SPECIAL COLLECTOR'S EDITION

SCIENTIFIC AMERICAN

Probing the Mysteries of the Cosmos

EXTREME PHYSICS

INSIDE

Gravitational waves are about to supercharge physics

At the edges of the quantum universe

String theory under fire

The oldest supermassive black holes

PLUS

Are mysterious particles the "glue" of reality?

SPRING 2019

COVER IMAGE: KENN BROWN AND CHRIS WREN *Mondolithic Studios;*
THIS PAGE: GETTY IMAGES

Scientists have only recently gathered definitive evidence of the existence of black holes.

FROM THE EDITOR

Physics at the Limits

To outsiders, the field of physics may seem like a neat and tidy affair. What fundamental discoveries are left to make, right? Wrong. In fact, physicists are pushing into the extreme ends of the universe as we know it—from invisible particles and colliding massive black holes to the most crushing gravitational forces ever detected and spooky quantum entanglement. The 14-billion-year-old tale of our universe is far from over.

Innovative projects are hunting the smallest units of matter. Scientists at CERN's Large Hadron Collider have set their sights on "virtual" particles that seem to pop in and out of existence and defy the laws of physics (*page 4*). And soon the largest experiment of its kind will beam minuscule neutrinos underground from Illinois to South Dakota, with luck demonstrating how these mysterious particles buck the Standard Model (*page 24*).

On the larger scales of physics, the detection in 2015 of two black holes crashing together some 1.3 billion light-years away launched a wave of discovery, and new data are rolling in at a frequency of months rather than years (*page 46*). Meanwhile the debate rages over whether invisible particles indeed make up the dark matter that fills the entire universe or whether we need to revise what we know about how gravity works altogether (*page 54*).

Thanks to new gravitational-wave detectors, we are finally getting a clearer picture of some of the densest places in existence; the core of neutron stars may prove to be a goopy superfluid (*page 68*). An alternative theory posits that rather than arising from dead stars, the earliest black holes materialized out of clouds of dust (*page 62*).

Quantum theory continues to captivate; theorists are trying to get their head around the ramifications of a multiverse—that is, do infinite worlds coexist in a statistical "probability space" (*page 84*), or do too many universes violate the laws of physics entirely (*page 108*)? The next phase of quantum research will attempt to put living organisms into superposition, probing for where precisely scientists can observe quantum effects (*page 100*).

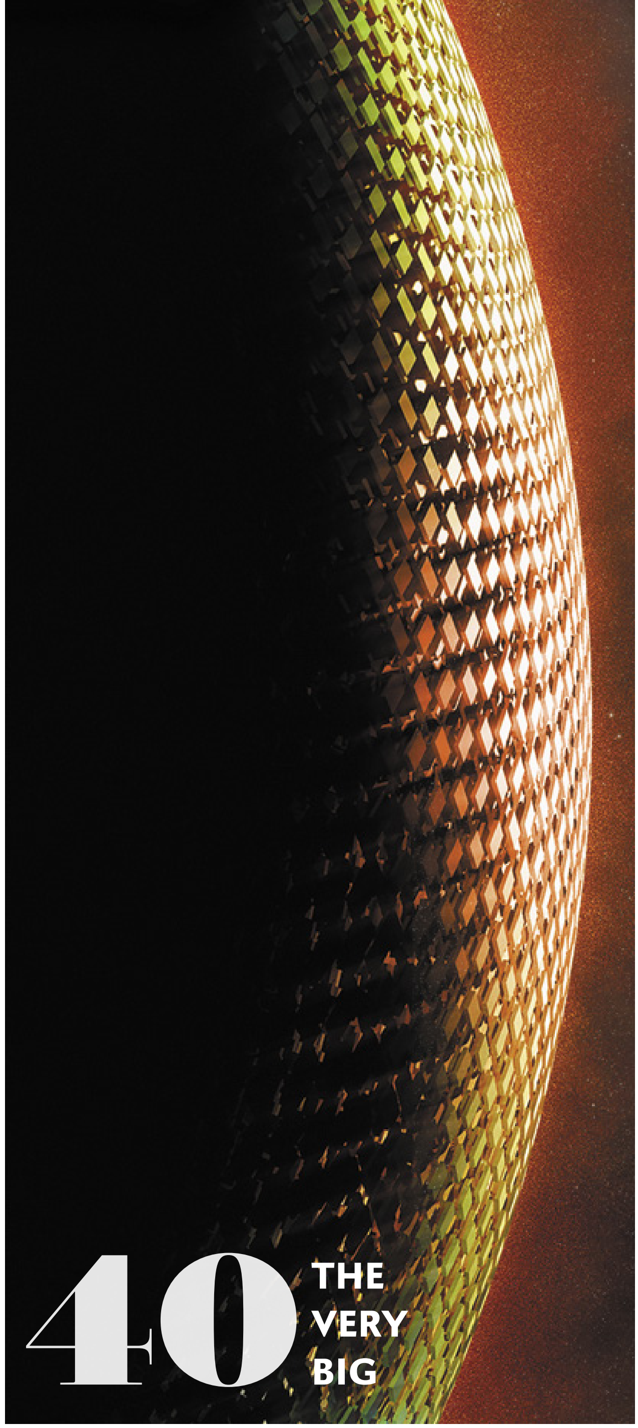
Perhaps more than any other field, physics requires a churning of old ideas, either modulating them to explain the latest observations or scrapping them altogether. Much of its work is done in the human mind as *Gedankenexperiments,* or "thought experiments," as Albert Einstein called them. For this reason, Sabine Hossenfelder writes on page 80, there is a fine, often blurry line between scientific intuition and fantasy. But there seems to be no finish line of discovery, even on the distant horizon.

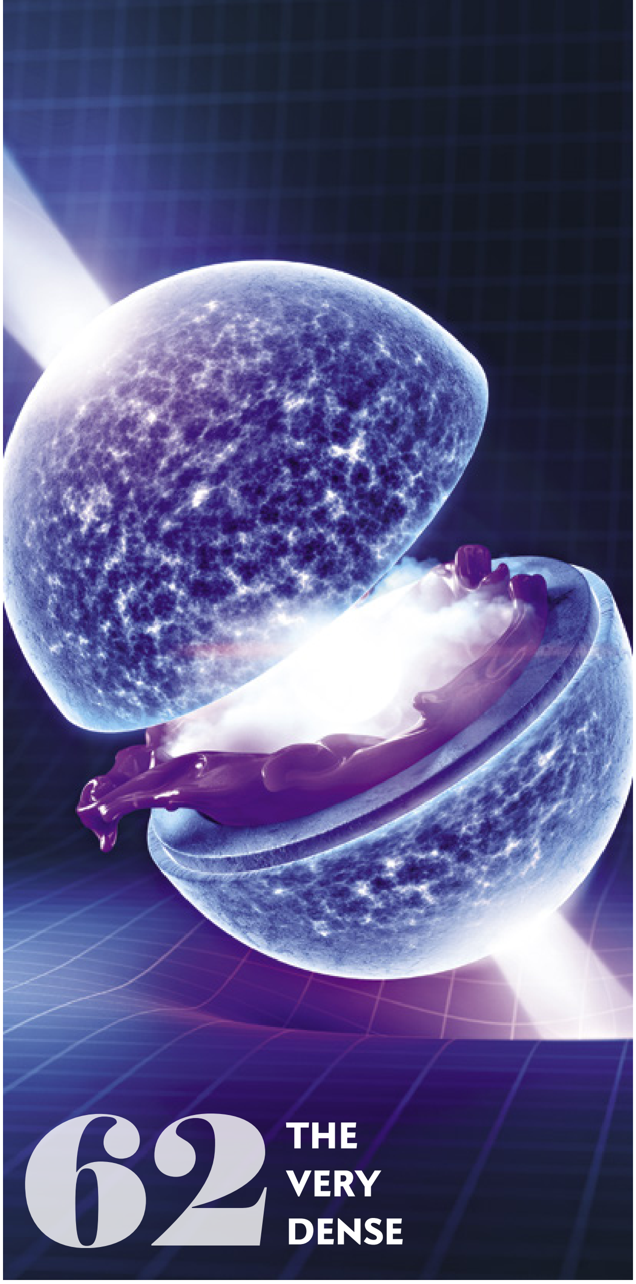
Andrea Gawrylewski
Collections Editor
editors@sciam.com

SPECIAL EDITION

SCIENTIFIC AMERICAN

Volume 28, Number 2, Spring 2019





DEPARTMENTS

Articles in this special issue are updated or adapted from previous issues of *Scientific American* and *Nature* and from ScientificAmerican.com.
 Scientific American Special (ISSN 1936-1513), Volume 28, Number 2, Spring 2019, published by Scientific American, a division of Springer Nature America, Inc., 1 New York Plaza, Suite 4600, New York, NY 10004-1562. Canadian BN No. 127387652RT; TVQ1218059275 TQ0001. To purchase additional quantities: U.S., $13.95 each; elsewhere, $17.95 each. Send payment to Scientific American Back Issues, 1 New York Plaza, Suite 4600, New York, NY 10004-1562. **Inquiries: fax 212-355-0408 or telephone 212-451-8415. Printed in U.S.A.**

VIEW INTO THE CAVERN housing the LHCb experiment, where protons collide to produce particles containing beauty quarks.

The Large Hadron Collider beauty experiment has seen hints of new particles that may point the way toward a higher theory of physics

MEASURING Beauty

By Guy Wilkinson

IN BRIEF

The LHCb experiment at CERN's Large Hadron Collider is searching for undiscovered particles that may illuminate new truths about how nature operates at its tiniest scales.

Instead of aiming to produce these new particles directly, LHCb scientists are hoping to detect the influence of "virtual" particles that pop briefly in and out of existence and influence conventional matter.

Already the experiment has shown hints of odd particle behavior that cannot easily be explained by current laws of physics. More research will determine if these are the first glimpses of new lands on the particle map.

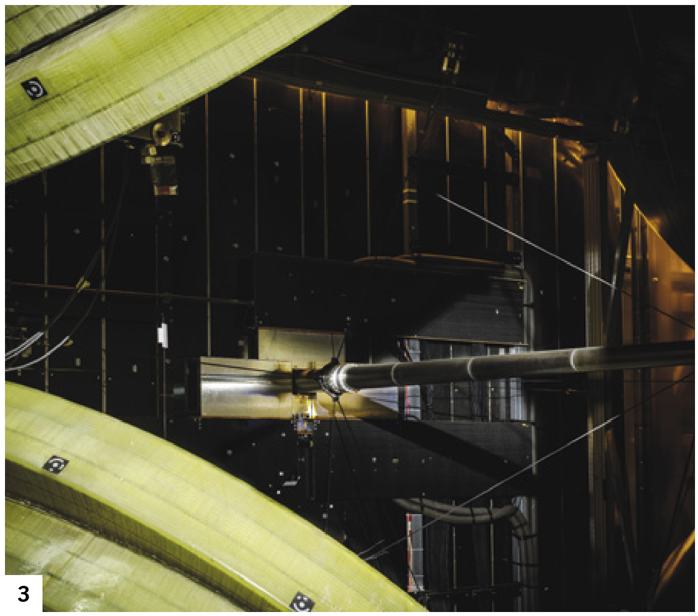
Photographs by Alastair Philip Wiper

IT IS UNUSUAL FOR TV NEWS to open with a story about physics, but it happened on July 4, 2012, when all around the world stations chose to devote prime time to breaking news from Geneva: a search of almost 50 years had ended with the discovery of the Higgs boson particle by the Large Hadron Collider (LHC) at the CERN physics laboratory. For experimentalists, the Higgs was the last and most important missing piece in the trophy cabinet of the Standard Model of particle physics—the theory describing all the known particles in the universe and the forces between them. Yet physicists believe there may be more elementary particles than those in the Standard Model, and a new and even more challenging hunt is on to find them.

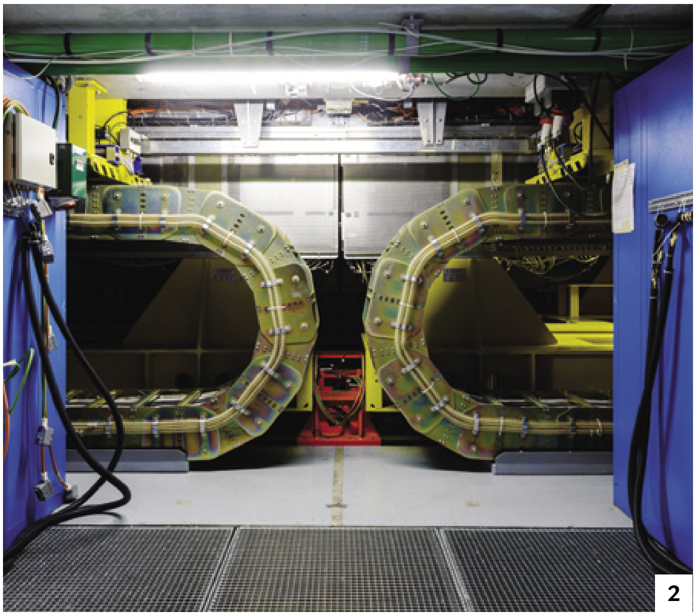
Like the quest for the Higgs, the race to discover hidden particles, thereby building a fuller picture of nature at its tiniest scales, is taking place at the LHC. The experiments that discovered the Higgs—ATLAS and CMS—will play an important role, but LHCb, a smaller and less well-known project operating at the same accelerator, brings guile and stealth to the chase. There is a real chance that this third experiment may be the first to bring home the prize.

LHCb follows a different game plan than most pursuits of new particles. Whereas ATLAS, CMS and many other efforts try to create undiscovered particles directly, the LHCb experiment on which I work uses so-called beauty hadrons to look for the signatures of unseen particles that we cannot directly produce

1

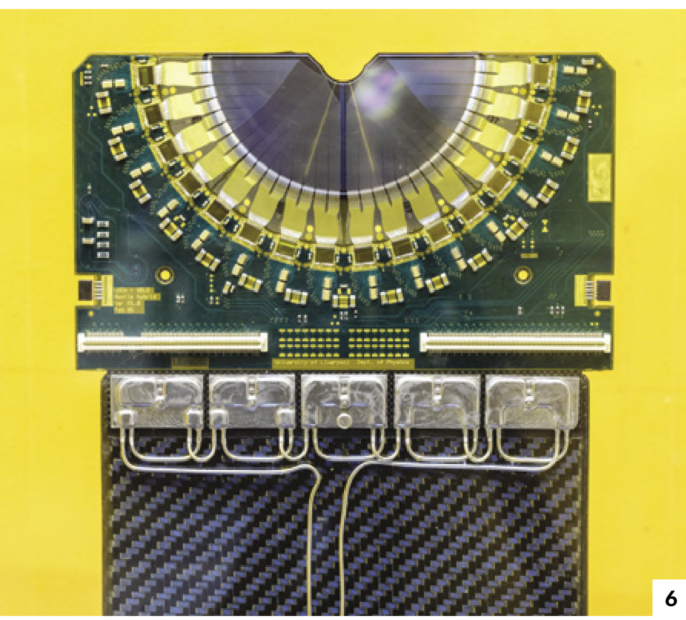

3


5



2

4



6

LHCb, seen from the side (*1*) and underneath (*2*), studies collisions of protons that travel through a beam pipe (*3*) into the experiment. Inside the control room (*4*), physicists monitor operations. Computer processors (*5*) determine which reactions to record for analysis. The collisions occur inside the delicate Vertex Locator (VELO), which uses silicon sensors (*6*) to detect beauty particles.

but that affect reactions behind the scenes. LHCb (the "b" stands for "beauty") studies what happens when beauty hadrons are created in the Large Hadron Collider and then decay into other particles. Beauty hadrons make excellent test subjects because they decay in a huge variety of ways, and physicists have very precise predictions about how these reactions should proceed. Any deviation from those predictions is a clue that we might be seeing interference from unknown particles.

This type of search is complex and requires great precision, but it has the potential to uncover particle species that are impossible for ATLAS and CMS to access. Already it has turned up several intriguing hints of phenomena that threaten to defy the laws of physics as they are currently written. We may be witnessing the actions of particles or forces in nature that physicists have never before observed and possibly never even imagined. If so, our investigations at LHCb could reveal the workings of the cosmos on a more fundamental level than humans have ever glimpsed before.

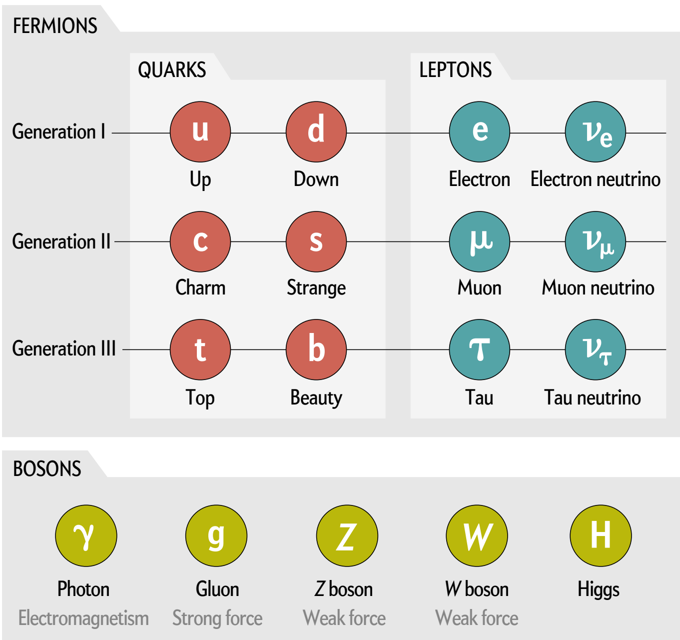
AN INCOMPLETE THEORY

THE STANDARD MODEL has been highly successful at describing the behavior of the elementary particles of nature and the forces that act on these particles. It divides the elementary particles into quarks and leptons. There are six quarks arranged in three groups, called generations: up and down, charm and strange, and beauty (also called bottom) and top. We never see these quarks in isolation; rather they cluster together in so-called hadrons—beauty hadrons, therefore, are particles containing beauty quarks. Likewise, there are three families of leptons: the electron and electron neutrino, the muon and muon neutrino, and the tau and tau neutrino. The up and down quark and the electron—all from the first generation—make up the atoms of everyday matter. The particles belonging to the other two generations tend to be more elusive; we must use particle accelerators to coax them into existence. The forces that act on these particles—excluding gravity, which is unimportant at the subatomic level—are electromagnetism, the weak force and the strong force. Each force is transferred by an additional particle: for example, the photon carries electromagnetism, and the *W* and *Z* bosons deliver the weak force. Alongside all of these, the Higgs boson sits alone, the manifestation of an underlying field that gives some particles mass.

And yet physicists know that the Standard Model must be wrong. "Wrong," though, is an extreme word; rather we prefer to say that the theory is incomplete. It succeeds very well in answering certain questions but has nothing to say about others. At the cosmic level, it cannot explain why the universe is overwhelmingly constituted of matter, whereas in the big bang, matter and antimatter must have been created in equal proportion. Nor can it tell us anything about the nature of dark matter, the extra mass in the universe that we cannot see but that we know must be there to

The Standard Model

The known particles and forces in the universe make up the Standard Model of particle physics. It includes six kinds of quarks and six types of leptons, as well as five bosons, which transfer the forces of nature. But physicists believe there are more particles out there than those in the Standard Model, and they aim to find them through projects such as the Large Hadron Collider beauty experiment.



drive the observed motion of the stars and galaxies. Indeed, the Standard Model does not include gravity, the dominant force on large scales, and all attempts to include it so far have failed.

And even in the world of the known subatomic particles, many puzzles remain. The Higgs boson happens to have a mass not much larger than the *W* and *Z* bosons, whereas the Standard Model suggests it should be about 10,000 trillion times heavier. There is no reason that we can discern for the three-generation arrangement of the matter particles. The generations appear to be copies of one another, except for the fact that there is a striking hierarchy of mass, from the up and down quarks, which "weigh" very little, to the top quark, which is almost as heavy as a gold nucleus. On these and many other questions, the model is silent. Hence, despite its long track record of success, the Standard Model must still be only an approximation, the visible facade of a higher theory that we hope will yield solutions to these puzzles. Our goal at LHCb, along with ATLAS, CMS and many other experiments around the world, is to discover elements of that higher theory in the form of particles that exist in nature but have not yet revealed themselves to us.

THE BEAUTY EXPERIMENT

THE LARGE HADRON COLLIDER, home to LHCb, is a 27-kilometer-long, ring-shaped accelerator in which two beams of high-energy protons circulate in opposite directions at close to the speed of light. Inside LHCb these beams collide up to 40 million times per second. The dense points of energy that are formed when the protons smash together and annihilate one another can condense into particles that are very different than the protons that collided—for example, particles containing beauty quarks. Even if they are very short-lived, these new particles spring into existence and then decay into products that LHCb can detect.

The LHCb experimental site sits approximately four kilometers from the main CERN lab, nestled against the perimeter fence of the Geneva Airport. The surface buildings are functional in design and mostly inherited from a previous experiment. A large, circular window, a sole concession to aesthetics, allows passengers looking out from planes on the nearby runway to easily spot the main hall. Inside one of these buildings, in a well-appointed control room, physicists sit day and night monitoring the status of the experiment, which is situated in a cavern 100 meters below.

Although modest in size compared with its bigger siblings around the LHC ring, the LHCb detector is still an imposing and impressive sight spanning around 20 meters in length and 10 meters in height. Its elongated design gives LHCb a very different appearance to the cylindrical geometries of ATLAS and CMS and allows it to record the signals of particles produced close to one wall of the cavern. This stretched geometry helps in the study of beauty hadrons, which are particles containing beauty quarks. Because of their relatively modest mass (around 5 GeV, or giga electron volts, which is only a little heavier than a helium nucleus), when beauty hadrons form at the LHC there is always plenty of surplus energy left over. This extra energy tends to throw the newly created beauty quarks forward from the collision point into the detector. Despite its unusual layout, LHCb has many of the same components as other experiments. These include a large magnet, tracking stations to reconstruct the trajectories of particles produced in the collisions and calorimeters to measure the particles' energies.

But several attributes are unique to LHCb and are designed specifically for beauty physics. For instance, a silicon-strip detector placed just eight millimeters from the LHC particle beams can reconstruct the position of a particle decay with great precision—a useful tool because beauty hadrons typically fly forward just a centimeter or so before decaying into a collection of lighter particles. LHCb also has a system of so-called RICH (*r*ing-*i*maging *Ch*erenkov) counters, which can determine the identities of the beauty hadron decay products based on the patterns of light many of them emit.

THE SEARCH FOR NEW PHYSICS

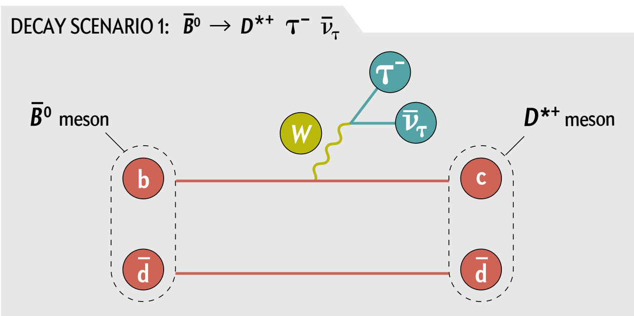
DURING THE LHC'S FIRST RUN, from 2010 to 2012, the accelerator produced almost a trillion beauty hadrons inside our experiment. These particles can decay in a huge number of ways, some of which are more interesting than others. We are looking for decays that may serve as signposts to "new physics"—behavior that the Standard Model cannot explain.

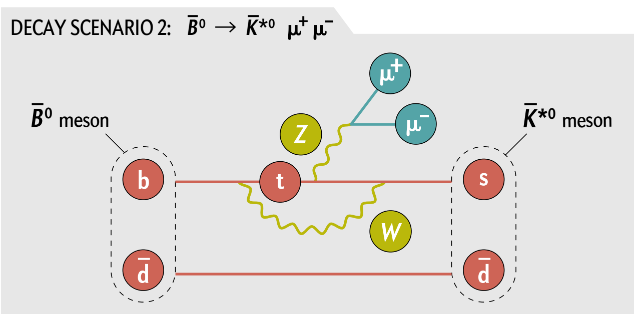
Illustrations by Jen Christiansen

Theoretical physicists have many hypotheses for what this theory could be, but most ideas involve new particles that are somewhat heavier than those we know of. This heaviness is one excellent reason the LHC is so well equipped to seek new physics: the high energy of its collisions means that it can produce and detect rather massive particles, up to a few thousand GeV in equivalent mass (by way of comparison, the Higgs boson weighs around 125 GeV and the humble proton 0.9 GeV). The ATLAS and CMS experiments have been designed to search directly for such massive particles through the distinctive signatures their decays would create. Yet there is another, more cunning way to look for new physics. We can detect the presence of new particles through their "virtual" effects on the decay of Standard Model particles.

To understand the idea of virtual particles, we must turn to Feynman diagrams [*see boxes below*]. The renowned 20th-century American theoretician Richard Feynman invented these diagrams as a way to visualize and calculate the decays and interactions of subatomic particles. Here we will examine the Feynman diagrams of two possible decay paths of beauty hadrons (particles that unfortunately tend to be called by rather ungainly conglomerations of Greek letters and symbols).

In both examples, we start with a so-called $\bar{B}^0$ (pronounced "b zero bar") meson, a hadron composed of a beauty quark and an anti-down quark (antimatter particles are denoted with the suffix "bar"). In the diagrams, time runs from left to right. In the first case, we can see that our starting meson decays into a D^{*+} meson (made of a charm and an anti-down quark), a negatively charged tau lepton (τ^-) and an anti-tau neutrino ($\bar{\upsilon}_\tau$); hence, the process is designated $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\upsilon}_\tau$. The other decay, $\bar{B}^0 \rightarrow \bar{K}^{*0}\mu^+\mu^-$, produces a $\bar{K}^{*0}$ meson (built of a strange quark and an anti-down quark), a muon and an anti-muon. The law of conservation of energy, as well as the equivalence of mass and energy as described in Albert Einstein's famous equation $E = mc^2$, requires that these final particles have a total mass that is less than that of the initial beauty meson. The difference in mass turns into the kinetic energy of the decay products.





Let us focus on what is happening at the heart of the diagrams, where the decay occurs. In the first case, we see a *W* boson, one of the particles that carries the weak force, appearing at the point where the beauty quark transforms into a charm quark. This *W* boson then decays into a tau and anti-tau neutrino. What is striking is that the *W* is around 16 times more massive than the initial $\bar{B}^0$ meson. Why does its appearance in the decay process not violate the rule of energy conservation? According to the mysterious accounting of quantum mechanics, such violation is actually allowed as long as it happens over a sufficiently short timescale! In this case, we say that the *W* boson is *virtual*. Now turning to the $\bar{B}^0 \rightarrow \bar{K}^{*0}\mu^+\mu^-$ decay, we see that the decay process is more complicated, involving a loop structure and three internal points of decay. But here, in addition to a *W*, several other virtual particles also participate: a virtual top quark (t) and a virtual *Z* boson, both much more massive than the initial meson. Virtual particles may sound fanciful, but the rules of quantum mechanics allow us to draw such diagrams, and these diagrams have proved correct time and time again at predicting the probability that these decays will occur. Indeed, it was by such methods that physicists first predicted the existence of the charm quark and the top quark and made the first estimates of their mass.

The diagrams we have discussed represent only two possibilities for how those particular decays can proceed. We can imagine others, some with particles we have never seen tracing the path between the internal decay points or even finding different ways to link the initial and final state particles. And what is amazing is that all these possibilities matter. The rules of quantum mechanics tell us that what happens in nature is driven by the net contribution of *all* the valid diagrams we can draw, although the simplest and most obvious have the greatest weight. Hence, all these possible decay paths should play a role, and we must account for them in the calculations we make predicting the rate of the decay, the trajectories of the products and other particulars. In other words, even when a particle decays in a normal process involving only conventional members of the Standard Model, it feels the effects of every possible particle out there. Therefore, if a measurement of a decay disagrees with our calculations based only on the Standard Model ingredients, we know that something else must be at work.

This fact is the guiding principle behind LHCb's strategy of indirect searches for new particles and new physics. Because these new particles would be virtual participants in every decay that we measure, the mass of the particles we can detect is not limited by the energy capacity of our accelerator. In principle, if we studied the right decay processes with enough precision, we could observe the effects of particles even heavier than those that can be created and detected within ATLAS and CMS.

CRACKS IN THE STANDARD MODEL

MY COLLEAGUES at LHCb and I have already seen hints that all might not be well with the Standard Model description of beauty hadron decays. The clues come from a variety of measurements, but they all share some common signatures. It is important to emphasize that with more data and a better understanding of the theory, we might find that the Standard Model *does* in fact agree with our findings. Even if this turns out to be the case, though, these early hints illustrate how cracks in the Standard Model edifice may develop and widen.

Exhibit A concerns the $\bar{B}^0 \to D^{*+}\tau^-\bar{\upsilon}_\tau$ decay that we discussed earlier and the possible violation of a rule called lepton universality. In the Standard Model, the *W* boson has the same probability of decaying into a tau lepton and its antineutrino as it has of decaying into the members of the muon and electron families (after we account for the different masses of the tau, muon and electron). In other words, the rules of *W* decay should be universal for all leptons. But at LHCb, after we counted the decays in each category, subtracted any processes that might fake the signals of these decays and corrected for the fact that not all decays are observed, we found that beauty hadrons appear to be decaying into taus rather more often than the Standard Model says they should.

Our results are not yet conclusive; the discrepancy we found has a strength of "two sigma," where "sigma" denotes uncertainty. Because of statistical fluctuations, one-sigma effects are not infrequent in experimental science, and physicists really only sit up and take notice when three-sigma deviations occur. Five sigma is the commonly adopted benchmark for announcing the discovery of a new particle or declaring that a prediction is wrong. Hence our two-sigma effect is not so remarkable—unless you consider what physicists are finding at other experiments.

Researchers have also looked for violations of lepton universality at BaBar and Belle, two beauty physics experiments in California and Japan, respectively, that collected data in the first decade of the millennium. The results from these experiments consistently favor taus in the same decays we measured as well as similar processes. Furthermore, at LHCb we made a new measurement of lepton universality in these decays earlier this year using a different technique, and once again we found that taus come in slightly above expectations. Altogether this ensemble of measurements gives a result that is separated by four sigma from conventional predictions. This is one of the most striking discrepancies in all of particle physics and constitutes a real problem for the Standard Model.

What could be going on? Theorists have some ideas. A new type of charged Higgs particle, for example, could be involved. Higgs bosons do not respect lepton universality, and they decay preferentially into particles of higher mass, hence favoring the production of tau particles. Yet the exact size and pattern of the discrepancies we see do not fit neatly into the simplest theories that predict such additional Higgs species. Another, even more exotic explanation would be a leptoquark, a hypothetical particle that can allow quarks and leptons to interact. Finally, of course, the results we are seeing could be an experimental effect caused by a misunderstood signal masquerading as the decays we are looking for. To sort through these possibilities, we need new, more precise measurements. We expect several in the coming years, from LHCb as well as from the new-generation Belle II experiment that began operation in April 2018.

Our next example showing hints of new physics comes from the decay $\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^-$, which we discussed earlier. Decay processes of this kind are an excellent place to search for signs of new physics for two reasons. First, the "loopy" structure at the heart of the Feynman diagram immediately tells us that elaborate gymnastics are necessary for the decay to occur in the Standard Model; however, new physics particles might have an easier time bringing the process about, and hence their presence may be more evident. Second, this decay has many properties that we can measure: we can note the rate at which the process occurs, as well as the angles and energies of the decay products and other types of information. We can then build these properties into various "observables"—quantities that we can compare directly with Standard Model predictions (but that, unfortunately, do not always equate to properties that are easy to picture).

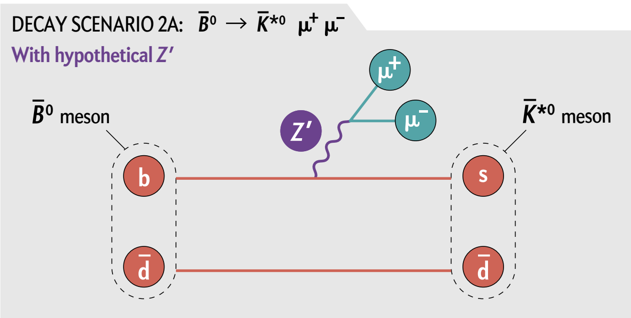
In many ways, $\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^-$ is the poster child of beauty physics, with its virtues evident by the huge body of theory papers that were written about it well before the LHC even turned on. The only thing that this decay lacks is a decent nomenclature, as the names used to label the different observables are rather underwhelming, such as "P_5'" (pronounced "p5 prime"), which is nonetheless the hero of our story.

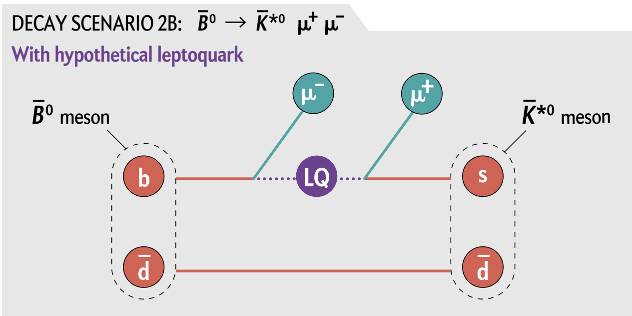
We made a first analysis of P_5' with some of the early LHCb data, measuring this observable for different categories of the decay characterized by the directions and energies of the pair of muons produced in the end. For certain configurations we found a significant discrepancy between predictions and our observations. Based on these first results, the physics community eagerly awaited the updated analysis we unveiled a couple of years later using the complete run-one data set. Would the discrepancy persist, or would it prove to be a statistical fluke? It remained. The size of the effect is now around 3.5 sigma, which is not large enough to justify ordering champagne but certainly sufficient to be taken seriously. And we find further encouragement from the fact that measurements of other observables in similar decay processes also exhibit intriguing discrepancies. Altogether the total disagreement with the Standard Model rises to as much as 4.5 sigma—a problem for the theory that we cannot ignore.

Theorists have come up with a whole swathe of potential new physics explanations for this effect. The leptoquark, already invoked in the $\bar{B}^0 \to D^{*+}\tau^-\bar{\upsilon}_\tau$ decay, is a possibility. Another is a Z' ("z prime") particle, which would be an exotic, heavier cousin of the well-known *Z* boson but one that decays into quarks and leptons in its own distinctive manner. Such speculation, however, must always respect the constraints that already exist from other measurements. For example, the mass and behavior of these hypothetical new particles must be such that it makes sense that they have not yet shown up in direct searches at ATLAS and CMS.

Theorists are nothing if not ingenious, and there are plenty of plausible scenarios that satisfy these criteria. But we must be cautious. Some physicists worry that the Standard Model predictions for these observables are not fully under control, meaning that the real discrepancy between measurement and theory may be much smaller than imagined. In particular, the repercussions of difficult-to-calculate but mundane effects associated with the strong force may be larger than first thought. The good news is that there are ways to test these ideas through additional measurements. These tests require detailed analysis and more data, but these data are arriving all the time.

The final puzzle LHCb has turned up involves a twin set of





measurements that has something in common with both our previous examples but that may turn out to be the most interesting of the three. Here we investigated a ratio, dubbed R_{K^*}("r k star"), that compares the rate of the process that we studied for P_5', where beauty hadrons decay into a $\bar{K}^{*0}$ meson and a muon-antimuon pair, to the rate of a similar decay that produces an electron and antielectron in place of the muon pair. We also examined a second ratio, R_K, comparing decays where the $\bar{K}^{*0}$ meson has been replaced with another kind of strange hadron called simply a K meson. Again, we are trying to test lepton universality, but in this case, between the first two generations of leptons—the electrons and muons.

Within the Standard Model the prediction is trivial—the two decays in each ratio should occur at the same rate, giving the two ratios R_K and R_{K^*} expected values of very nearly one. Again we expected that lepton universality would hold. And the measurements, though far from straightforward, have fewer experimental challenges than in the previously discussed lepton universality analyses and therefore constitute an extremely clean and crisp test of the Standard Model.

We performed the R_K analysis first and found that it came in low, with a value of 0.75, with a precision that put it 2.6 sigma away from predictions. This deviation was sufficiently intriguing that we were all very eager to know the value for R_{K^*}, which we finally published earlier this year. The wait was well worthwhile because, for the same conditions where we examined R_K, R_{K^*} showed remarkably similar behavior. We measured a ratio of 0.69, lying 2.5 sigma below the Standard Model prediction. Although it is quite possible that these undershoots are statistical fluctuations, the fact that we found them in two different measurements, as well as the pristine nature of the tests, means that this anomaly is getting a great deal of attention.

If the R_K and R_{K^*} measurements are a true representation of reality, they indicate that something in nature favors decays that produce electrons over those that create muons, with leptoquarks or a Z' boson again being likely culprits. It seems as if muons, in fact, are being underproduced, whereas electrons are sticking more closely to the Standard Model script. If so, whatever mechanism is responsible would not only explain the R_K and R_{K^*} oddities but would also neatly account for the muon-based P_5' measurement. For good measure, some more ambitious theorists have even proposed solutions that would also make sense of the $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$ puzzle, but conceiving of a particle with the necessary characteristics to explain all three measurements looks to be a tall order.

What is clear is that we will know more very soon. We are analyzing data from the LHC's second run now, and our knowledge of the values of R_K and R_{K^*} will rapidly improve. Either the significance of the discrepancies will grow, and then these anomalies will become the biggest story in physics, or they will diminish, and the caravan will move on.

GALILEO'S MOTTO

THE RESULTS WE HAVE discussed are only the most prominent examples of a host of interesting measurements that have recently emerged in beauty physics. They rightly excite many in the particle physics community, but the older and wiser scientists among us have seen such effects come and go in previous experiments, so we are content to wait and see.

