial t}.$$

The second element is the groundwater flow equation, a form of the diffusion equation that governs water dynamics in aquifers. Freeze and Harlan framed it as

$$\left(\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}+\frac{\partial^2\varphi}{\partial z^2}\right)+2\rho g\beta\left[\left(\frac{\partial\varphi}{\partial x}\right)^2+\left(\frac{\partial\varphi}{\partial y}\right)^2+\left(\frac{\partial\varphi}{\partial z}\right)^2-\frac{\partial\varphi}{\partial z}\right]=\frac{\rho g}{k}\left[(1-\theta)\alpha+\theta\beta\right]\frac{\partial\varphi}{\partial t}.$$

The third piece is the one-dimensional Saint-Venant equations, which describe the surface movement of water, including flow overland and in river channels:

$$y\frac{\partial v}{\partial x}+v\frac{\partial y}{\partial x}+\frac{\partial y}{\partial t}=q-i$$

and

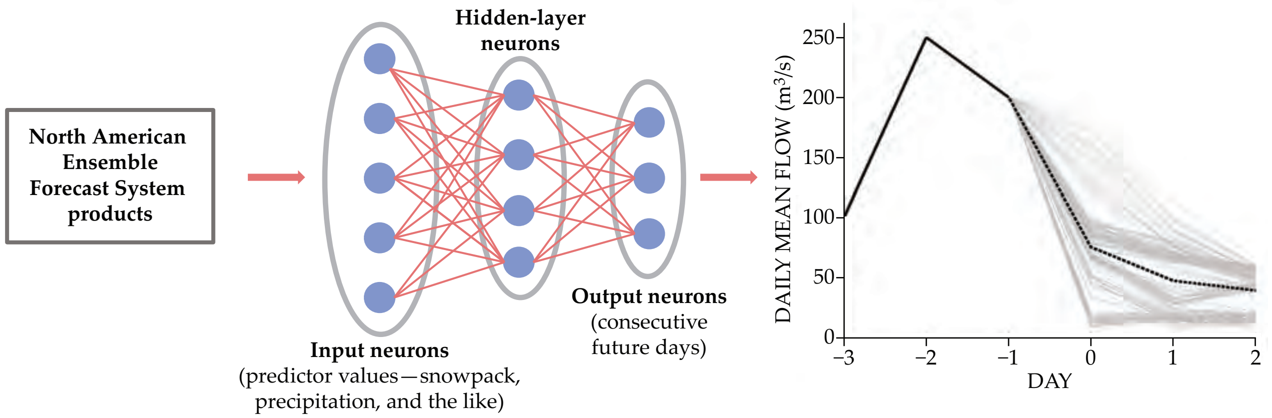
$$\frac{\partial v}{\partial t}+v\frac{\partial v}{\partial x}+g\frac{\partial y}{\partial x}=g\,(S_o-S_f)-\frac{v}{y}\,(q-i).$$

All of the equations are related to Navier–Stokes and continuity. For example, the groundwater flow equation is derived from the combination of continuity and Darcy's law; although the law, which describes flow through a porous medium, was originally discovered empirically, it can also be derived from Navier–Stokes.

BOX 2. APPLYING ARTIFICIAL INTELLIGENCE TO RIVER PREDICTION

River forecasters were early adopters of machine learning. In the mid 1990s, they began publishing papers that applied artificial neural networks to flood forecasting. Although that research continued, practical applications—particularly to operational forecasts—were few. Roadblocks included an apparent lack of interpretability, an unproven track record, and the deterministic nature of most artificial intelligence (AI) forecasts. However, the use of AI is now becoming routine. Machine-learning algorithms are more accessible, and methods are available for making probabilistic predictions and facilitating the integration of experiential knowledge into machine-learning models. Views are therefore changing toward using AI for practical river-prediction applications.

A hybrid prediction system was recently operationally tested by a flood-forecasting agency in the Pacific Northwest.[18] Several neural-network river-prediction models formed an ensemble; each represented a slightly different geophysical conceptualization of the dominant flood-generating mechanisms in the mountain watershed, which received precipitation as both rain and snow. Each of the neural networks was driven in turn by an ensemble of downscaled and bias-corrected numerical weather predictions from the North American Ensemble Forecast System, a joint project of the US, Canadian, and Mexican national weather services, and by observational data on antecedent streamflow and snowpack conditions. The resulting Monte Carlo super-ensemble, run with a one-day timestep, enabled probabilistic forecasts, including the likelihood of the river topping its banks during the next three days. (Figure adapted from refs. 4 and 18.)



incremental improvements in WSF accuracy can contribute additional public value of more than $100 million annually for a single river basin.[13] The accuracy and lead times of flood forecasts are also becoming ever more crucial: Flood risks are escalating with the increased development of floodplains, more extreme rainfall events under climate change, and urbanization-induced losses in the landscape's capacity to absorb rainfall.

Moreover, two billion people currently live without adequate access to drinking water, and UNESCO expects global water demand to increase by 55% in the next few decades due to population and economic growth. Avoiding lethal and socioeconomically destabilizing global failures to meet basic water-supply needs will require better water-management approaches based on improved understanding and prediction of river dynamics across a range of space and time scales.

Several directions for future work are apparent.[2,4,14] Predictive skill needs to improve in difficult environments like deserts, alpine watersheds, and dense cities. Renewed attention should be paid to complex systems science, a field that has attracted great interest in statistical mechanics, ecology, and sociophysics. River networks are a classic example of fractal geometry, and chaos theory limits weather's predictability to an approximately two-week theoretical window. In general, neither of those modern mathematical concepts appears explicitly in river-prediction models; a complex systems approach could incorporate them.

Continuing to capitalize on the data revolution will drive progress in hydrology. Developing novel data types,[15] discovering predictive climatic information,[12] and exploring new analytical directions to support environmental monitoring and prediction[16] all offer opportunities for growth. Substantial forward movement on those fronts will be crucial to managing rivers and water resources in an increasingly uncertain future.

REFERENCES

1. M. W. Doyle, *J. Am. Water Resour. Assoc.* **48**, 820 (2012).
2. S. W. Fleming, *Where the River Flows: Scientific Reflections on Earth's Waterways*, Princeton U. Press (2017).
3. A. T. Wolf, *Water Policy* **1**, 251 (1998).
4. S. W. Fleming, "Machine learning, soft computing, and complex systems analysis: emerging approaches…," presentation at the Center for Nonlinear Studies, Los Alamos National Laboratory, 20 August 2018.
5. R. A. Freeze, R. L. Harlan, *J. Hydrol.* **9**, 237 (1969).
6. S. L. Painter et al., *Water Resour. Res.* **52**, 6062 (2016).
7. S. J. Kollet et al., *Water Resour. Res.* **46**, W04201 (2010).
8. V. N. Livina et al., *Phys. Rev. E* **67**, 042101 (2003).
9. D. C. Garen, *J. Water Resour. Plan. Manag.* **118**, 654 (1992).
10. K. Hsu, H. V. Gupta, S. Sorooshian, *Water Resour. Res.* **31**, 2517 (1995).
11. S. W. Fleming, A. G. Goodbody, *IEEE Access* **7**, 119943 (2019).
12. A. M. Kennedy, D. C. Garen, R. W. Koch, *Hydrol. Process.* **23**, 973 (2009).
13. A. F. Hamlet, D. Huppert, D. P. Lettenmaier, *J. Water Resour. Plan. Manag.* **128**, 91 (2002).
14. M. P. Clark et al., *Hydrol. Earth Syst. Sci.* **21**, 3427 (2017).
15. T. H. Painter et al., *Remote Sens. Environ.* **184**, 139 (2016).
16. M. J. Halverson, S. W. Fleming, *Hydrol. Earth Syst. Sci.* **19**, 3301 (2015).
17. T. Roy et al., *J. Hydroinformatics* **19**, 911 (2017).
18. S. W. Fleming et al., *J. Am. Water Resour. Assoc.* **51**, 502 (2015). PT