What would it mean if one or more of these anomalies move from the category of "intriguing hint" to "clear contradiction of the Standard Model"? For sure, it would be the most important development in particle physics for many decades, giving us a window onto the landscape that lies beyond our current understanding of the laws that govern the universe. At that point we would need to discover exactly what is responsible for this breakdown in the Standard Model. Depending on the nature of the new physics particle—whether it be an exotic Higgs, a leptoquark, a Z' or something else entirely—its effects should appear in other beauty hadron decays, giving us more clues. Moreover, unless it is very heavy, this new particle could also appear directly in collisions at the LHC's ATLAS or CMS or at some future accelerator of even higher energy.

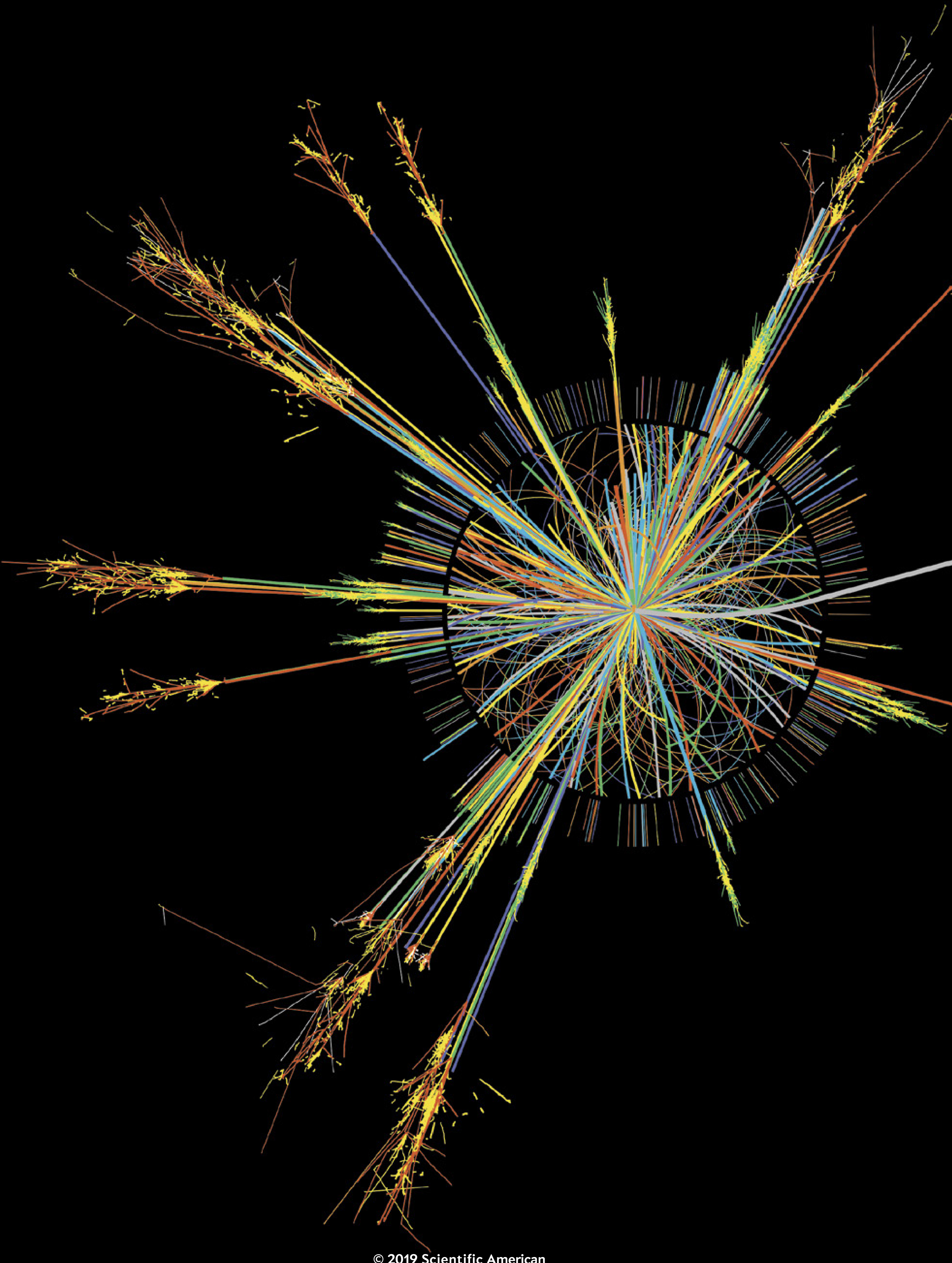
Regardless of how the future unfolds, LHCb's exquisite sensitivity and the excellent prospects for significant improvement in the coming years are undeniable. We do not know if the road to new physics through indirect searches will be short or long, but most of us feel sure that we are heading in the right direction. After all, it was Galileo who is said to have instructed us to "measure what is measurable, and make measurable what is not so." We could have no finer motto for LHCb. SA

Guy Wilkinson is a particle physicist at the University of Oxford and a former spokesperson for the Large Hadron Collider beauty experiment at CERN.

MORE TO EXPLORE

A Challenge to Lepton Universality in *B*-meson Decays. Gregory Ciezarek et al. in *Nature*, Vol. 546, pages 227–233; June 8, 2017.
Flavour-Changing Neutral Currents Making and Breaking the Standard Model. F. Archilli et al. in *Nature*, Vol. 546, pages 221–226; June 8, 2017.
Large Hadron Collider beauty experiment (LHCb): **http://lhcb-public.web.cern.ch/lhcb-public**

scientificamerican.com/magazine/sa



THE PARTICLE CODE

Scientists are creating mathematical tools to identify novel particles and phenomena at the world's largest particle accelerator

By Matthew von Hippel

THE LARGE HADRON COLLIDER, OR LHC, IS THE biggest machine humans have ever built. Pooling the resources of more than 100 countries, it accelerates protons to within a millionth of a percent of the speed of light. When they collide, the protons break into their component parts (quarks and the gluon particles that glue them together) and create particles that were not there before. This is how, in 2012, the LHC achieved the first detection of a Higgs boson, the final missing particle predicted by the Standard Model of particle physics. Now physicists hope the LHC will find something genuinely new: particles not already in their current theory—particles that explain the mystery of dark matter, for instance, or offer solutions to other lingering questions. For such a discovery, scientists must pore through the 30 petabytes a year of data the machine produces to identify tiny deviations where the results do not quite match the Standard Model.

Of course, all of that effort will be useless if we do not know what the Standard Model predicts.

IN BRIEF

To search for new phenomena at the Large Hadron Collider, physicists must be able to precisely calculate the odds of different particle collisions and reactions.

Scientists in a field called amplitudeology are designing cutting-edge mathematical techniques to deal with these difficult computations. In particular, they are building an "alphabet" of logarithms they can combine in different ways to complete previously impossible calculations.

A recent advance in the alphabet could enable the precision necessary for physicists to identify never before seen particles that open the door to a deeper theory of physics.

ATLAS EXPERIMENT © 2016 CERN

That is where I come in. The questions we want to ask about the LHC come in the form of probabilities. "What is the chance that two protons bounce off each other?" "How often will we produce a Higgs boson?" Scientists compute these probabilities with "scattering amplitudes," formulas that tell us how likely it is that particles "scatter" (essentially, bounce) off each other in a particular way. I am part of a group of physicists and mathematicians who work to speed up these calculations and find better tricks than the old, cumbersome methods handed down by our scientific forebears. We call ourselves "amplitudeologists."

Amplitudeologists trace our field back to the research of two physicists, Stephen Parke and Tomasz Taylor. In 1986 they found a single formula that described collisions between any number of gluons, simplifying what would ordinarily be pages of careful case-by-case calculations. The field actually kicked off in the 1990s and early 2000s, when a slew of new methods promised to streamline a wide variety of particle physics computations. Nowadays amplitudeology is booming: the Amplitudes 2018 conference had 160 participants, and 100 attended the summer school the week before, aimed at training young researchers in the tricks of the field. We have gotten some public attention, too: physicists Nima Arkani-Hamed and Jaroslav Trnka's Amplituhedron (a way to describe certain amplitudes in the language of geometry) made the news in 2013, and on television *The Big Bang Theory*'s Sheldon Cooper has been known to dabble in amplitudeology.

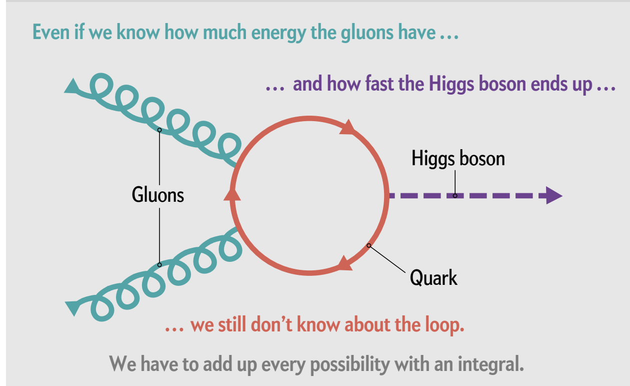
Lately we have taken a big step forward, moving beyond the basic tools we have already developed into more complex techniques. We are entering a realm of calculations sensitive enough to match the increasing precision of the LHC. With these new tools we stand ready to detect even tiny differences between Standard Model predictions and the reality inside the LHC, potentially allowing us to finally reveal the undiscovered particles physicists dream of.

LOOPS AND LINES

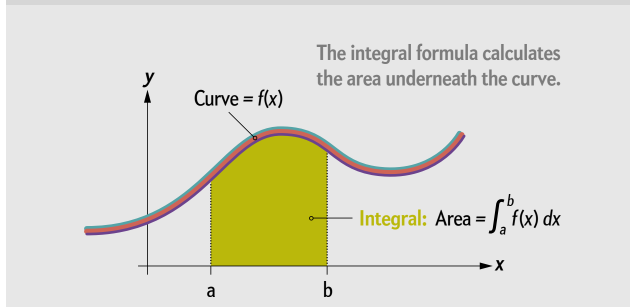
TO ORGANIZE OUR CALCULATIONS, scientists have long used pictures called Feynman diagrams. Invented by physicist Richard Feynman in 1948, these figures depict paths along which particles travel. Suppose we want to know the chance that two gluons merge and form a Higgs boson. We start by drawing lines representing the particles we know about: two gluons going in and one Higgs boson coming out. We then have to connect those lines by drawing more particle lines in the middle of the diagram, according to the rules of the Standard Model. These additional particles may be "virtual": that is, they are not literally particles in the way the gluons and Higgs are in our picture. Instead they are shorthand, a way to keep track of how different quantum fields can interact.

Feynman diagrams are not just pretty pictures—they are instructions, telling us to use information about the particles we draw to calculate a probability. If we know the speed and energy of the gluons and Higgs boson in our diagram, we can try to work out the properties of the virtual particles in between. Sometimes, though, the answer is uncertain. Trace your finger along the particle paths, and you might find a closed loop: a path that ends up back where you started. A particle traveling in a loop like that is not "input" or "output": its properties never get measured. We do not know how fast it is going or how much energy it has. Though counterintuitive, it is a consequence of the fundamental uncertainty of quantum mechanics, which prevents us from measuring two traits of a particle, such as speed and position, at the same time. Quantum mechanics tells us how to deal with this uncertainty—we have to add up every possibility, summing the probabilities for any speed and energy the virtual particles could have, using a technique you might remember from high school calculus: an integral.

Feynman Diagram: Two gluons in, one Higgs boson out



Integral: The area under the function's curve

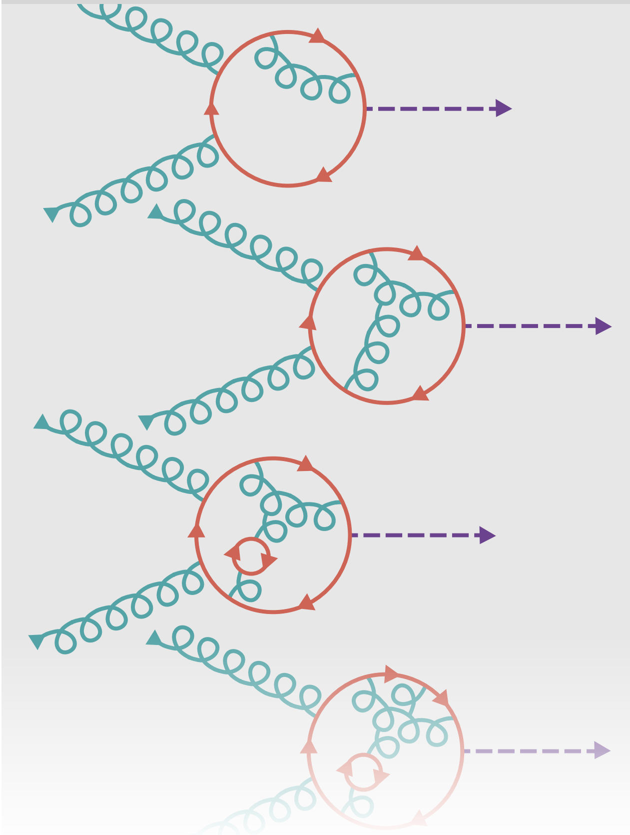


In principle, to calculate a scattering amplitude we have to draw every diagram that could possibly connect our particles, every way the starting ingredients could have turned into the finished products (here the pair of gluons and the Higgs boson). That is a lot of diagrams, an infinite number, in fact: we could keep drawing loops inside loops as far as we like, requiring us to calculate more and more complicated integrals each time.

In practice, we are saved by the low strength of most quantum forces. When a group of lines in a diagram connect, it depicts an "interaction" among different types of particles. Each time this happens we have to multiply by a constant, related to the strength of the force that makes the particles interact. If we want to draw a diagram with more closed loops, we have to connect up more lines and multiply by more of these constants. For electricity and magnetism, the relevant constants are small: for each loop you add, you divide by roughly 137. This means that the diagrams with more and more loops make up a smaller and smaller piece of your final answer, and eventually that piece is so small that the experiments cannot detect it. The most careful experiments on electricity and magnetism are accurate up to an astounding 10 decimal places, some of the

Graphics by Jen Christiansen

Loops inside Loops



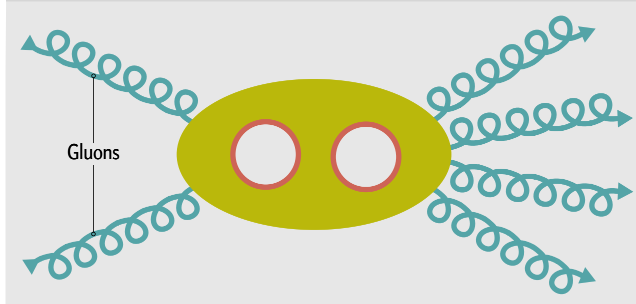
most precise measurements in all of science. Getting that far requires "only" four loops, four factors of $^1/_{137}$ before the number you are calculating is too small to measure. In many cases, these numbers have actually been calculated, and all 10 decimal places agree with experiments.

The strong nuclear force is a tougher beast. It is the force that glues together protons and neutrons and the quarks inside them. It is quite a bit stronger than electricity and magnetism: for calculations at the LHC, each loop means dividing not by 137 but by 10. Getting up to 10 digits of precision would mean drawing 10 loops.

The LHC is not as precise as those electricity and magnetism experiments. At the moment, measurements from the machine are just starting to match the precision of two-loop calculations. Still, those results are already quite messy. For example, a two-loop calculation in 2010 by physicists Vittorio Del Duca, Claude Duhr and Vladimir Smirnov computed the chance that two gluons collide and four gluons come out. They made their calculation using a simplified theory, with some special shortcuts, and the resulting formula still clocked in at 17 pages of complicated integrals. That length was not too surprising; everyone knew that two-loop calculations were hard.

Until a few months later, when another group managed to write the same result on two lines. That group was a collaboration among three physicists—Marcus Spradlin, Cristian Vergu and Anastasia Volovich—and a mathematician, Alexander B. Goncharov. The trick they used was extraordinarily powerful, and it exposed amplitudeologists to an area of mathematics that most of us had not seen before, one that has driven my career to this day.

Two-Loop Example: Two gluons in, four gluons out



PERIODS AND LOGS

SHOW A MATHEMATICIAN like Goncharov one of the integrals we get out of Feynman diagrams, and the first thing you will hear is, "That's a period."

Periods are a type of number. You might be familiar with the natural numbers (1, 2, 3, 4 ...) and the rational numbers (fractions). The square root of 2 is not rational—you cannot get it by dividing two natural numbers. What it is, though, is algebraic: you can write an algebraic equation, say $x^2 = 2$, where the square root of 2 is the solution. Periods are the next step up: although you cannot always get them from an algebraic equation, you *can* always get them from an integral.

Why call them periods? In the simplest cases, that is literally what they are: the distance before something repeats. Thinking back to high school, you might remember grappling with sines and cosines. You might even remember that you can put them together with imaginary numbers (the square roots of negative numbers—in other words, numbers that would not normally exist) using Euler's formula: $e^{ix} = \cos(x) + i \sin(x)$ (here e is a constant, and i is the square root of -1). All three of these—$\sin(x)$, $\cos(x)$, and e^{ix}—*have* period 2π: if you let x go from 0 to 2π, the function repeats, and you get the same numbers again.

2π is a period because it is the distance before e^{ix} repeats, but you can also think of it as an integral. Draw a graph of e^{ix} in the complex plane: imaginary numbers on one axis; real numbers on the other. It forms a circle. If you want to measure the length of that circle, you can do it with an integral, adding up each little segment all the way around. In doing so, you will find exactly 2π.

What happens if you go partway around the circle, to some point z? In that case, you must solve the equation $z = e^{ix}$. Thinking back again to high school, you might remember what you need to solve that equation: the natural logarithm, $\ln(z)$.

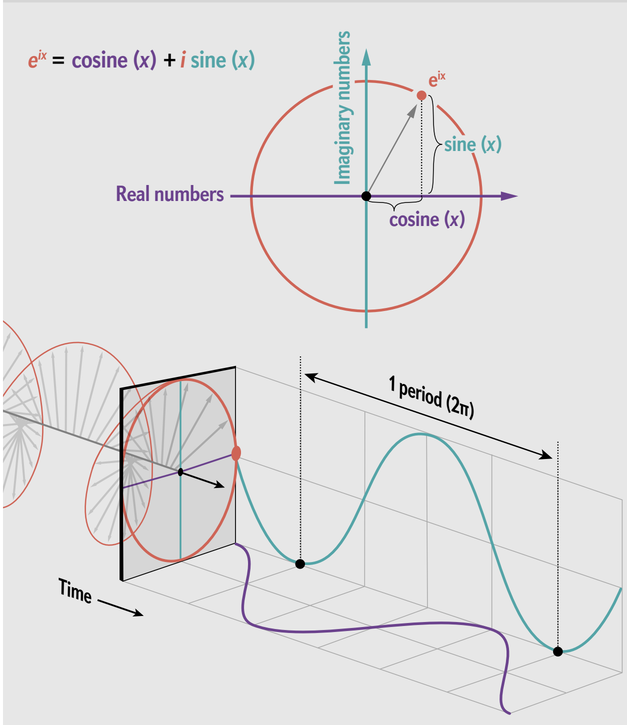
Euler's Formula

$$e^{ix} = \text{cosine}\,(x) + i\ \text{sine}\,(x)$$

Logarithms might not look like "periods" in the way 2π does, but because you can get them from integrals, mathematicians call them periods as well. Besides 2π, logarithms are the simplest periods.

The periods mathematicians and physicists care about can be much more complicated than this scenario, of course. In the mid-1990s physicists started classifying periods in the integrals

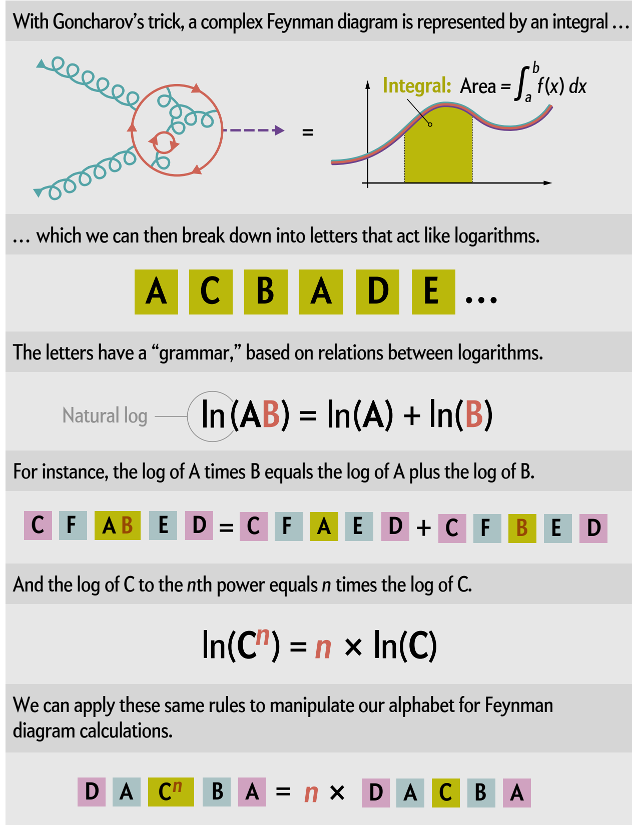
Euler's formula visualized as a circle, then projected through time



that come out of Feynman diagrams and have since found a dizzying array of exotic numbers. Remarkably, though, the high school picture remains useful. Many of these exotic numbers, when viewed as periods, can be broken down into logarithms. Understand the logarithms, and you can understand almost everything else.

That was the secret that Goncharov taught Spradlin, Vergu and Volovich. He showed them how to take Del Duca, Duhr and Smirnov's 17-page mess and chop it up into a kind of "alphabet" of logarithms. That alphabet obeys its own "grammar" based on the relations between logarithms, and by using this grammar, the physicists were able to rewrite the result in terms of just a few special "letters," making a messy particle physics calculation look a whole lot simpler.

To recap, physicists calculate scattering amplitudes using Feynman diagrams, which require doing integrals. Those integrals are always periods, sometimes complicated ones, but we can often break those complicated periods apart into simpler periods (logarithms) using Goncharov's trick, which was what ignited my area of the amplitudes field. We can divide many of the integrals we use into an alphabet of letters that behave like logarithms. And the same rules that apply to logarithms, such as basic laws like $\ln(xy) = \ln(x) + \ln(y)$ and $\ln(x^n) = n \times \ln(x)$, work for the alphabet.



WORD JUMBLE

GONCHAROV'S ALPHABET TRICK would not be nearly as impressive if all it did was save space in a journal. Once we know the right alphabet, we can also do new calculations, ones that would not have been possible otherwise. In effect, knowing the alphabet lets us skip the Feynman diagrams and just guess the answer.

Think about that newspaper mainstay, the word jumble. The puzzle tells you which letters you need and how long the word is supposed to be. If you were lazy, you could have a computer write down the letters in every possible order, then skim through the list. Eventually you would find a word that made sense, and you would have your solution.

The list of possibilities can be quite long, though. Luckily in physics, we start with hints. We begin with an alphabet of logarithms that describe the properties our particles can have, such as their energy and speed. Then we start writing words in this alphabet, representing integrals that might show up in the final answer. Certain words do not make physical sense: they describe particles that do not actually exist or diagrams that would be impossible to draw. Others are needed to explain things we already know: what happens when a particle gets

Word Jumble: Unscramble the letters

LCPASIRET

MORHIAGLT

LDRLCEOI

very slow or very fast. In the end, we can pare things down from what might have been millions of words to thousands, then tens, and finally just one unique answer. Starting with a guess, we end up with the only possible word that can make sense as our scattering amplitude.

Lance J. Dixon, James M. Drummond and Johannes Henn used this technique to find the right "word" for a three-loop calculation in 2011. I joined the team in 2013, when I snuck away from graduate school on Long Island to spend the winter working for Dixon at SLAC National Accelerator Laboratory at Stanford University. Along with then grad student Jeffrey Pennington, we got the result into a form we could compare with the old two-loop calculation from Del Duca, Duhr and Smirnov. Now instead of 17 pages, we had a formula that was 800 pages long—and all without drawing a single Feynman diagram.

Since then, we have pushed to even more loops, and our collaboration has grown, with Duhr, Andrew McLeod, Simon Caron-Huot, Georgios Papathanasiou and Falko Dulat joining the team. We are at seven loops, and I do not know how many pages the new formulas will take to write out. Goncharov's trick is not enough to simplify the result when the calculation is this complicated. Here we are just happy it makes the calculation *possible!* We store our results in computer files now, big enough that you would think they were video files, not text.

THE ELLIPTIC FRONTIER

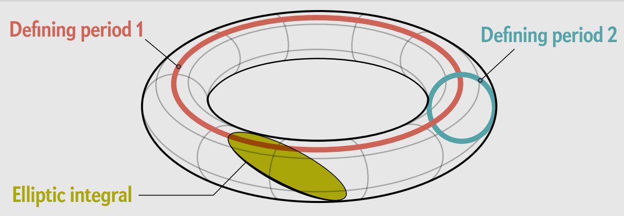
RECALL THAT THE MORE LOOPS you include in your scattering amplitude calculation, the more precise your prediction will be. Seven loops would be more precise than the two or so loops the LHC can measure, more precise than the four-loop state of the art in quantum electromagnetism. I say "would be" here, though, because there is a catch: our seven-loop calculations use a "toy model"—a simpler theory of particle interactions than any that can describe the real world. Upgrading our calculations so they describe reality will be difficult, and there are numerous challenges. For one, we will need to understand something called elliptic integrals.

The toy model we use is very well behaved. One of its nicer traits is that for the kind of calculations we do, Goncharov's method always works: we can always break the integral up into an alphabet of logarithms, of integrals over circles. In the real world, this tactic runs into problems at two loops: two integrals can get tangled together so they cannot be separated.

Think about two hooked rings that cannot be pulled apart. If you move one ring around the other, you will draw a doughnut shape, or a torus. A torus has two "periods," two different ways you can draw a line around it, corresponding to the two different rings. Integrate around a circle by itself, and you get a logarithm. Try to draw a ring around a torus, and you will not always get a circle: instead you might get an ellipse. We call such integrals around a torus elliptic integrals—integrals over an elliptic curve.

Understanding elliptic curves involves some famously complex mathematical problems. Some of these problems are so difficult to solve that organizations such as the National Security Agency use them to encode classified information, on the assumption that no one can solve them fast enough to crack the code. The problems we are interested in are not quite so intractable, but they are still tricky. With the LHC's precision increasing, though, elliptic integrals are becoming more and more essential, spurring on groups around the world to tackle the new mathematics. The machine shut down in late 2018 for upgrades, but scientists still have hordes of data to sort through; it will start up again in 2021 and will go on to produce 10 times more collisions than before.

Elliptic Integral around a Torus



At times the speed at which the field is moving leaves me breathless. Two winters ago I holed up at Princeton University with a group of collaborators: McLeod, Spradlin, Jacob Bourjaily and Matthias Wilhelm. Within two weeks we went from a sketched-out outline to a full paper, calculating a scattering amplitude involving elliptic integrals. It was the fastest I have ever written a paper, and the entire time we worried that we were going to be scooped, that another group would do the calculation first.

We did not end up getting scooped. But not long after, we received a bit of an early Christmas present: two papers by Duhr, Dulat, Johannes Broedel and Lorenzo Tancredi that explained a better way to handle these integrals, building on work by mathematicians Francis Brown and Andrey Levin. Those papers, along with a later one with Brenda Penante, provided us with the missing piece we needed: a new alphabet of "elliptic letters."

With an alphabet like that, we can apply Goncharov's trick to more complicated integrals and start to understand two-loop amplitudes, not just in a toy model but in the real world as well.

If we can do two-loop calculations in the real world, if we can figure out what the Standard Model predicts to a new level of precision, we will get to see if the LHC's data match those predictions. If it does not, we will have a hint that something genuinely new is going on, something our theories cannot explain. It could be the one piece of data we need to move particle physics to the next frontier, to unlock those lasting mysteries we cannot seem to crack. SA

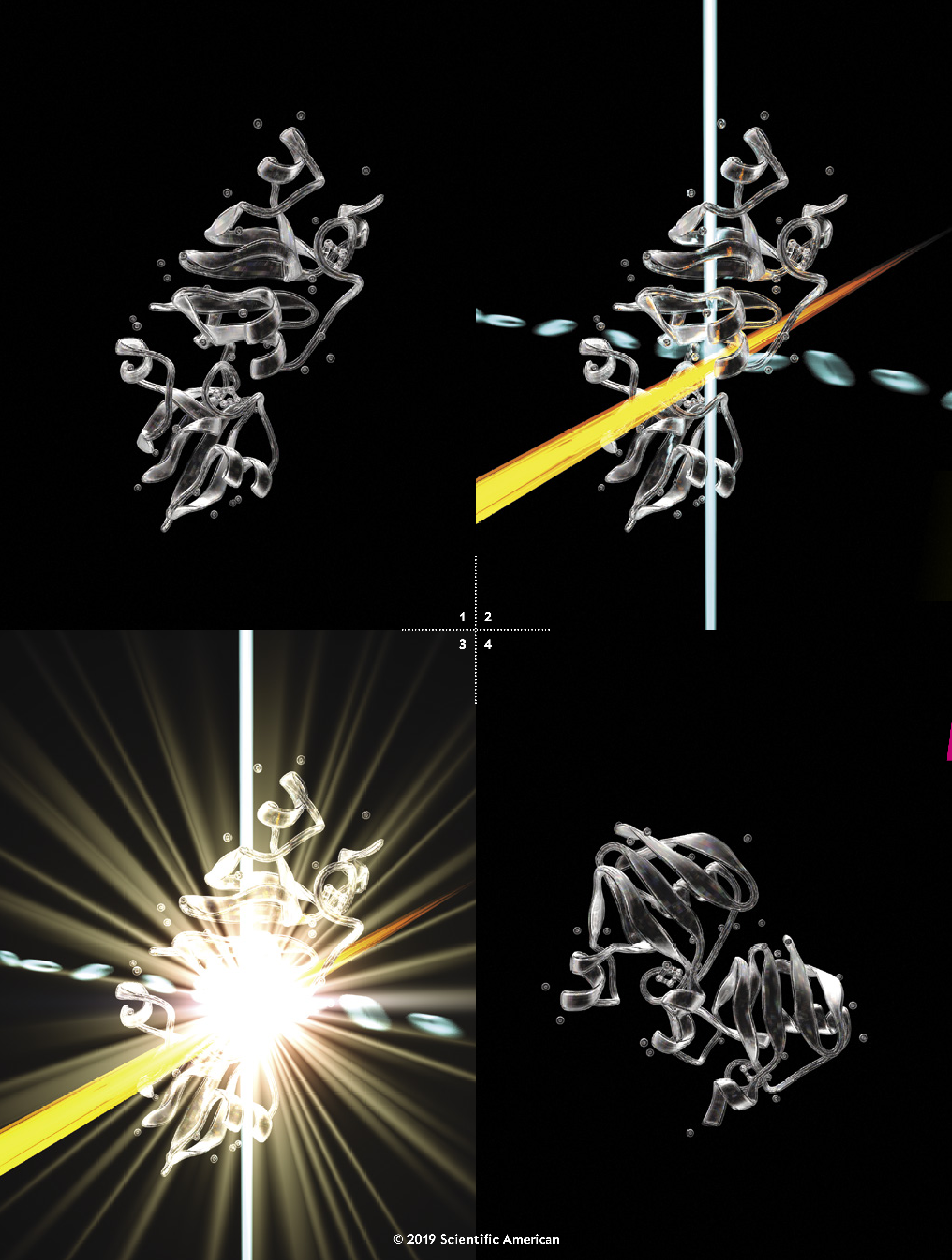
Matthew von Hippel is a postdoctoral scholar at the Niels Bohr International Academy in Copenhagen. He has been working on amplitudes since he stumbled into his adviser's office in graduate school, looking for a summer project. He has also been doing science outreach since he got into a discussion with *Ars Technica*'s science editor about the definition of "theory." He blogs at https://4gravitons.wordpress.com

MORE TO EXPLORE

Classical Polylogarithms for Amplitudes and Wilson Loops. A. B. Goncharov et al. in *Physical Review Letters,* Vol. 105, Article No. 151605; October 7, 2010. Preprint available at https://arxiv.org/abs/1006.5703

Bootstrapping Six-Gluon Scattering in Planar N = 4 Super-Yang-Mills Theory. Lance J. Dixon et al. Presented at the 12th DESY Workshop on Elementary Particle Physics, Weimar, Germany, April 27–May 2, 2014. Preprint available at https://arxiv.org/abs/1407.4724

scientificamerican.com/magazine/sa


1
2
3
4

SPLIT-SECOND REACTIONS

New movies of drug proteins or photosynthesis in action, shot in millionths of a billionth of a second, show how the molecules work—or fail

By Petra Fromme and John C. H. Spence

IN BRIEF

Proteins are in constant motion, carrying out the reactions that make life possible. These movements happen on a scale too small, and too fast, to be seen with microscopes.

Using x-ray laser pulses lasting just millionths of a billionth of a second, researchers have created "molecular movies" that show how proteins change structure when they interact.

These movies can reveal biological reactions in unprecedented detail and demonstrate why drugs sometimes do not hit target proteins and how plant photosynthesis creates clean energy.

Illustration by Bryan Christie

BURROWED DEEP UNDER THE FOOTHILLS NEAR PALO ALTO, CALIF., SCIENTISTS SCURRIED THROUGH AN UNDERGROUND LABORATORY, making final preparations for a series of explosions. ***THEIR PLAN:*** blow up tiny crystals of proteins that could reveal one of nature's best-kept secrets—how plant photosynthesis turns light into chemical energy. The potential payoff: a step toward unlimited clean power.

It was December 2009, and a sleep-deprived team of researchers and students at SLAC National Accelerator Laboratory had been working nonstop for days to set up this experiment at the world's most powerful x-ray laser, the Linac Coherent Light Source (LCLS), which accelerates electrons to nearly the speed of light. One group feverishly adjusted injectors that would shoot crystals of proteins into the x-ray beam. Another locked and loaded the injector with fresh crystals of a protein complex called photosystem I, which is key to photosynthesis.

At the end of the two-mile accelerator tunnel, the crystals began their march into the intense laser light. But before each of them exploded, its snapshot was taken with a newly developed scientific technique. Today that method promises to reshape our understanding of biology on the tiniest scale because we can now assemble a rapid sequence of such images—shot in femtoseconds, or millionths of a billionth of a second—into movies.

Physicist Richard Feynman once said, "Everything that living things do can be understood in terms of the jigglings and wigglings of atoms." But never before have we been able to directly see the wiggling of atoms and molecules within living things at this speed. Our method, called serial femtosecond crystallography (SFX), lets us watch high-speed molecular dances that determine how medicines affect diseased cells and how chemical reactions convert energy to different forms.

Already research teams around the world have used SFX to reveal fine details of how an experimental drug regulates blood pressure—paving the way to better hypertension medications. SFX has also shown the structure of the enzyme that destroys red blood cells in sleeping sickness, a fatal disease caused by parasites. And it has yielded the first look at the initial steps during photosynthesis that split water into hydrogen and oxygen.

Back in that underground lab in 2009, the stakes were high as x-ray pulses began to annihilate our carefully formed crystals. Many scientists had said SFX would never work and rejected our requests for funding. But then beautiful images of scattered x-rays began to emerge on computer screens. We still remember our cheers erupting around the room as we watched what would become proof that a new field of x-ray science had been born.

X-RAY VISION

BEFORE SFX, scientists made amazing advances in detecting the changes of certain chemical structures, but they could not actually observe the most delicate and complex biological structures in action. In the 1980s, for instance, the late chemist Ahmed H. Zewail invented a way to watch atoms move during chemical reactions using ultrafast pulses of visible laser light. Yet the light's wavelength was too long to distinguish the tiniest details of protein structure. More recently, dramatic advances in microscope technology have produced near-atomic-resolution images of proteins and viruses. But they are not quick enough to capture rapid reactions such as photosynthesis.

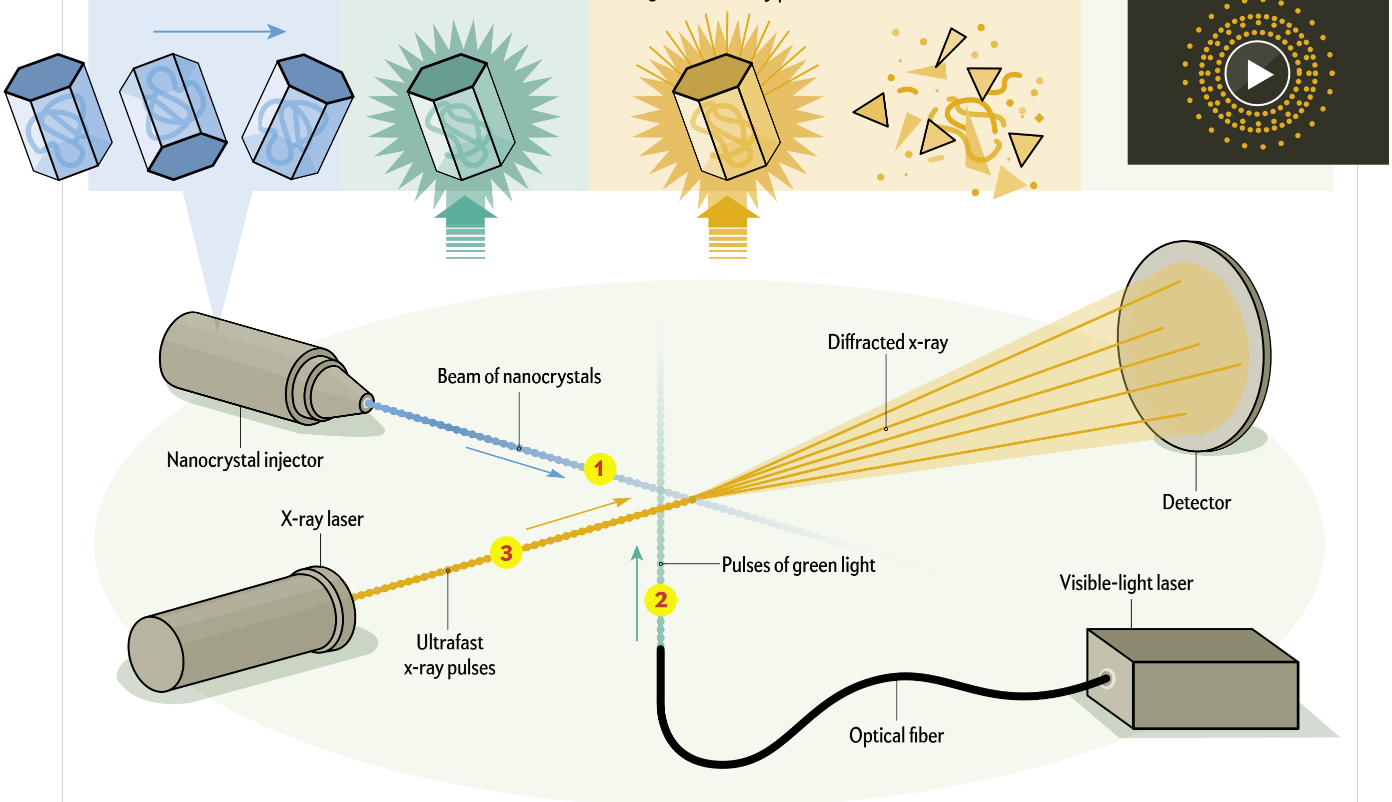
We decided to use x-rays, which have the necessary speed and resolution to record biological reactions in action. Key to our work was developing a technology that would allow x-rays to form pictures of molecules in the instant before destroying them. Traditionally scientists who do this work painstakingly grow large crystals of proteins and other molecules to map the positions of atoms within them. Then they bounce x-rays off the crystals and record the pattern of x-ray scattering, or diffraction. In a crystal, molecules are held in place in an orderly arrangement, so the x-rays scatter in predictable ways, allowing scientists to interpret the position and identity of atoms. This method is called x-ray crystallography, and our serial femtosecond crystallography uses the same principle to see atomic structure but does so far faster.

X-rays ultimately destroy the molecules we are trying to see, however. It was commonly believed that the x-ray laser, which concentrates high-energy x-rays into a powerful beam, would only make matters worse. The laser's bright light alone can punch a hole through steel. A fragile biomolecule, one would think, would not stand a chance. We needed to outrun the x-rays' damage and capture an image in femtoseconds. For perspective, the difference between one femtosecond and one full second is equiv-

Moviemaking on a Molecular Scale

Photosynthesis makes life on earth possible by converting sunlight into chemical energy. A new kind of molecular movie has given scientists their first glimpse of the process in action. Researchers use visible light, simulating sunlight on a leaf, to spur proteins to begin photosynthesis, then use a powerful x-ray laser to take snapshots of changes in these proteins in the fractions of a second before they are destroyed. Snapshots are made in five steps (*shown*) and combined into a movie.

1. A light-responsive protein (called a photosystem) is formed into tiny crystals, in which the protein's orderly arrangement makes it possible to determine its structure. Hundreds of thousands of crystals flow through the injector per second.
2. Pulses of green light simulate sunlight on a leaf, triggering changes in molecules within the nanocrystals. This first step in photosynthesis happens in femtoseconds, or just millionths of a billionth of a second.
3. The crystals are then hit by a powerful x-ray pulse. The x-rays scatter in a distinctive pattern when they hit the nanocrystals, creating a "snapshot" of molecular structure at that instant. To capture the next frame in the movie, the experiment is repeated with a longer delay between the green light and the x-ray pulse.
4. The x-ray pulse lasts just 50 femtoseconds but is so strong that it destroys the protein.
5. Software assembles tens of thousands of 2-D snapshots to create one 3-D view of the protein's structure. More structural images are captured throughout the reaction process and are then stitched into a movie sequence.



alent to the difference between a second and 32 million years.

The key to the SFX technique lies in that imperceptible sliver of time between the molecule being struck by the x-ray laser pulse and electrons being ripped off its atoms by x-ray energy. Stripped of electrons, the positively charged remnants repel one another, causing the molecules to expand and ultimately explode.

Here is how it works: First, we prompt molecules to interact to form a tiny crystal. Then we shoot a powerful x-ray beam at the crystal in an extremely short pulse, just long enough for some of the x-rays to scatter off the crystal before the beam's energy rips the molecules apart. Finally, a detector captures the bounced x-rays, whose pattern reveals the type and position of atoms in the protein. By capturing images of a stream of protein crystals as they tumble through the x-ray beam at different angles, we can re-create the structure in 3-D. Finally, we can collect images at different time points in a reaction and put the pictures together in sequence, like images in a film strip.

Illustration by Lucy Reading-Ikkanda

CRYSTALLIZED VIEW

THE FIRST STEP toward making these molecular movies came in 2000, when biophysicists Janos Hajdu and Richard Neutze, both then at Uppsala University in Sweden, calculated that it would take roughly 10 femtoseconds for molecules to begin exploding after being hit by x-rays. Thus, scientists needed to take a snapshot faster than that. In 2006 Henry Chapman, now at the German Electron Synchrotron (DESY), and his colleagues were able to do just that using a “diffract then destroy” approach to capture a low-resolution image of two tiny stick figures and the sun etched into a silicon nitride membrane.

But would this work for delicate biological molecules? Much of the scientific community was skeptical when we proposed to try. Our first 10 grant proposals were all rejected. Doubters said that the x-ray laser pulses would not be short enough, or the protein crystals would be too small to give any detectable signal, or we would never be able to figure out the crystal’s orientation when it was struck by the x-ray pulse, information needed to determine its structure.

But we thought that if other kinds of molecules could be imaged, as Chapman had proved, then biomolecules could be, too. One of us (Fromme) and her team sought to prove SFX using one of the most difficult tests imaginable: photosystem I. Consisting of 36 proteins and more than 300 light-capturing green and orange pigments, it is among the most complex protein structures analyzed with x-rays to date.

Fromme knew photosystem I intimately, having worked for years to crystallize it and determine its structure using other methods. We also thought that the biomolecular complex’s large size could actually be an advantage because with even a small number of diffraction patterns, we could get a low-resolution image that would be recognizable as photosystem I. And this is what we were able to do in that underground lab in 2009.

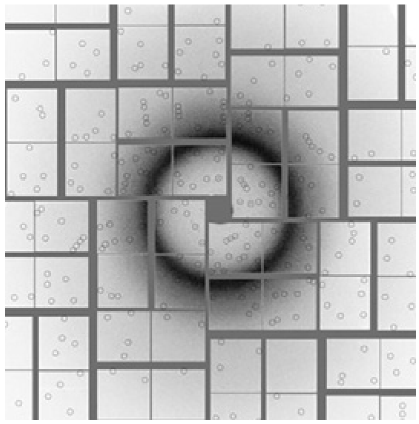
SMALL IS BEAUTIFUL

TO GET OUR SNAPSHOT, we first needed crystals of photosystem I. In typical crystallography, scientists grow large crystals, which have long been necessary to create enough x-ray scattering to reconstruct a protein’s structure. But it can take years of experimentation to grow large, well-ordered crystals of some proteins. Several have proved nearly impossible, and photosystem I was one of them.

Instead SFX uses nanometer-sized crystals, which are much easier to grow in the lab. Using nanocrystals meant new challenges, however. Not only would we have to get a strong enough signal from such a small crystal, but we faced some basic physical challenges: How do you detect nanocrystals too small to see under a microscope, much less position them in front of x-ray pulses and do so consistently 120 times each second?

First, we had to invent new ways to see our nanocrystals. One of the methods we applied is called SONICC (second-order nonlinear imaging of chiral crystals), in which the crystals convert two ultrafast pulses of infrared light into one green photon—this lights up the nanocrystals like fireflies in the night so that we can detect them.

Another invention shoots the crystals into the x-ray laser pulse at a consistent clip. One of us (Spence), along with physicists Uwe Weierstall of Arizona State University and R. Bruce Doak, now at the Max Planck Institute of Medicine in Heidelberg, Germany, came up with a device that functions much like an inkjet printer, firing a stream of nanocrystal-containing solution across the x-ray beam. This injector fires nanocrystals so precisely that they march into the beam in a single-file line.



SPOT ON: **With serial femtosecond crystallography, the gray dots in these panels show patterns of x-rays after they collide with protein crystals, revealing their structure.**

FROM “SERIAL FEMTOSECOND CRYSTALLOGRAPHY DATASETS FROM G PROTEIN-COUPLED RECEPTORS,” BY THOMAS A. WHITE ET AL., IN *SCIENTIFIC DATA*, VOL. 3, ARTICLE NO. 160057. PUBLISHED ONLINE AUGUST 1, 2016

To keep the injector from clogging—which could shut down the stream of nanocrystals—Weierstall had to design a wide nozzle that still had the ability to produce a narrow stream. He did this by surrounding the outer end of the nozzle with a stream of helium gas, focusing the stream of crystals to a tiny fraction of a human hair even though the nozzle itself was more than 10 times larger.

Once we had all the machinery in order, we faced one more problem: how to master a Mount Everest of data. A single experiment can generate up to 100 terabytes of data, enough to fill 25 top-of-the-line desktop computer hard drives. And to construct a 3-D view, we have to find, then merge, the correct orientation of each of the crystals in tens of thousands of snapshots. So we developed special software in collaboration with Richard Kirian and Thomas White, both then members of Chapman’s team at DESY. With the new software, we can turn our tsunami of data into accurate 3-D images of a molecule.

Step by step, we improved our technique. And by 2014 our work gave us the first real-time glimpse of the transfer of electrons between two key players in photosynthesis: the large sunlight-catcher photosystem I and a protein called ferredoxin.

When light hits photosystem I, it is converted into electrons, which ferredoxin then carries away to be used for converting CO_2 into biological molecules. When ferredoxin leaves, the protein crystals quickly dissolve, making the reaction difficult to follow. Only the superfast process of SFX can see the rapid change.

The next challenge in this line of research is a big focus of Fromme’s work as a biochemist: unraveling how a plant splits water into hydrogen and oxygen using just sunlight and the earth’s abundant metals. Splitting water the way that a plant does could provide cheap, clean-burning hydrogen as fuel for cars and power generators, a long-held dream for developing a renewable energy economy.

We have gathered the first low-resolution snapshots of the water-splitting process and have seen an initial hint of significant structural changes to the protein complex involved:

WITH SFX, THE FILMS WE PRODUCE COULD LEAD NOT ONLY TO FUTURE BREAKTHROUGHS BUT MORE IMMEDIATELY TO NEW AND BETTER MEDICATIONS.

photosystem II. Just recently Jian-Ren Shen's group working at Okayama University in Japan has applied the SFX technique to show the same snapshot of the process at even more detail. Next, we seek to make high-resolution movies that can reveal all stages of the process at the atomic level and to discover the secret of photosynthesis.

DESIGNER DRUGS

NOW THAT SCIENTISTS have begun making movies using SFX, the films we produce could lead not only to future breakthroughs but also more immediately to new and better medications. We saw this potential when we studied angiotensin II receptor blockers (ARBs). These drugs interfere with a cell receptor for the hormone angiotensin II, which constricts blood vessels. ARBs are used to treat high blood pressure (hypertension), the leading cause of stroke and heart failure in the U.S. Whereas the first generation of these drugs has proved useful, the drugs bind to their targets only weakly and must be used in high doses, worsening their side effects, which can include headaches and dizziness and occasionally more serious problems such as swelling in the face and throat.

Our research has revealed the reason behind the poor binding: the drugs really do not fit the receptor as well as they should, so many of their molecules fall away. More accurate structures of the receptors could lead to new ARBs that will more effectively control blood pressure. And in fact, one drug called ZD7155 is already being evaluated.

These refinements could improve many other drugs, too. Angiotensin II receptors belong to a larger and extremely important group of cell receptors called G-protein-coupled receptors. These cell-surface molecules allow a cell to sense and respond to its environment. The scientists who first uncovered the structure and actions of this receptor class won the 2012 Nobel Prize in Chemistry for the breakthrough. The vital role that G-protein-coupled receptors play in cell survival and growth makes them crucial targets for new drugs. Being able to see how their structures change will help pharmaceutical chemists design drugs that fit the receptors precisely and in their active state, thereby reducing the chance of side effects.

"We have shown that in all the previous molecular models, the best guesses for how receptor and drug fit together were wrong in many important details," says Vadim Cherezov of the University of Southern California, who conducted the angiotensin II experiment. For example, SFX has revealed differences in the structures of G-protein-coupled receptors at room temperature compared with the cryogenically cold temperatures traditionally used in crystallography—meaning that drugs designed for receptors at frozen temperatures will not fit properly when used in the warm human body. (Sometimes drugs hit too broad a target. This is the problem for drugs used to treat sleeping sickness. Our motion pictures have shown that the drugs interact in similar ways with proteins from the parasite that causes the disease and with proteins from human cells. Our more precise images give chemists a chance to make drugs that affect only the parasite protein, not people.)

EYES HAVE IT

WE HAVE ALSO BEEN THRILLED to see how other researchers are using our SFX techniques to answer different questions. Marius Schmidt of the University of Wisconsin–Milwaukee and his colleagues, for example, have used molecular movies to help explain how our eyes can see. Although we do not typically think of bacteria as being capable of sight, they have light-responsive proteins that are the evolutionary precursors of those in our own visual system. By capturing snapshots faster than ever achieved before, the team made an ultraslow-motion video of extremely rapid events, revealing how a protein in bacteria senses and responds to light.

The group used SFX to capture images of the crystallized protein as it reacted to light in increments of less than a trillionth of a second. Specifically, the team mapped the protein's atoms in motion as a dye molecule buried within the protein turned yellow in response to light. For the first time, the structure of the yellow dye was captured immediately after it absorbed the light and before it could react; this state is fundamental to light perception in all living organisms, including bacteria and plants, and is the first event in human vision.

Seeing how this protein responds to light not only helps us understand how vision arose but also gives us an unprecedented look at how a biological reaction unfolds on chemistry's ultrafast timescale. "This puts us dramatically closer to understanding the chemistry necessary for all of life," Schmidt says.

We are convinced that the future of protein crystallography—as well as our knowledge of nature—lies with the SFX method. And who knows—perhaps within 10 years, half of all known protein structures will not be static images on a textbook page but 3-D movies. SA

Petra Fromme is Paul V. Galvin Professor and director of the Center for Applied Structural Discovery at Arizona State University.

John C. H. Spence is Richard Snell Professor of Physics at Arizona State and director of science at the National Science Foundation's BioXFEL Science and Technology Center.

MORE TO EXPLORE

Femtosecond X-ray Protein Nanocrystallography. Henry N. Chapman et al. in *Nature,* Vol. 470, pages 73–77; February 3, 2011.

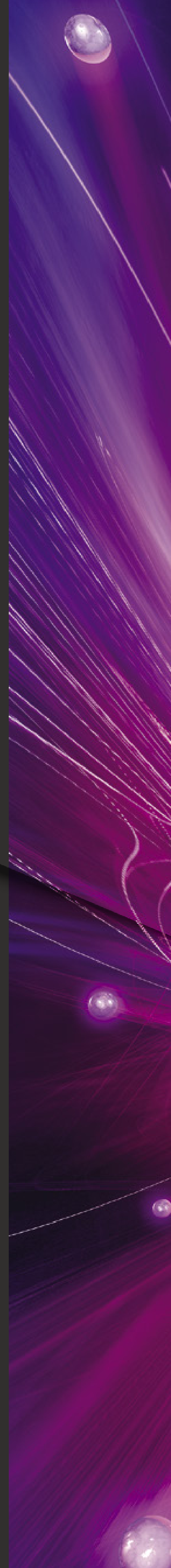
X-ray Lasers for Structural and Dynamic Biology. J.C.H. Spence, U. Weierstall and H. N. Chapman in *Reports on Progress in Physics,* Vol. 75, No. 10, Article No. 102601; October 2012.

X-ray Science: The Big Guns. M. Mitchell Waldrop in *Nature,* Vol. 505, pages 604–606; January 30, 2014.

XFELs Open a New Era in Structural Chemical Biology. Petra Fromme in *Nature Chemical Biology,* Vol. 11, No. 12, pages 895–899; December 2015.

scientificamerican.com/magazine/sa

The Neutrino Puzzle



The largest experiment ever to probe these mysterious particles could point the way to new physics

By Clara Moskowitz

Illustration by Mark Ross

I'M STANDING ON A CATWALK IN A GIANT CAVE CRAMMED WITH INDUSTRIAL equipment, and I'm told that trillions of neutrinos are flying through every inch of my body each second. I reach out my arms as if to heighten the sensation, but of course, I can't feel a thing. Nearly massless, traveling close to the speed of light, the ghostly particles traverse the empty space between my atoms without a trace. They also move mostly unimpeded through the hulking metal box that dominates the cavern. But a few times a day one will collide with an atom inside the school bus–size contraption, liberating charged particles that leave light trails visible to scientists. And these trails, physicists hope, will lead them into unknown territory.

The apparatus is part of the NuMI Off-Axis Electron Neutrino Appearance experiment, or NOvA, here at Fermi National Accelerator Laboratory (Fermilab) in Batavia, Ill. A similar but larger detector is buried 800 kilometers away in Minnesota, where it catches neutrinos that have passed through this one and all the ground in between. NOvA, which has been operating since 2014, is the world's longest-distance neutrino experiment, but it is laying the groundwork for something much larger—the Deep Underground Neutrino Experiment (DUNE). DUNE will start at Fermilab, where an accelerator will speed up and smash protons into graphite to create a beam of neutrinos. Those neutrinos will then fly through 1,300 kilometers of earth from Illinois to South Dakota. The additional 500 kilometers of travel should make it more likely that the neutrinos will display some of their trademark odd behavior.

DUNE is the most ambitious particle physics experiment to be attempted on U.S. soil since the failed Superconducting Super Collider (SSC) of the 1990s. The $1.5-billion project is scheduled to start up in the 2020s and should run for at least 20 years. But it is not just Americans who are excited—the project involves more than 1,000 researchers from 31 countries and counting. It will be the biggest neutrino experiment on the planet. It will also mark the first time that Europe's major particle physics laboratory, CERN, has ever invested in a project outside the continent. Just as the Large Hadron Collider (LHC) discovered the famed Higgs boson in 2012, revealing the presence of a hidden field that fills the cosmos, scientists hope DUNE can use neutrinos to understand the universe on a deeper level. "We want to do for neutrinos what the LHC did for Higgs," says DUNE's former co-spokesperson Mark Thomson, an energetic Brit from the University of Cambridge, who helped lead the charge on the experiment. (He is currently executive chair of the U.K.'s Science and Technology Facilities Council.) "We believe we are on the verge of launching the next major revolution in particle physics."

Neutrinos stoke such extravagant hopes because they are the first particles to break from the so-called Standard Model, physicists' best description of nature's fundamental particles and the rules that govern them. The Standard Model, which explains the behavior of every other known particle with extraordinary precision, predicts that neutrinos should be massless. And that's what scientists thought until around 20 years ago, when experiments in Japan and Canada discovered that neutrinos *do* have the slightest bit of

IN BRIEF

Neutrinos may be the least understood fundamental particles that we know of. Chargeless and insubstantial, neutrinos rarely interact with other particles and were originally predicted to be massless. Now physicists know that they do have a small amount of mass, but the reason why is a mystery.

An ambitious project under construction called the Deep Underground Neutrino Experiment (DUNE) will beam neutrinos 1,300 kilometers from Illinois to South Dakota.

As they make the journey, the particles are likely to morph from one type, or flavor, to another, a phenomenon known as neutrino oscillation. By studying this peculiar behavior, physicists hope to elucidate the origin of neutrino mass and other quandaries.

mass. But neutrinos don't seem to acquire mass the way other particles do. Instead, it appears, they come by their heft through so-called new physics—some particle, force or phenomenon that scientists have not yet found.

Over the past few years neutrinos have come to look like an ever more promising bridge to the future of physics because other attempts to reach that frontier have come up short. So far the LHC has failed to produce any particles not predicted by the Standard Model. Experiments designed to reveal the particles that make up dark matter, the invisible stuff that dominates the cosmos, have also come up empty. "We know the Standard Model is not complete—there are other things going on, but we don't know what," says Fermilab neutrino physicist Stephen Parke. "Some people are betting on the LHC with their careers. Others of us are betting on neutrinos."

MASSIVE MYSTERY

THE DAY AFTER my visit to the NOvA cave, I find myself sitting in an empty office on the third floor of Robert Rathbun Wilson Hall, Fermilab's main building. Parke, who is here along with theorist André de Gouvêa of Northwestern University, says he chose this room for our meeting because it was once the office of the late Leon Lederman, the former director of Fermilab, who developed a way to create a beam of neutrinos with a particle accelerator. That work, the bedrock of DUNE, revealed the existence of one of the three known types of neutrinos in 1962 and later won Lederman a Nobel Prize. Parke and de Gouvêa admit that although the field has come a long way since Lederman's day, scientists are still puzzled. "The thing about neutrinos is, the more you understand, the more questions you have," Parke says. "They're very mischievous particles."

Parke, a native of New Zealand, got hooked on neutrinos shortly after coming to the U.S. for graduate school in the 1970s. In the subsequent decades, neutrinos lost their reputation as massless, boring particles. "There have been these revolutions one after the other," he says. "The question is, Are there more revolutions out there?" He and de Gouvêa are betting yes. "We've only just begun to measure neutrino properties at a level comparable to other particles," de Gouvêa says. "We don't know their masses, there could be new [types of neutrinos], the neutrinos could talk to other particles that don't talk to anybody else."

DUNE will focus on neutrinos' bizarre tendency to swap identities, a process called oscillation. The particles come in three varieties, or flavors: electron neutrinos, muon neutrinos and tau neutrinos. Researchers can tell them apart because when they interact with atoms in detectors, they produce different end products—electron neutrinos create electrons, muon neutrinos produce muons and tau neutrinos make tau particles (muons and taus are heavier cousins of electrons). Strangely, these three flavors are mutable. The particles might leave Fermilab as muon neutrinos and arrive in South Dakota as electron neutrinos. Or they might show up as tau neutrinos. As far as physicists know, neutrinos are the only particles that undergo this bizarre act of identity transformation.

When physicists discovered the shape-shifting tendency of neutrinos almost two decades ago, it solved a long-standing mystery. In the 1960s, when scientists began studying neutrinos streaming out of the sun, they measured only about a third of the output predicted by theory. Oscillation explained why: the missing two thirds were morphing from electron neutrinos into muon and tau neutrinos as they traveled to Earth, but the instruments were set up to see only electron neutrinos. Although the discovery put to bed the so-called solar neutrino problem, it exposed another mystery: according to theory, the only way for neutrinos to switch flavors is for them to have mass—and that is something that the Standard Model did not predict.

The reason physicists know neutrinos must have mass is a head-scratcher that comes from quantum theory. For neutrinos to change flavors, each flavor must be made up of different "mass states." Weirdly, each neutrino flavor does not appear to have a definitive mass; instead the flavors are a mix of three possible masses. (If that sounds strange, blame quantum mechanics, which tells us that particles are not definite entities but uncertain hazes of probability.) As neutrinos fly through space, the parts associated with each mass state travel at slightly different rates, a consequence of Einstein's special theory of relativity, which established that the velocity of a particle traveling near the speed of light depends on its mass. Over time this difference is thought to cause the mixture of masses in each neutrino to change, so a particle that starts out as, say, a muon neutrino, defined by its precise mass mixture, can turn into an electron or tau neutrino.

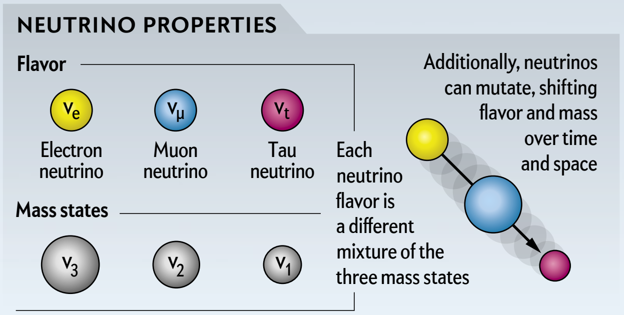
Scientists still do not know what the precise neutrino mass states are—only that they are different and nonzero. But by counting how many neutrinos oscillate during the journey from Illinois to South Dakota, DUNE aims to determine how the different neutrino masses compare with one another. Theory suggests that the three possible neutrino masses might be ordered so that two are very lightweight and one is heavy or, alternatively, that two of the masses are heavy and one is smaller. The first of these two options is known as the normal hierarchy, whereas the second arrangement is called the inverted hierarchy. DUNE should be able to distinguish between the two because the matter inside Earth is thought to affect neutrino oscillations; if the normal hierarchy were correct, scientists would expect to see different ratios of the three flavors than if the inverted hierarchy were right. "By firing neutrinos through matter, you can determine that difference very easily, and the farther you fire your neutrinos, the clearer your signal is," Thomson says. "That's a bit of physics that DUNE is absolutely guaranteed to nail within a few years."

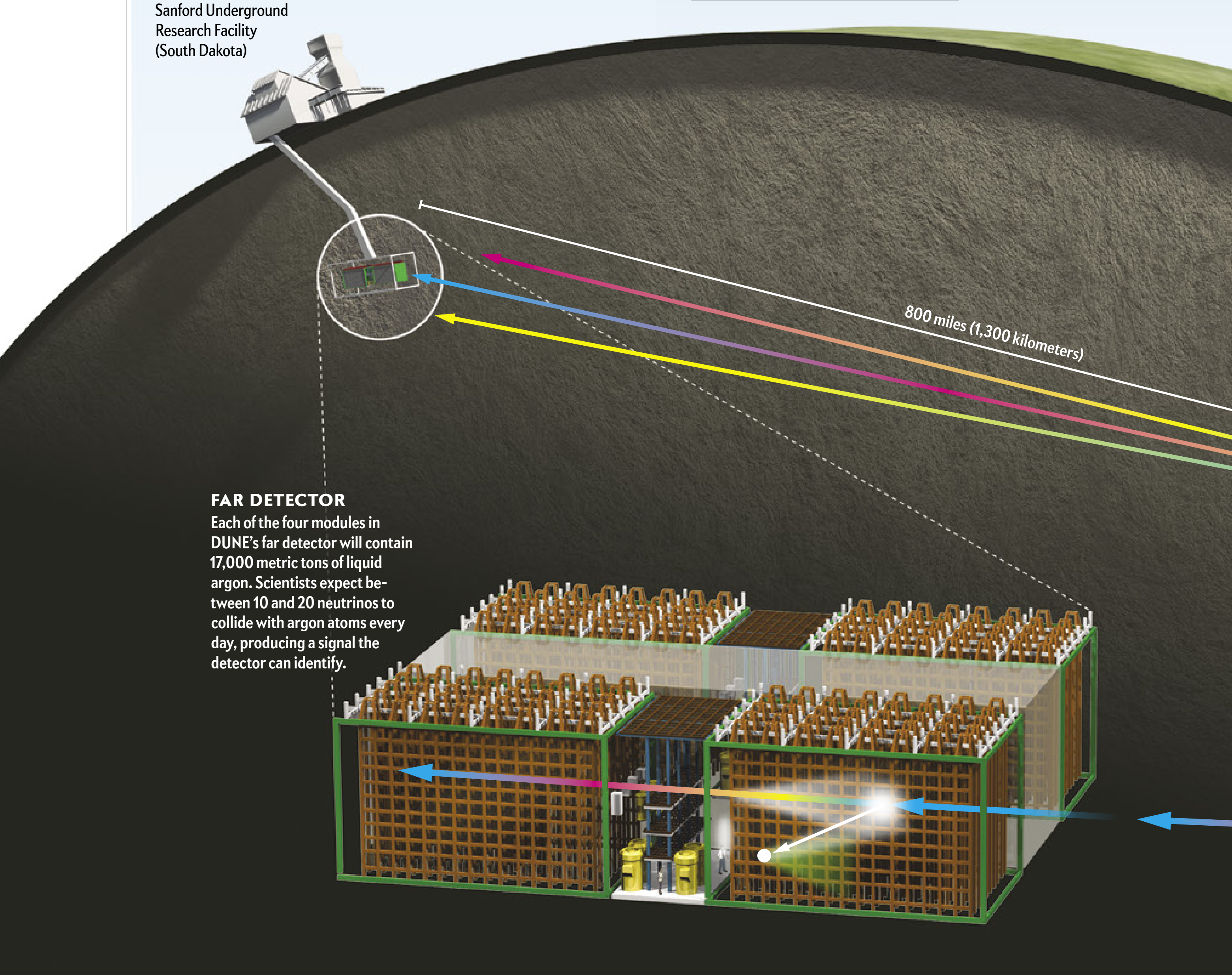
Perplexing Particles

Neutrinos are tiny particles that fly through matter at near light speed. They come in three types, called flavors. Weirdly, as they travel through space neutrinos that started out as one flavor can switch, or "oscillate," into another. Scientists aim to investigate this strange behavior in the Deep Underground Neutrino Experiment (DUNE), the most ambitious neutrino project ever undertaken, due to start operating in the 2020s. Physicists will shoot a stream of neutrinos from Fermi National Accelerator Laboratory (Fermilab) in Illinois to the Sanford Underground Research Facility in South Dakota and watch how many oscillate between flavors over the journey. Through this phenomenon scientists hope neutrinos will lead to a deeper understanding of physics.

NEUTRINO PRIMER

The three neutrino flavors—electron neutrino, muon neutrino and tau neutrino—are named after the particles they interact with—electrons, muons and taus. Neutrinos are not, as scientists once thought, massless. Because of the oddities of quantum mechanics, the flavors do not have definite masses; rather each flavor is a unique mix of three different "mass states." The precise values of the mass states remain a mystery.

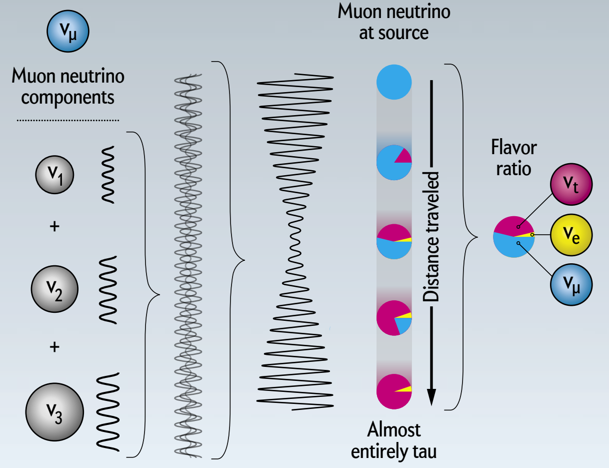




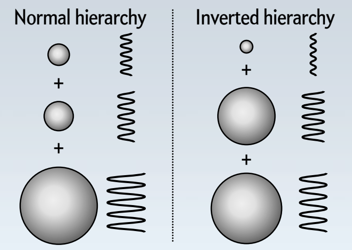
Illustrations by Don Foley (DUNE schematic) and Jen Christiansen (neutrino primer)

FLAVOR OSCILLATIONS AND THE ROLE OF MASS

As a neutrino moves through space, the different mass states of which it is composed travel at slightly different rates. Over time this lag causes the mix of mass states within a neutrino to change, and its flavor shifts accordingly. In this way, a neutrino that starts out as muon-flavored may turn into a tau or electron neutrino.

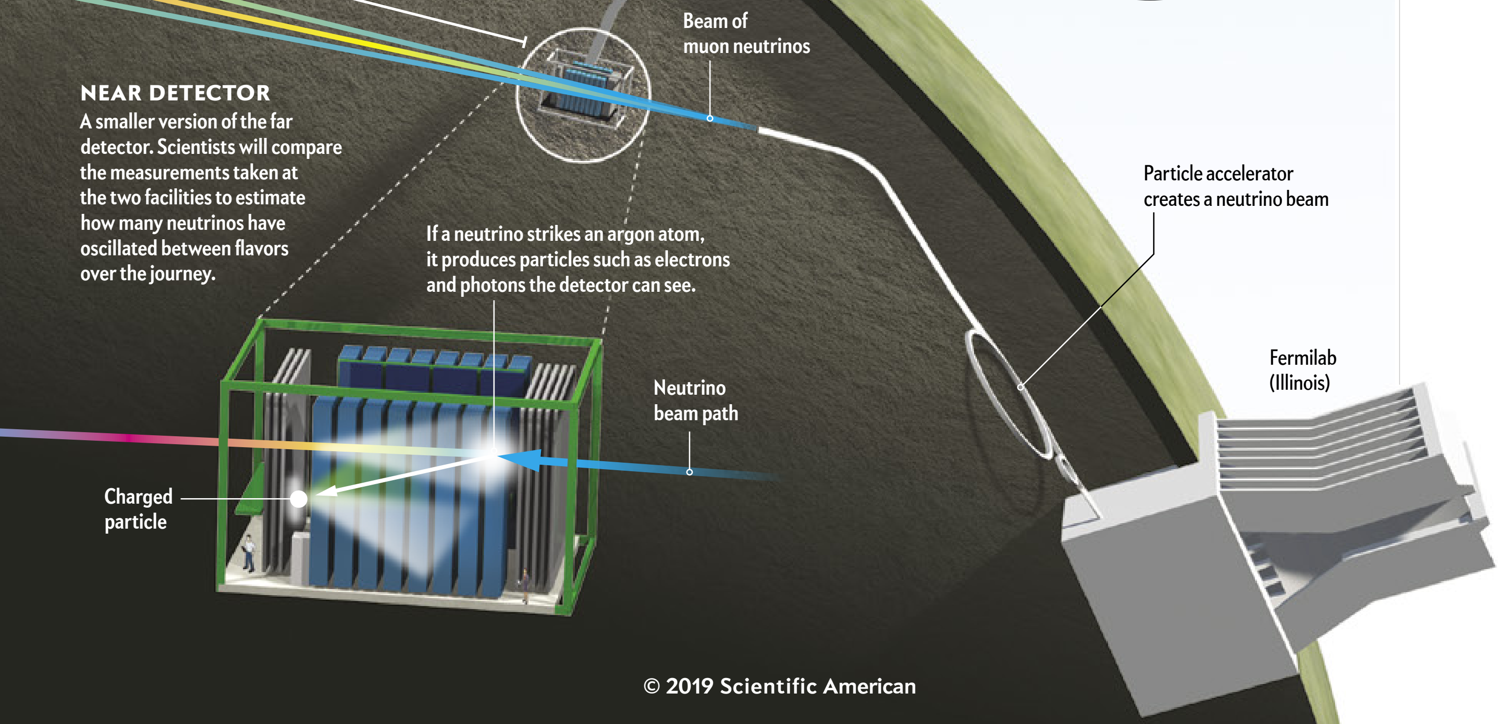


Scientists do not know the values of the three mass states, but theory suggests either that two are lightweight and one is relatively heavy (a configuration known as the normal hierarchy) or that one is light and two are heavy (the inverted hierarchy). DUNE should be able to determine which hierarchy is correct.



GOING THE DISTANCE

DUNE will send neutrinos over 1,300 kilometers from Fermilab in Batavia, Ill., to the Sanford Underground Research Facility in Lead, S.D. This stretch, the longest yet for a neutrino experiment on the earth, should allow ample time for neutrinos to oscillate.



NEAR DETECTOR

A smaller version of the far detector. Scientists will compare the measurements taken at the two facilities to estimate how many neutrinos have oscillated between flavors over the journey.

FERMILAB'S main injector, an underground particle accelerator ring, ramps up protons to create beams of neutrinos to be studied by the DUNE experiment.

DEPARTMENT OF ENERGY AND FERMILAB

THE ORIGIN OF MASS

ONCE THEY KNOW the ordering of the neutrino masses, researchers can tackle the larger question of how neutrinos get their mass. Most particles, such as the protons and neutrons inside atoms, acquire mass by interacting with the Higgs field; this field, which pervades all of space, is associated with the Higgs boson found at the LHC. But the Higgs mechanism works only on particles that come in both right-handed and left-handed versions, a fundamental difference related to the orientation of their spin relative to their direction of motion. So far neutrinos have been seen only in left-handed form. If they got mass from the Higgs field, then right-handed neutrinos must also exist. But right-handed neutrinos have never been observed, which suggests that if they are real they do not interact at all with any other forces or particles in nature—and that prospect strikes some physicists as far-fetched. Furthermore, if the Higgs field did work on neutrinos, theorists would expect them to have similar masses to the other known particles. Yet neutrinos are inexplicably light. Whatever the mass states are, they are less than one hundred-thousandth of the mass of the already puny electron. "Very few people think it's the Higgs mechanism that gives mass to the neutrinos," says Fermilab's director Nigel Lockyer. "There's probably a completely different mechanism, and therefore there should be other particles associated with how that happens."

One possibility that excites physicists is that neutrinos could be Majorana particles—particles that are their own antiparticles. (This is possible because neutrinos have no electric charge, and it is a difference in charge that distinguishes a particle from its antimatter counterpart.) Theorists think Majorana particles have a way of getting mass without involving the Higgs field—perhaps by interacting with a new, undiscovered field. The mathematics behind this scenario also requires the existence of a very heavy set of neutrinos that has yet to be discovered; these particles would have up to a trillion times the mass of some of the heaviest known particles and would, in a sense, counterbalance the light neutrinos. For particle physicists, the prospect of discovering a new mass scale is enticing. "Historically we've always made progress by exploring nature at different scales," de Gouvêa says. And if some new field gives mass to neutrinos, maybe it affects other particles as well. "If nature knows how to do it to neutrinos, where else does it do it?" Lockyer speculates. "Theorists are asking: Could dark matter be a Majorana mass?"

DUNE will not directly test whether neutrinos are Majorana particles, but by measuring the mass hierarchy, it will help scientists interpret the results of experiments that do, which are going on now in Japan, Europe, the U.S. and elsewhere. Plus, DUNE should help elucidate the origin of neutrino mass by providing details about how neutrinos switch between mass combinations during oscillation. "We want to do the best possible neutrino oscillation experiment," de Gouvêa says, "because that's the one place where we know we're going to learn something about neutrino masses."