LIBRARY OF CONGRESS, NEW YORK WORLD-TELEGRAM AND SUN COLLECTION. COURTESY AIP EMILIO SEGRÈ VISUAL ARCHIVES

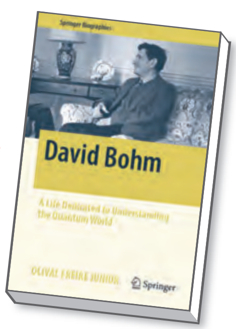
DAVID BOHM IN 1949, shortly after he had refused to testify before the House Un-American Activities Committee.

The man who explained quantum mechanics

David Bohm (1917–92) was a man of principle. He refused to denounce fellow members of the Communist Party in the US during the McCarthy era, and he insisted that physicists should search for an explanation of quantum phenomena when almost all physicists believed that none existed. Olival Freire Jr examines his life and work in a new biography, *David Bohm: A Life Dedicated to Understanding the Quantum World*.

Bohm's was an eventful life filled with significant scientific achievements. His 1943 doctoral dissertation on the scattering of protons was classified because of its potential military value for the Manhattan Project, and his subsequent work on plasma physics is renowned for its use of collective variables and for the discovery of what is now called Bohm diffusion. But Bohm is perhaps best known for his work on quantum mechanics. In 1951 he published a textbook on the topic in which he defended the orthodox Copenhagen interpretation. He gave a copy to Albert Einstein, and in the ensuing conversation, Einstein convinced him that there was something fundamentally wrong with the Copenhagen view.

David Bohm
A Life Dedicated to Understanding the Quantum World
Olival Freire Jr
Springer, 2019.
$84.99



So Bohm broke up with Copenhagen. He immediately went to work on an alternative approach and came up with what is still the simplest and most baffling solution of the paradoxes of quantum mechanics. He proposed that electrons had trajectories and suggested an equation of motion. From that proposal, he developed a full theory that he called the causal or ontological interpretation of quantum mechanics, now often called Bohmian mechanics.

At the time, almost all physicists believed that the paradoxes of quantum mechanics simply could not be solved. Most of Bohm's contemporaries opposed his theories and considered him a heretic. To make matters worse, Bohm also got into political trouble. After he refused to name other communists to the House Un-American Activities Committee, he was indicted for contempt of Congress, whereupon Princeton University terminated his contract. They refused to renew it even after Bohm was acquitted a few months later. As no other US university would hire him, he was effectively forced to leave the country. He moved first to Brazil, then a few years later to Israel, and from there to the UK.

Near the time of his move to the UK, he discovered the Aharonov–Bohm effect jointly with his student Yakir Aharonov; the effect demonstrates that the electromagnetic vector potential can influence the interference pattern of an electron's wavefunction even if the field strength vanishes. Also around that time, despite the personal price he had paid for his loyalty to the Communist Party, he abandoned Communism after learning of the crimes of the Soviet government under Joseph Stalin.

Freire, a physicist and historian of physics at the Federal University of Bahia in Brazil, has written about the history behind quantum mechanics before: In his previous book, *The Quantum Dissidents: Rebuilding the Foundations of Quantum Mechanics (1950–1990)*, published in 2015, he studied the growing interest in alternatives to the quantum orthodoxy, with chapters on Hugh Everett, Eugene Wigner, John Bell's nonlocality theorem, decoherence, and also Bohm.

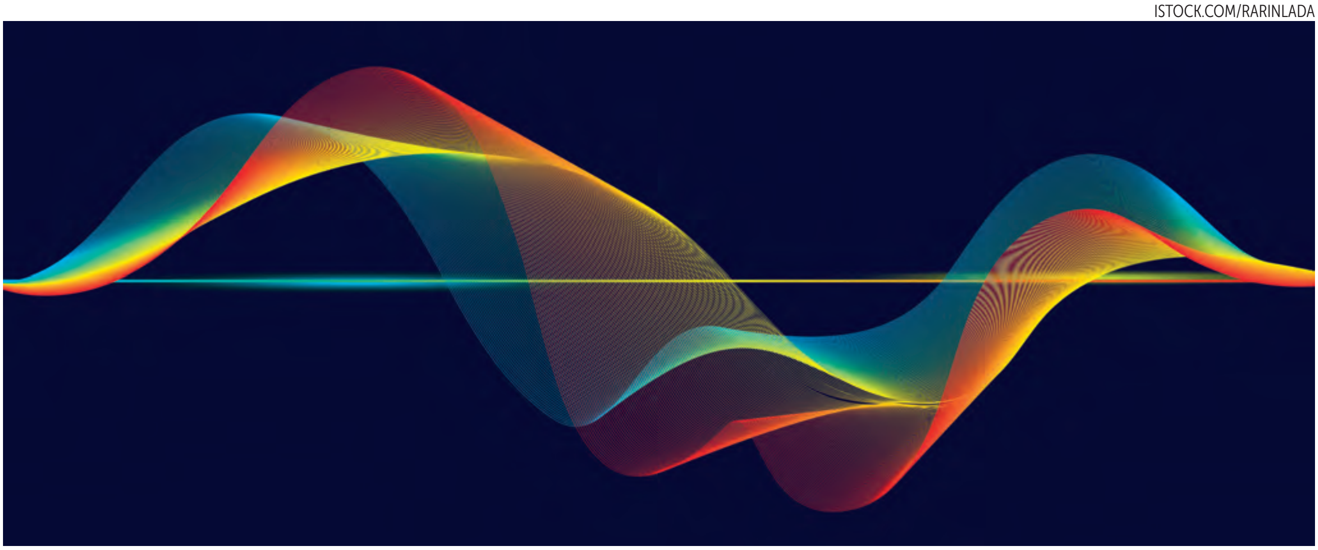
In contrast to an earlier biography by F. David Peat, *Infinite Potential: The Life and Times of David Bohm* (1996), which focused more on Bohm's life, Freire is primarily interested in the content, development, and reception of Bohm's scientific ideas. As a physicist himself, Freire dives into the scientific details of Bohm's arguments with and reactions to his contemporaries, such as his responses to Bell's famous theorem and the experimental tests of Bell's inequality.

Freire meticulously works his way through a wealth of historical materials. The book includes quotations from letters and interviews and plots citation counts to help the reader visualize the impact of Bohm's papers over the years. Freire's declared goal is to depict his subject not as a hero or saint, but with all his contradictions, changes, and doubts. And that he does. For example, he describes how Bohm's enthusiasm for Bohmian mechanics waned over the years, how lackluster his defenses of it were by the 1960s, and how his interest returned to it in the 1980s. Freire avoids passing judgment on the value of Bohm's various ideas and approaches.

Readers will not learn Bohm's theories from this book. Those who want to understand Bohmian mechanics in its modern formulation and its status compared to the orthodox Copenhagen interpretation would do well to consult Detlef Dürr and Stefan Teufel's *Bohmian Mechanics: The Physics and Mathematics of Quantum Theory* (2009) or Jean Bricmont's *Making Sense of Quantum Mechanics* (2016). For those curious about the story behind the theory and the struggles and breakthroughs of its pioneer, *David Bohm* is a rich resource.

Roderich Tumulka
Eberhard Karls University of Tübingen
Tübingen, Germany

ISTOCK.COM/RARINLADA

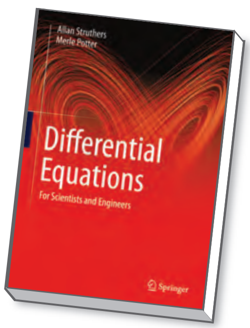


Testing the waters of differential equations

The textbook *Differential Equations: For Scientists and Engineers* by Allan Struthers and Merle Potter is an excellent starting point for undergraduates with a background in multivariable calculus who want to take the next step in their mathematical education. The book is framed as an ongoing discussion of the theory of differential equations. The authors frequently reference concepts and examples introduced earlier in the book to motivate further learning. That strategy makes the book feel like a continuing dialog between peers in which questions are asked and answers are proposed. The writing style is accessible to undergraduates, and students without much knowledge of differential equations will be able to jump right in.

Struthers and Potter choose not to prove the theorems they use throughout the book. Instead, they merely state them and then use them as tools to solve differential equations from physics, engineering, and biology. That practical approach makes *Differential Equations* an excellent resource for students who may not be mathematics majors. However, it also limits the scope of the book for advanced undergraduates, particularly mathematics majors who might have benefited from seeing more detailed proofs of some seminal theorems.

Differential Equations
For Scientists and Engineers
Allan Struthers and Merle Potter
Springer, 2019 (2nd ed.). $59.99



Chapter 1 is a detailed overview of concepts and techniques from linear algebra that are relevant to differential equations, and it includes plenty of explanations and examples. Much of its material is referenced and built on later in the book. As a probabilist, I was impressed by the way Markov chains were introduced as an example to motivate matrix–vector multiplication. Fundamental ideas from probability theory, including the Markov transition matrix and steady states of Markov chains, are explained effortlessly. The chapter concludes with an interesting, albeit short, discussion about how Google's algorithms determine page rankings for search results by solving an eigenvalue problem for a huge matrix.

Chapter 2 introduces simple methods for solving first-order differential equations. Most of the material will be a review for advanced undergraduate students, since they would have seen all the methods in one form or another in a calculus course. The last section discusses nonlinear first-order differential equations that are either separable or in exact form.

The authors point to chapter 3, titled "Linear Systems of Differential Equations," as the heart of the textbook. It begins with the long-term behavior of 2 × 2 systems in terms of eigenvalues and eigenvectors, then introduces methods of solving systems of linear first-order equations. That is a novel approach; most textbooks first discuss general techniques for solving systems and only examine long-term behavior after arriving at the solution. Struthers and Potter do a good job of starting with simple examples of linear systems, showing how to analyze those systems, and then generalizing the analytical techniques to arrive at concrete theorems.

A detailed discussion of beats of resonances is the highlight of chapter 4, "Higher-Order Differential Equations." The authors analyze damped-harmonic-oscillator equations with periodic forcing functions through a series of intelligently crafted examples. That section caters nicely to students with backgrounds in physics and engineering.

Laplace transforms are introduced in chapter 5 as tools that allow us to solve differential equations whose forcing functions are discontinuous. They also can be used, as the authors show, to deal with impulses. That approach gives instructors and students a clear motivation for studying Laplace transforms. The authors build up to solving second-order linear equations with Dirac delta force through a series of examples.

Chapter 6 is a good synopsis of the most commonly used numerical methods—including Euler's, Taylor's, and the Runge–Kutta—for solving differential equations and systems. In chapter 7, titled "Series Solutions for Differential Equations," the authors spend considerable time on the Hermite, Laguerre, and Bessel equations. However, all three are introduced rather abruptly, and the chapter could have benefited from a discussion of why those equations are important.

Except for the first chapter, all chapters contain a section called "For Further Study." That material deserves a special mention. Each section reads like a single long exercise problem but is in fact a step-by-step analysis of interesting problems coming from different scientific fields. The problems tie together techniques and concepts developed throughout the chapter, while at the same time teaching some new material. I will soon begin teaching an undergraduate class on differential equations and I think those sections will come in handy as a source of assignments. For example, the "For Further Study" section in chapter 3 leads students to the formal definition of matrix exponentials and gives them the opportunity to explore how the exponentials can be used to solve systems of first-order differential equations.

I am very happy to have been introduced to this excellent textbook. I plan to use it as an additional resource in the upcoming semester.

Pratima Hebbar
Duke University
Durham, North Carolina

Debating astronomy in Victorian newspapers

In 1881, in a series of essays titled *The Poetry of Astronomy*, astronomer and popularizer Richard Proctor argued for the value of imagination in the practice of astronomy. He wrote that no one "who studies aright the teachings of the profoundest students of nature will fail to perceive that [they] have been moved in no small degree by poetic instincts, and that their best scientific work has owed as much to their imagination as to their reasoning and perceptive faculties." Comparing astronomy to poetry was no mere rhetorical flourish for Proctor–it was imagination, in his view, that transformed dry scientific data into knowledge. Imagination gave the astronomer access to causes and meanings that were not physically evident.

Proctor had no small stake in debates over how astronomy should be practiced and who had authority to produce astronomical knowledge claims. As historian Joshua Nall recounts in his book *News from Mars: Mass Media and the Forging of a New Astronomy, 1860–1910*, Proctor made his living publishing astronomical texts for public consumption. When writing for the public, Proctor drew authority from his own bona fides as a practicing astronomer. When addressing his peers in the scientific community, he argued that his ability to reach wide audiences and support himself with his scientific writing made him a true professional.

Mars was at the center of many of Proctor's debates

News from Mars
Mass Media and the Forging of a New Astronomy, 1860–1910
Joshua Nall
U. Pittsburgh Press, 2019. $50.00

PD-US, VIA WIKIMEDIA COMMONS

PERCIVAL LOWELL at his observatory in Flagstaff, Arizona in 1914.

with other astronomers. At the end of the 19th century, astronomers and the public were fascinated with the red planet, and both knowledge of and speculation about Mars seemed to be growing by leaps and bounds. In 1877, the same year American astronomer Asaph Hall discovered that Mars had two moons, Italian astronomer Giovanni Schiaparelli published hand-drawn maps reflecting his observations of the planet's surface. Schiaparelli's maps indicated that the surface of Mars was criss-crossed by a system of channels that he labeled *canali*. A public controversy ensued about whether those canali were built by an intelligent civilization.

Proctor's writings enthusiastically described canals and civilizations on the surface of Mars. He and his fellow canal enthusiast Percival Lowell, founder of the Lowell Observatory in Arizona, have been dismissed in hindsight as out of sync with professional astronomy during that period. Proctor and Lowell published in newspapers and in cheaply produced publications, not in the professional journals of the "real" astronomers, and that choice has led historians to mark them as amateurs or outsiders. But in characterizing Proctor and Lowell as discredited popularizers, historians have mostly relied on the words of their critics, the victors in the Mars debates. Nall invites us to reconsider Proctor's position through a careful examination of his writings. Nall's analysis suggests that historians should rethink debates about what constituted professional astronomy in the late 19th century and reexamine astronomy's relationship with the emerging mass media of the day. Newspaper and book publishers were eager to cash in on public desire for news about Mars. Meanwhile, astronomers were interested in the popular press's ability to spread their ideas and bolster their authority.