MATTER VS. ANTIMATTER

PROBING THE ODDITIES of these minuscule particles could also help solve a mystery of cosmic proportions: why the universe is made of matter and not antimatter.

Cosmologists predict the two should have existed in equal amounts after the big bang. Somehow, after most of the matter annihilated with most of the anti-

"THE THING ABOUT NEUTRINOS IS, THE MORE YOU UNDERSTAND, THE MORE QUESTIONS YOU HAVE."
—Stephen Parke, Fermilab

matter (as the two do on contact), there was a slight excess of matter left over. That matter makes up the galaxies, stars and planets that we see today.

To account for this asymmetry, scientists are on the lookout for a type of particle that behaves differently from its antimatter counterpart, and various clues, including hints seen at other experiments, point to neutrinos. DUNE will search for signs of so-called CP (charge parity) violation—in other words, evidence that antineutrinos oscillate from flavor to flavor at different rates than neutrinos. For example, theory suggests that DUNE might see antimatter muon neutrinos turning into electron neutrinos at anywhere between half to twice the rate at which matter neutrinos make this transition—a difference that Parke calls "enormous" and that could explain why matter won out in that initial battle. (Bizarrely, neutrinos could still oscillate differently from antineutrinos even if the two turn out to be same thing—in other words, if neutrinos are Majorana particles. In that case, the only thing separating neutrinos from antineutrinos would be their handedness, related to their direction of spin. Matter neutrinos, being left-handed, could act differently from antimatter neutrinos, which would be right-handed.)

DUNE will also be able to determine whether neutrinos come in only three flavors or whether there are more waiting to be discovered, as some theories speculate. The additional neutrino flavors would be so-called sterile neutrinos because they would not interact with normal matter at all. Earlier experiments, including the Liquid Scintillator Neutrino Detector at Los Alamos National Laboratory and the Mini Booster Neutrino Experiment (MiniBooNE) at Fermilab, saw inconclusive signs that an extra type of neutrino was interfering with oscillations, suggesting that sterile neutrinos exist that are heavier than the regular three. Researchers hope DUNE will either confirm or rule out that possibility. "Sterile neutrinos can change the pattern of oscillations we see at DUNE by quite a large amount," Thomson says.

BETTING BIG

TO ADDRESS ALL THESE QUANDARIES, scientists designed DUNE to collect far more data at far greater levels of precision than every previous neutrino experiment. The project will use a beam of neutrinos about twice as powerful as the strongest existing high-energy neutrino stream, and it will blast it at a detector that is more than 100 times larger than the biggest of its kind.

The centerpiece of the experiment will be the far detector to be installed in the Sanford Underground Research Facility in Lead, S.D. That machine will consist of four detector modules, each as long as an Olympic pool but six times as deep, that will be filled with 17,000 metric tons of liquid argon. When a neutrino strikes the nucleus of an argon atom in either the far or near detector, it will become, depending on its flavor, an electron, a muon or a tau particle. Muons will travel through the liquid argon in straight lines, kicking electrons out of argon atoms as they go, leaving a trail of electrons the detector can see. If the neutrino produces an electron, on the other hand, the process will create a photon that will then spawn two electrons, and then more photons, and so on, in a cascade of new particles. Tau neutrinos, likewise, would result in tau particles but only if the initial neutrino was energetic enough; taus, being more massive than electrons or muons, take more energy to create. Scientists at CERN began testing a miniature version of DUNE's far detector, called ProtoDUNE, in 2018. "These detectors, it's kind of like a space mission in that once you turn them on you really can't stop them and take them apart to fix things," says Joseph Lykken, Fermilab's deputy director. "Once you put the 17,000 tons of liquid argon in, it's just too hard to get it out."

To succeed, DUNE will have to overcome the political and funding hurdles that have killed large physics projects before. In July 2017 scientists and officials held a groundbreaking ceremony at the Sanford facility to mark the start of major excavation, which is still ongoing. Of course, plenty of excavation took place for the SSC, which was planned to be even bigger than the LHC. The SSC probably would have discovered the Higgs boson, but it was canceled in 1993 because of cost overruns and changing political tides. "You can go back in history and look at the Supercollider, and, boy, is that a sad story," Lockyer says. "The international nature of DUNE is such a step forward." Having commitments and funding from more than just one country should help DUNE avoid the SSC's fate. "I'll say it's definitely happening," Lockyer says. And then he catches himself: "But could it not happen? Yes." SA

Clara Moskowitz is *Scientific American*'s senior editor specializing in space and physics. She has a bachelor's degree in astronomy and physics from Wesleyan University and a graduate degree in science journalism from the University of California, Santa Cruz.

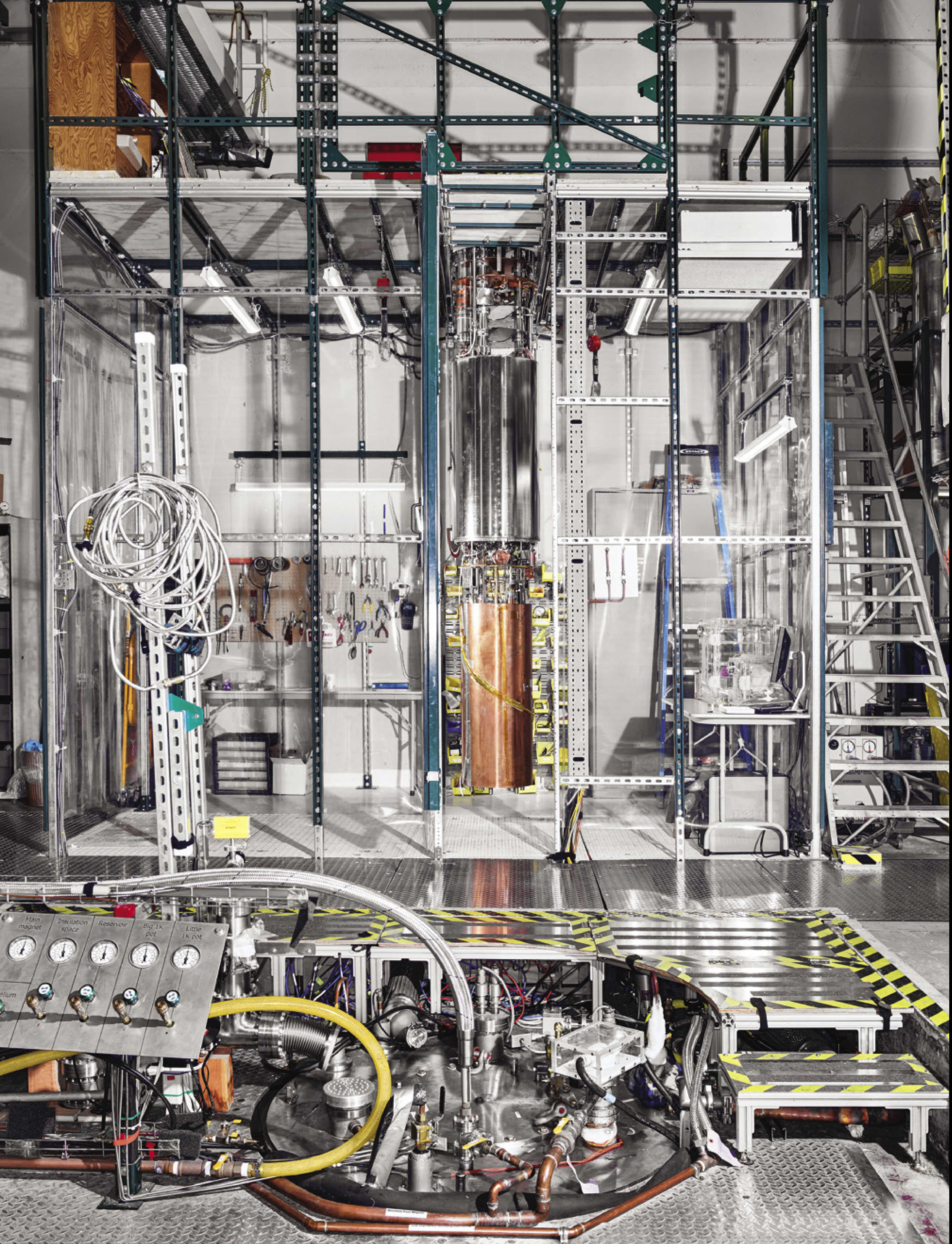
MORE TO EXPLORE

Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE) Conceptual Design Report, Vol. 1: **The LBNF and DUNE Projects.** DUNE Collaboration. Preprint revised January 20, 2016. Preprint available at **https://arxiv.org/abs/1601.05471v1**

Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE) Conceptual Design Report, Vol. 2: **The Physics Program for DUNE at LBNF.** DUNE Collaboration. Preprint revised January 22, 2016. Preprint available at **https://arxiv.org/abs/1512.06148v2**

Deep Underground Neutrino Experiment: **www.dunescience.org**

scientificamerican.com/magazine/sa


Main magnet
Insulation space
Reservoir
Big 1K pot
Little 1K pot

SEARCHING FOR THE DARK

The Axion Dark Matter Experiment just entered the most sensitive phase yet in its search for invisible particles to explain the universe's hidden mass

By Leslie Rosenberg

IN THE BACKGROUND, the insert containing the heart of the ADMX experiment sits in a clean room. It will soon be lowered into a hole (*covered in this image*) in the foreground to begin a new run.

Photographs by Ian Allen

THE COSMOS IS MOSTLY MADE OF SOMETHING WE CANNOT SEE.

That was the conclusion astronomers started to reach in the 1930s by looking at galaxy clusters, which should have blown apart unless some "dark matter" was binding them together. Scientists started taking the idea more seriously in the 1970s, when astronomers studying how fast galaxies rotated found the same thing. Soon researchers realized that dark matter was unlikely to be made up of normal matter and radiation. By now it seems nearly inescapable that more than 90 percent of the stuff in the universe that clumps together under gravity is some exotic material, perhaps a new particle left over from the big bang.

For a long time the most popular dark matter candidate was the theoretical weakly interacting massive particle (WIMP), which fits into the much loved but speculative theory of supersymmetry. Yet sensitive terrestrial WIMP detectors have found no signs of such particles despite decades of searching. It is certainly too early to write off WIMPs, but these null results have raised the profile of non-WIMP dark matter candidates.

A less well-known candidate is the axion, another theorized particle that would weigh much less than the WIMP but would have a similar tendency to ignore normal matter. If axions are dark matter, they would abound everywhere—tens or even hundreds of trillions of them could be floating around in every cubic centimeter around you. Their only effects on the rest of the universe would be felt through gravity—their accumulated mass would be substantial enough to tug on the orbits of stars in galaxies and of galaxies in clusters.

For more than 20 years I have been part of the Axion Dark Matter Experiment (ADMX) search for these particles. Although we have not found them yet, we have been steadily improving our technology. In 2016 ADMX began a new phase. It is now sensitive enough that it should be able to detect axions or to rule out the most plausible versions of them over the next five to 10 years. We stand at an important threshold, and exciting news is coming soon, either way.

THE ORIGIN OF AXIONS

I WAS A GRADUATE STUDENT in the 1980s shortly after the idea of axions first arose from a problem with a theory called quantum chromodynamics (QCD). QCD governs the strong force, which holds together atomic nuclei. It has been remarkably consistent with experiments, except when it comes to something called the strong CP problem. (CP stands for "charge parity.") QCD suggests that if you were to flip a particle's charge parity—that is, flip its electric charge and view it in a mirror—it would no longer follow the same rules of physics. Yet researchers have found no evidence that this is the case. This conflict between theory and experiment presents a serious conundrum—a crack in our best model of particle physics. The crack is the strong CP problem, and it suggests we are missing something big.

In 1977, when physicists Helen Quinn and Roberto Peccei were both at Stanford University, they realized that they could attack the strong CP problem in a simple, elegant way using the idea of broken symmetries. This concept, one of the recurring ideas in physics, goes like this: Sometimes nature is not sym-

IN BRIEF

Scientists are searching for unseen particles to explain the "dark matter" that seems to be exerting a gravitational pull on everything else in the universe.

An underdog candidate is the axion, a theoretical particle that could explain dark matter *and* solve a mystery about the "strong force," which binds atomic nuclei together.

The Axion Dark Matter Experiment recently became sensitive enough to either detect the most plausible versions of axions or rule them out.

metrical when it seems that it should be. For instance, if you stand a pencil on its end, there is a rotational symmetry whereby it is equally likely to fall in any direction. But what if it always falls in one direction? We would say that nature has made a choice and has "broken" the symmetry. When this happens in the context of particle physics, a new particle arises to maintain the underlying symmetry even though it appears, on the surface, to be broken. (The symmetry does not have to be obvious; it can be some abstract symmetry of the underlying mathematics.)

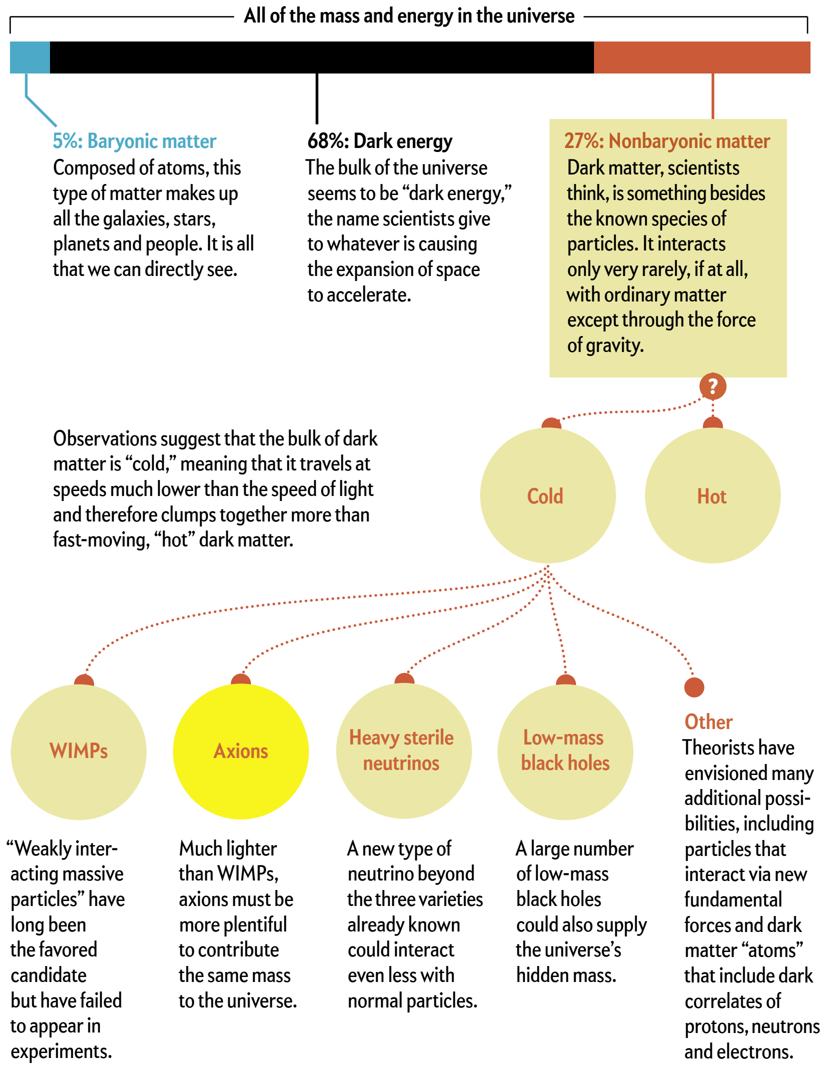
In what I thought was a brilliant insight, Quinn and Peccei applied this idea to the strong force. They speculated that a hidden type of symmetry related to this force has been broken. If this were the case, it would nullify the expected CP difference that theory predicted but that experiments failed to see. Problem solved. Shortly thereafter, in another brilliant insight, Steven Weinberg, now at the University of Texas at Austin, and Frank Wilczek, now at the Massachusetts Institute of Technology, realized this so-called Peccei-Quinn mechanism would result in a new particle: the axion. (Physics folklore says that the name was borrowed from that of a washing detergent because it "cleaned up" the strong CP problem.) By the mid-1980s theorists concluded that the big bang could have produced enough axions to account for dark matter.

The theory did not tell us how heavy axions would be or how likely they would be to interact with normal matter. We knew, though, that they had to be pretty inert because so far particle colliders and other experiments had not seen them. If they were extremely inert, they would also likely be very lightweight. And in 1987 a major cosmic event further limited the possibilities for the axion's mass. At that time a supernova exploded in the Large Magellanic Cloud, a nearby dwarf galaxy. Almost the entire gravitational binding energy of the star that collapsed escaped in the form of neutrinos, some of which made it to underground detectors here on Earth. If axions had a mass of even a few milli electron volts divided by the speed of light squared (meV/c^2) (somewhat more than one billionth the mass of the electron), they would have been produced in the explosion and distorted the escape time of the neutrinos on their way to Earth. Because experiments observed no such distortions, we knew the axions must have a smaller mass. Such light axions have extraordinarily feeble interactions with normal matter and radiation. For instance, a relatively mundane particle called the neutral pion decays into two photons roughly once every 10^{-16} second. A light axion would decay into two photons once every 10^{45} years—and that is many, many orders of magnitude longer than the age of the universe. The axion would be by far the least interactive particle known.

Interestingly, if the axion mass is too small, we have new problems. Because of the intricacies of the process by which we think axions were created near the beginning of the universe, the lower the axion mass, the greater the mass density of axions that results. Should the axion mass be too small, the big bang

Dark Matter Contenders

Something unseen appears to be exerting a gravitational pull on the normal matter in galaxies and clusters throughout the cosmos, but what is it? Scientists have theorized several potential explanations for the "dark matter" they think makes up about a quarter of the total mass and energy in the universe. These possibilities fall into various categories, as outlined.

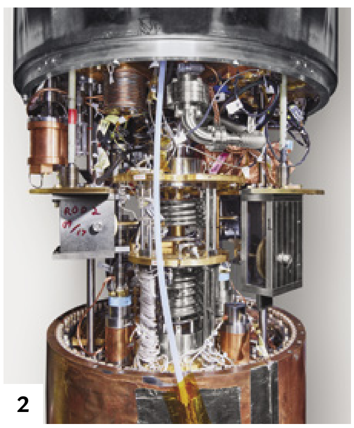


Graphic by Jen Christiansen

1

would have produced way more axions than necessary to account for dark matter. There are substantial uncertainties about this mechanism, and theorists have come up with clever ways to evade the issue, but to me, it becomes increasingly implausible to have axions with masses much below one micro electron volt divided by c^2 (μeV/c^2).

To recap, axions cannot be too heavy, or else we would have seen them already, either through particle colliders or through their effects on the evolution of other stars. Moreover, axions cannot be too light, or else there would be too much dark matter. Determining exactly what these mass limits are is very challenging, but a reasonable range of allowed dark matter axion masses is in the neighborhood of around 1 μeV/c^2 to 1 meV/c^2. This range is the "sweet spot" for the axion mass, but such particles would be so unreactive to normal matter and radiation that they have been dubbed "invisible axions."



2

SCIENTISTS attach sensors to the experiment insert (*1*). Above the insert's copper-plated cavity is a liquid-helium reservoir surrounding electronics (*2*).

SIKIVIE'S GREAT IDEA

WHEN QUINN AND PECCEI first theorized the existence of axions, physicists at Stanford and elsewhere began searching for them in the explosions produced at particle colliders. Yet the very properties that make the axion an attractive dark matter candidate—its feeble interactions with ordinary matter and radiation—made these searches feel hopeless. It was frustrating to know that we may be bathed in a dense sea of particles—about 10 trillion axions or more per cubic centimeter—that are impossible to conjure up in the laboratory.

Then Pierre Sikivie of the University of Florida had a clever idea: rather than trying to create axions in accelerators, we could look for the cosmic axions that make up the vast, pervasive sea of dark matter around us. Sikivie imagined a magnetic field filling a cylindrical cavity that was devoid of everything except, presumably, the cosmic axions that flood all of space. When an axion interacted with the magnetic field, its total energy would be almost completely converted into a photon. This interaction would be much more likely to occur if the cavity was tuned to resonate with the same frequency as the photon produced by the axion. Because axions' mass is very small, and the cosmic ones near us are presumably moving at speeds similar to the rest of the Milky Way, their energy is tiny, so the resulting photon would be somewhere in the microwave-frequency range. Exactly where, though, is a mystery until we know the precise axion mass. For this reason, experimenters would need to continually adjust the resonant frequency of the experiment's cavity to "scan" the pos-

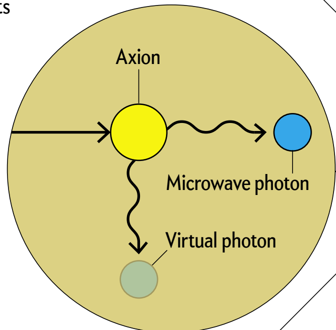
The Hardware

If axions are all around us, the Axion Dark Matter Experiment (ADMX) could find them on the rare occasions that they decay into microwave photons. To make this decay more likely, the experiment has a large magnetic field and a radio-frequency cavity that, if tuned to the same frequency as the photons produced by the axions, should encourage the transformation. In 2016 the project entered a new phase and began its most sensitive search yet.

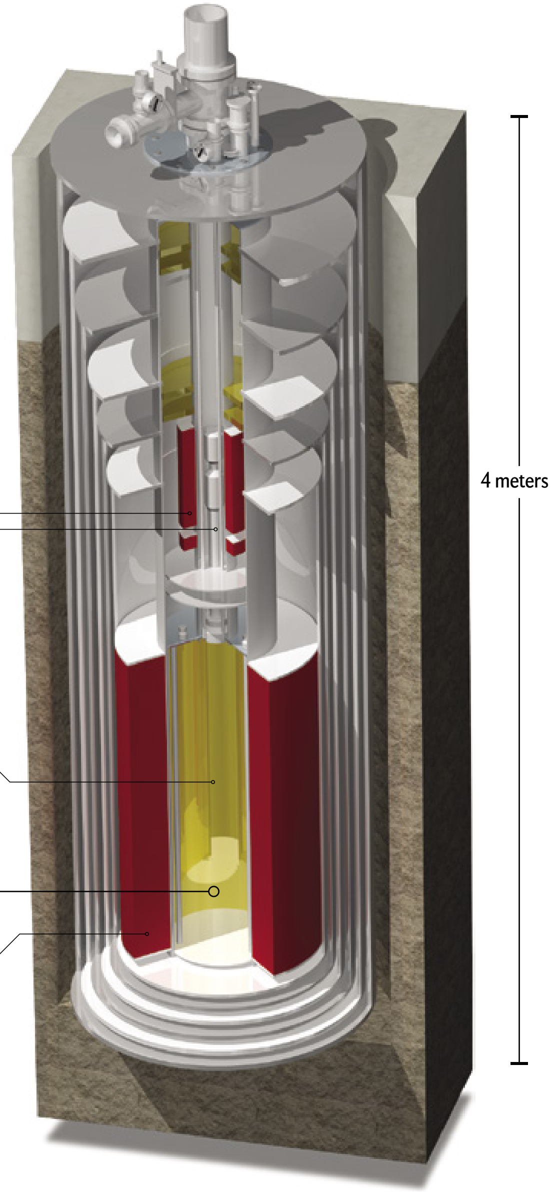
Bucking Magnet
This smaller magnet cancels out, or "bucks," the magnetic field of the main magnet in the vicinity of the SQUID amplifier, which relies on a tiny magnetic field created by the photons to detect a signal.

SQUID Amplifier
This device uses quantum-mechanical effects to detect and amplify the minute signal created when an axion decays into a photon.

Microwave Cavity
The heart of the experiment, this empty cavity is where scientists expect ambient axions, which should be present throughout space if they constitute dark matter, to transform into microwave photons under the right conditions.



8-Tesla Magnet
The main magnet in the experiment fills the cavity with a magnetic field that encourages the axions to decay into photons.



sible range in hopes of hitting on the right match for the axion.

The resulting signal would be very small, perhaps 10^{-21} watt or less, with accompanying noise of around the same amount. But very sensitive microwave detectors, collecting a signal for a sufficiently long time, should be up to the job. Two of my great loves are radio electronics and particle physics, so in my mind, Sikivie's ideas fit together powerfully.

ADMX IS BORN

I RECEIVED MY PH.D. while at Stanford in the 1980s, when the influence of Quinn and Peccei was still there, and axions made a big impression on me. They appeared to solve two huge mysteries in physics—the strong CP problem and the dark matter issue. And after Sikivie's paper, it seemed that there might be a way to detect them.

From Stanford, I moved to the University of Chicago, where I was privileged to work under the late James W. Cronin as an Enrico Fermi Fellow. There I became aware of the first attempts to put Sikivie's idea into practice, including the Rochester-Brookhaven-Fermilab experiment and a project at the University of Florida. Both lacked the sensitivity to detect axions in the plausible mass range, but they developed the basic hardware used by all subsequent experiments.

While in Chicago, I got to talking with Karl van Bibber, then at Lawrence Livermore National Laboratory, and David Tanner of the University of Florida, and we realized that we could improve on these efforts. We could begin by deploying a large cavity volume with a strong magnetic field—that would bring us partway to the sensitivity we wanted. To go the rest of the way, we knew we would need better microwave amplifiers. They

Illustration by Don Foley

1

2

EQUIPMENT RACKS house ADMX's room-temperature microwave electronics (*1*). Engineers study sensor data from the experiment (*2*).

were the key to being able to pick up and boost the extremely weak microwave signal we expected axions to produce—yet the transistor microwave amplifiers available at the time were too noisy by far. We wanted an amplifier that was limited only by the unavoidable noise produced by quantum-mechanical uncertainty, but they did not yet exist in our frequency range.

Thus was the ADMX program conceived: We would start with a large magnet, the best available microwave amplifiers and liquid helium to cool the experiment to 4.2 kelvins to reduce noise. In the intermediate term, we would focus on developing quantum-limited microwave amplifiers. In the long term, we would add a "dilution refrigerator"—a system that would cool the cavity and amplifiers to temperatures around 100 millikelvins, thus reducing noise. It was an ambitious program—each phase would take about a decade. Fortunately, we had the backing of the Department of Energy's High Energy Physics division and a vision to carry us through.

THE EARLY YEARS

IN 1993 I MOVED TO M.I.T. to be an assistant professor, and once I was there, we formed a collaboration to begin the experiment. Lawrence Livermore supplied a large superconducting magnet and the experiment site. The gifted Lawrence Livermore physicist Wolfgang Stoeffl made the initial cryogenic design, and we are still using much of his ingenious system today. Tanner largely conceived and developed the innards of the experiment based on the early University of Florida project, and our group at M.I.T. built an ultralow-noise microwave receiver to pick up the signal. In 1998 we published the initial results from this early ADMX "phase 0"—the first experiment sensitive to plausible dark matter axions. We had not found the elusive particles, but we were off to a good start.

Meanwhile we pushed forward on the quest for an amplifier that would be sensitive to the faint microwave signals we expect axions to produce. Around then, I heard a talk by John Clarke, a brilliant quantum-device physicist at the University of California, Berkeley, on quantum amplification. He had been working on so-called superconducting quantum-interference devices (SQUIDs), which take advantage of the phenomenon of quantum tunneling—the ability of a particle to pass through walls or traverse barriers that a macroscopic object cannot. If a photon arose in the experiment, it would induce a small magnetic field in the SQUID that would disrupt this tunneling in a measurable way. The devices were exquisitely sensitive, but they did not yet exist for use on microwave-frequency signals. For that application, Clarke developed what is called a microstrip DC SQUID amplifier. This gadget has a clever geometry that allows the SQUID to operate at higher frequencies.

The plan was promising, but there were still some issues. The tiny signal magnetic fields of the SQUID would be swamped by the larger field inside the ADMX cavity. The DOE reviewed our plan and flagged the SQUID issue as "high risk." At this point, in early 2002, I moved to Lawrence Livermore, and my collaborators and I decided to divide ADMX into two sequential phases: "phase 1a" would demonstrate that SQUIDs can work in the experiment's large magnetic field. A later "phase 1b" would then add the dilution refrigerator we needed to get the experiment down to the low temperatures we required.

We began phase 1a by developing a system to protect the SQUID's sensitive magnetic field from the huge field of the experiment. We did this with a series of nested shields and magnets surrounding a large magnet called a bucking coil that would cancel out, or "buck," the main magnetic field. By the mid-2000s we had demonstrated that this system works, and we began work on the dilution refrigerator—the major element needed for ADMX's phase 1b.

THE EXPERIMENT GROWS UP

AROUND THIS TIME, I MOVED to the University of Washington, and the ADMX experiment came with me to a new and substantially upgraded site. Meanwhile the DOE and the National Science Foundation were conceptualizing "Generation 2" dark matter detectors meant to be major improvements on the sensitivity of existing searches. Most of the experiments they had in mind sought WIMPs, but they were also interested in axions. Our ADMX phase 1b plans slotted closely into the Generation 2 program, and ADMX Gen 2 was born. Scheduled to begin operations in 2016 and to run into 2021, ADMX Gen 2 finally adds the dilution refrigerator into our experiment. It also more than doubles our effective data-taking rate. We have added substantial features to improve the experiment's sensitivity, and it can now conduct what we call a "definitive search"—a sweep over a broad range of axion masses, from around 1 to 40 μeV/c^2, that includes the sweet spot for predicted dark matter axions.

ADMX IS NOW SENSITIVE ENOUGH THAT IT SHOULD BE ABLE TO DETECT AXIONS OR TO RULE OUT THE MOST PLAUSIBLE VERSIONS OF THEM OVER THE NEXT FIVE TO 10 YEARS.

ADMX has many complicated parts that must all work in concert, but most of its systems are now highly refined and reliable. The collaboration has grown to include Lawrence Livermore, U.C. Berkeley, the University of Florida, the University of Washington, Washington University in St. Louis, Pacific Northwest National Laboratory, Los Alamos National Laboratory, Fermi National Accelerator Laboratory, the National Radio Astronomy Observatory and the University of Sheffield in England. A new ADMX leadership team has emerged, with co-spokespersons Gray Rybka of the University of Washington and Gianpaolo Carosi of Lawrence Livermore.

Although we are now surveying the most likely mass range for dark matter axions, there is always a chance nature could surprise us. Searching in a slightly lower mass range is not very difficult, but outfitting our experiment to look for even higher masses is a challenge. As the axion mass increases, the cavity's resonant frequency needs to increase as well, and thus the diameter of the cavity must decrease, thereby reducing the available volume to search for axions. We could pack a large number of cavities inside a single big magnet to maintain a large volume, but doing so becomes a "Swiss watch problem": the complexity of such a system is daunting. We may also be able to live with a small cavity as long as we can increase the strength of the magnetic field to compensate. Such an increase is expensive, but research into this possibility is underway. Perhaps within five to 10 years increased magnetic field strength—to 32 or even 40 tesla—could expand the mass range of our search. At much higher axion masses—those approaching 1 meV/c^2—we may even be able to see a signal from space. If axions exist in this range and form dark matter halos around galaxies, radio telescopes could spot very weak emission lines.

Eventually ADMX and other projects will be able to fully explore the theoretical window of possible dark matter axion masses. The fact that the full plausible mass range is totally accessible to experiments makes axions an attractive candidate for dark matter, compared with some alternatives that we may never be able to test completely.

As our experimental work marches on, theorists are also making progress on trying to understand the nature of dark matter. Sophisticated cosmological models running on supercomputers are working on more reliable predictions of the axion mass. It is also possible, for instance, that axions would clump together throughout the universe in a different pattern than WIMPs would, in ways both subtle and dramatic. Future astrophysical experiments, such as the Large Synoptic Survey Telescope due to begin observations in 2019, may be able to map out the large-scale structure in the universe accurately enough to allow scientists to discriminate among the dark matter candidates.

Another possibility is that the axions predicted by quantum chromodynamics are just a reflection of some greater theory of physics existing on a higher energy scale. One such theory contender, string theory, predicts axions with much smaller masses than those probed by ADMX. String theory, however, is highly speculative, as are its predictions.

Twenty years ago the comfortable consensus was that dark matter is made up of WIMPs. Since then, the appeal of axions has increased. In the not too distant future, we should know whether or not they are the solution to the mystery of the invisible side of the cosmos. SA

Leslie Rosenberg is a professor of physics at the University of Washington. He has been hunting for axion dark matter for more than two decades.

MORE TO EXPLORE

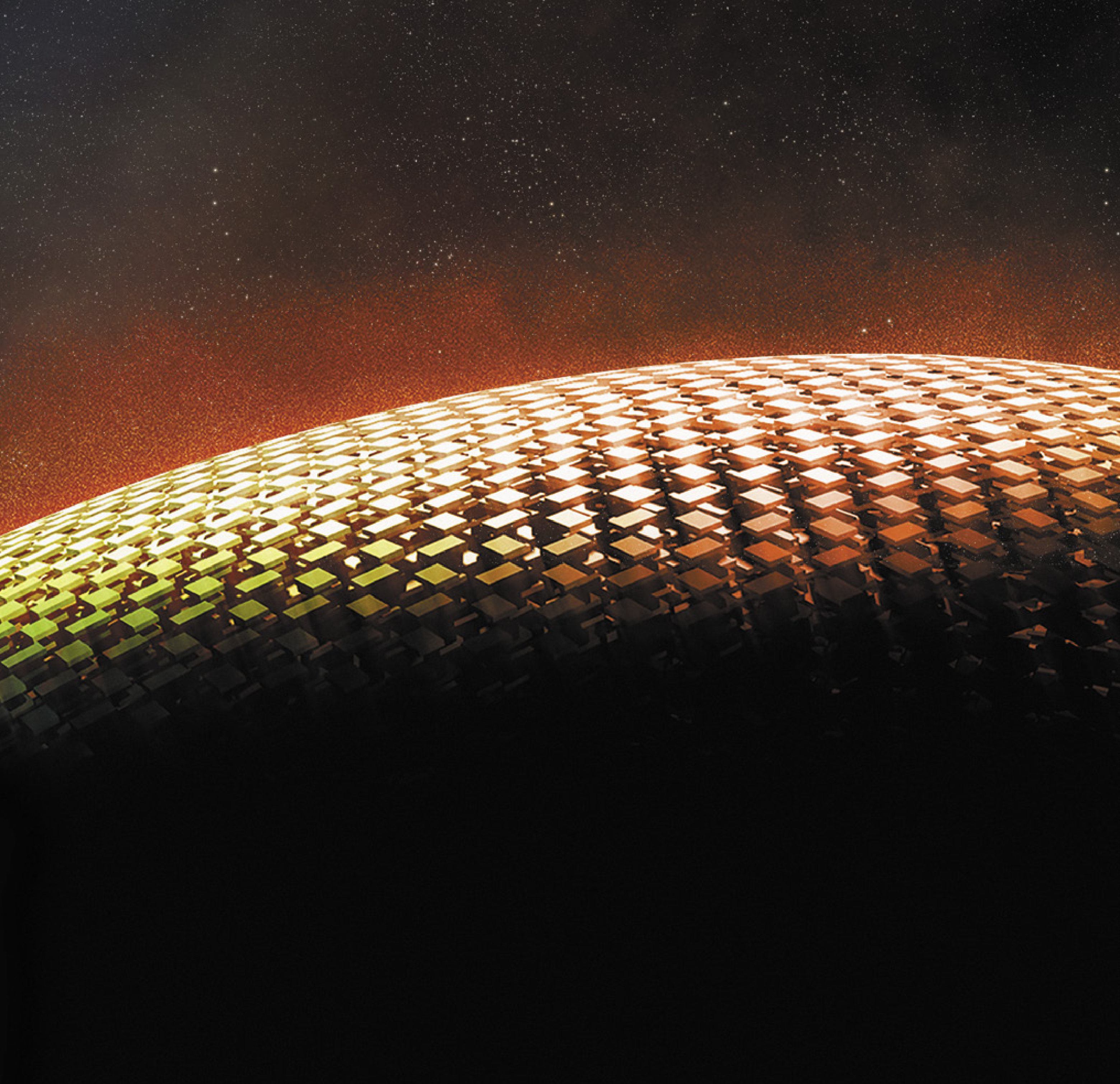
A New Light Boson? Steven Weinberg in *Physical Review Letters,* Vol. 40, No. 4, pages 223–226; January 23, 1978.

Problem of Strong *P* and *T* Invariance in the Presence of Instantons. F. Wilczek in *Physical Review Letters,* Vol. 40, No. 5, pages 279–282; January 30, 1978.

Axions, Domain Walls, and the Early Universe. P. Sikivie in *Physical Review Letters,* Vol. 48, No. 17, pages 1156–1159; April 26, 1982.

scientificamerican.com/magazine/sa

BLACK from the

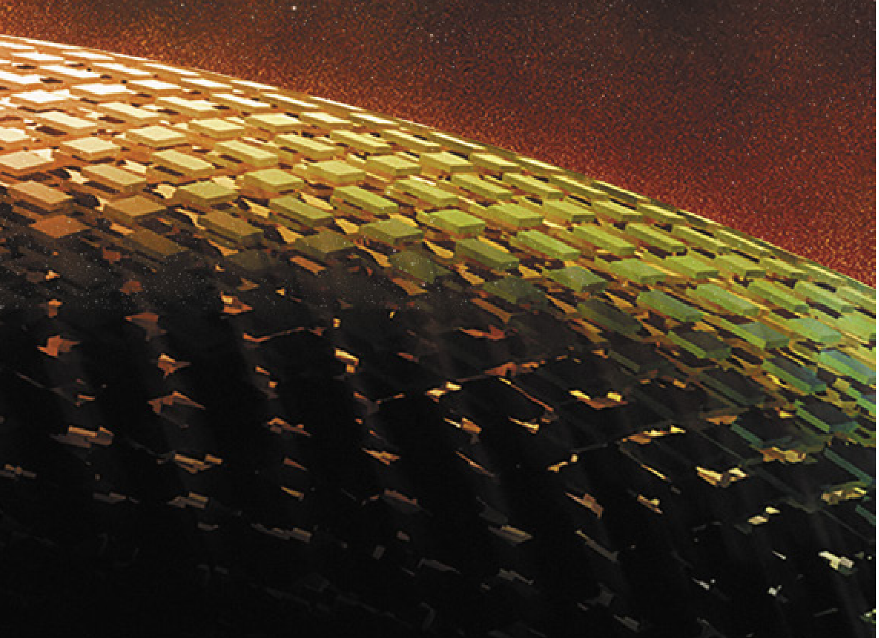


HOLES Beginning of Time

A hidden population of black holes born less than one second after the big bang could solve the mystery of dark matter

By Juan García-Bellido and Sébastien Clesse

Illustration by Kenn Brown, Mondolithic Studios



IN BRIEF

The nature of dark matter—the invisible material that holds galaxies together by its gravity—is a deep cosmic enigma.