Proctor harnessed the press to sell his vision of scientific practice. His approach was multidisciplinary, anti-elitist, and populist in character, but was not "amateur," as modern readers understand that term. He used what Nall describes as an "imaginative astronomy," a set of methods that employed analogy to link his readers' familiar, lived experience with the concepts he described and with the scientific evidence gained through astronomical observation of the planets. In Proctor's view, his approach to astronomy was entirely professional.

Proctor and Lowell's opponents, on the other hand, advocated a very different kind of astronomical professionalism. Chief among them were the astronomers of California's Lick Observatory, whose first director, Edward Holden, attempted to frame their speculation about life on Mars as newspaper sensationalism. Holden and his peers also sought to make a living from their science and to wrest scientific authority from gentlemen astronomers whose reputation stemmed to no small degree from their social station. The new generation of astronomers distinguished themselves from the older generation by their physical instruments; they incorporated spectroscopy and other new methods to make claims about the composition of stars and planets.

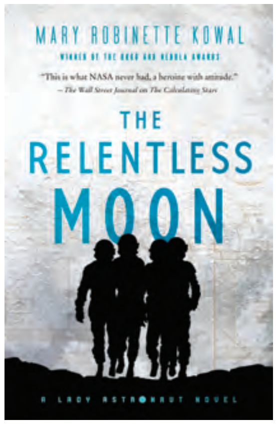
The so-called new astronomy upset the field's existing power structure at exactly the moment when the rise of mass media created a new venue where astronomers could fight for disciplinary authority. Proctor and Lowell represented threats to the new astronomy, not because they were wrong, but because they were popular. Proctor had a talent for writing for the public, and Lowell frequently issued press releases about his latest maps of Martian canals. Holden's opposition was based in a fear that not only did the canal headlines bring bad science to the public, but they were crowding out the reports about Mars that were coming from his own observatory.

Nall's nuanced account of how astronomers attempted to discredit and compete with Proctor and Lowell shows that the disciplinary norms of professional astronomy emerged during, not before, the canal controversy. The victors in the debate won by forging alliances between observatories and the press and thus establishing which astronomers and observatories would be considered reliable sources of news from Mars. Nall shows that Proctor and Lowell were not led to wrong conclusions because they were amateurs, but that they were marked as amateurs because they failed to fall in line with the emerging norms of the new astronomy.

Matthew Shindell
Smithsonian National Air and Space Museum
Washington, DC

NEW BOOKS & MEDIA

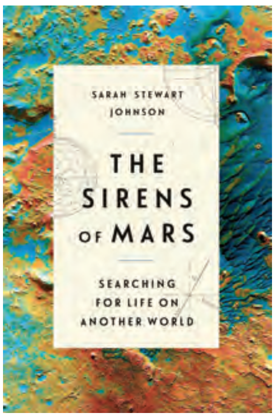
The Relentless Moon



Mary Robinette Kowal
Tor Books, 2020. $17.99 (paper)

Mary Robinette Kowal's Lady Astronaut series takes place in an alternate universe where the catastrophic effects of a meteor impact in 1952 force humans to throw their full weight behind spaceflight in an attempt to escape their dying planet. The first two books, *The Calculating Stars* and *The Fated Sky*, followed the story of Elma York, a mathematician and pilot who becomes the first woman in space. In the third book, *The Relentless Moon*, Kowal brings in a new narrator: Nicole Wargin, an astronaut who is married to the governor of Kansas. Wargin's story loosely parallels the fascinating life of Janey Hart, an activist, pilot, and senator's wife who hoped to become an astronaut but was thwarted by NASA's men-only policy. Kowal also brings some real-life historical figures into *The Relentless Moon*, including physicist Otto Frisch. —MB

The Sirens of Mars

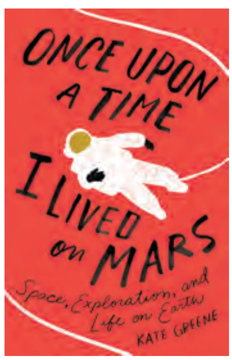


Searching for Life on Another World

Sarah Stewart Johnson
Crown, 2020. $28.99

Planetary scientist Sarah Stewart Johnson answers the siren call of the red planet with this retrospective of Mars missions and the search for extraterrestrial life. Hers is a very personal account, in which she intersperses stories of various spacecraft and some of the key scientists involved with reminiscences of her childhood, development as a scientist, and experiences working on such projects as NASA's *Curiosity* rover and traveling to some of the most remote and hostile environments on Earth in pursuit of exotic life forms. Through her vivid observations and descriptions, Johnson aptly conveys the seemingly quixotic search for the answer to one of the most fundamental questions facing humanity: Are we alone in the universe? —CC

Once Upon a Time I Lived on Mars

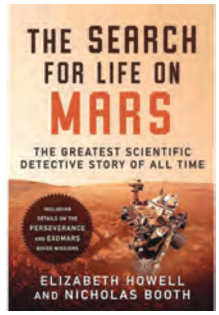


Space, Exploration, and Life on Earth

Kate Greene
St Martin's Press, 2020. $27.99

Between April and August 2013, science journalist Kate Greene joined five other pseudo-astronauts to live in a two-story geodesic dome on Hawaii's Mauna Loa as part of the first Hawaii Space Exploration Analog and Simulation (HI-SEAS). The mission's focus was to study the psychological effects of food and whether the opportunity to cook meals and prepare favorite dishes, instead of consuming only rehydrated fare, could improve astronaut morale and appetite during an extended stay on Mars. In her thoughtful, well-written account of the mission, Greene not only discusses what it was like to spend several months cooped up indoors with five strangers and limited resources but also reflects on what this and other space missions can teach us about ourselves and life on Earth. —CC

The Search for Life on Mars

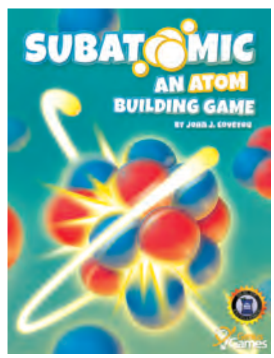


The Greatest Scientific Detective Story of All Time

Elizabeth Howell and Nicholas Booth
Arcade, 2020. $27.99

Hardly a traditional whodunit, *The Search for Life on Mars* does have elements of detective fiction: forensic evidence of Mars's surface and interior obtained by numerous unmanned orbiters and landers, principal investigators analyzing it for vital clues, and plenty of red herrings and false leads. In this ambitious collaboration, journalists Elizabeth Howell and Nicholas Booth attempt to present more than five decades of Martian exploration and make it readable for people with no technical background. Although there is no big reveal at the end—we still don't know if there is, or ever was, life on Mars—scientists hope to glean a more definitive answer to that mystery from a plethora of upcoming missions, including NASA's *Perseverance* rover, which is scheduled to launch later this month. —CC

Subatomic



An Atom Building Game

John J. Coveyou
Genius Games, 2019 (2nd ed.). $39.99

The goal of this innovative board game is to build an atom from photons and quarks. Players are dealt cards of subatomic particles and must combine them to form real elements in order to win the game. To help build their elements, players can hire famous physicists like J. J. Thomson, Maria Goeppert Mayer, and Ernest Rutherford to aid their efforts. For ages 10 and up. —MB PT

NEW PRODUCTS

Focus on photonics, spectroscopy, and spectrometry

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

Lasers for biomedical and quantum technologies

Hübner Photonics has expanded its Cobolt 06-01 series of plug-and-play modulated lasers. The series now features 12 additional wavelengths covering the 405–975 nm range and higher powers on several existing wavelengths: 405 nm with 365 mW, 445 nm with 400 mW, 457 nm with 400 mW, and 515 nm with 150 mW. The wide spectral coverage, compact form factor, direct modulation capability, and true "off" during modulation make the lasers suitable for laboratory and research applications in the life sciences and quantum technologies. All Cobolt 06-01 series lasers are manufactured using proprietary HTCure technology, which, according to the company, provides high reliability and a high level of immunity to varying environmental conditions. ***Hübner Photonics Inc***, *2635 N 1st St, Ste 228, San Jose, CA 95124, www.hubner-photonics.com*

Compact, aberration-free spectrograph

Teledyne Princeton Instruments has added a new aberration-free imaging spectrograph to its IsoPlane spectrograph portfolio. The IsoPlane 81, formerly FERGIE, features an advanced, proprietary spectrograph design integrated with a deeply cooled, research-grade CCD camera. The new model has been improved to further reduce readout noise, increase scanning speed, and create easier access to gratings. External updates enhance the system's ruggedness. According to the company, the IsoPlane 81 provides superior resolution and signal-to-noise ratio compared with cameras that have much larger footprints and focal lengths. Accessories focused on Raman spectroscopy and Raman microscopy include a dedicated module for fast, accurate microspectroscopy, light sources for calibration, and various prealigned cubes for configuring experiments. ***Teledyne Princeton Instruments***, *3660 Quakerbridge Rd, Trenton, NJ 08619, www.princetoninstruments.com*

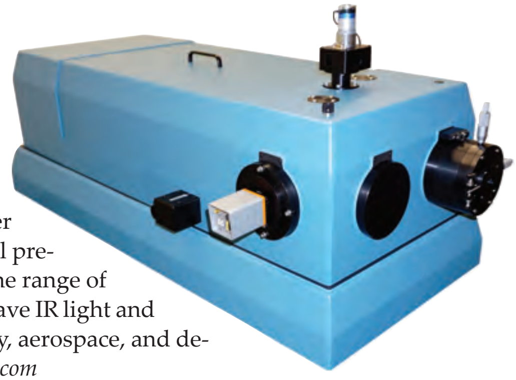
Reflective spectrofluorometer

Horiba Scientific has unveiled the fourth generation of its Fluorolog modular spectrofluorometer. The Fluorolog-QM is a lens-free, reflective spectrofluorometer that provides sharp focus at all wavelengths from the deep UV (180 nm) to the near-IR (5500 nm). Using the company's first standard deviation (FSD) method, it delivers high sensitivity at a 32 000:1 signal-to-noise ratio for the Raman band of water. (FSD denotes the difference of peak signal minus background signal, divided by the square root of the background signal.) For optimal stray light rejection, it offers the industry's longest focal length at 350 mm for single monochromators and 700 mm for double monochromators. The Fluorolog-QM delivers steady-state, spectral, and time-resolved photoluminescence from 180 nm to 5500 nm. Combined with up to four light sources, up to six detector options, and sample handling accessories, it can be enhanced to suit a broad range of luminescence research applications. ***Horiba Scientific***, *20 Knightsbridge Rd, Piscataway, NJ 08854, www.horiba.com*

Laser-cavity-tuning amplifier

Vescent Photonics has introduced its SLICE-DHV high-voltage amplifier for piezoelectric control of laser tuning mirrors and cavity length stretching. According to the company, it offers high-bandwidth control over high-voltage-driven piezoelectric transducers while maintaining extremely low noise. Depending on the capacitance of the load, bandwidths greater than 500 kHz are possible over a control range of ±10 V superposed on a bias voltage of 0–200 V. The resulting fast control over the cavity length of a frequency comb oscillator supports repetition-rate matching for dual-comb spectroscopy or a tight lock on f_{opt} to operate the comb as a frequency ruler. A high-gain mode allows easy control over the full 0–200 V output range. The bias plus fast servo-control mode can be used to lock external cavity laser diodes and frequency combs that are lead zirconate titanate tuned. To be more economical and leave a smaller footprint, each SLICE-DHV offers two independently operating channels. ***Vescent Photonics LLC***, *14998 W 6th Ave, Ste 700, Golden, CO 80401, www.vescent.com*

High-resolution IR-imaging spectrometers



McPherson has announced that it has made its dispersive spectrometers easier to use in the IR. Ample, accessible space and a mechanically robust platform for mounting IR arrays now permit devices such as uncooled microbolometer arrays to be readily coupled to the exit focal plane. A bolometer array IR digital camera mounted on a 1-m-focal-length McPherson spectrometer can sort 50 IR bands simultaneously. For more sensitive spectral imaging applications, larger IR cameras with thermoelectric or cryogenic cooling can be adapted with optical precision to the McPherson spectrometer. The new sensor compatibility expands the range of wavelengths in which the company's spectrometers can detect mid- and long-wave IR light and increases utility for spectroscopy and imaging applications in chemistry, biology, aerospace, and defense. ***McPherson Inc****, 7A Stuart Rd, Chelmsford, MA 01824, https://mcphersoninc.com*

Nanoscale FTIR spectroscopy system

According to Bruker, its nanoIR3-s Broadband is the most advanced nanoscale Fourier-transform IR spectroscopy system currently available. It combines two techniques: scattering-type scanning near-field optical microscopy and advanced IR laser technology. Using a broadband light source based on a novel femtosecond laser with a distributed optical parametric oscillator feedback, the system provides nanoscale imaging and spectroscopy over the 2.5–15 μm (4000–670 cm^{-1}) spectral range. The source allows for switching its linewidth for imaging or spectroscopy. The nanoIR3-s Broadband is suitable for advanced research on organic and inorganic materials and for nanoscale optical imaging of 2D materials, including plasmonic and nanophotonic processes and structures. ***Bruker Corporation****, 40 Manning Rd, Billerica, MA 01821, www.bruker.com*

PT

OBITUARIES

Freeman J. Dyson

Freeman J. Dyson died on 28 February 2020 of complications after a fall in the dining hall of the Institute for Advanced Study in Princeton, New Jersey. That lunch discussions with young colleagues were his frequent pleasure at age 96 says much about this brilliant, gracious, and sometimes contrarian man.

Dyson was born on 15 December 1923 in Crowthorne, UK, into a family at once artistic and establishment; his father, George Dyson, was a noted composer and educator. A mathematical prodigy, Freeman easily won a scholarship to Cambridge University. His mentor there, celebrated mathematician G. H. Hardy, told him, “Young men should prove theorems; old men should write books.” That became the pattern of Dyson’s life. His “theorems” illuminated physics. His first book of artful and influential essays, *Disturbing the Universe*, was published when he was 56. The first half of his life established his mathematical brilliance; the second half, his unabashed intellectual boldness and literally cosmic imagination.

Immediately after World War II, the effort to turn nonrelativistic quantum mechanics into a relativistic quantum field theory (QFT) was in disarray. Starting with Dirac’s equations for quantum electrodynamics, “old” versions of QFT existed, but they did not lead to finite results for radiative corrections. The calculation of observed physics—the Lamb shift or the anomalous magnetic moment of the electron—stumbled on the effects of particles’ self-interactions, with divergent results.