Many researchers suspect dark matter is made of weakly interacting massive particles and have been seeking them in experiments. But to date, no such "WIMPs" have been found.

"Primordial" black holes that may have formed shortly after the big bang are an alternative candidate for dark matter. Yet these, too, have so far eluded detection.

More evidence for primordial black holes may emerge in new data from gravitational-wave detectors and other observatories. If confirmed to exist, these objects could solve the mystery of dark matter and several other cosmic conundrums.

More than a billion years ago two black holes in the distant universe spiraled around each other in a deathly dance until they merged. This spiraling collision was so violent that it shook the fabric of spacetime, sending perturbations—gravitational waves—rippling outward through the cosmos at the speed of light. In September 2015, after traveling more than a billion light-years, those ripples washed over our planet, registering as a "chirp" in the sensors of the Advanced Laser Interferometer Gravitational-wave Observatory (LIGO).

This was the first direct detection of gravitational waves, and the observation confirmed Albert Einstein's century-old prediction of their existence. Yet the chirp revealed that each of the merger's progenitor black holes was 30 times heavier than the sun. That is, their masses were two to three times larger than ordinary black holes born from supernova explosions of massive stars. These black holes were so heavy, it is hard to explain how they formed from stars at all. Furthermore, even if two such black holes did independently form from the deaths of very massive stars, they would have then had to find each other and merge—an event with an exceedingly low probability of occurring within the current age of the universe. It is thus reasonable to suspect that these massive black holes formed via some other, more exotic pathway that might not involve stars at all. Beyond its detection of gravitational waves, it may be that LIGO has unveiled something even more extraordinary: black holes that predate the formation of the stars themselves.

Although such "primordial" black holes have never before been seen, some theoretical models suggest they could have formed in astronomical numbers from the hot, dense plasma that filled the cosmos less than one second after the big bang. This hidden population could solve several outstanding mysteries in modern cosmology. In particular, primordial black holes could constitute some, if not all, of dark matter—the invisible 85 percent of the matter in the universe that acts as gravitational glue to hold galaxies and galaxy clusters together. Further studies with LIGO and other facilities will soon test these ideas, potentially unleashing a new revolution in our understanding of the cosmos.

THE FALL OF MACHOS, THE RISE OF WIMPS

BLACK HOLES would initially seem to be ideal candidates for dark matter because they emit no light. Indeed, along with other dark objects such as planets and brown dwarfs, they make up one long-proposed solution to the dark matter problem: MACHOs, short for *ma*ssive *c*ompact *h*alo *o*bjects. Found both in spherical halos surrounding each galaxy and near each galaxy's luminous center, MACHOs would create the gravitational pull responsible for the otherwise anomalous motions of stars and gas that astronomers observe in the outskirts of galaxies. Simply put, galaxies seem to be rotating too fast to be held together by the visible mass in stars that we observe. Dark matter provides the extra pull to prevent spinning galaxies from flinging off their stars.

If MACHOs make up most of the universe's dark matter, they must also account for other observations. Whatever dark matter is, it shapes the universe's largest structures, determining the origin and growth of galaxies as well as clusters and superclusters of galaxies. These objects coalesce from the gravitational collapse of clumps of gas inside dark matter halos. Cosmologists have precisely mapped the spatial distribution of these clumps through deep and wide galaxy surveys and correlated them with tiny temperature fluctuations present in the cosmic microwave background (CMB), the big bang's all-sky afterglow. The diffuse mass of dark matter in large galaxies and clusters also bends space to distort the light from far distant background objects—a phenomenon known as gravitational lensing.

The MACHO hypothesis, however, fell from favor a decade ago when MACHOs did not turn up in tentative, indirect searches for their existence. Most notably, astronomers looked for them via microlensing, a variety of gravitational lensing in which a black hole, a brown dwarf or even a planet passes in front of a background star and temporarily magnifies the star's light. Several multiyear microlensing surveys of millions of stars in the Large and Small Magellanic Clouds, the main satellite galaxies of the Milky Way, found no evidence that MACHOs made up the entirety of our galactic halo. These results were conclusive enough to rule out MACHOs up to around 10 solar masses as the primary constituent of dark matter. As these surveys took place, theorists built the case for an alternative hypothesis—WIMPs, or weakly interacting massive particles.

WIMPs are predicted by certain extensions of the Standard Model of particle physics, but they remain at least as elusive as MACHOs. To date, no evidence of their existence has been found despite decades of searches using particle accelerators, underground detectors and space telescopes. As null results piled up in the search for WIMPs, some researchers began reconsidering the MACHO hypothesis, focusing particularly on primordial black holes. But what process could have seeded these strange objects throughout the observable universe, and how could they have eluded detection for so long?

BLACK HOLES FROM THE BIG BANG

PHYSICISTS BERNARD CARR and the late Stephen Hawking proposed the idea of primordial black holes in the 1970s. They considered black holes with masses smaller than that of a mountain. Such minuscule black holes would have already evaporated and vanished within the age of our nearly 14-billion-year-old universe, via a quantum-mechanical process discovered by Hawking and appropriately called Hawking radiation. But they also investigated the possibility that more massive, nonevaporating black holes could constitute the missing matter in galaxy clusters. The possibility that massive primordial black holes could actually be most or even all of the dark matter hinges on an idea known as cosmic inflation, first proposed by physicist Alan Guth in the early 1980s.

Inflation is a hypothetical phase of prodigious expansion immediately after the big bang. In 10^{-35} second, two points separated by less than an atomic radius would have become separated by four light-years, a distance comparable to that of the closest stars. Moreover, during inflation tiny quantum fluctuations are magnified to macroscopic scales by the rapid expansion, seeding the growing universe with underdense and overdense regions of matter and energy from which all cosmic structures subsequently emerge. As bizarre as it may seem, the theory of inflation is strongly supported by observations of such density fluctuations in the CMB.

In 1996 one of us (García-Bellido), together with Andrei Linde of Stanford University and David Wands of the University of Portsmouth in England, discovered a way for inflation to form sharp peaks in the spectrum of density fluctuations in the early universe [*see box on opposite page*]. That is, we showed how quantum fluctuations enormously magnified by inflation would naturally produce particularly dense regions that would collapse to form a population of black holes less than one second after inflation ends. Such black holes would then behave as dark matter and would dominate the matter content of the present-day universe. This model generated a population of black holes with the same mass, determined by the amount of energy within the collapsing region. Many other groups then started exploring these ideas within different models of inflation.

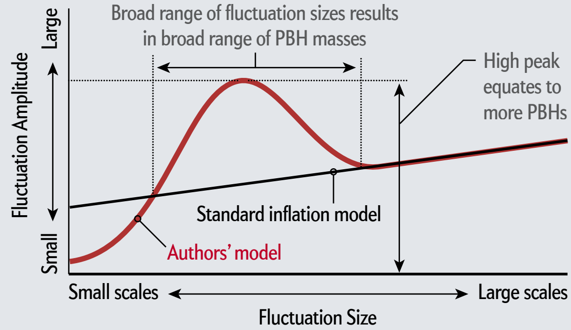
In 2015 the two of us (Clesse and García-Bellido) proposed a scenario, similar to that of 1996, in which these primordial fluctuations exhibit a broad peak in their energy densities and spatial sizes, giving rise to primordial black holes with a wide range of masses. A key consequence of this scenario is the fact that large density fluctuations collapse in close spatial proximity to one another, generating clusters of black holes of different masses—from one 100th to 10,000 times the mass of our sun. Within half a million years of the big bang, each growing, evolving cluster could contain millions of primordial black holes in a volume just hundreds of light-years across.

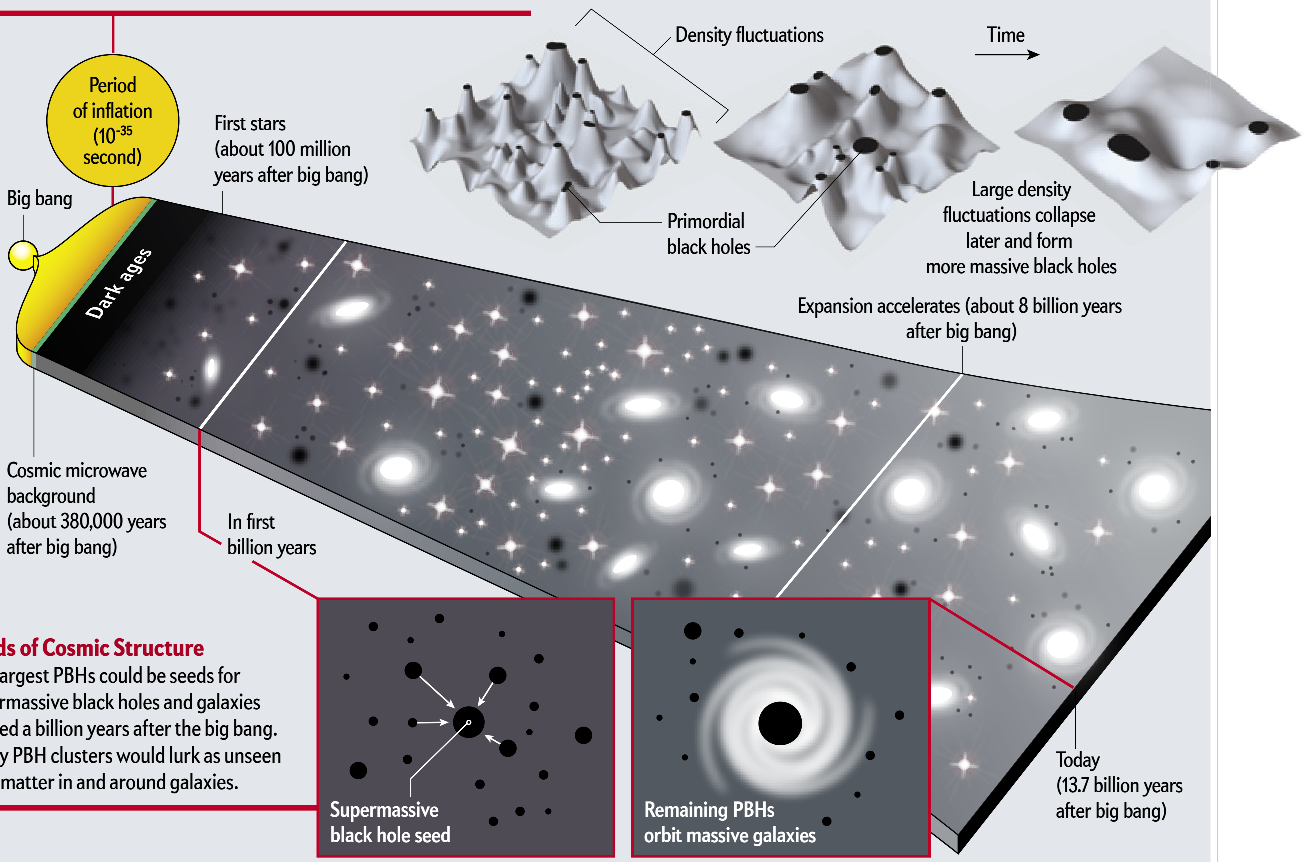
Black Holes Birthed by the Big Bang

The universe's first black holes may have been born in the earliest moments of cosmic time, when all was a seething, thick fog of fundamental particles. In the 1970s theorists realized that dense regions of that fog could collapse under their own gravity just a second after the big bang, forming so-called primordial black holes (PBHs) that would then shape the structure of the evolving, expanding universe. Emitting no light, PBHs would be a natural—albeit difficult to detect—candidate for dark matter.

Primordial Black Holes Form in Clusters

Inflation—a proposed acceleration to the universe's expansion less than a second after the big bang—would form PBHs by magnifying quantum fluctuations to immense scales. As inflation ended, these fluctuations would create density perturbations that then form PBHs. Larger, more powerful fluctuations would create more massive and numerous PBHs. The authors' inflationary model predicts a broad peak of magnified fluctuations and a range of density perturbations, producing PBHs in clusters, with each PBH ranging from one 100th to 10,000 times the mass of our sun. Half a million years after the big bang, a cluster might span hundreds of light-years and contain millions of PBHs. As the PBHs within such clusters merged together, scattered apart, and fed on ordinary gas and dust, they would guide the growth of galaxies and galactic clusters.





Seeds of Cosmic Structure

The largest PBHs could be seeds for supermassive black holes and galaxies formed a billion years after the big bang. Today PBH clusters would lurk as unseen dark matter in and around galaxies.

Such clusters of primordial black holes would be sufficiently dense to explain LIGO's mysterious black hole mergers, which one would not otherwise expect to occur with regularity. From time to time, the trajectories of two primordial black holes within a cluster can cross, so that both objects become gravitationally bound to each other. They would then spiral closer together for up to millions of years, radiating gravitational waves until they merged. In January 2015 we actually predicted that LIGO would detect gravitational waves from such massive mergers—waves identical to those LIGO then detected later that year. Our estimates for the rate of merger events within primordial black hole clusters fit perfectly within the limits set by LIGO. If LIGO and other similar facilities detect many more mergers within the next few years, it may be possible to deter-

Illustration by Jen Christiansen and George Retseck (density-fluctuation panels)

Is Dark Matter Made of Primordial Black Holes?

The quest for observational evidence continues

Seven observations point toward the hypothesis that primordial black holes constitute dark matter in the universe:

1. Mass, spin and merging rates of black holes. Gravitational-wave detectors are finding black hole mergers at a rate suggesting an abundance of black holes similar to the dark matter. In June 2017 LIGO announced the detection of black hole merger GW170104, for which the axis of rotation of one of the black hole progenitors was very likely not aligned with the axis of the orbital plane. This would be a sign that the black hole binary probably formed by a capture process, as expected for primordial black holes rather than from a preexisting star binary. Ultimately if we found a black hole with a mass lower than the so-called Chandrasekhar limit (1.4 solar masses, below which stars cannot produce a black hole), it would undeniably have a primordial origin. Fortunately, LIGO may very soon reach the sensitivity to detect such a black hole if its companion is more massive (greater than 10 solar masses). Finally, on cosmological scales, abundant black hole binaries should induce a diffuse background of gravitational waves, which could be detected by the future space-based Laser Interferometer Space Antenna (LISA) and by ground-based pulsar timing arrays.

2. Microlensing of stars in M31 and distant quasars. A broad mass distribution of primordial black holes, centered on a few solar masses, would perfectly explain not only the LIGO/VIRGO gravitational-wave detections but also the microlensing of stars in the Andromeda galaxy (M31) and of distant quasars if they constitute dark matter. Such a model passes all the other cosmological and astrophysical constraints.

3. Ultrafaint dwarf galaxies. In 2015 astronomers using data from the Dark Energy Survey discovered dozens of ultrafaint dwarf galaxies in the galactic halo, suggesting that hundreds of such dark matter–dominated dwarf galaxies could orbit around the Milky Way. Strangely, none of them had a radius smaller than 50 light-years. Regular close encounters of primordial black holes with stars would make such dwarf galaxies unstable. On the other hand, a rapid accretion episode in the densest halos provides a new mechanism to explicate why some dwarf galaxies are strongly dominated by dark matter and others are not. Future space-based facilities such as the European Space Agency's Euclid mission and NASA's Wide-field Infrared Survey Telescope (WFIRST) should detect many more ultrafaint dwarf galaxies and so will test this hypothesis.

4. Galactic dark matter profiles. New observations suggest the dark matter distribution is more homogeneous near the center of galaxies of all sizes, which would hardly be explained by any known astrophysical processes such as supernovae feedback. Primordial black holes of several solar masses would naturally explain the existence of core dark matter profiles, thanks to gravitational scattering that homogenizes their distribution in the densest regions.

5. Correlations between x-ray and infrared backgrounds. Alexander Kashlinsky of NASA's Goddard Space Flight Center suggested that correlations between x-ray and infrared cosmic background fluctuations that cannot be explained by any known population of galaxies would naturally arise if dark matter is made of massive primordial black holes. They would emit x-rays when they accrete matter within black hole clusters where a first generation of stars could also ignite and shine in infrared.

6. Billions of supermassive black holes. Recent observations of Chandra suggest that billions of supermassive black holes already existed one billion years after the big bang, shortly after the formation of the first stars. The required black hole seeds could have a primordial origin.

7. Unexpectedly cool universe at the time of star formation. In February 2018 a modest radio telescope built by a small team of astronomers at the Massachusetts Institute of Technology and at Arizona State University, called EDGES, might have detected the light from the epoch of formation of the first star generation, 200 million years after the big bang. Yet the amplitude of the signal was totally unexpected and suggests that the universe was half as warm—or colder—as in any realistic scenario. We are exploring if primordial black hole dark matter could explain this observation, for instance, as an indirect effect of accretion of the neutral hydrogen gas.

mine the range of masses and spins for all the progenitor black holes. Such a statistical analysis of black hole mergers would provide crucial information for testing their potentially primordial origins.

A key aspect of this scenario is that it evades the constraints on MACHOs previously set by gravitational microlensing experiments—constraints that ruled out black holes of up to about 10 solar masses as the main constituent of dark matter. If primordial black holes exist and possess a wide range of masses, only a small fraction would be visible to these microlensing experiments, with the bulk remaining invisible. And if primordial black holes are grouped in clusters, this arrangement suggests a probability of less than one in 1,000 that a cluster would happen to be along the line of sight of the stars in the nearby satellite galaxies monitored for microlensing events.

To avoid this effect, one could search for microlensing events elsewhere in the sky, looking for the magnified light from stars in the Milky Way's neighboring Andromeda galaxy or even from quasars in far distant galaxies. In this way, one could probe a much larger volume of galactic halos for signs of MACHOs—that is, for primordial black holes. Recent observations suggest that whereas MACHOs of up to 10 solar masses may not make up the entirety of an average galaxy's halo, MACHOs between one tenth and a few solar masses could easily account for about 20 percent of the mass in a typical galactic halo. This value is consistent with our broad-mass primordial black hole scenario.

Simply put, we cannot yet rule out the possibility that dark matter is mostly made up of primordial black holes. Indeed, this proposed scenario could decipher several other cosmic mysteries related to dark matter and galaxy formation.

MANY PROBLEMS, ONE SOLUTION

CLUSTERS OF PRIMORDIAL BLACK HOLES could clear up the so-called missing satellite problem—the apparent lack of "dwarf" satellite galaxies that should form around massive galaxies such as our Milky Way. Current simulations modeling the cosmic distribution of dark matter accurately replicate the universe's observed large-scale structure, in which halos of dark matter pull galaxy clusters into giant filaments and sheets surrounding great voids of lower density. On smaller scales, however, these simulations predict the existence of numerous subhalos of dark matter orbiting around massive galaxies. Each of these subhalos should host a dwarf galaxy, and hundreds should surround the Milky Way. Yet astronomers have found far fewer dwarf galaxies than predicted.

Many potential explanations for the missing satellite problem exist, mainly the notion that simulations fail to fully account for the influence of ordinary matter (hydrogen and helium in stars) on the formation and behavior of the predicted dwarf galaxies. Our scenario suggests that if clustered primordial black holes made up most dark matter, they would dominate the subhalos surrounding the Milky Way, absorbing a fraction of ordinary matter and reducing the rate of star formation in the subhalos. Moreover, even if these subhalos vigorously formed stars, these stars could easily be ejected by close encounters with massive primordial black holes. Both effects would greatly reduce the brightness of the satellites, making them very hard to detect without wide-field cameras of exquisite sensitivity. Fortunately, such cameras now exist, and astronomers have already used them to discover dozens of ultrafaint dwarf galaxies surrounding the Milky Way. These objects appear to host up to hundreds of times more dark matter than luminous stars, and our model predicts that thousands more should orbit our galaxy.

Simulations also predict a population of galaxies intermediate in size between dwarf galaxies and massive galaxies. Such objects are said to be too big to fail because they would be sufficiently large to readily form stars and be easily seen. Still, they have not turned up in astronomers' searches of the Milky Way's vicinity. This too-big-to-fail problem has a solution similar to that of the missing satellite problem: massive primordial black holes in the cores of intermediate-sized galaxies could eject stars and star-forming gas from these objects, rendering them effectively invisible to most surveys.

Primordial black holes could also resolve the origin of supermassive black holes (SMBHs). These monsters weigh from millions to billions of solar masses and are observed at the centers of quasars and massive galaxies very early in the universe's history. Yet if these SMBHs formed and grew from the gravitational collapse of the universe's first stars, they should not have acquired such gigantic masses in such a relatively short time—less than a billion years after the big bang.

In our scenario, although most primordial black holes have just tens of solar masses, a very small fraction will be far heavier, ranging from hundreds to tens of thousands of solar masses. Born less than a second after the big bang, these monstrous objects would then act as giant seeds for the formation of the first galaxies and quasars, which would rapidly develop SMBHs at their centers. Such seeds could also account for the existence of intermediate-mass black holes possessing 1,000 to a million solar masses observed orbiting SMBHs and at the centers of globular clusters of stars. In short, primordial black holes may be the missing link between conventional stellar-mass black holes and SMBHs. The observational case for this scenario is building rapidly: recent detections of unexpectedly abundant x-ray sources in the early universe are most easily explained by large numbers of primordial black holes producing x-rays as they gorge on gas less than one billion years after the big bang.

SEEING IN THE DARK

EVEN THOUGH massive primordial black holes could solve the mystery of dark matter, as well as many other long-standing problems of cosmology, the game is not yet over. Other models and explanations are still possible, and future observations should allow us to distinguish among the alternatives. Indeed, there already exist seven kinds of observations to test the primordial black hole scenario [*see box on opposite page*]. They include the detection of ultrafaint dwarf galaxies, the influence of massive primordial black holes on the positions of stars in the Milky Way, the mapping of neutral hydrogen during the first epoch of star formation and future observations of distortions in the cosmic microwave background.

Beyond these experiments, we also now possess a completely new tool to unravel the mysteries of the universe in the form of Advanced LIGO and other gravitational-wave detectors. If indeed LIGO has detected merging members of a hidden population of massive primordial black holes, we should expect many more to be detected in coming years. Scientists from the Advanced LIGO and VIRGO collaborations have presented so far 10 confirmed detections of gravitational waves emitted during the merging of black hole binaries, and many new detections are expected in the next run, starting in the spring of 2019. These detections suggest that binary black holes are much more frequent than expected and that they are broadly distributed in mass, in agreement with our scenario of massive primordial black holes.

Taken together these new experiments and observations could confirm the existence of primordial black holes and their possible linkage to the universe's missing matter. Soon we may no longer be in the dark about dark matter. SA

Juan García-Bellido is a theoretical physicist and professor at the Institute for Theoretical Physics in Madrid. His research interests are the early universe, dark energy, black holes and quantum gravity. García-Bellido is an active member of the Dark Energy Survey and the European Space Agency missions Euclid and Laser Interferometer Space Antenna (LISA).

Sébastien Clesse is a Belgian cosmologist and an F.R.S.-FNRS postdoctoral researcher at the University of Louvain and University of Namur in Belgium. His work covers cosmic inflation, modified gravity and primordial black holes. Clesse is an active member of the Euclid mission, the Square Kilometer Array collaboration and the LISA collaboration.

MORE TO EXPLORE

Density Perturbations and Black Hole Formation in Hybrid Inflation. Juan García-Bellido, Andrei Linde and David Wands in *Physical Review D,* Vol. 54, No. 10, pages 6040–6058; November 15, 1996.

LIGO Gravitational Wave Detection, Primordial Black Holes, and the Near-IR Cosmic Infrared Background Anisotropies. A. Kashlinsky in *Astrophysical Journal Letters,* Vol. 823, No. 2, Article No. L25; June 1, 2016.

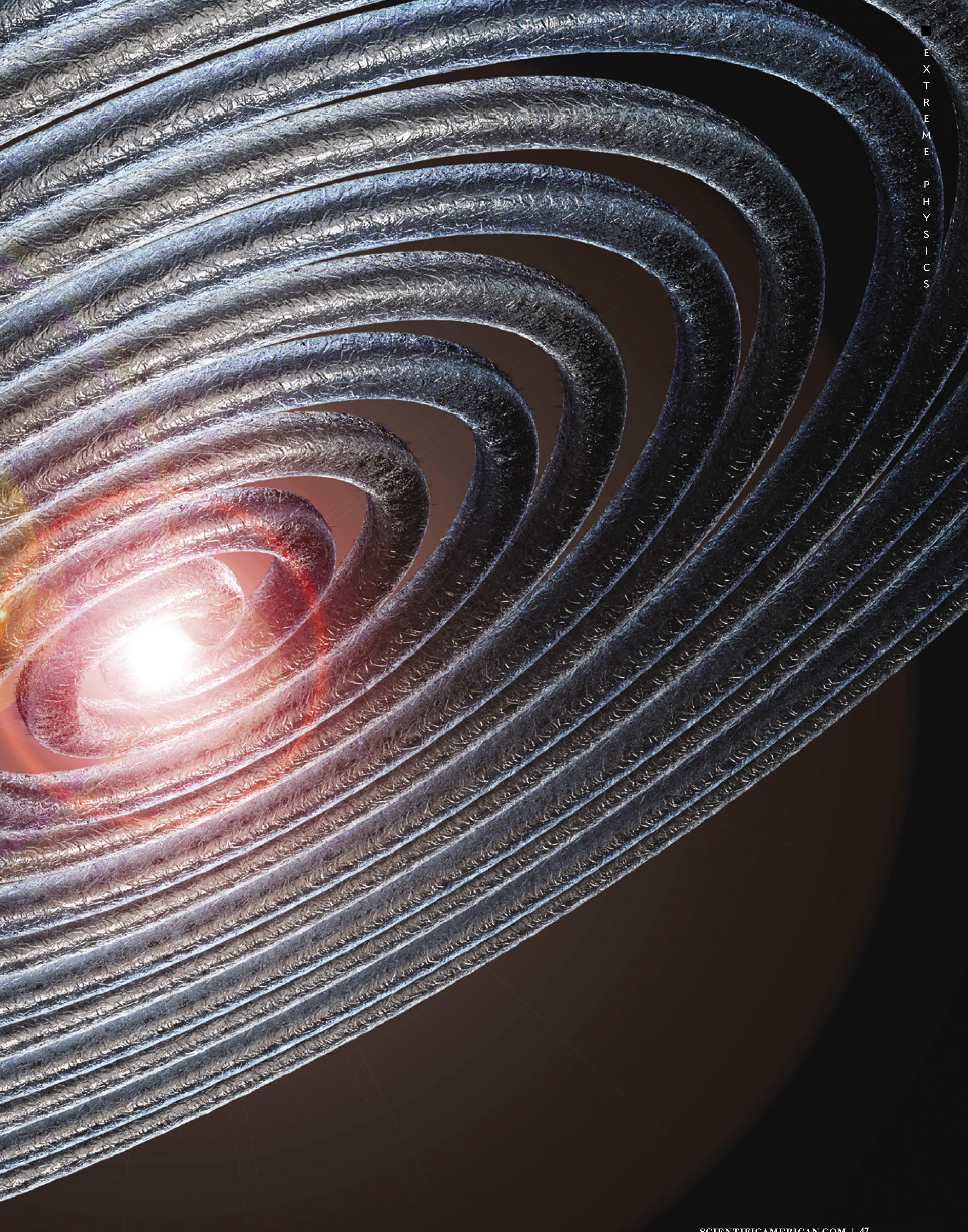
The Clustering of Massive Primordial Black Holes as Dark Matter: Measuring Their Mass Distribution with Advanced LIGO. Sébastien Clesse and Juan García-Bellido in *Physics of the Dark Universe,* Vol. 15, pages 142–147; March 2017.

scientificamerican.com/magazine/sa

HERE COME THE WAVES

After a clutch of historic detections, gravitational-wave researchers have set their sights on some ambitious scientific quarry

By Davide Castelvecchi



IN THE MID-1980S BERNARD SCHUTZ CAME UP WITH A NEW SOLUTION TO ONE of astronomy's oldest problems: how to measure the distance from the earth to other objects in the cosmos. For generations, researchers have relied on an object's brightness as a rough gauge for its distance. But this approach carries endless complications. Dim, nearby stars, for example, can masquerade as bright ones that are farther away. Schutz, a physicist at Cardiff University in Wales, realized that gravitational waves could provide the answer. If detectors could measure these ripples in spacetime emanating from interacting pairs of distant objects, scientists would have all the information needed to calculate how strong the signal was to start with—and so how far the waves must have traveled to reach the earth. Thus, he predicted, gravitational waves could be unambiguous markers of how quickly the universe is expanding.

His idea was elegant but impractical: nobody at the time could detect gravitational waves. But in August 2017 Schutz finally got the opportunity to test this concept when the reverberations of a 130-million-year-old merger between two neutron stars passed through gravitational-wave detectors on the earth. As luck would have it, the event occurred in a relatively nearby galaxy, producing a much cleaner first measure than Schutz had dreamed. With that one data point, Schutz was able to show that his technique could become one of the most reliable for measuring distance. "It was hard to believe," he says. "But there it was."

More mergers like that one could help researchers to resolve an ongoing debate over how fast the universe currently is expanding. But cosmology is just one discipline that could make big gains through detections of gravitational waves in the coming years. With a healthy bunch of discoveries already under their belts, gravitational-wave scientists have a long list of what they expect more data to bring, including insight into the origins of the universe's black holes; clues about the extreme conditions inside neutron stars; a chronicle of how the universe structured itself into galaxies; and the most stringent tests yet of Albert Einstein's general theory of relativity. Gravitational waves might even provide a window into what happened in the first few moments after the big bang.

IN BRIEF

Since the first gravitational-wave detection in 2015, the field has yielded an unprecedented quarry of discoveries. But scientists have a long list of expectations they hope the newest data will fulfill.

Several new detector facilities are under construction or close to launch. With increased sensitivity and sophistication, the totality of black hole detections will soon be able to delineate a map of the universe.

Wave detections, in combination with other data, have led to a new type of cosmological indicator, a standard siren; it is a new measurement of distance between the earth and the object being detected.

Future projects may be able to detect signals from events much farther away, perhaps from the entire observable universe.

PRECEDING PAGES: GETTY IMAGES

Researchers will soon start working down this list with the help of the U.S.-based Laser Interferometer Gravitational-wave Observatory (LIGO), the Virgo observatory near Pisa, Italy, and a similar detector in Japan that could begin making observations in late 2019. They will get an extra boost from space-based interferometers and from terrestrial ones that are still on the drawing board—as well as from other methods that could soon start producing their own first detections of gravitational waves.

Like many scientists, Schutz hopes that the best discoveries will be ones that no theorist has even dreamed of. "Any time you start observing something so radically new, there's always the possibility of seeing things you didn't expect."

SPINNING CLUES

FOR A FIELD of research that is not yet five years old, gravitational-wave astronomy has delivered discoveries at a staggering rate, outpacing even the rosiest expectations. In addition to the discovery in 2017 of the neutron-star merger, LIGO has recorded 10 pairs of black holes coalescing into larger ones since 2015 [*see box on page 53*]. The discoveries are the most direct proof yet that black holes truly exist and have the properties predicted by general relativity. They have also revealed, for the first time, pairs of black holes orbiting each other.

Researchers now hope to find out how such pairings came to be. The individual black holes in each pair should form when massive stars run out of fuel in their cores and collapse, unleashing a supernova explosion and leaving behind a black hole with a mass ranging from a few to a few dozen suns.

There are two leading scenarios for how such black holes could come to circle each other. They might start as massive stars in each other's orbit and stay together even after each goes supernova. Alternatively, the black holes might form independently but be driven together later by frequent gravitational interactions with other objects—something that could happen in the centers of dense star clusters.

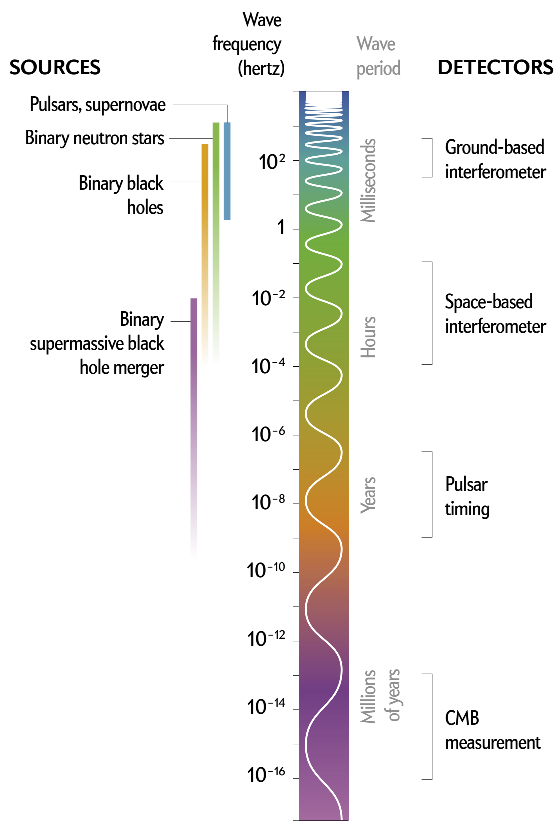
Either way, the objects' energy gradually disperses in the form of gravitational waves, a process that pulls the pair into an ever tighter and faster spiral, eventually fusing into one more massive black hole. Ilya Mandel, a LIGO theorist at the University of Birmingham in England, says that for LIGO and Virgo to see such pairs merge, typical black holes need to have started their mutual orbit separated by a distance of less than one quarter that between the earth and the sun. "If you start out with the two black holes any farther apart, it will take longer than the age of the universe" for them to merge, Mandel says.

The black hole mergers discovered so far are not sufficient to determine which formation scenario dominates. But in an analysis of the first three detections, a group including Mandel and Will M. Farr, a theoretical astrophysicist and LIGO member at the University of Birmingham, suggested that just five more observations could provide substantial evidence in favor of one scenario or the other. This would involve scrutinizing the gravitational waves for clues about how black holes rotate: those that pair up after forming independently should have randomly oriented spins, whereas those with a common origin should have spin axes that are parallel to each other and roughly perpendicular to the plane in which they orbit.

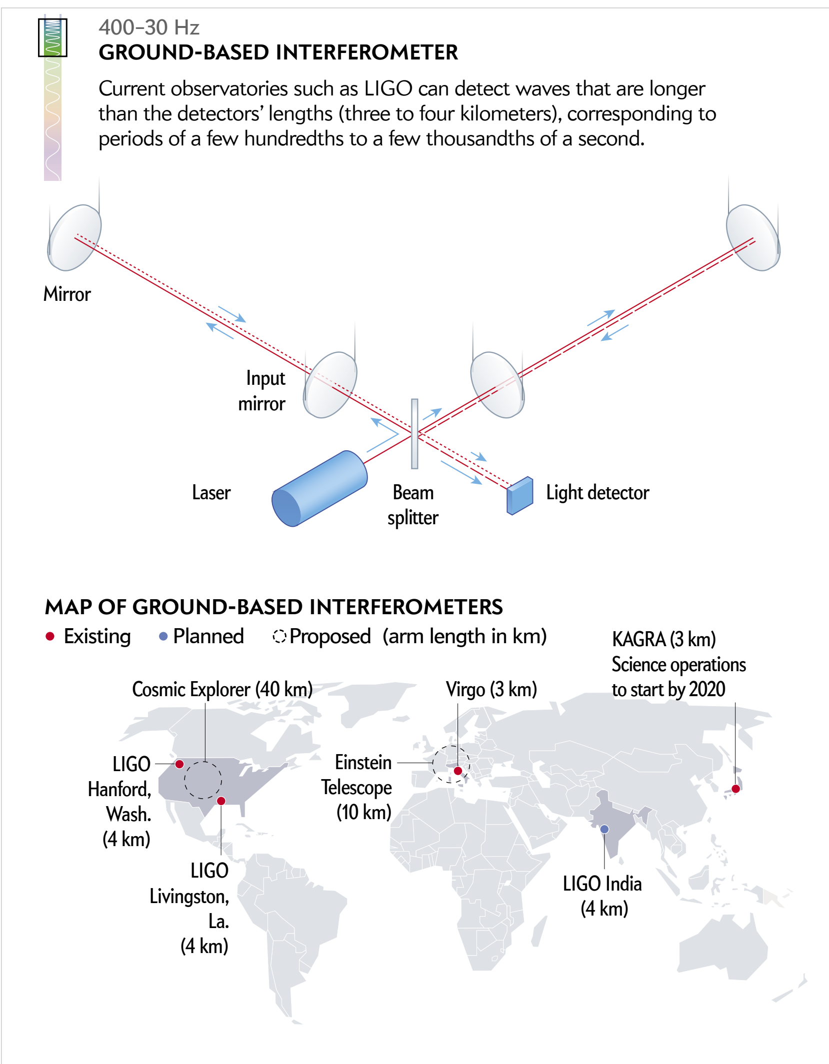
Further observations could also provide insight into some of the fundamental questions about black hole formation and stellar evolution. Collecting many measurements of masses should reveal gaps—ranges in which few or no black holes exist, says Vicky Kalogera, a LIGO astrophysicist at Northwestern University. In particular, "there should be a paucity of black holes at the low-mass end," she says, because relatively small supernovae tend to leave behind neutron stars, not black holes, as remnants. And at the high end—around 50 times the mass of the sun—researchers expect to see another cutoff. In very large stars, pressures at the core are thought eventually to produce antimatter, causing an explosion so violent that the star simply disintegrates without leaving any remnants at all. These events, called pair-instability supernovae, have been theorized, but so far there has been scant observational evidence to back them up.