By the spring of 1948, however, the landscape had changed, and three theories seemed viable. Julian Schwinger, then a prodigy of 30, had a so-called Green’s function method involving prodigious calculations of which only he seemed capable. Richard Feynman, also 30, had what we now call Feynman diagrams. And Sin-itiro Tomonaga’s QFT was just becoming known in the West. Feynman also had the “sum-over-paths” idea, now called path-integral formulation. He asserted that it was the same theory as his diagrams, but no one fully understood how that could be.

Dyson, who never received a PhD, was exactly the right person in temperament, place, and time. In June 1948, in Ann Arbor, Michigan, Schwinger gave a series of lectures that most, including Dyson, found incomprehensible. Contrary to Schwinger’s reputation, Dyson found him approachable and willing to tutor the 25-year-old acolyte in the arcane art. In the same month, Dyson and Feynman made their now-legendary road trip from Cleveland, Ohio, to Albuquerque, New Mexico, with Feynman talking nonstop. Later in the summer, sleep-deprived on a Greyhound bus somewhere in Kansas, Dyson understood, virtually in a flash, that all three (or four) formalisms were actually the same theory. Schwinger’s Green’s functions were the commutators of old QFT but cleverly manipulated to avoid the infinities; Feynman propagators were time-ordered Green’s functions and were indeed a recognizable perturbation expansion of his path-integral formulation. Dyson’s 1949 paper “The radiation theories of Tomonaga, Schwinger, and Feynman” made a Nobel Prize virtually inevitable; it was awarded in 1965 to Feynman, Schwinger, and Tomonaga—not to Dyson. The prize is allowed to be split at most three ways.

In the 1960s the floodlight of Dyson’s curiosity shifted from mathematical physics to nuclear energy: reactors, bombs, disarmament, public affairs, and other areas. A member of the JASON group, he advised the federal government on defense issues. Project Orion was definitely “other,” a serious attempt at designing and building a spaceship the size of a small mountain, with a crew of 200, that would be powered by small nuclear bombs and capable of exploring the solar system in a matter of months. Orion was overtaken and doomed by the atmospheric nuclear test ban. More down-to-earth, Dyson led, at General Atomics in San Diego, the design of small, intrinsically safe nuclear reactors; many of those TRIGA reactors still exist.

At the same time, he also began to imagine the grand cosmic schemes for which he is now best known to the public: Dyson spheres, in which the mined metals of a single planet (Earth, say) were sufficient to entirely surround its star (the Sun, say) and capture a billion times more solar energy than the planet


Freeman J. Dyson

HEKA DAVIS/COURTESY AIP ESVA/PHYSICS TODAY COLLECTION

alone; or comets as a source of carbon, oxygen, and hydrogen for the growth, by photosynthesis, of genetically engineered giant sequoia trees—more for the delight of it than for any practical purpose. Dyson’s imagination embodied joy and occasional silliness.

One cannot read Dyson’s works without sensing a muted Enlightenment deism beneath the scientifically rigorous inventiveness. He credited to his mother his use of the phrase “world soul” and expressed several variants of the thought that intelligence, human or otherwise, played a special role in giving purpose to the universe. It was a controversial view for a physicist. His late-in-life contrarian stance on climate change—that it might be a good thing—made a kind of sense in the context of his seeming belief that the universe is not just purposeful but good. Always respectful of organized religion, Dyson was a recipient of the 2000 Templeton Prize, which celebrates “scientific and spiritual curiosity.”

Dyson, whom we remember for his lifelong technological optimism, ended life soberly, writing that today’s world—with persistent poverty, brutal dictators, and small wars “presaging worse horrors to come”—most reminded him of the year 1936, when he was 12. He did not live to see the effects of the coronavirus pandemic.

William H. Press
University of Texas at Austin

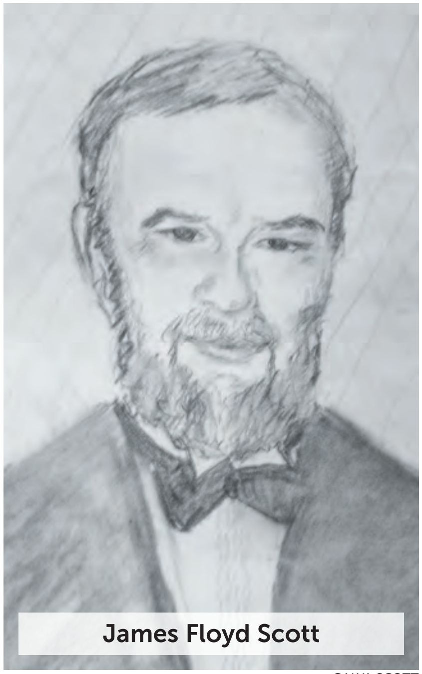
James Floyd Scott

With the passing of James Floyd Scott, the international ferroelectrics community is mourning the loss of its father figure. Just as ferroelectric materials might be polarized up or down and stay that way, Jim was unfailingly decisive and possessed an encyclopedic knowledge of his subject. Combining those qualities with great creativity, he had a remarkable career that was productive, distinguished, and richly decorated by an incredible gift for storytelling, which can be appreciated from the transcript of his 2018 IEEE interview (https://ethw.org/Oral-History:James_F._Scott).

Jim's work led to the integration of ferroelectrics with semiconductors, a demonstration of fatigue-free ferroelectric switching, the canonical 2000 book *Ferroelectric Memories*, and an army of "fellow electricians" who now carry the torch. Jim was a warm, generous, and friendly person to the majority, but he did not gladly suffer most fools or political correctness, as reflected in his eloquent demonstration that electrically leaky materials cannot be identified as ferroelectric because their hysteresis loops resemble those of bananas.

Born, in his words, "on the wrong side of the tracks" in Beverly, New Jersey, on 4 May 1942, Jim went on missions to the USSR and China in the 1970s, when those countries were closed. While visiting Pyotr Kapitsa in the USSR, he met Galya, whom he married in 1982. Many in the ferroelectrics community know her well and will be moved to see her sketch of Jim reproduced here. It is one of her early works, and yet it captures something of his essence.

Jim's early work did not focus on ferroelectrics, but each stage was increasingly relevant: Physics formed only part of his 1963 BA from Harvard University; high-resolution molecular spectroscopy was the subject of his 1966 PhD from the Ohio State University, under K. Narahari Rao; and at Bell Labs he was then well



James Floyd Scott

GALYA SCOTT

placed to perform five years of Raman spectroscopy and explain the soft mode in strontium titanate. Ferroelectrics began for Jim in 1970 with William Cochran at the University of Edinburgh in Scotland. Shortly thereafter, Jim started a 20-year career at the University of Colorado Boulder, where he performed groundbreaking research on ferroelectrics, served as assistant vice chancellor for research, won a teaching prize, and twinned Boulder with the city of Dushanbe, Tajikistan. Somewhat like head-to-head domains, those twins faced the iron curtain in opposite directions.

To integrate ferroelectrics with semiconductors for technological purposes, Jim cofounded Ramtron International Corp in 1984, and then in 1986 he cofounded Symetrix Corp, both in Colorado Springs. Within the next few years, Symetrix was working with the Matsushita Electric Corp in Japan to exploit Jim's clever idea that bismuth oxide planes can act as oxygen reservoirs for fatigue-free switching in strontium–bismuth tantalate. Jim subsequently exploited that material with Sony in Japan, but not before he had moved base to Australia in 1991. As the new millennium dawned, FeRAMs (ferroelectric random-access memories) began to enjoy commercial applications, first in Sony's Playstation 2 and later, notably, in Japanese railway fare cards.