The Gravitational-Wave Spectrum

Much like electromagnetic waves, gravitational waves are emitted by many different objects over a wide range of frequencies. Terrestrial interferometers such as the Laser Interferometer Gravitational-wave Observatory (LIGO) and Virgo are sensitive to only a subset of those frequencies, which limits their ability to "see" certain cosmic phenomena. They will not detect collisions of supermassive black holes found in the hearts of galaxies, for example. But space-based interferometers and other approaches for picking up gravitational waves could extend physicists' reach.



NIK SPENCER *Nature*



NIK SPENCER *Nature*

Eventually the black hole detections will delineate a map of the universe in the way galaxy surveys currently do, says Rainer Weiss, a physicist at the Massachusetts Institute of Technology, who was the principal designer of LIGO. Once the numbers pile up, "we can actually begin to see the whole universe in black holes," he says. "Every piece of astrophysics will get something out of that."

To ramp up these observations, LIGO and Virgo have plans to improve their sensitivity, which will reveal not only more events but also more details about each merger. Among other things, physicists are eager to see the detailed "ringdown" waves that a postmerger black hole emanates as it settles into a spherical shape—an observation that could potentially reveal cracks in the general theory of relativity.

"WE CAN ACTUALLY BEGIN TO SEE THE WHOLE UNIVERSE IN BLACK HOLES."
—Rainer Weiss

Having more observatories spread around the globe will also be crucial. KAGRA, a detector under construction deep underground in Japan, might start gathering data by late 2019. Its location—and in particular its orientation with respect to incoming waves—will complement LIGO's and Virgo's and enable researchers to nail down the polarization of the gravitational waves, which encodes information about the orientation of the orbital plane and the spin of the spiraling objects. And India is planning to build another observatory in the next decade, made in part with spare components from LIGO.

An even bigger trove of discoveries could come from observing neutron star mergers. So far researchers have announced only one such detection, called GW170817. That signal, seen in August 2017, was almost certainly the most intensely studied event in astronomy's history. And it solved a number of longstanding mysteries in one stroke, including the origin of gold and other heavy elements in the universe, as well as the cause of some gamma-ray bursts.

Further observations could allow scientists to explore the interiors of these objects. Neutron stars are thought to be as dense as matter can possibly be without collapsing into a black hole, but exactly how dense is anybody's guess. No laboratory experiment can study those conditions, and there are dozens of proposals for what happens there. Some theories predict that quarks—the subatomic components that make up protons and neutrons—should break free from one another and roam about, perhaps in superconducting, superfluid states. Others posit that heavier "strange" quarks form and become part of exotic cousins of the neutron.

Pinning down the radii of neutron stars might allow physicists to evaluate the theories because they predict different "equations of state"—formulas that link pressure, temperature and density of matter. Such equations determine to what extent matter can be compressed and thus how wide or narrow a neutron star will be for a given mass and how massive such stars can get.

The 100-second-long signal in August 2017 eventually became too high in pitch for LIGO and Virgo to detect, which prevented the observatories from seeing the two neutron stars' final moments, when they should have deformed each other in ways that would have revealed their size and hardness or resistance to compression. Still, says Bangalore S. Sathyaprakash, a LIGO theoretical physicist at the Pennsylvania State University, from that one event, "we can rule out equations of state that allow neutron star sizes larger than 15 kilometers in radius"—a figure that is consistent with other measurements and favors "softer" matter.

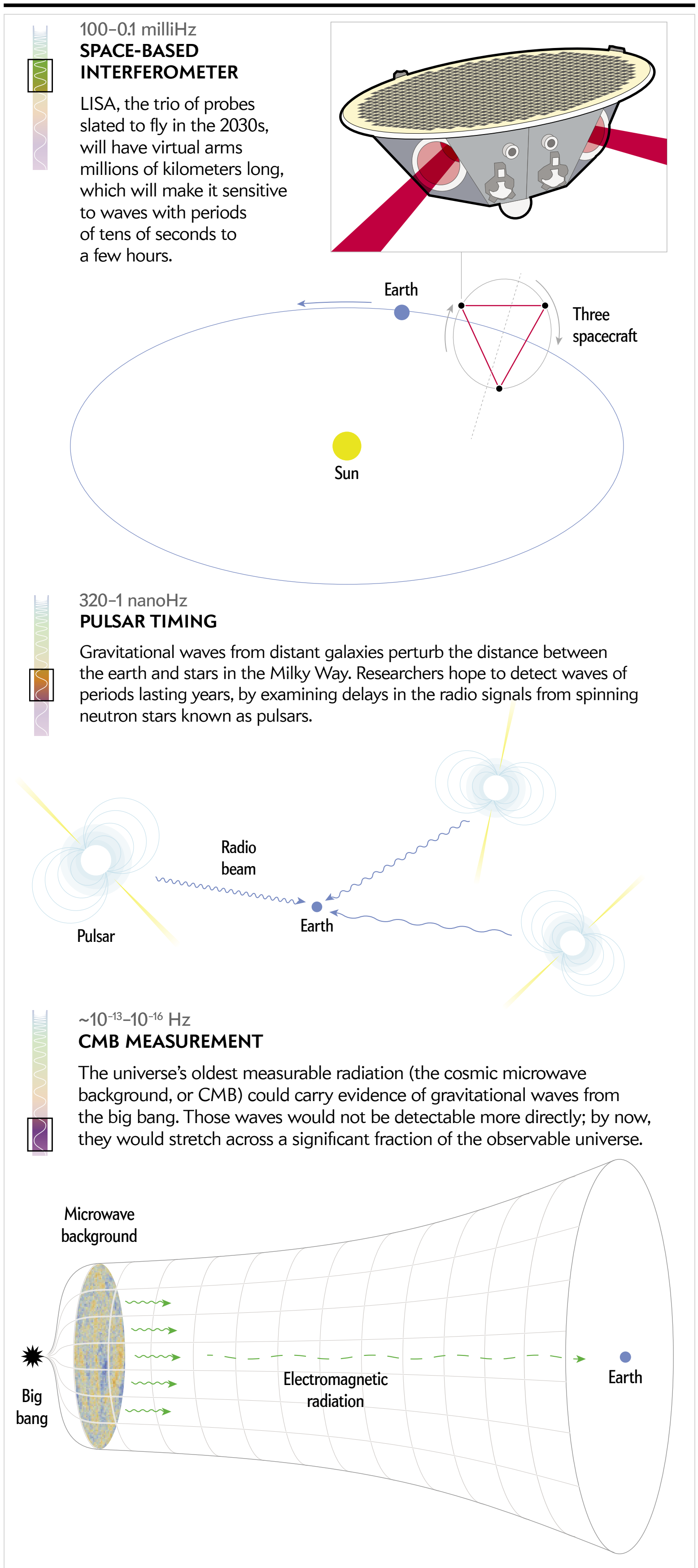
Future detections—and detectors—will give much more detail. Sathyaprakash says that the Einstein Telescope, a possible next-generation observatory dreamed up by a team in Europe, could take physicists far beyond an upper limit. "We want to be able to pin down the radius to the level of 100 meters," he says—a precision that would be astounding, given that these objects are millions of light-years away.

SIREN CALLS

SIGNALS SIMILAR to GW170817, which was observed through both gravitational waves and light, could have dramatic implications for cosmology. Schutz calculated in 1985 that the frequency, or pitch, of waves from spiraling objects, together with the rate at which that pitch increases, reveals information about the objects' collective mass. That determines how strong their waves should be at the source. By measuring the strength of the waves that reach the earth—the amplitude of the signal actually picked up by interferometers—one can then estimate the distance that the waves have traveled from the source. All other things being equal, a source that is twice as far, for example, will produce a signal half as strong. This type of signal has been dubbed a standard siren, in a nod to a common method of gauging distances in cosmology: stars called standard candles have a well-known brightness, which allows researchers to work out their distance from the earth.

By coupling the distance measurement of GW170817 with an estimate of how fast the galaxies in that region are receding from the earth, Schutz and his collaborators made a new and completely independent estimate of the Hubble constant—the universe's current rate of expansion. The result, part of a crop of papers released by LIGO, Virgo and some 70 other astronomy teams in October 2017, "ushers in a new era for both cosmology and astrophysics," says Wendy L. Freedman, an astronomer at the University of Chicago, who has made highly precise measurements of the Hubble constant using time honored but less direct techniques.

As a direct and independent measure of this constant, standard sirens could help to resolve a disagreement among cosmologists. State-of-the-art techniques, refined over nearly a century of work that started with Edwin Hubble himself, now give estimates that differ by a few percent. This first standard-siren measurement does not resolve the tension: the expansion rate it predicts falls somewhere in the middle of the range and, because it is based on just one merger event, has a large error bar. But in the future,



NIK SPENCER *Nature*

researchers expect standard sirens to nail down the Hubble constant with an error of less than 1 percent. So far standard candles have done it with precisions of 2 to 3 percent.

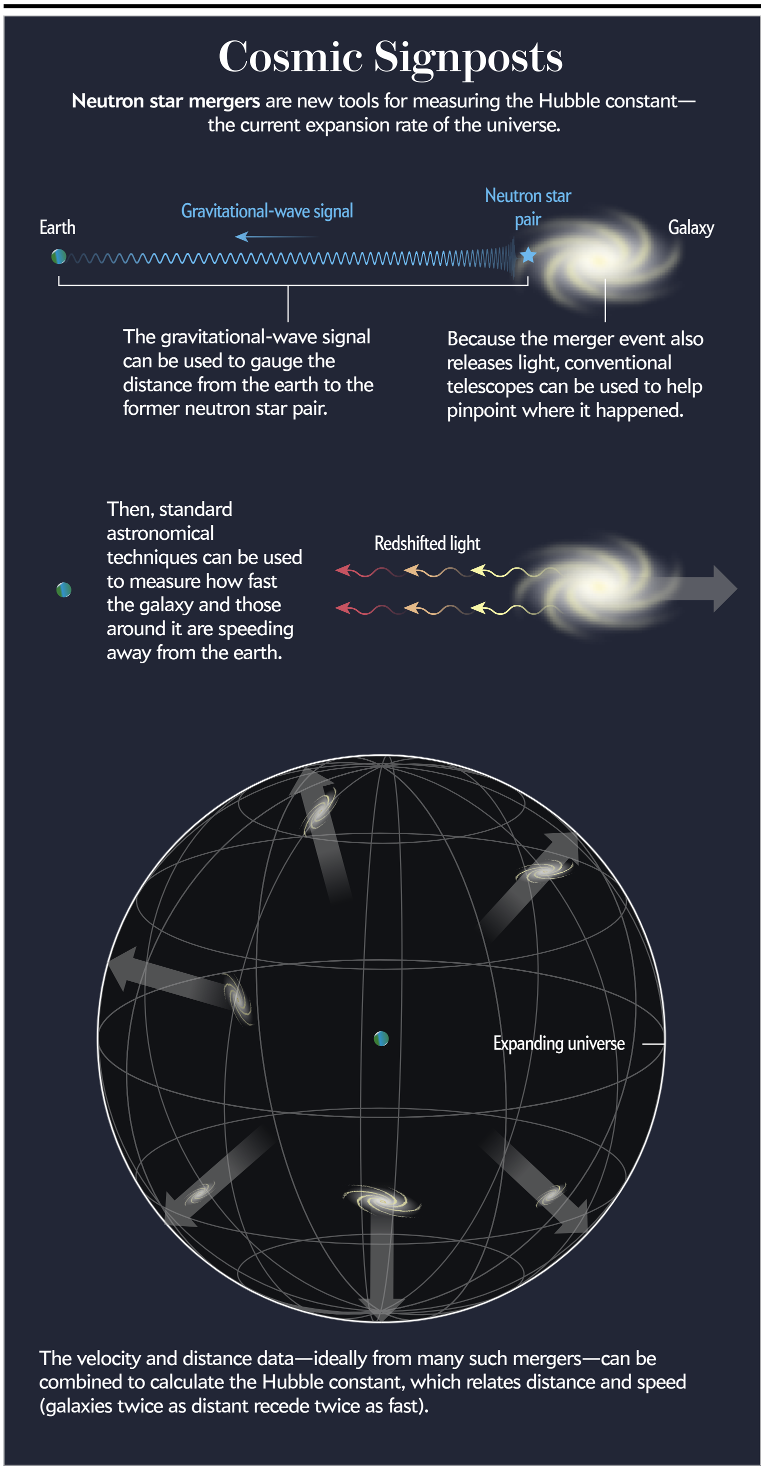
Standard sirens could become even more powerful tools with space-based interferometers such as the Laser Interferometer Space Antenna (LISA), a trio of probes that the European Space Agency, which is leading the mission, plans to launch in the 2030s. LISA is designed to be sensitive to low-frequency waves that ground-based observatories cannot detect. This would give it access to more massive systems, which radiate stronger gravitational waves. In principle, LISA could pick up sirens from across the universe and, with the help of conventional telescopes, measure not just the current rate of cosmic expansion but also how that rate has evolved through the eons. Thus, LISA could help address cosmology's biggest puzzle: the nature of dark energy, the as yet unidentified cosmic component that is driving the universe's expansion to accelerate.

Whereas ground-based interferometers detect events that are brief and far between, LISA is expected to hear a cacophony of signals as soon as it turns on, including a constant chorus of tight binary white dwarfs—the ubiquitous remnants of sun-sized stars—in our own galaxy. "It's as if we lived in a noisy forest and we had to single out the sounds of individual birds," says astrophysicist Monica Colpi of the University of Milan–Bicocca in Italy, who is part of a committee setting the mission's science goals.

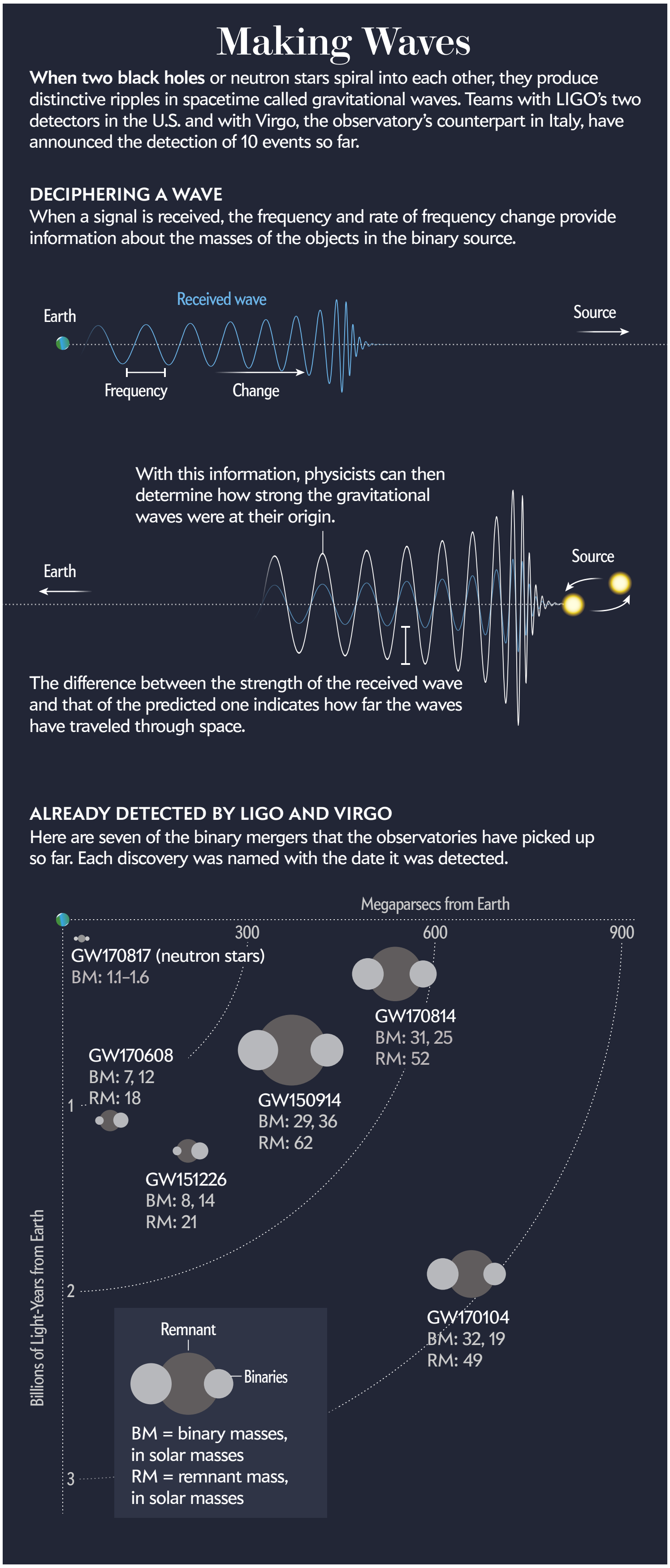
Occasionally, LISA should see black hole mergers such as the ones LIGO does, but on a much grander scale. Most galaxies are thought to harbor a central supermassive black hole that weighs millions or even billions of solar masses. Over a scale of billions of years, galaxies might merge several times; eventually their central black holes might merge, too. These events are not frequent for individual galaxies, but because there are trillions of galaxies in the observable universe, a detectable merger should occur somewhere at least a few times per year. Scientists are also pursuing a separate way of detecting gravitational waves from pairs of these behemoths at earlier stages of their orbits. Using radio telescopes, they monitor pulsars inside the Milky Way and look for small variations in their signals caused by the passage of gravitational waves through the galaxy. Today there are three "pulsar-timing arrays," in Australia, Europe and North America and a fourth forming in China.

Thanks to LISA's planned sensitivity and the strong signals produced by spiraling supermassive black holes, the observatory should be able to pick up gravitational waves from pairs of supermassive black holes months before they merge and see the merger in enough detail to test general relativity with high precision. After years of operation, LISA could accumulate enough distant events for researchers to reconstruct the hierarchical formation of galaxies—how small ones combined to form larger and larger ones—in the universe's history.

On the ground, too, physicists are beginning some "grand new ventures," Weiss says. A U.S. team envisions a Cosmic Explorer with 40-kilometer detecting arms—10 times as long as LIGO's—that would be sensitive to signals from



NIK SPENCER *Nature*



NIK SPENCER *Nature*

events much farther away, perhaps across the entire observable universe.

The concept for the Einstein Telescope in Europe calls for a detector with 10-kilometer arms arranged in an equilateral triangle and placed in tunnels 100 meters or so underground. The quiet conditions there could help to broaden the observatory's reach to frequencies one tenth those detectable by current machines. That might allow scientists to find black holes beyond the range thought to be prohibited by pair-instability supernovae; at high enough masses, stars should have a different collapse mechanism and be able to form black holes of 100 solar masses or more.

If scientists are lucky, gravitational waves might even let them access the physics of the big bang itself at epochs that are not observable by any other means. In the first instants of the universe, two fundamental forces—the electromagnetic force and the weak nuclear force—were indistinguishable. When these forces separated, they might have produced gravitational waves that, today, could show up as a "random hiss" detectable by LISA, Schutz says. This hypothetical signal is distinct from a much longer-wavelength one from even earlier on, which might appear in the universe's oldest visible radiation: the cosmic microwave background. In 2014 a team reported that it had observed this effect with the BICEP2 telescope at the South Pole, but the researchers later acknowledged problems with that interpretation.

With the reopening of both LIGO and Virgo this year, the next big discovery on Weiss's wish list is the signal from a collapsing star—something that astronomers might also observe as a type of supernova. But he has high hopes for what else might be on the horizon. "If we don't see something that we hadn't thought of," Weiss says. "I'd be disappointed." SA

Davide Castelvecchi is a senior reporter at *Nature* in London covering physics, astronomy, mathematics and computer science.

MORE TO EXPLORE

Observation of Gravitational Waves from a Binary Black Hole Merger. B. P. Abbott et al. (The LIGO Scientific Collaboration and Virgo Collaboration) in *Physical Review Letters,* Vol. 116, No. 16, Article 061102; February 12, 2016.
GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. B. P. Abbott et al. (The LIGO Scientific Collaboration and Virgo Collaboration) in *Physical Review Letters,* Vol. 119, No. 16, Article 161101; 20 October 2017.
Europe Shows First Cards in €1-billion Quantum Bet. Davide Castelvecchi on Nature.com; October 29, 2018.
Holy Cow! Astronomers Agog at Mysterious New Supernova. Davide Castelvecchi on ScientificAmerican.com; November 7, 2018.
Physicists Lay Out Plans for a New Supercollider. Davide Castelvecchi on ScientificAmerican.com; January 15, 2019.

scientificamerican.com/magazine/sa

IS DARK MATTER REAL?

Astrophysicists have piled up observations that are difficult to explain with dark matter. It is time to consider that there may be more to gravity than Einstein taught us

By Sabine Hossenfelder and Stacy S. McGaugh

Illustration by Kenn Brown and Chris Wren, Mondolithic Studios

THE STARS STILL HAVE SECRETS. WE KNOW WHY THEY SHINE, AND WE KNOW why they twinkle, but we still do not know why they move the way they move. The problem has been with us for the better part of a century. In the 1930s Swiss astronomer Fritz Zwicky observed that some galaxies in a cluster of about 1,000 fly surprisingly fast around their common center of mass. Even with generous estimates of the individual galaxies' masses, they did not add up enough to account for this motion. Zwicky fixed the mismatch by conjecturing the existence of a new kind of matter: "dark matter."

In the 1970s American astronomer Vera Rubin, who died in 2016, saw the same thing happening in single galaxies. The velocities of stars far out from the center of a galaxy remained roughly the same as those closer in, when astronomers would have expected them to slow down because of the dwindling gravity at the galaxy's far reaches. Again, the visible mass alone was not sufficient to explain the observations. Rubin concluded that in galaxies, too, dark matter must be present.

Since then, even more evidence has accumulated that we must be missing something. The tiny temperature fluctuations in the cosmic background radiation astronomers see pervading space, as well as the gravitational bending of light around galaxies and galaxy clusters and the formation of the cosmic web of large-scale structure throughout space, confirm that normal matter alone cannot explain what we see.

For many decades the most popular hypothesis has been that dark matter is composed of new, so far undetected particles that do not interact with light. The alternative explanation that we have the right particles but the wrong laws of gravity has received little attention.

Thirty years ago this stance was justified. The idea of particle dark matter gained traction because back then physicists had other reasons to believe in the existence of new particles. Around the 1950s and 1960s physicists realized that the protons, neutrons and electrons that make up atoms are not the only particles out there. Over the next decades particle accelerators started turning up new particles left and right; these came to make up the Standard Model of particle physics and opened theorists' minds to even more possibilities. For instance, efforts to unify the fundamental forces of nature into a single force required theorizing a set of new particles, and the concept of supersymmetry, developed in the 1970s, predicted a mirror particle for every known particle in the universe. Some of these theorized particles would make good dark matter candidates. Another suspect for the role was a particle called the axion, invented to explain the smallness of a parameter in the Standard Model. But after three decades of failed

IN BRIEF

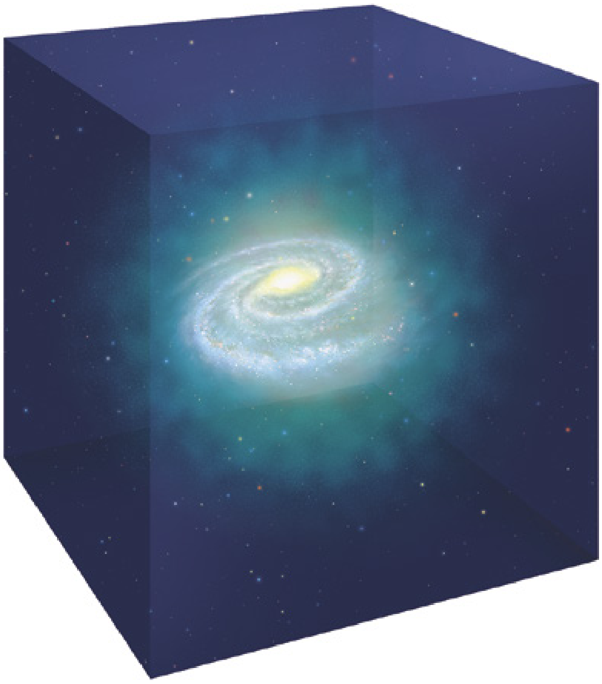
Scientists have long assumed that some invisible "dark matter" particles must accompany the normal matter in the universe to explain how stars orbit in galaxies and how galaxies orbit in clusters. An alternative idea that there is no extra matter and that our equations of gravity need updating has received much less attention.

But numerous experiments have failed to find evidence for dark matter particles, and the possibility remains that gravity must be modified.

Lately, in fact, some astrophysical evidence, such as recent observations of gravitation in galaxies, favors modified gravity theories over dark matter. It is time that physicists let go of their prejudices and reexamined this underdog idea.

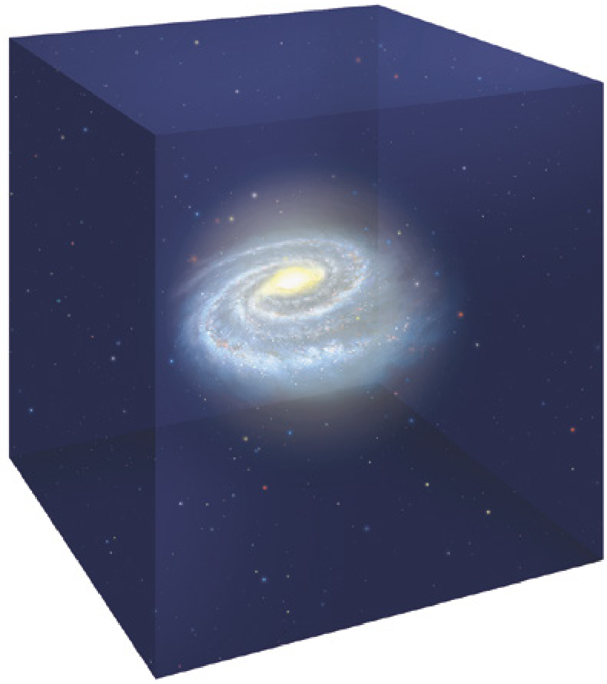
Dark Matter vs. Modified Gravity

Astronomers noticed long ago that the universe seems to be missing mass. Theorists suggested that some kind of hidden particles, dubbed "dark matter," must inhabit the universe to explain how stars move in galaxies and galaxies move in clusters. But experiments have failed to find dark matter. An alternative idea—that our equations of gravity need modifying—deserves a second look.



With Hidden Particles

The dark matter hypothesis suggests that invisible particles swarm around galaxies and clusters, far outweighing the visible matter. Around every galaxy, for instance, a spherical "halo" of dark particles would engulf the visible stars and gas, contributing a huge bulk of extra mass that would explain why stars at the edges of galaxies move nearly as fast as those toward the centers.



Without Hidden Particles

If dark matter does not exist, then scientists could tweak the laws of gravity to explain the speeds of stars at the edges of galaxies. Modified gravity theories revise Einstein's equations of general relativity to account for what we observe. Instead of an invisible dark matter halo surrounding galaxies, the visible objects are all there is.

attempts to detect any of these particles, ignoring alternative hypotheses is no longer reasonable.

Meanwhile the idea that dark matter is made of particles has come under pressure from an entirely different direction. New astrophysical data gathered and analyzed by one of us (McGaugh), as well as others, conflict with particle dark matter predictions. It is also becoming increasingly clear that some old problems with the dark matter paradigm persist even after many attempts to resolve them.

Updating the equations of gravity is still a viable way forward. Rather than adding particles to the universe to account for the extra gravity that seems to exist in galaxies and clusters, we can instead stick with the known particles but increase the force they exert on one another. Often dismissed and overlooked, modified gravity, as these theories are called, has never been ruled out. Now is a good opportunity to reconsider the option that we have been looking for the wrong thing in the wrong places. It is time to have a closer look at modified gravity.

TWEAKING GRAVITY

FIRST PUT FORWARD BY Israeli physicist Mordehai Milgrom in 1983, modified gravity changes the mathematical rules that govern how the force of gravity arises from mass. In most cases (that is, in nonextreme situations where Newtonian gravity is a good approximation), we describe this force by the inverse square law: the strength of gravity between two objects depends on their masses and decreases with the inverse square of the distance between them. This law is a classic and shows up all over physics, from equations describing how light intensity drops off with distance to rules describing sound pressure. But what if gravity does not always follow the inverse square law? What if the equations, in certain circumstances, should be tweaked?

Illustration by Lynette Cook

Milgrom's first proposal—modified Newtonian dynamics (MOND)—dealt only with the Newtonian laws of gravity. But Einstein's general theory of relativity taught us that gravity is not a force and is instead caused by the curvature of space and time. This limitation of the original MOND was likely a key reason many physicists did not take the idea seriously. But we now know several ways to make MOND compatible with general relativity, each using different types of fields that behave slightly differently to describe how gravitational attraction arises from mass. It is these 10 or so more complete theories that we collectively refer to as modified gravity. Dismissing them on purely theoretical grounds is no longer warranted. Another objection to modified gravity is that its mathematical expression appears inelegant from the perspective of particle physics. Not only does it look unfamiliar, it is also more difficult to deal with than particle dark matter, which employs techniques taught as part of the standard curriculum. Although these factors help to explain the idea's unpopularity, they are not scientific grounds for discounting it.

IT IS BECOMING INCREASINGLY CLEAR THAT SOME OLD PROBLEMS WITH THE DARK MATTER PARADIGM PERSIST.

Despite the potential of modified gravity, however, scientists have put almost all their energy on this front into searching for dark matter. Since the mid-1980s dozens of projects have sought the rare interactions predicted between dark matter particles and normal matter. Such experiments place large tanks of liquefied noble gases or carefully prepared solids, kept at extremely low temperatures, in well-shielded environments such as underground mines to avoid contamination from cosmic radiation. Sensitive detectors patiently wait for telltale signs of a dark matter particle bouncing off an atomic nucleus in the liquid or solid target.

The most recent round of dark matter searches just concluded. The very sensitive Large Underground Xenon (LUX) experiment in South Dakota and PandaX-II (for Particle and Astrophysical Xenon Detector) in Sichuan Province in China, like all other dark matter detection experiments before them, recently reported no evidence for particles that could make up dark matter. The first results from XENON1T at Gran Sasso National Laboratory in Italy (an upgrade of XENON100, which was itself an upgrade of XENON10) were also negative. Neither has Super-Kamiokande in Japan seen any signal of protons decaying, which would be evidence for a unification of the fundamental forces and give credibility to the idea that unseen particles must exist. At the same time, scientists at the Large Hadron Collider (LHC) at CERN near Geneva have been looking for novel particles with the right properties for dark matter and have seen no signs of them. Besides the expected Higgs boson, the LHC has seen no new particles at all.

Of course, these negative results do not rule out dark matter. Theories for particle dark matter have become increasingly sophisticated, not to say contrived. To evade conflict with experimental null results, theorists now assume the particles interact with normal matter even less than originally thought. Some researchers have begun to conjecture new forces and additional particle species to go with the original new particles. This proliferation of unseen particles has become so common in the literature that they have been given a collective name: the "hidden sector."

COMPARING THE THEORIES

IN THE ABSENCE OF any signs of new particles, we should ask how well the theories of dark matter and modified gravity, respectively, explain the evidence we do have from nature.

For the most part, the hypothesis that the universe contains about five times as much dark matter as normal matter works well to explain the cosmos around us. Although dark matter's microscopic properties can be complicated, it follows simple equations in bulk. We can describe dark matter as behaving like a fluid without internal pressure, its one variable being the average density of particles in space.

Treating dark matter as a pressureless fluid suffices to reproduce the patterns we observe in the cosmic microwave background. It also does a good job with the formation of large-scale cosmic structures. As the early universe expanded and matter cooled, particle dark matter, because it cannot build up internal pressure, would have begun to clump under the pull of gravity faster than normal matter. Only later would the normal matter collect in the clouds of dark matter to form galaxies. This scenario fits well with some aspects of our observations.

Particle dark matter explains the motions of stars within galaxies when we distribute suitable amounts where needed; clusters of galaxies work out in much the same way. Because theorists can sprinkle dark matter so flexibly, they can make all current observations fit with the predictions of general relativity.

But this flexibility of particle dark matter is also its greatest shortcoming. Galaxies are not particles, and no two are alike. Each galaxy has its own history; each came about in its own delicate dance of billions of stars following the pull of gravitational attraction. Some young galaxies collide and form larger galaxies. Some do not. Some galaxies end up as spinning disks, some as elliptic puff balls. Sometimes dark matter catches a lot of normal matter in its gravitational pull; sometimes it does not. Because of these many variations, you would expect a ratio of dark matter to normal matter that differs from one galaxy to the next. You would expect variety, not strict rules. But the data beg to differ.

In 2016 McGaugh and his colleagues made thousands of measurements in more than 150 galaxies and compared the gravitational pull expected from the normal matter in them with the observed gravitational pull that presumably resulted from the dark matter and normal matter combined. What they found was surprising: a strong correlation between the two. In fact, a simple equation relates the apparent amount of dark matter to the amount of normal matter in each galaxy; deviations from the curve are small and few [*see box on next page*].

This correlation is difficult to reproduce with computer simulations that treat the two types of matter as independent components. Scientists can make the simulations fit the data, but they must insert many parameters that have to be carefully chosen. Modified gravity, in stark contrast, simply predicts this correlation. Because this scenario involves only one type of matter—normal matter—of course the total gravity closely follows the gravity caused by the visible matter. Milgrom even predicted this observation in the early 1980s.

UNUSUAL GALAXIES

THERE ARE OTHER PROBLEMS with the dark matter hypothesis—for instance, "low-surface-brightness galaxies." In these dim galaxies, visible matter is spread more thinly than in galaxies similar to the Milky Way.

The dark matter hypothesis originally led us to expect that galaxies with low surface brightness—that is, low amounts of visible matter—should also generally have low amounts of dark matter. Scientists assumed stars orbiting at large distances from the galactic center would move slower in these galaxies than in normal galaxies of the same size because there was less total gravity pulling the stars along their orbits. But when the data came in, this expectation turned out to be wrong. The outer stars in these unusual galaxies were moving just as fast as they do in normal ones, suggesting that there was actually quite a lot of matter in low-surface-brightness galaxies, despite the sparseness of the stars. It turns out that in these objects, the ratio of dark matter to normal matter must be much higher than originally expected. But why should that be?

Originally, the dark matter hypothesis offered no explanation. But as we noted already, it is a very flexible hypothesis, so when theorists sought ways to explain this odd situation, they found them.

To square the findings with the theory, scientists had to fine-tune the amount of dark matter in each

NASA/CXC/CfA/M. MARKEVITCH ET AL. *(x-ray)*; NASA/STScI AND MAGELLAN/UNIVERSITY OF ARIZONA/D. CLOWE ET AL. *(optical)*; NASA/STScI, ESO WFI AND MAGELLAN/UNIVERSITY OF ARIZONA/D. CLOWE ET AL. *(lensing map)*

A Collision Offers Clues

The Bullet Cluster is a pair of galaxy clusters that crashed together long ago. It is a rare instance of a high-speed head-on collision. Images taken in visible and x-ray light (*red*), along with measurements of how gravity is bending light (gravitational lensing) (*blue*), reveal that in each cluster the center of the total mass and gravity is misplaced from the center of the visible mass.

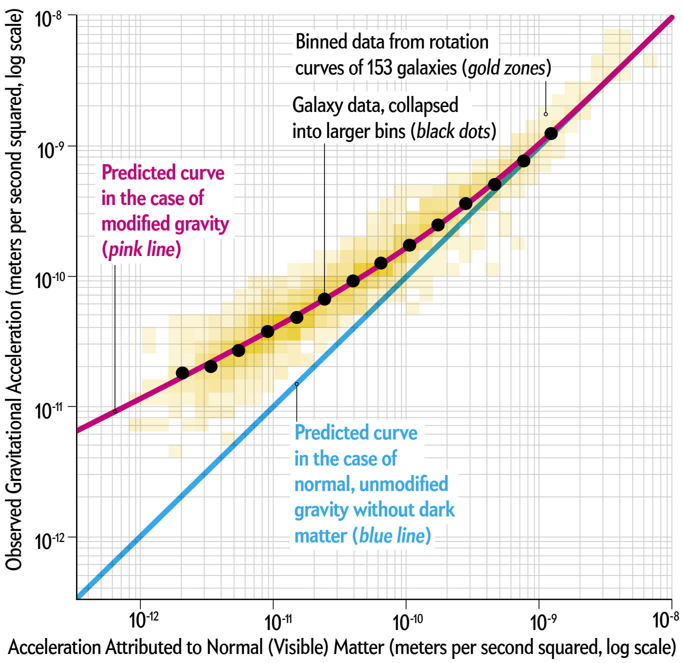
Scientists often claim the Bullet Cluster is evidence for particle dark matter. Because such particles would interact less than normal matter, the collision would have allowed the clusters' dark matter clouds to pass one another while the visible mass interacted with itself and lagged behind. This story matches what we observe, but it is crudely oversimplified.

In modified gravity, too, the point where gravitational attraction focuses can be displaced from the normal mass. This can occur because all forces, including gravity, are thought to be transmitted by a special type of particle. These particles have their own dynamical laws to fulfill. When modified gravity takes into account potential repercussions from these carrier particles, it can also predict what we see in the Bullet Cluster.

More important, this cluster is an extreme event and a statistical outlier. Its mere existence is difficult to explain both with particle dark matter and with modified gravity. Using it as evidence for or against either approach is an exercise in confirming our own biases.

A Problem for Dark Matter

A 2016 study examined stars' movements in galaxies and found that the total gravity present (*y-axis*) is directly proportional to the amount of gravity caused by the visible matter (*x-axis*). This extreme proportionality would be quite surprising if dark matter exists because the number of invisible particles should not depend solely on the amount of visible matter—the different shapes, sizes and gas content of galaxies should cause some variation. Modified gravity theories, however, predict just this relation.



galaxy to depend on the surface brightness of the stars: the dimmer the system, the more dark matter. Doing so required some mechanism to rid these galaxies of luminous matter while they formed, so that the matter ratio tilted in favor of dark matter. Currently the most popular method is to add "stellar feedback" to the computer simulations. Stellar feedback is caused by the pressure created when massive stars irradiate their surrounding gas with highly energetic photons, blow strong stellar winds and ultimately go supernova. These giant explosions can blow matter out of galaxies. And because dark matter interacts so weakly, this blowout would affect normal matter more than dark matter. Galaxies that happen to have many supernovae would thus end up with an increased ratio of dark matter.

But although we know stellar feedback plays an important role in the formation of stars and stellar clusters, its role during galaxy formation is less clear. To solve the problem with low-surface-brightness galaxies, supernovae's energy must go almost entirely into pushing matter out of galaxies. Such a high level of efficiency, however, is strikingly implausible for a naturally occurring process. Modified gravity, on the other hand, predicts the observed outcome without involving feedback, just as it predicted the observed rotation speeds of stars in normal galaxies.

MORE PROBLEMS

THE ISSUE WITH low-surface-brightness galaxies is far from the only shortcoming of particle dark matter. The theory predicts, for instance, a highly peaked density of matter in the cores of galaxies, in contrast with what we measure. It predicts many more small dwarf galaxies than we observe and fails to predict the way that galaxies and their satellite galaxies align along a single plane. These are just the most prominent disagreements. Modified gravity does better in all these areas.

The lack of density peaks in galactic cores, in particular, fits so badly with the dark matter story that when the data were new, many astrophysicists doubted they were correct. First, the theorists asserted that the resolution of the measurements was inadequate. When subsequent data settled the issue of resolution, they blamed other systematic errors. But after several more generations of observations obtained by multiple groups, the conclusion remains the same: dark matter does a bad job of explaining what we see at the centers of galaxies.

It is true that incorporating stellar feedback and other astrophysical effects into the computer simulations alleviates these issues. Because these extra processes add more parameters to the simulations, researchers can coax the software into producing galaxies that resemble what we observe reasonably well. These simulated galaxies can then also reproduce the observed correlation between the amount of particle dark matter and normal matter. What the computer simulations do not offer, however, is any explanation for the origin of this correlation.

And modified gravity has another advantage. In contrast to dark matter simulations, modified gravity can explain how small galaxies behave when trapped in the gravitational field of larger galaxies. For instance, its calculations have been enormously successful in predicting how a bunch of recently discovered dwarf galaxies swirl around our large neighbor galaxy, Andromeda. These tiny dwarfs are subject to a gravitational pull from their giant host that is stronger than their internal gravity. In such a situation, modified gravity offers a different prediction than it would if the dwarf galaxies were isolated, and it is this unique prediction that we find realized in the observations. Fitting this aspect of the data with particle dark matter, however, requires adding yet more assumptions to the computer simulations.

But let us be fair: Despite these many predictive successes, modified gravity has serious problems. Although it works across a huge range of different galaxy types, it cannot explain the motion of galaxy clusters very well. And on the behavior of the cos-

SOURCES: FEDERICO LELLI *European Southern Observatory*; "RADIAL ACCELERATION RELATION IN ROTATIONALLY SUPPORTED GALAXIES," BY STACY S. McGAUGH, FEDERICO LELLI AND JAMES M. SCHOMBERT, IN *PHYSICAL REVIEW LETTERS*, VOL. 117, ARTICLE NO. 201101; NOVEMBER 11, 2016

Graphic by Jen Christiansen

mos as a whole, modified gravity is mute. In these cases, particle dark matter works better. It accounts for the properties of the cosmic microwave background and the distribution of galaxies throughout the universe, where modified gravity has no answers. Yet discarding modified gravity because it does not address these situations misses the point. The theory has made successful predictions. Even if we do not understand why, ignoring it will not help.

MOVING FORWARD

AT THIS POINT, both particle dark matter and modified gravity have advantages and shortcomings. Some recent theoretical developments suggest that maybe the truth is in between: a type of particle dark matter that can masquerade as modified gravity.

In 2015 Justin Khoury of the University of Pennsylvania and his colleagues found that some types of particle dark matter can become superfluids—fluids that flow with no resistance, in which quantum effects are dominant. When the superfluid dark matter collects in galaxies, its quantum properties can generate a long-range force that resembles modified gravity. The superfluid itself has a gravitational pull, but according to Khoury's hypothesis, most of the effect we now assign to dark matter comes not from gravity but from the superfluid's direct interaction with normal matter. This phenomenon would explain why the force we witness acting on normal matter in galaxies is hard for gravity to account for: it is not caused by gravity.

The idea that dark matter is a type of superfluid that mimics modified gravity also clarifies why modified gravity does not work well for galaxy clusters. Throughout most clusters, gravity is not strong enough to make the particles superfluid. In these situations, they behave like a normal fluid—that is, they behave like particle dark matter.

And as one of us (Hossenfelder) noticed by accident, the superfluid concept matches another line of research, pioneered by Erik Verlinde of the University of Amsterdam. Verlinde uses ideas from string theory to argue that the impression that the universe contains more matter than we can see is an illusion caused by the reaction of space to the presence of mass. Although this notion sounds entirely different from Khoury's superfluid hypothesis, the key equation in both cases is almost the same.

This line of research is young and might turn out to be a dead end. But it exemplifies how having a closer look at modified gravity may help overcome the current phase of stagnation in the search for dark matter.

And new data should be available soon that will help determine the truth. Traditional particle dark matter, modified gravity and superfluid dark matter all make different predictions for low-surface-brightness galaxies that may become testable in the near future. The Dark Energy Survey currently identifies such galaxies, and the Large Synoptic Survey Telescope should find them by the hundreds when it comes online this year. The theories also differ when talking about the early universe, when the first galaxies were forming. These galaxies should be observable by the James Webb Space Telescope, which is set to launch in 2021, and future long-wavelength radio observations will probe the dark ages at still earlier epochs.

The advent of gravitational-wave astronomy is also giving us new clues. The Laser Interferometer Gravitational-wave Observatory (LIGO) recently detected gravitational waves caused by the collision of two neutron stars. At the same time, various telescopes observed light in different wavelengths emitted by the same event. These observations show, to excellent precision, that gravitational waves travel at the same speed as light. This finding has ruled out some, but certainly not all, variants of modified gravity.

MAYBE THE TRUTH IS IN BETWEEN: A TYPE OF DARK MATTER THAT CAN MASQUERADE AS MODIFIED GRAVITY.

Right now a few dozens of scientists are studying modified gravity, whereas several thousand are looking for particle dark matter. Perhaps modified gravity is wrong, but perhaps the scientific community is not putting in a good faith effort to know for sure. The universe has had a habit of surprising us; we should be prepared to greet what future data reveal with open minds. The stars may still have secrets, but they are under close surveillance. SA

Sabine Hossenfelder is a theoretical physicist at the Frankfurt Institute for Advanced Studies in Germany, who researches physics beyond the Standard Model. She is author of the physics blog Backreaction and the book *Lost in Math: How Beauty Leads Physics Astray* (Basic Books, 2018).

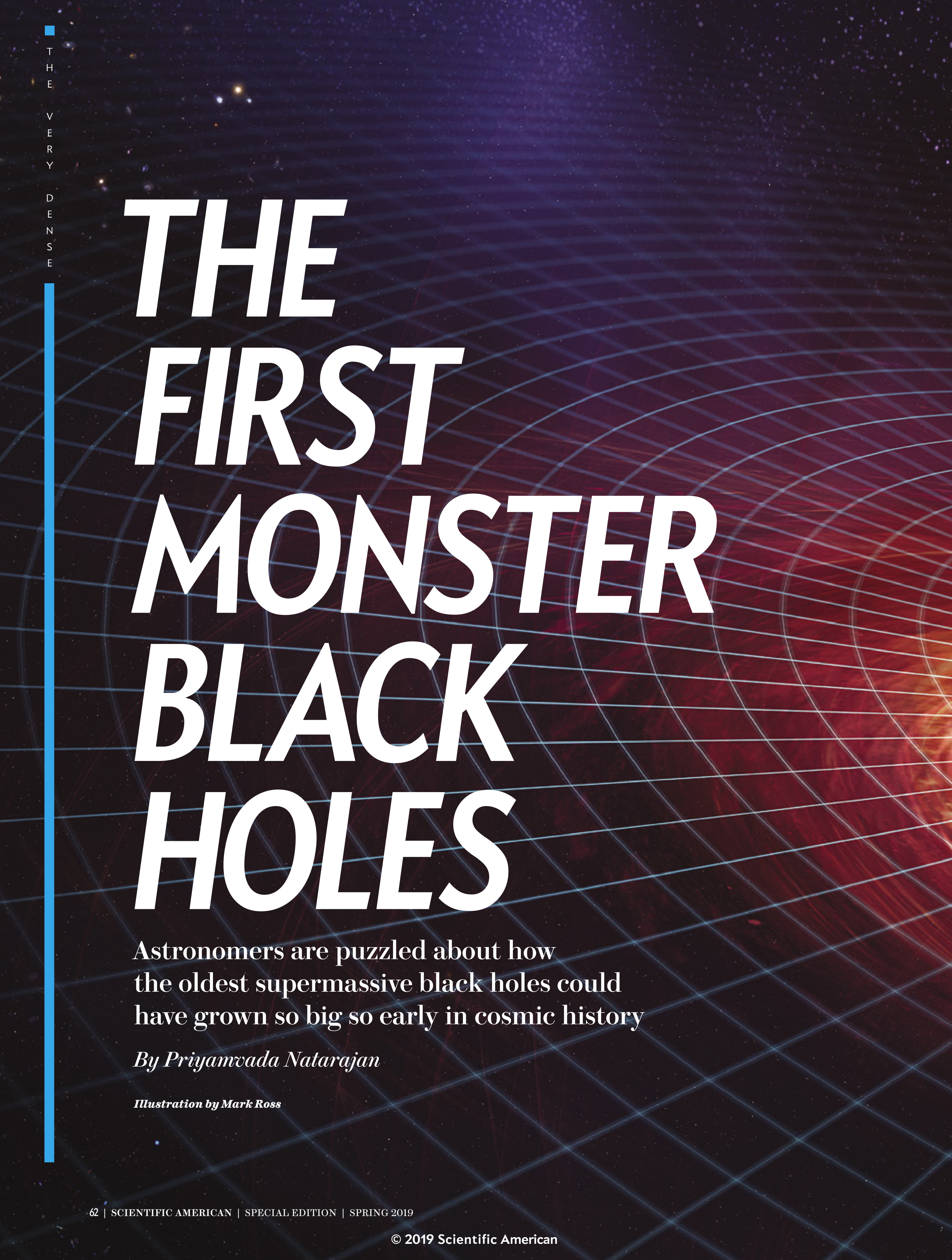
Stacy S. McGaugh is an astrophysicist at Case Western Reserve University. His research focuses on low-surface-brightness galaxies, which provide strong tests of modified gravity and dark matter.

MORE TO EXPLORE

Radial Acceleration Relation in Rotationally Supported Galaxies. Stacy S. McGaugh , Federico Lelli and James M. Schombert in *Physical Review Letters,* Vol. 117, No. 20, Article No. 201101; November 11, 2016. Preprint available at **https://arxiv.org/abs/1609.05917**

Covariant Version of Verlinde's Emergent Gravity. Sabine Hossenfelder in *Physical Review D,* Vol. 95, No. 12, Article No. 124018; June 15, 2017. Preprint available at **https://arxiv.org/abs/1703.01415**

scientificamerican.com/magazine/sa

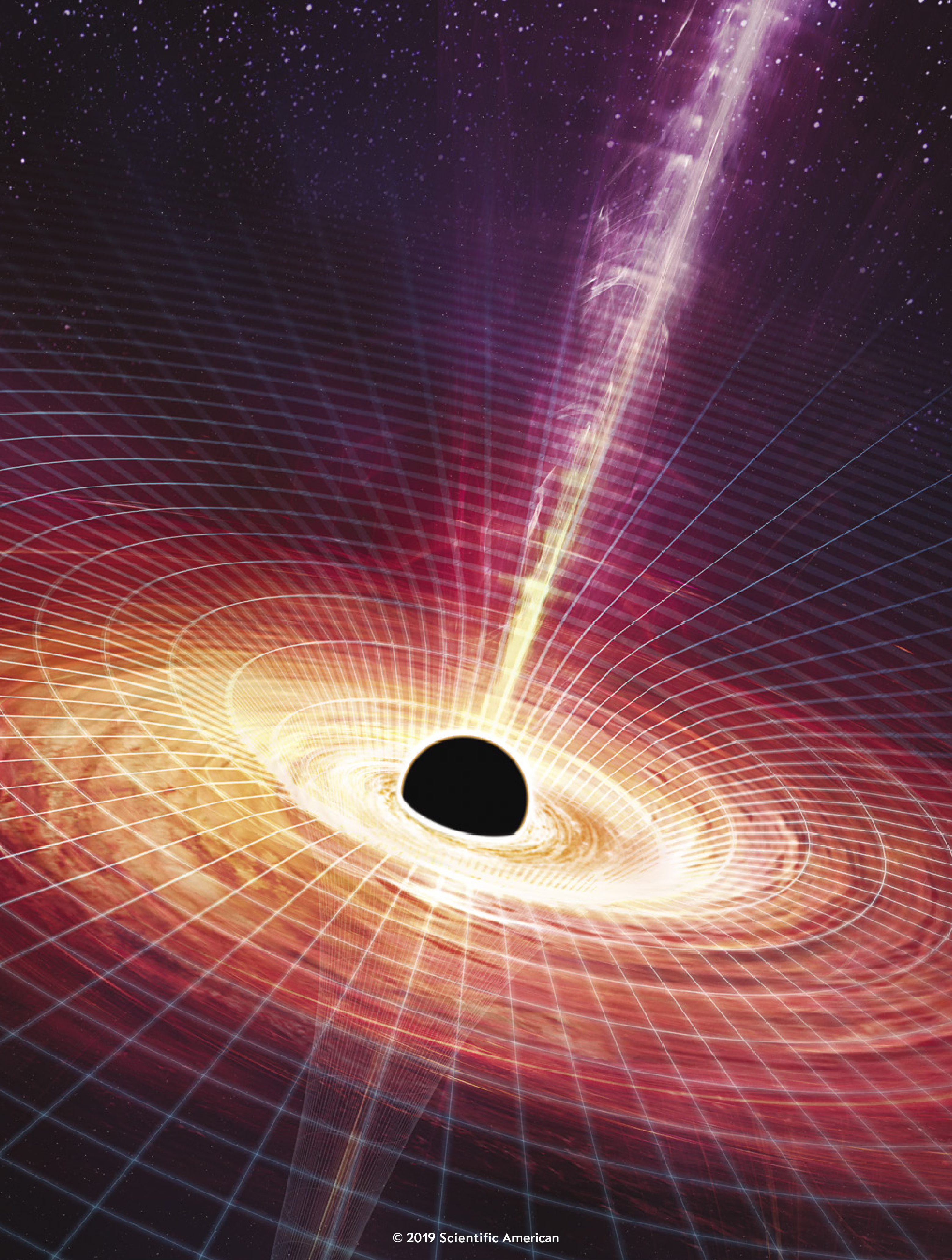


THE FIRST MONSTER BLACK HOLES

Astronomers are puzzled about how the oldest supermassive black holes could have grown so big so early in cosmic history

By Priyamvada Natarajan

Illustration by Mark Ross



IMAGINE THE UNIVERSE IN ITS INFANCY. MOST SCIENTISTS THINK space and time originated with the big bang. From that hot and dense start the cosmos expanded and cooled, but it took a while for stars and galaxies to start dotting the sky. It was not until about 380,000 years after the big bang that atoms could hold together and fill the universe with mostly hydrogen gas. When the cosmos was a few hundred million years old, this gas coalesced into the earliest stars, which formed in clusters that clumped together into galaxies, the oldest of which appears 400 million years after the universe was born. To their surprise, scientists have found that another class of astronomical objects begins to appear at this point, too: quasars.

IN BRIEF

In the very distant, ancient universe, astronomers can see quasars—extremely bright objects powered by enormous black holes. Yet it is unclear how black holes this large could have formed so quickly after the big bang.

To solve the mystery, scientists proposed a novel mechanism for black hole formation. Rather than being born in the deaths of massive stars, the seeds of the most ancient supermassive black holes might have collapsed directly from gas clouds.

Astronomers may be able to find evidence for direct-collapse black holes using the James Webb Space Telescope, due to launch in 2021, which should see farther back in space and time than any instrument before it.

Quasars are extremely bright objects powered by gas falling onto supermassive black holes. They are some of the most luminous things in the universe, visible out to the farthest reaches of space. The most distant quasars are also the most ancient, and the oldest among them pose a mystery.

To be visible at such incredible distances, these quasars must be fueled by black holes containing about a billion times the mass of the sun. Yet conventional theories of black hole formation and growth suggest that a black hole big enough to power these quasars could not have formed in less than a billion years. In 2001, however, with the Sloan Digital Sky Survey, astronomers began finding quasars that dated back earlier. The oldest and most distant quasar known, which was reported in December 2017, existed just 690 million years after the big bang. In other words, it does not seem that there had been enough time in the history of the universe for quasars like this one to form.

Many astronomers think that the first black holes—seed black holes—are the remnants of the first stars, corpses left behind after the stars exploded into supernovae. Yet these stellar remnants should contain no more than a few hundred solar masses. It is difficult to imagine a scenario in which the black holes powering the first quasars grew from seeds this small.

To solve this quandary, a decade ago some colleagues and I proposed a way that seed black holes massive enough to explain the first quasars could have formed without the birth and death of stars. Instead these black hole seeds would have formed directly from gas. We call them direct-collapse black holes (DCBHs). In the right environments, direct-collapse black holes could have been born at 10^4 or 10^5 solar masses within a few hundred million years after the big bang. With this head start, they could have easily grown to 10^9 or 10^{10} solar masses, thereby producing the ancient quasars that have puzzled astronomers for nearly two decades.

The question is whether this scenario actually happened. Luckily, the James Webb Space Telescope (JWST) is due to launch in 2021, and we should be able to find out.

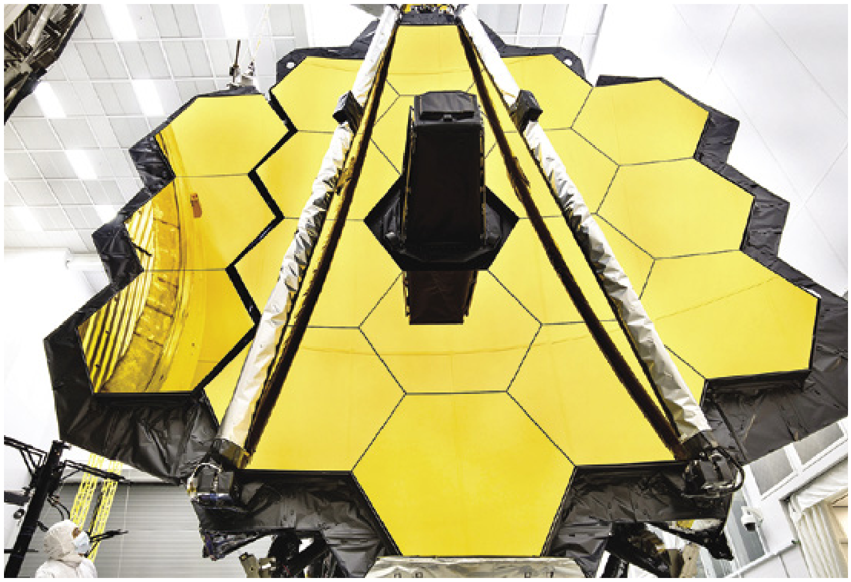
THE FIRST SEEDS

BLACK HOLES ARE ENIGMATIC astronomical objects, areas where the gravity is so immense that it has warped spacetime so that not even light can escape. It was not until the detection of quasars, which allow astronomers to see the light emitted by matter falling into black holes, that we had evidence that they were real objects and not just mathematical curiosities predicted by Einstein's general theory of relativity.

Most black holes are thought to form when very massive stars—those with more than about 10 times the mass of sun—exhaust their nuclear fuel and begin to cool and therefore contract. Eventually gravity wins, and the star collapses, igniting a cataclysmic supernova explosion and leaving behind a black hole. Astronomers have traditionally assumed that most of the black holes powering the first quasars formed this way, too. They could have been born from the demise of the universe's first stars (Population III stars), which we think formed when primordial gas cooled and fragmented about 200 million years after the big bang. Population III stars were probably more massive than stars born in the later universe, which means they could have left behind black holes as hefty as several hundred solar masses. These stars also probably formed in dense clusters, so it is likely that the black holes created on their deaths would have merged, giving rise to black holes of several thousand solar masses. Even black holes this large, however, are far smaller than the masses needed to power the ancient quasars.

Theories also suggest that so-called primordial black holes could have arisen even earlier in cosmic history, when spacetime may have been expanding exponentially in a process called inflation. Primordial black holes could have coalesced from tiny fluctuations in the density of the universe and then grown as the universe expanded. Yet these seeds would weigh only between 10 and 100 solar masses, presenting the same problem as Population III remnants.

As an explanation for the first quasars, each of these pathways for the formation of black hole seeds has the same problem: the seeds would have to grow extraordinarily quickly within the first billion years of cosmic history to create the earliest quasars. And what we know about the growth of black holes tells us that this scenario is highly unlikely.



DUE TO LAUNCH in 2021, the James Webb Space Telescope will be powerful enough to find evidence for direct-collapse black holes, if they exist.

CHRIS GUNN NASA

FEEDING A BLACK HOLE

OUR CURRENT UNDERSTANDING of physics suggests that there is an optimal feeding rate, known as the Eddington rate, at which black holes gain mass most efficiently. A black hole feeding at the Eddington rate would grow exponentially, doubling in mass every 10^7 years or so. To grow to 10^9 solar masses, a black hole seed of 10 solar masses would have to gobble stars and gas unimpeded at the Eddington rate for a billion years. It is hard to explain how an entire population of black holes could continuously feed so efficiently.

In effect, if the first quasars grew from Population III black hole seeds, they would have had to eat faster than the Eddington rate. Surpassing that rate is theoretically possible under special circumstances in dense, gas-rich environments, and these conditions may have been available in the early universe, but they would not have been common, and they would have been short-lived. Furthermore, exceptionally fast growth can actually cause "choking," where the radiation emitted during these super-Eddington episodes could disrupt and even stop the flow of mass onto the black hole, halting its growth. Given these restrictions, it seems that extreme feasting could account for a few freak quasars, but it cannot explain the existence of the entire detected population unless our current understanding of the Eddington rate and black hole feeding process is wrong.

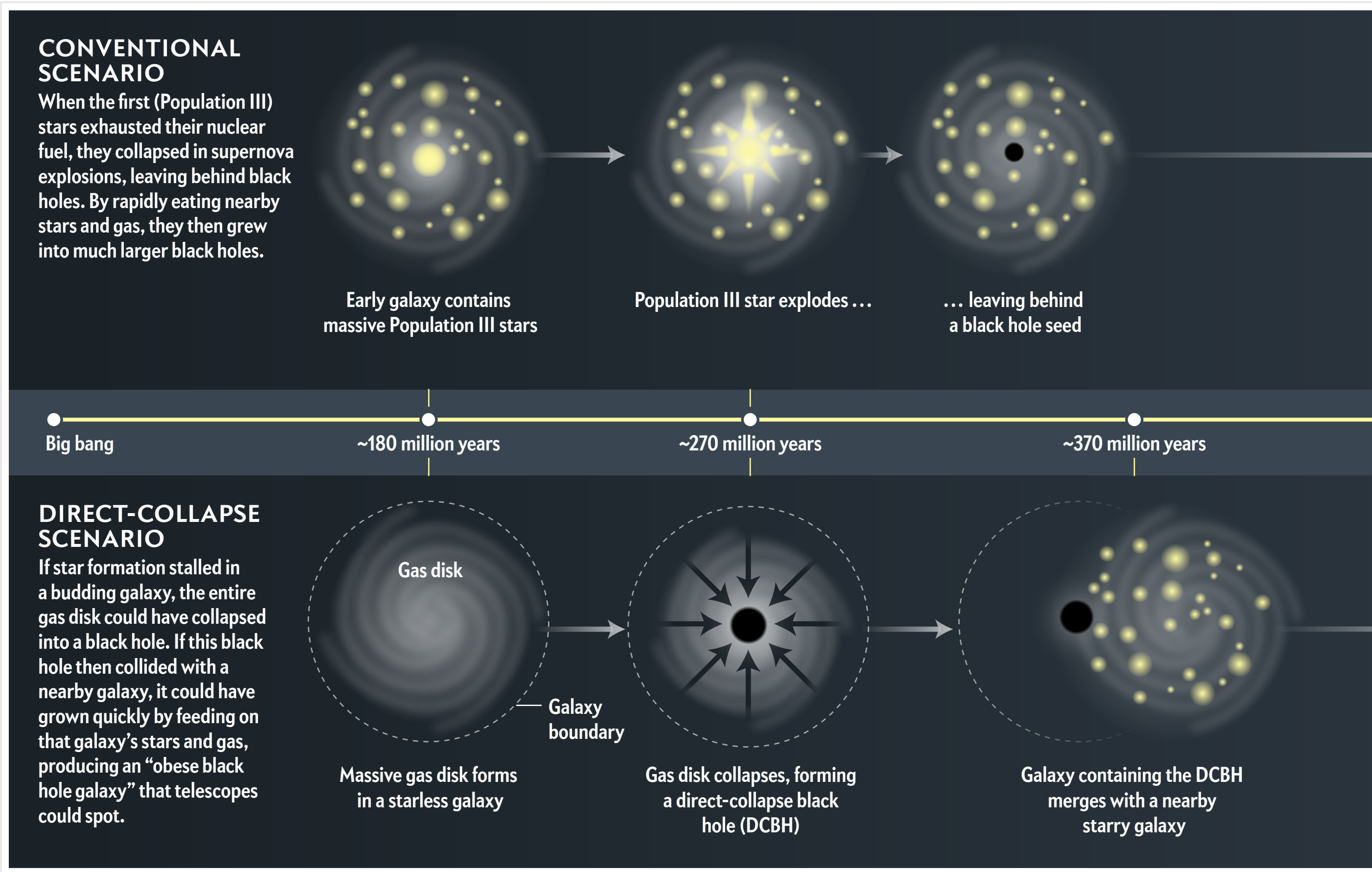
Thus, we must wonder whether the first black hole seeds could have formed through other channels. Building on the work of several other research groups, my collaborator Giuseppe Lodato and I published a set of papers in 2006 and 2007 in which we proposed a novel mechanism that could have produced more massive black hole seeds from the get-go. We started with large, pristine gas disks that might otherwise have cooled and fragmented to give rise to stars and become galaxies. We showed that it is possible for these disks to circumvent this conventional process and instead collapse into dense clumps that form seed black holes weighing 10^4 to 10^6 solar masses. This outcome can occur if something interferes with the normal cooling process that leads to star formation and instead drives the entire disk to become unstable, rapidly funneling matter to the center, much like water flowing down a bathtub drain when you pull the plug.

Disks cool down more efficiently if their gas includes some molecular hydrogen—two hydrogen atoms bonded together—rather than atomic hydrogen, which consists of only one atom. But if radiation from stars in a neighboring galaxy strikes the disk, it can destroy molecular hydrogen and turn it into atomic hydrogen, which suppresses cooling, keeping the gas too hot to form stars. Without stars, this massive irradiated disk could become dynamically unstable, and matter would quickly drain into its center, rapidly driving the production of a massive, direct-collapse black hole. Because this scenario depends on the presence of nearby stars, we expect DCBHs to typically form in satellite galaxies that orbit around larger parent galaxies where Population III stars have already formed.

Simulations of gas flows on large scales, as well as the physics of small-scale processes, support this model for DCBH formation. Thus, the idea of very large initial seeds appears feasible in the early universe. And starting with seeds in this range alleviates the timing problem for the production of the supermassive black holes that power the brightest, most distant quasars.

LOOKING FOR PROOF

BUT JUST BECAUSE DCBH SEEDS are feasible does not mean they actually exist. To find out, we must search for observational evidence. These objects would appear as bright, miniature quasars shining through the early universe. They should be detectable during a special



phase when the seed merges with the parent galaxy—and this process should be common, given that DCBHs probably form in satellites orbiting larger galaxies. A merger would give the black hole seed a copious new source of gas to eat, so the black hole should start growing rapidly. In fact, it would briefly turn into a special kind of quasar that outshines all the stars in the galaxy.

These black holes will not only be brighter than their surrounding stars, they will also be heavier—a reversal of the usual order of things. In general, the stars in a galaxy outweigh the central black holes by about a factor of 1,000. After the galaxy hosting the DCBH merges with its parent galaxy, however, the mass of the growing black hole will briefly exceed that of the stars. Such an object, called an obese black hole galaxy (OBG), should have a very special spectral signature, particularly in the infrared wavelengths between one and 30 microns where the JWST's Mid-Infrared Instrument (MIRI) and Near-Infrared Camera (NIRCam) will operate. This telescope will be the most powerful tool astronomers have ever had for peering into the earliest stages of cosmic history. If the telescope detects these obese black hole galaxies, it will provide strong evidence for our DCBH theory. Traditional black hole seeds, on the other hand, which derive from dead stars, are likely to be too faint for the JWST or other telescopes to see.

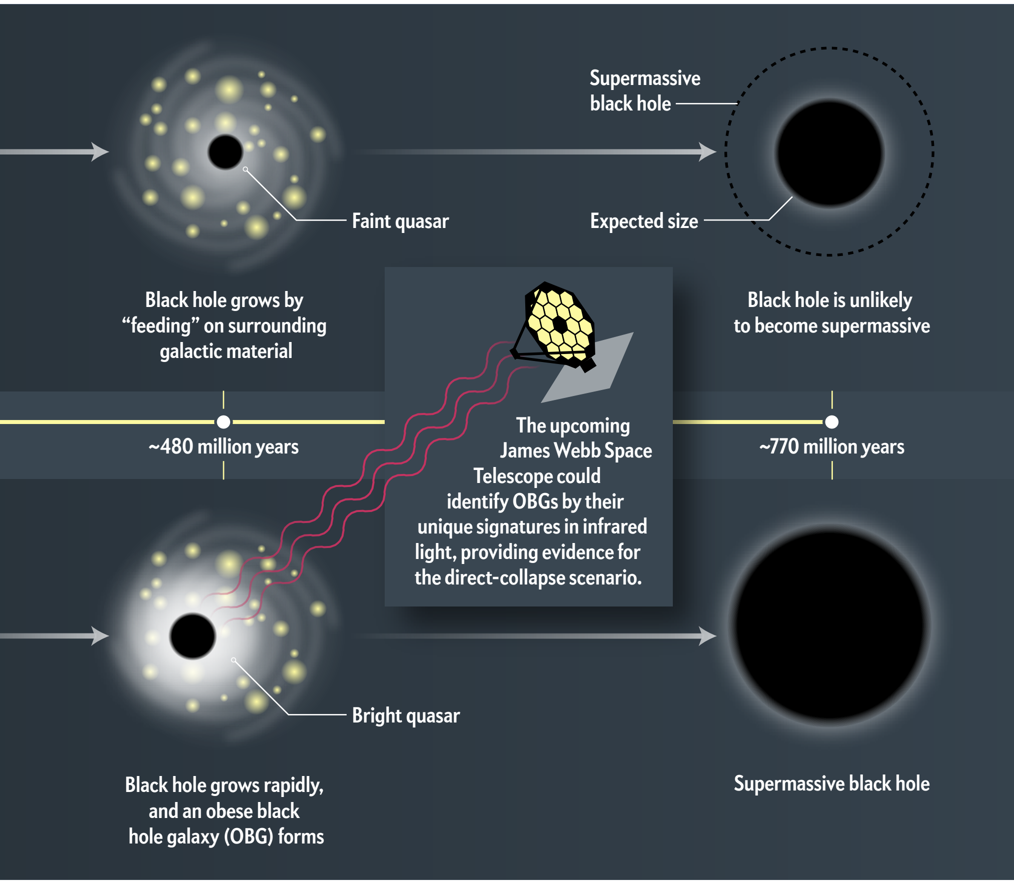
It is also possible that we might find other evidence for our theory. In the rare case that the parent galaxy that merges with the DCBH also hosts a central black hole, the two holes will collide and release powerful gravitational waves. These waves could be detectable by the Laser Interferometer Space Antenna (LISA), a European Space Agency/NASA mission expected to fly in the 2030s.

A FULLER PICTURE

IT IS ENTIRELY POSSIBLE that both the DCBH scenario and small seeds feeding at super-Eddington rates both occurred in the early universe. In fact, the initial black hole seeds probably formed via both these pathways. The question is, Which channel created the bulk of the bright ancient quasars that astronomers see? Solving this mystery could do more than just clear up the time line of the early cosmos. Astronomers also want to understand more broadly how supermassive black holes affect the larger galaxies around them.

Data suggest that central black holes might play an important role in adjusting how many stars form in the galaxies they inhabit. For one thing, the energy produced when matter falls into the black hole may heat up the surrounding gas at the center of the galaxy, thus preventing cooling and halting star formation. This energy may even have far-reaching effects outside the galactic center by driving energetic outflows. I predicted the existence of ghostly shadows that these quasar outflows

Illustration by Amanda Montañez



Two Ways to Make a Black Hole

The conventional picture of the process that formed the first supermassive black holes (*top*), which begins with the deaths of the first stars, seems incapable of explaining the enormous size of the black holes that must power the most ancient quasars. A newer idea (*bottom*) suggests that some gas disks collapsed directly into black holes rather than evolving into stars and galaxies.

would imprint on the relic radiation from the big bang in 1999; radio astronomers reported the first detection in December 2018. These winds likely heat up gas in outer regions and shut down star formation there. These effects are complex, however, and they stand to corroborate our current picture of black hole formation and growth. Finding the first seed black holes could help reveal how the relation between black holes and their host galaxies evolved over time.

These insights fit into a larger revolution in our ability to study and understand all masses of black holes. When the Laser Interferometer Gravitational-wave Observatory (LIGO) made the first detection of gravitational waves in 2015, for instance, scientists were able to trace them back to two colliding black holes weighing 36 and 29 solar masses, the lightweight cousins of the supermassive black holes that power quasars. The project continues to detect waves from similar events, offering new and incredible details about what happens when these black holes crash and warp the spacetime around them. Meanwhile a project called the Event Horizon Telescope, which uses radio observatories scattered around Earth to image the supermassive black hole at the center of the Milky Way, is due to report its first findings very soon. Scientists hope to spot a ringlike shadow around the black hole's boundary that general relativity predicts will occur as the hole's strong gravity deflects light. Any deviations the Event Horizon Telescope measures from the predictions of general relativity have the potential to challenge our understanding of black hole physics. In addition, experiments looking at pulsing stars called pulsar timing arrays could also detect tremors in spacetime caused by an accumulated signal of many collisions of black holes. And eventually the JWST will open up an entirely new window on the very first black holes to light up the universe.

Many revelations are in store in the very near future, and our understanding of black holes stands to be transformed. SA

Priyamvada Natarajan is a theoretical astrophysicist at Yale University whose research focuses on cosmology, gravitational lensing and black hole physics.