Jim maintained his research output in Australia, first at RMIT University in Melbourne, where he started PhD programs in 14 departments and served as dean of applied science, and then at the University of New South Wales in Sydney, where he integrated the optometry research laboratory onto campus and was dean of science.

In 1999 Jim joined the department of Earth sciences at Cambridge University in England. With a loyal team, a small laboratory, and an ornate coffee table in the common room, he continued to make huge contributions to his core field of ferroelectrics while finding time to work with me on magnetoelectrics after what for him was a 30-year hiatus. As he told it, our serendipitous discovery of magnetoelectric effects in commercially manufactured multilayer capacitors arose from waterborne discussions while crossing the maritime city of Kiel, Germany. Our review article eventually garnered even more citations than Jim's paper on fatigue-free switching, and if anyone was keen on citation numbers, it was Jim. His prolific output was bolstered by working with a range of collaborators around the world, and he continued in that vein after moving in 2009 to the Cavendish Laboratory, where he worked with the locals on quantum criticality.

Jim's ferroelectric odyssey ended, as it began, in Scotland—the country from which his name is fittingly derived. He spent five productive years at the University of St Andrews before returning home to Cambridge in poor health at the dawn of 2020, and he passed away on 6 April. For decades he had referred to various medical problems while noting with gallows humor that none were trivial. This attitude lent him an aura of physical invincibility that nicely complemented his intellectual mettle. Jim lived through an amazing period in history and witnessed horse-drawn carriages dominating the streets of Beijing through to the advanced electronics in which he played a part.

I fondly recall the good times spent with Jim over coffee in the Earth sciences common room and how he was the most wonderful traveling companion and roommate on long-haul economy trips. Jim enabled all of my current research on electrocalorics and magnetoelectrics, and I didn't even work for him.

Neil D. Mathur
University of Cambridge
Cambridge, UK

PT

TO NOTIFY THE COMMUNITY
about a colleague's death, visit
https://contact.physicstoday.org
and send us a remembrance to post.
Select submissions and, space permitting, a list of recent postings will appear in print.

Ralph Lorenz is a planetary scientist at the Johns Hopkins University Applied Physics Laboratory in Laurel, Maryland. He studies surface–atmosphere interactions, including sand dunes and air–sea exchange, on the terrestrial planets and Titan. His work on dust devils is supported by NASA's *InSight* Participating Scientist Program.

Dust devils on Mars

Ralph D. Lorenz

Whirlwinds, familiar on Earth, can be informative probes of the Martian surface environment.

Airplane pilots and parachutists have been killed by them; Mars rovers have been saved by them. Dust devils, whirlwinds that stalk the deserts of Earth and Mars—and perhaps Venus and Titan as well—are the producers of those unlikely results. They are convective vortices, driven by solar heating of the ground and made visible by lofted surface material.

Unlike tornados and hurricanes, dust devils are dry, powered by the sensible heat from the ground rather than by the latent heat of water. So they occur most frequently on hot summer afternoons. As air is drawn into a narrow column rising through surrounding cooler air, its rotation speed is boosted by the conservation of angular momentum. Unlike hurricanes, dust devils are too small for a planet's rotation to influence their circulation: Equal numbers of randomly cyclonic and anticyclonic devils generally exist, regardless of hemisphere.

Although dust devils cause occasional damage, injury, and even death—usually by light structures such as aircraft, barn doors, or sheds—they tend to be considered idle curiosities of terrestrial weather. But they are a prominent feature of Mars's meteorology, and their prevalence has stimulated study of their counterparts on Earth.

Pressure drops

A typical dust devil on Earth spans a few meters to tens of meters in diameter, with a height 5 to 20 times as large. Its height is capped by the top of the planetary boundary layer (PBL), usually between 0.2 km and 2 km thick, where air can mix in contact with the ground. Martian devils on average are about three times as large as their terrestrial counterparts, in part because the Martian PBL is deeper.

The word "typical" must be used cautiously, because dust-devil diameters and the core-pressure drops that define a whirlwind's intensity have highly skewed distributions. Much like in earthquakes, those properties follow power-law statistics. But while you might see a hundred times more 5 m dust devils than 50 m ones, the single largest (and usually longest-lived) devil in the set may lift more dust than the rest put together. Many vortices can exist on a landscape but be invisible as "dustless devils" if fine surface particulates are not abundant or too sticky.

Pressure data can easily detect vortices (see figure 1). A pronounced dip in the pressure record on Earth or Mars gives them away. Occasionally, the dip reaches 1–2% of ambient pressure; by comparison, the most violent terrestrial tornadoes and hurricanes experience pressure drops of 10–20%. Other signatures include infrasound emissions and electric fields caused by charged dust. The swirled dust and sand can produce atmospheric oxidants, such as ozone, either via catalysis on abraded surfaces or through electrical discharges. On Mars, the production of oxidants may be important in the destruction of methane and other organic compounds.

Prevalence and practicality

Like London buses, dust devils appear quasiperiodically. High-resolution numerical simulations show that in the con-

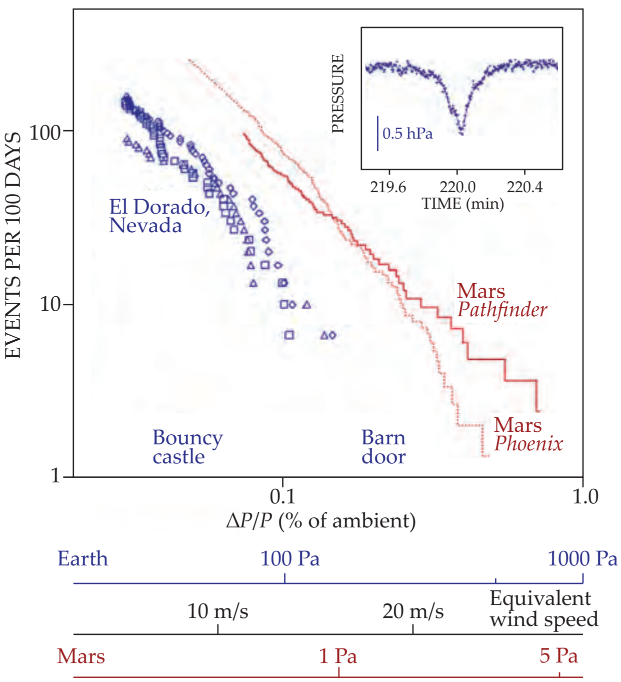


FIGURE 1. PRESSURE DATA. Vortices are most easily identified in pressure (P) records, in which a dust devil appears as a transient dip (inset). Deeper pressure drops, associated with stronger winds, occur less often, roughly according to a power-law distribution. On Earth, the pressure drops occasionally reach the weight per unit area of light structures, such as a barn door or an inflatable bouncy castle. The absolute pressure drops on Mars, where the ambient pressure is one-hundredth as large, are much smaller. Because the pressure drop provides the centripetal acceleration that moves the air in a circle, the fractional pressure drop relates to the wind speed on both planets. (Figure from Ralph Lorenz.)

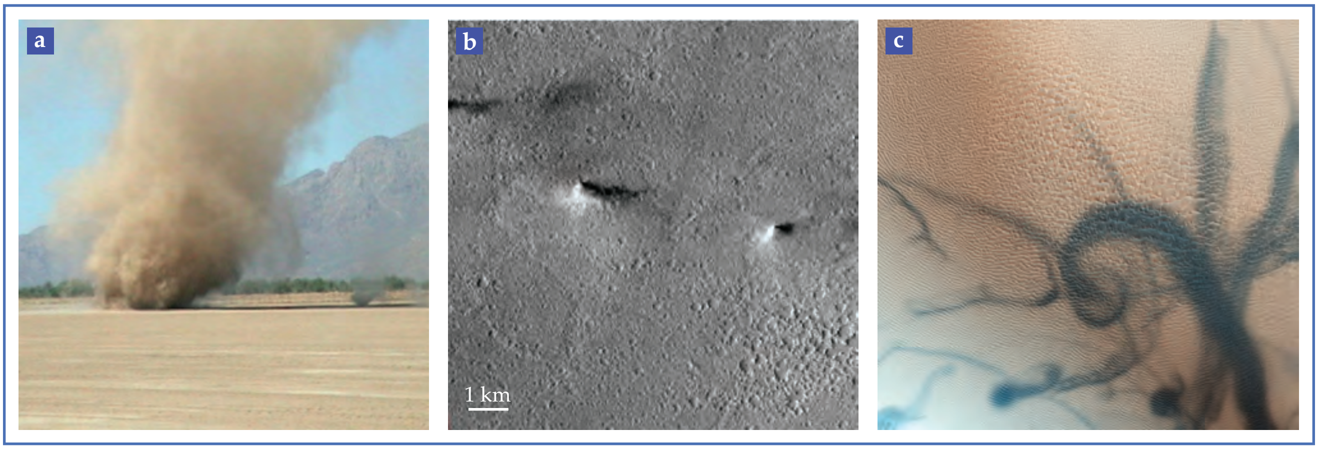


FIGURE 2. TRACKS IN THE WIND. (a) This desert dust devil is typical on Earth. It has a width of a few meters and sits next to a smaller devil at right. **(b)** Four Martian dust devils and their shadows are photographed from an orbiting spacecraft. Their regular spacing suggests a planetary boundary layer depth of 6 km. **(c)** In another orbital view of Mars, swirling cycloidal dark tracks appear from dust devils moving in light winds. The sand ripples are 1-4 m apart. (Images courtesy of NASA.)

vecting PBL, atmospheric cells circulate with broad, downwelling centers and upwelling sheets of warm air at their edges. Vortices form preferentially where those sheets join (at the corners of the cells), and the width of the cells, approximately equal to the PBL depth, can define a characteristic spacing of vortices (see figure 2b). Indeed, pressure records sometimes show periodic signatures as the cellular pattern drifts across a fixed meteorological station. Dust devils are sometimes abundant enough to mark out the regular spacing of the cell edges.

Dust-devil "whisperers" like me have learned how to remotely diagnose the wind conditions and composition of surface material from the shape and path of individual dust devils. When a mixture of sand and dust is present, a conical "sand skirt" may form, as sand too heavy to be lofted skips around the base of the devil. And over a deposit of uniform fine dust, a tall, perfect cylinder can appear with a clear center—the suspended dust centrifuged to the perimeter as it rises in the updraft.

A strong ambient wind can cause the dust column to tilt downwind and migrate in a nearly straight line. In light winds over a horizontal surface, a dust devil may wander randomly—sometimes curling like honey drizzled from a spoon—whereas on a slope, it tends to move uphill. Those migration patterns are frequently revealed on Mars in images from orbit, as shown in figure 2c, and they allow astronomers to remotely discern wind patterns and vortex activity over previous days to months.

Dust devils and the vortices that beget them are important for both climatological and practical reasons. On dry Mars, roughly 30% of the dust that hangs in the atmosphere is put there by dust devils. That dust, in turn, warms the atmosphere and regulates its ability to hold water vapor. It can thus influence the important exchange of volatiles between the atmosphere and polar caps. Furthermore, while the bright red dust settles, dust devils can scour it away to expose darker sands beneath. The effect of that change in albedo may be enough to modify the local climate.

The local removal of dust is a fortunate side effect for solar-powered Mars rovers and landers. In 1997 the *Sojourner* rover saw steady dust accumulation cut its power by 0.3% per day, an experience that conditioned NASA's expectations for the life span of later rovers *Spirit* and *Opportunity*. Although the power of those rovers similarly declined, it unexpectedly recovered from time to time. Some phenomenon had cleared dust from the panels and allowed the missions to persist for several thousand days—years beyond the rovers' expected 90-day lifetime.

The dust-clearing events occurred when solar heating of Mars's surface was greatest and with a frequency of once every few hundred days. Those conditions are consistent with the prevalence of dust devils, but direct attribution was impossible until this past year. In February 2019 the *InSight* mission's meteorological instruments recorded a direct hit with a strong vortex at the precise moment some of its solar panels increased their power output by 1–2%.

Dust devils at the *InSight* landing site are proving valuable in another respect. A vortex is a low-pressure system, and when one integrates the pressure field across the footprint of a typical devil, the weight can amount to that of a truck. Seismometers on the lander are sensitive enough at low frequencies to record the tiny tilt the ground makes as it deforms elastically under that (negative) load. Although atmospheric signals of all kinds influence seismometers, the localized and symmetrical nature of dust-devil loads makes them particularly convenient probes of the near-surface stiffness of Mars. Better the devil you know. . . .

Additional resources

▶ L. K. Fenton, R. Lorenz, "Dust devil height and spacing with relation to the Martian planetary boundary layer thickness," *Icarus* **260**, 246 (2015).

▶ R. D. Lorenz, *Exploring Planetary Climate: A History of Scientific Discovery on Earth, Mars, Venus and Titan*, Cambridge U. Press (2019).

▶ R. D. Lorenz et al., "History and applications of dust devil studies," *Space Sci. Rev.* **203**, 5 (2016).

▶ W. B. Banerdt et al., "Initial results from the InSight mission on Mars," *Nat. Geosci.* **13**, 183 (2020).

BACK SCATTER

Burning branches

Molecular hydrogen is highly flammable, even when it's mixed with air in concentrations as low as 4%. That property makes it an attractive alternative energy source to fossil fuels. But compared with other hydrocarbons, hydrogen requires one-tenth the energy to ignite, which makes it more prone to dangerous explosions, especially in locations with poor ventilation. To better investigate how hydrogen flames propagate through small spaces, Fernando Veiga-López and Mario Sánchez-Sanz, both with the University Carlos III of Madrid in Spain, collaborated with Mike Kuznetsov of the Karlsruhe Institute of Technology in Germany and other colleagues.

The researchers piped a mixture of hydrogen and air between two transparent plates separated by a few millimeters. The image here shows the various paths the flames took to travel downward before the hydrogen fuel was depleted. The flames themselves are hard to see; the visible branching pattern is the result of the condensed water trail that formed as combustion oxidized the hydrogen. In two additional experiments, the flames fueled by methane and by dimethyl ether, both mixed with air, quenched more quickly than hydrogen. The lightest element's exceptional mass diffusivity explains why its flames persist: In narrow spaces, hydrogen diffuses quickly enough to counteract the conductive heat losses and sustain the temperature necessary for continued combustion. (F. Veiga-López et al., *Phys. Rev. Lett.* **124**, 174501, 2020; image courtesy of Fernando Veiga-López.) —AL

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