MORE TO EXPLORE

New Observational Constraints on the Growth of the First Supermassive Black Holes. E. Treister, K. Schawinski, M. Volonteri and P. Natarajan in *Astrophysical Journal,* Vol. 778, No. 2, Article No. 130; December 1, 2013.
Seeds to Monsters: Tracing the Growth of Black Holes in the Universe. Priyamvada Natarajan in *General Relativity and Gravitation,* Vol. 46, No. 5, Article No. 1702; May 2014.
Mapping the Heavens: The Radical Scientific Ideas That Reveal the Cosmos. Priyamvada Natarajan. Yale University Press, 2016.
Unveiling the First Black Holes with *JWST:* Multi-wavelength Spectral Predictions. Priyamvada Natarajan et al. in *Astrophysical Journal,* Vol. 838, No. 2, Article No. 117; April 1, 2017.

scientificamerican.com/magazine/sa

E X T R E M E   P H Y S I C S

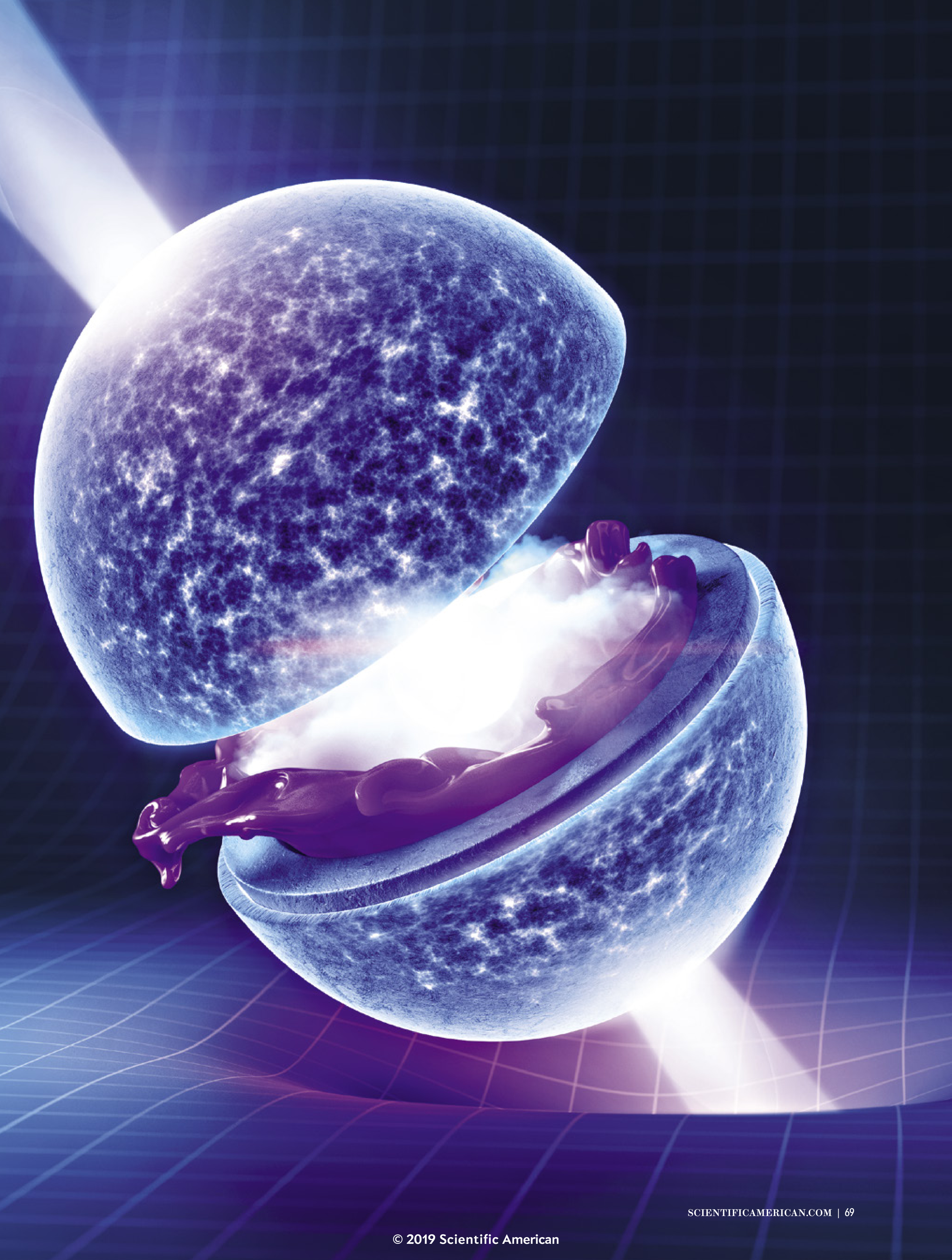
THE INNER LIVES OF NEUTRON STARS

The insides of neutron stars—the densest form of matter in the universe—have long been a mystery, but it is one that scientists are starting to crack

By Clara Moskowitz

Illustration by FOREAL

When a star the size of 20 suns dies, it becomes, in the words of astrophysicist Zaven Arzoumanian, "the most outrageous object that most people have never heard of"—a city-size body of improbable density known as a neutron star. A chunk of neutron star the size of a Ping-Pong ball would weigh more than a billion metric tons. Below the star's surface, under the crush of gravity, protons and electrons melt into one another to form a bulk of mostly neutrons—hence the name. At least, that is what we think. The issue is far from settled. Astronomers have never seen a neutron star up close, and no laboratory on Earth can create anything even approaching the same density, so the inner structure of these objects is one of the greatest mysteries in space. "They are matter at the highest stable density that nature allows, in a configuration that we don't understand," says Arzoumanian, who works at NASA's Goddard Space Flight Center. They are also the most strongly gravitating form of matter known—add just a bit more mass, and they would be black holes, which are not matter at all but rather purely curved space. "What goes



IN BRIEF

Neutron stars are born when stars within a certain mass range run out of fuel and collapse, leaving extremely compact remnants behind. They are the densest form of matter in the universe.

Scientists know that inside a neutron star, gravity crushes protons and electrons together to form neutrons, but they do not know what forms these neutrons take. Do they link up to create a viscosity-free "superfluid" or break down further into the quarks and gluons that constitute them?

Detectors capable of measuring gravitational waves from neutron star collisions and other new experiments promise to provide insight into these enigmatic objects.

on at that threshold," Arzoumanian says, "is what we're trying to explore."

There are several competing theories about what goes on at that threshold. Some ideas suggest that neutron stars really are just full of regular neutrons and maybe a few protons here and there. Others propose much stranger possibilities. Perhaps the neutrons inside neutron stars dissolve further into their constituent particles, called quarks and gluons, which swim untethered in a free-flowing sea. And it is possible that the interiors of these stars are made of even more exotic stuff, such as hyperons—weird particles composed not of regular "up" and "down" quarks (the kind found in atoms) but their heavier "strange quark" cousins.

Short of cutting open a neutron star and looking inside, there is no easy way to know which of these theories is right. But scientists are making progress. A big break came in August 2017, when terrestrial experiments detected gravitational waves—undulations in spacetime produced by the acceleration of massive objects—from what looked like a head-on collision of two neutron stars. These waves carried information about the masses and sizes of the stars right before the crash, which scientists have used to place new limits on the properties and possible compositions of all neutron stars.

Clues are also coming from the Neutron Star Interior Composition Explorer (NICER), an experiment that started at the International Space Station in June 2017. NICER watches pulsars, which are highly magnetic, furiously rotating neutron stars that emit sweeping beams of light. As these beams pass over Earth, we see pulsars blink on and off at more than 700 times a second. Through these experiments and others, the prospect of understanding what is inside a neutron star finally looks possible. If scientists can do that, they will have a handle not just on one class of cosmic oddity but on the fundamental limits of matter and gravity as well.

SUPERFLUID SEAS

NEUTRON STARS ARE FORGED in the cataclysms known as supernovae, which occur when stars run out of fuel and cease generating energy in their cores. Suddenly gravity has no opposition, and it slams down on the star like a piston, blowing the outer layers away and smashing the core, which at this point in a star's life is mostly iron. The gravity is so strong it quite literally crushes the atoms, pushing the electrons inside the nucleus until they fuse with protons to create neutrons. "The iron is compressed by a factor of 100,000 in each direction," says Mark Alford, a physicist at Washington University in St. Louis. "The atom goes from being a tenth of a nanometer across to just a blob of neutrons a few femtometers wide." That is like shrinking Earth down to the size of a single city block. (A femtometer is a millionth of a nanometer, which is itself a billionth of a meter.) When the star has finished collapsing, it contains about 20 neutrons for every proton. It is much like a single giant atomic nucleus, says James Lattimer, an astronomer at Stony Brook University—with an important difference. "A nucleus is held together by nuclear interactions," Lattimer says. "A neutron star is held together by gravity."

Astronomers Walter Baade and Fritz Zwicky proposed neutron stars in 1934 as an answer to the question of what might be left over after a supernova—a term they coined at the same time for the extra-bright explosions being spotted across the sky. It had only been two years since British physicist James Chadwick discovered the neutron. Initially some scientists were skeptical that such extreme objects could exist, and it was not until Jocelyn Bell Burnell and her colleagues observed pulsars in 1967—and researchers over the next year determined they must be spinning neutron stars—that the idea was widely accepted.

Physicists think that neutron stars can range from roughly one to two and a half times the mass of the sun and that they probably consist of at least three layers. The outer layer is a gaseous "atmosphere" of hydrogen and helium a few centimeters to meters thick. It floats atop a kilometer-deep outer "crust" made of atomic nuclei arranged in a crystal structure, with electrons and neutrons between them. The third, interior layer, which makes up the bulk of the star, is a bit of a mystery. Here nuclei are crammed in as tight as the laws of nuclear physics will allow, with no separation between them. As you move inward toward the core, each nucleus holds ever larger numbers of neutrons. At some point, the nuclei cannot contain any more neutrons, so they spill over: now there are no nuclei anymore, just nucleons (that is, neutrons or protons). Eventually in the innermost core, these may break down as well. "We are in the hypothetical regime where we do not know what happens at these insane pressures and densities," Alford says. "What we think might happen is that the neutrons actually get crushed together, and they overlap so much you can't really talk about it as being a fluid of neutrons anymore but a fluid of quarks."

What form that fluid takes is an open question. One possibility is that the quarks form a "superfluid," which has no viscosity and, once set in motion, will theoretically never stop moving. This bizarre state of matter is possible because quarks feel an affinity for other quarks, and if they are pushed close enough together, they can form bound "Cooper pairs." By itself, a quark is a fermion—a particle whose spin has the quantum-mechanical value of half an integer. When two quarks pair up, together they act as a single boson—a particle with spin equal to zero or one or another integer. After this change, the particle follows new rules. Fermions are bound by the Pauli exclusion principle, which says that no two identical fermions can occupy the same state—but bosons have no such restrictions. When they were fermions, the quarks were forced to take on higher energies to stack on top of one another in the crowded neutron stars. As bosons, however, they can stay in the lowest-possible energy state—any particle's preferred position—and still cram in together. When they do this, the quark pairs form a superfluid.

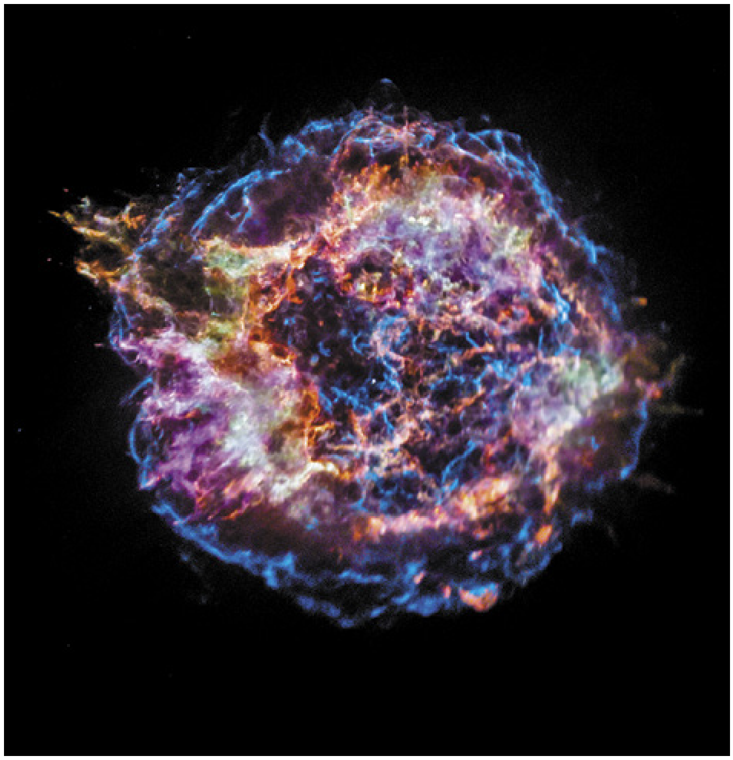
Outside the densest part of the core, where neutrons are likely to be intact, neutrons can also pair up to make a superfluid. In fact, scientists are fairly sure neutrons in the crust of the star do this. The evidence comes from observations of pulsar "glitches," episodes in which a spinning neutron star rapidly speeds up. Theorists think that these glitches occur when the rotation speed of the star as a whole grows out of sync with the rotation of the superfluid inside its crust. Overall, the star's rotation naturally slows with time; the superfluid, flowing without friction, does not. When the difference between these rates gets too great, the superfluid transfers angular momentum to the crust. "It's like an earthquake," Lattimer says. "You get a hiccup and a burst of energy, and the spin frequency increases for a brief time and then settles back down again."

In 2011 Lattimer and his colleagues suggested they had also found evidence of a superfluid in a neutron star's core, but he admits that this is still open to debate. To find that evidence, Lattimer's team, led by Dany Page of the National Autonomous University of Mexico, studied 15 years of x-ray observations of Cassiopeia A, the remnant of a supernova that first became visible on Earth in the 17th century. The scientists found that the pulsar at the center of the nebula is cooling faster than traditional theory suggests it should. One explanation is that many of the neutrons inside the star are pairing up to become a superfluid. The pairs break and re-form, emitting neutrinos, which causes the neutron star to lose energy and cool off. "This is something we never thought we would see," Lattimer says. "But lo and behold, there is this one star with the right age for us to see this. The proof in the pudding is going to come in another 50 or so years, when it should start to cool more slowly because once the superfluid is made, there is no more extra energy to be lost."

WEIRD QUARKS

SUPERFLUIDS ARE ONLY ONE of the exotic possibilities waiting behind the mystery doors of neutron stars. It is also possible that they are home to rare "strange quarks."

Quarks come in six kinds, or flavors—up, down, charm, strange, top and bottom. Only the lightest two, up and down, are found in atoms. The rest of the flavors are so massive and unstable that they usually appear only as short-lived detritus from high-energy particle collisions inside atom smashers such as the Large Hadron Collider at CERN near Geneva. But in the extremely dense interior of neutron stars, the up and down quarks inside neutrons might sometimes transform into strange quarks. (The other unusual flavors—charm, top and bottom—are so massive that



NASA, CXC AND SAO

CASSIOPEIA A is the remnant of an ancient supernova. At its center is a neutron star whose core may contain "superfluid."

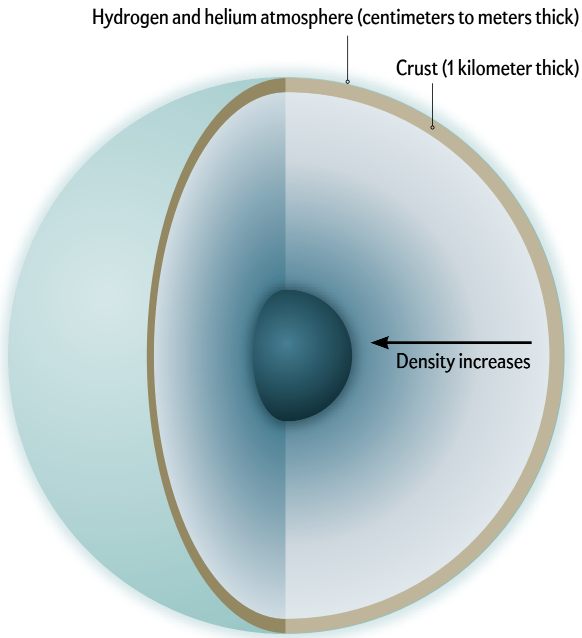
they likely would not form even there.) If strange quarks appear and remain bound to other quarks, they would make the mutant neutrons called hyperons. It is also possible that these quarks are not contained in particles at all—they might roam freely in a kind of quark soup.

Each of these possibilities should change the size of neutron stars in a measurable way. Intact neutrons inside the core would, in Arzoumanian's words, act "like marbles and make a hard, solid core." The solid core would tend to push on the outer layers and increase the size of the entire star. On the other hand, if the neutrons dissolved into a stew of quarks and gluons, they would make a "softer, squishier" and smaller star, he says. Arzoumanian is a co-principal investigator and science lead for the NICER experiment, which aims to determine which of these alternatives is true: "One of NICER's key objectives is to make a measurement of [neutron stars'] mass and radius that will help us pick out or exclude certain theories of dense matter."

NICER is a washing-machine-size box mounted to the exterior of the International Space Station. It steadily monitors several dozen pulsars spread across the sky, detecting x-ray photons from them. By measuring the photons' timing and energy, as well as how the stars' gravitational fields bend their light, NICER allows scientists to calculate the masses and radii of a collection of pulsars and compare them. "If NICER

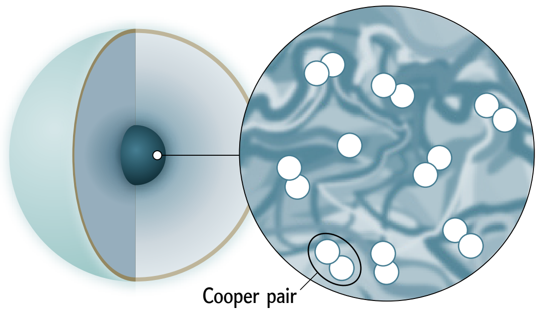
Inside a Neutron Star

Neutron stars are a puzzle. Scientists know they have a slight gaseous atmosphere above a thin crust layer made of heavy atomic nuclei and some floating electrons. But inside these outer layers lies the core—an unknown substance that is likely mostly neutrons. But what form these neutrons take and whether they break down into their ingredients, quarks and gluons, inside the densest inner core is an open question.



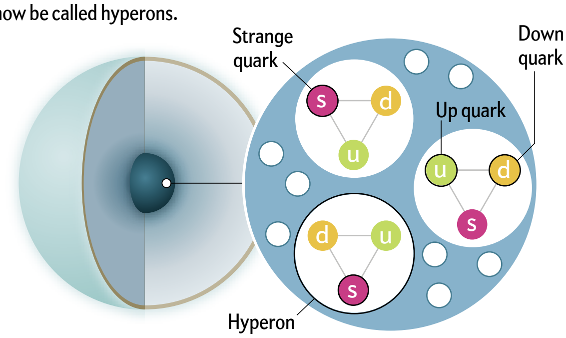
Core Hypothesis 1: Superfluid Seas

One possibility is that particles in the inner core are squeezed in so tight that some of them join to form new particles, called Cooper pairs. This can happen with protons, neutrons or, if these particles have dissolved, quarks. The new particles create a "superfluid" that flows without resistance.



Core Hypothesis 2: Weird Quarks

The incredible density could also prompt quarks in the inner core to transform from their usual type, "up" or "down," into exotic "strange quarks." If the quarks are still inside neutrons, these neutrons would now be called hyperons.



finds stars with roughly the same mass but very different radii, that would mean there's something funny going on," Alford says, "some new form of matter that, when it appears, makes the stars shrink down." Such a transition could occur, for instance, when neutrons break apart into quarks and gluons.

Measuring the sizes of neutron stars is a useful way to narrow the range of possible forms that matter inside neutron stars can take. Scientists once thought half the neutrons in any given neutron star would turn into hyperons that contained strange quarks; theoretical calculations suggested that such a hyperon-rich star could not exceed 1.5 times the mass of the sun. In 2010, however, astronomers led by Paul Demorest of the National Radio Astronomy Observatory measured the mass of one neutron star at 1.97 solar masses, eliminating a number of theories about the interior of a neutron star. Now physicists estimate that hyperons cannot make up more than 10 percent of a neutron star.

CRASH SITE DETECTIVES

STUDYING INDIVIDUAL neutron stars can tell us a lot, but we can learn much more when two of them slam together. For years telescopes have detected blasts of light, called gamma-ray bursts, that researchers suspected came from a crash of two neutron stars. In the August 2017 detection of gravitational waves, astronomers saw the first confirmed neutron star merger.

Specifically, on August 17, 2017, two experiments—the Laser Interferometer Gravitational-wave Observatory, or LIGO (based in Washington State and Louisiana), and Virgo (a European project based near Pisa, Italy)—simultaneously detected gravitational ripples produced as two neutron stars spiraled toward each other and merged to form either a single neutron star or a black hole. This was not the first detection of gravitational waves, but all the previous sightings were created by the collisions of two black holes. Before this date, scientists had never observed waves coming from neutrons stars, and this was also the first time that telescopes had responded to a gravitational-wave detection and seen light coming from the same place in the sky at the same time. The light and waves together provided a bounty of information about where and how the crash happened that proved a boon for neutron star physics. "I was quite flabber-

Illustration by Nigel Hawtin

gasted," Lattimer says of the lucky observation. "I thought this is just too good to be true."

Astrophysicists traced the waves back to a pair of neutron stars about 130 million light-years from Earth. The details of the waves—their frequency and strength and the pattern they followed over time—allowed researchers to estimate that each weighed roughly 1.4 solar masses and stretched between 11 and 12 kilometers in radius before the crash. This knowledge will help scientists to formulate an essential descriptor for understanding neutron stars—their equation of state. The equation describes the density matter will take under different pressures and temperatures and should apply to all neutron stars in the universe. Theorists have come up with several possible formulations of the equation of state that correspond to different configurations of matter inside neutron stars, and the new measurements offered a chance to rule some out. The discovery that the neutron stars' radii were relatively small, for instance, was a surprise. Some theories run into difficulty when they try to fit both these compact neutron stars and known heavy stars, such as the 1.97-solar-mass behemoth, into the same fundamental equation of state. "It's starting to make our equation of state thread a needle path through these different observations," says Jocelyn Read, an astrophysicist at California State University, Fullerton, and co-leader of LIGO's Extreme Matter team. "Trying to make compact stars, as well as supporting massive stars, is getting to be challenging to the theory. It's definitely interesting and might get more interesting."

So far LIGO and Virgo have seen only this one neutron star collision, but another such observation could come any day now. "I've been working in this field long enough," Read says, "that it's just so fantastic to move from an era of what-ifs: 'If we could see gravitational waves, then we might be able to do this.' Now we're actually getting a chance to do this, and it hasn't gotten old yet."

THE LIMITS OF MATTER

IN TIME, AS GRAVITATIONAL-WAVE DETECTORS improve in sensitivity, the payoffs could be huge. For instance, one test of what is inside a neutron star involves looking for gravitational waves emitted by any swirling liquid in its middle. If the liquid has very low or no viscosity—as a superfluid would—it might begin flowing in patterns, called r-modes, that release gravitational waves. "These gravitational waves would be much weaker than from a merger," Alford says. "It is matter quietly sloshing versus being ripped apart." Alford and his collaborators determined that the currently running Advanced LIGO detector would not be able to see these waves, but future upgrades to LIGO, as well as planned observatories such as the ground-based Einstein Telescope under consideration in Europe, might.

Cracking the case on neutron stars would give us a picture of matter at its barely comprehensible extremes—a form so removed from the atoms that make up our world that it stretches the bounds of what is possible. It might turn imagined curiosities such as sloshing quark matter, superfluid neutrons and outlandish hyperon stars into reality. And understanding neutron stars could do something more: physicists' deeper goal is to use these squashed stars to tackle larger open questions, such as the laws that govern nuclear interactions—the complicated dance among protons, neutrons, quarks and gluons—as well as the biggest mystery of all—the nature of gravity.

Neutron stars are just one way of investigating nuclear forces, and simultaneous work is going on at particle accelerators around the world, which act like microscopes to peer inside atomic nuclei. When more of the nuclear problem is nailed down, scientists can turn their focus to gravity. "Neutron stars are a mixture of gravitational physics and nuclear physics," says Or Hen, a physicist at the Massachusetts Institute of Technology. "Right now we are using neutron stars as a lab to understand nuclear physics. But because we have access to nuclei here on Earth, we should be able to constrain the nuclear aspect of the problem well enough eventually. Then we can use neutron stars to understand gravity, which is one of the biggest challenges in physics."

CRACKING THE CASE ON NEUTRON STARS WOULD GIVE US A PICTURE OF MATTER AT ITS BARELY COMPREHENSIBLE EXTREMES.

Gravity as currently understood—through Einstein's general theory of relativity—does not get along with the theory of quantum mechanics. Ultimately one of the theories must budge, and physicists do not know which it will be. "We will get there," Hen says, "and that is a very exciting prospect." SA

Clara Moskowitz is a senior editor at *Scientific American,* specializing in space and physics.

MORE TO EXPLORE

Rapid Cooling of the Neutron Star in Cassiopeia A Triggered by Neutron Superfluidity in Dense Matter. Dany Page et al. in *Physical Review Letters*, Vol. 106, Article No. 081101; February 22, 2011.

GW170817: Measurements of Neutron Star Radii and Equation of State. B. P. Abbott et al. (The LIGO Scientific Collaboration and the Virgo Collaboration) in *Physical Review Letters,* Vol. 121, Article No. 161101; October 15, 2018.

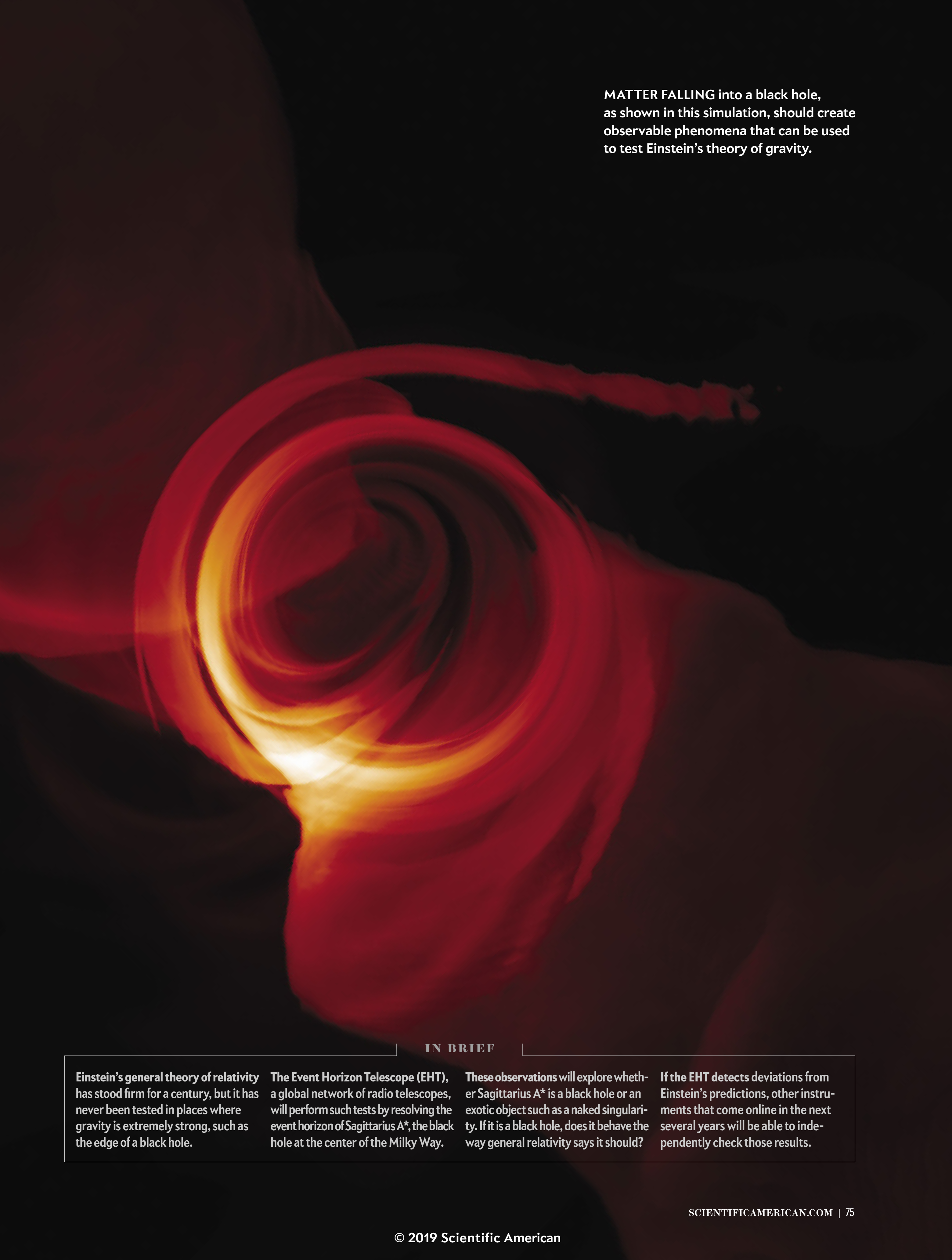
scientificamerican.com/magazine/sa

EXTREME PHYSICS

General relativity has never been tested in places where the effects of gravity become truly extreme—for example, at the edge of a black hole. That will soon change

By Dimitrios Psaltis and Sheperd S. Doeleman

THE BLACK HOLE TEST



MATTER FALLING into a black hole, as shown in this simulation, should create observable phenomena that can be used to test Einstein's theory of gravity.

IN BRIEF

Einstein's general theory of relativity has stood firm for a century, but it has never been tested in places where gravity is extremely strong, such as the edge of a black hole.

The Event Horizon Telescope (EHT), a global network of radio telescopes, will perform such tests by resolving the event horizon of Sagittarius A*, the black hole at the center of the Milky Way.

These observations will explore whether Sagittarius A* is a black hole or an exotic object such as a naked singularity. If it is a black hole, does it behave the way general relativity says it should?

If the EHT detects deviations from Einstein's predictions, other instruments that come online in the next several years will be able to independently check those results.

Scientists have been trying unsuccessfully to poke holes in Albert Einstein's general theory of relativity for a full century. So far, however, Einstein's theory has had it easy. Every assessment to date has been conducted in rather weak gravitational fields. To put general relativity to its greatest test, we need to see whether it holds up where gravity is extremely strong. And nowhere in the universe today is gravity stronger than at the edge of a black hole—at the event horizon, the boundary beyond which gravity is so overwhelming that light and matter that pass through can never escape.

The interior of a black hole is unobservable, but the gravitational field surrounding these objects causes matter close to the horizon to produce huge amounts of electromagnetic radiation that telescopes can detect. Near the black hole, the crushing force of gravity compresses inflowing matter, known as the accretion flow, into ever smaller volumes. This causes the infalling matter to reach temperatures of billions of degrees—which, ironically, makes the vicinity immediately surrounding a black hole one of the brightest spots in the cosmos.

If we could observe a black hole with a telescope with enough magnifying power to resolve the event horizon, we could follow matter as it spirals down toward the point of no return and see whether it behaves as general relativity says it should. There is, of course, a catch: developing a telescope that can resolve a black hole horizon poses several challenges. Notably we have to contend with the black hole's tiny size when viewed from Earth. Even the supermassive black holes now thought to inhabit the centers of most galaxies, which weigh in at millions or billions of our sun's mass and in some cases have diameters larger than our solar system, are so far away from Earth that they subtend incredibly tiny angles on the sky. The nearest example is Sagittarius A*, the four-million-solar-mass black hole at the center of the Milky Way; its event horizon would appear to be only 50 microarcseconds across, or roughly the size of a DVD seen on the moon. To resolve an object so small, a telescope must have an angular resolution more than 2,000 times finer than that achieved by the Hubble Space Telescope.

What is more, such black holes are obscured from our view in two ways. First, they occur at the very centers of galaxies, deep within dense clouds of gas and dust that block most of the electromagnetic spectrum. Second, even material that emits the light we want to detect—that glowing whirlpool of crushed matter spiraling in toward the horizon—is itself opaque to most wavelengths of light. Consequently, there are only a few wavelengths of light that can escape from the black hole's edge to be observed by us on Earth.

The Event Horizon Telescope (EHT) project is an international effort to overcome these hurdles and perform detailed observations of a black hole. To achieve the highest angular resolutions possible from the surface of Earth, the EHT exploits a technique known as very long baseline interferometry (VLBI), in which astronomers at radio dishes across the globe observe the same target simultaneously, record the data they collect on hard drives, and then later combine all those data using a supercomputer to form a single image. By doing so, many telescopes located on different continents can form one virtual Earth-sized telescope. The resolving power of a telescope is given by the ratio of the wavelength of light it observes to its size, and so VLBI routinely makes images of the radio sky with detail that far surpasses the magnifying power of any optical telescope.

By advancing the technologies used in VLBI so that observations can be made at the shortest radio wavelengths, the EHT will soon be able to meet all the challenges of black hole imaging. At these wavelengths (close to one millimeter in size), the Milky Way is largely transparent, enabling the EHT to observe Sagittarius A* with a minimum of blurring from the intervening gas. These same wavelengths are also able to pierce the matter falling toward the black hole, allowing access to the innermost regions surrounding Sagittarius A*'s event horizon. And in a true Goldilocks coincidence, the magnifying power of a globe-spanning VLBI array at millimeter wavelengths is well suited to resolving the event horizons of the nearest supermassive black holes.

In a parallel development, theoretical astrophysicists have developed mathematical models and computer simulations to explore a wide range of possible outcomes of these observations and to develop tools to interpret them. Using novel supercomputer algorithms, they have simulated the churn of matter just outside the black hole's event horizon, and in all simulations they have found that the black hole casts a "shadow" on the light coming off the accretion flow.

University of Washington physicist James Bardeen predicted the existence of a black hole shadow in 1973. By definition, any light that crosses the event horizon can never return. Bardeen identified the point *outside* the horizon where a photon will orbit the black hole. If a light ray crosses this orbit heading inward, it is caught forever and spirals inward to the event horizon. Light rays originating between the event horizon and this orbit *can* escape, but they have to be pointed almost radially outward, or they, too, risk being caught by the black hole's gravity and having their trajectories bent backward toward the event horizon. We call this boundary the photon orbit.

As far as light is concerned, the black hole acts like an opaque object, with the photon orbit defining its boundary. The contrast between the bright ring of the photon orbit and the dimmer interior is what is known as the shadow. The apparent size of this shadow as seen by observers on Earth is actually predicted to be quite a bit larger than the photon orbit. This occurs because the intense gravitational field surrounding the black hole "magnifies" the shadow through gravitational lensing.

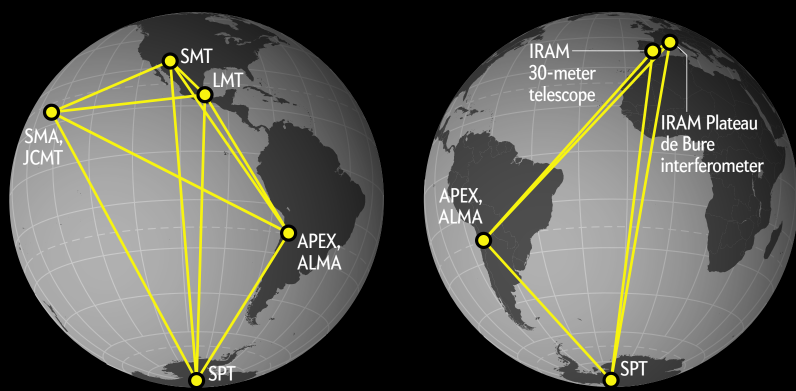
The EHT is now poised to observe this shadow and other features of black holes. In 2007 and 2009 observations verified that the technological approach was sound—and that the ultimate science goal was within reach—by targeting Sagittarius A* and another supermassive black hole at the heart of the galaxy Virgo A (also known as M87). These early observations linked together sites in Hawaii, Arizona and California to successfully measure the extent of radio emission at a 1.3-millimeter wavelength

PRECEDING PAGES: CHI-KWAN CHAN *University of Arizona* from both sources. In both cases, the measurements matched the expected size of the black hole shadow.

Observations planned with the full, planet-spanning web of dishes will yield enough data to allow us to construct complete images of these black holes. An additional, equally important set of observations will use VLBI data to search for and trace the trajectories of localized active regions ("hotspots") as they circle the black hole. Because general relativity predicts both what these black holes should look like and how matter should orbit them, these observations will allow us to perform a series of tests of Einstein's theory of relativity in the place where its most extreme predictions become manifest.

A Telescope the Size of Earth

At least nine radio telescopes and arrays around the globe will together form the Event Horizon Telescope (EHT). Every telescope is located at high altitude to minimize the absorption of the signals in Earth's atmosphere. By spanning the globe and operating at millimeter wavelengths, the array will achieve an effective angular resolution that is comparable to a few millionths of an arc second—good enough to spot a DVD on the moon.



CHECKING COSMIC CENSORSHIP

THE EHT WILL ENABLE US to answer a basic question: Is Sagittarius A* a black hole? All available evidence suggests that the answer is yes, but no one has ever directly observed a black hole, and other possibilities are consistent with general relativity. For example, Sagittarius A* could be something called a naked singularity.

A singularity in physics is a place where the solution to an equation is undefined and where the laws of nature as we understand them no longer operate. General relativity predicts that the universe began in a singularity—an initial moment when all the contents of the cosmos were concentrated into a single point of infinite density. The theory also tells us that a singularity, where gravity becomes infinite and matter is compressed to infinite density, lies at the center of every black hole.

In a black hole, the event horizon hides the singularity from our universe. General relativity does not require all singularities to be "clothed" by a horizon, however. There are an infinite number of solutions to Einstein's equations in which the singularities are "naked." Some of these solutions describe normal black holes spinning so fast that their horizons have "opened up" to reveal the singularity within; others describe black holes that have no event horizon.

Naked singularities, unlike black holes, remain highly theoretical: nobody has come up with a real-world recipe that would lead to their formation. Every astrophysically plausible computer simulation of the gravitational collapse of a star leads to the formation of a black hole with a horizon. Indeed, in 1969 Roger Penrose introduced the cosmic censorship hypothesis: the idea that physics somehow censors the nakedness of singularities by always enshrouding them with a horizon.

In September 1991 California Institute of Technology physicists John Preskill and Kip Thorne made a bet with University of Cambridge physicist Stephen Hawking that the cosmic censorship hypothesis is false and that naked singularities do exist. Even after Hawking's death 27 years later the bet is still standing, begging for an experiment that will settle it. Proving that Sagittarius A* has an event horizon would not conclusively disprove the existence of naked singularities elsewhere. Yet determining that the black hole in the center of our Milky Way is a naked singularity would allow us to directly observe phenomena at conditions where modern physics breaks down.

LOOKING FOR HAIR

DISCREDITING COSMIC CENSORSHIP would not be a death blow to general relativity; after all, its equations allow for naked singularities. Yet we also expect the EHT to test a long-standing idea about black holes called the no-hair theorem. And if the no-hair theorem is false, general relativity will, at minimum, have to be modified; the mathematical proof of this theorem leaves no wiggle room.

The theorem says that any black hole that is surrounded by an event horizon can be completely described using just three properties: mass, spin and electrical charge. In other words, any two black holes with the same mass, spin and electrical charge are entirely identical, just as any two elecrons are indistinguishable. Black holes, the theorem states, have no "hair"—no geometric irregularities or distinguishing characteristics.

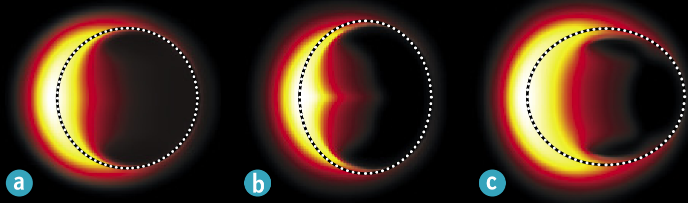
When we first started to think about imaging black holes using VLBI, we thought we could use the shapes and sizes of black hole shadows to learn the spins and orientations of the black holes that produced them. But our simulations presented us with an unexpected and, ultimately, very pleasant surprise. No matter how fast we let the black holes spin in our simulations, and no matter where we placed our mock observers, the black hole shadows always appeared nearly circular with an apparent size equal to about five times the radius of the event horizon. Because of some lucky coincidence—and if there is a

Maps by Terra Carta

Testing Einstein with Black Holes

Astrophysicists have created sophisticated models based on Einstein's general theory of relativity that predict how matter should behave in the vicinity of a black hole. Soon, Event Horizon Telescope observations of the black hole in the center of the Milky Way will tell us whether reality matches those predictions. If it does not, Einstein's theory may need to be modified.

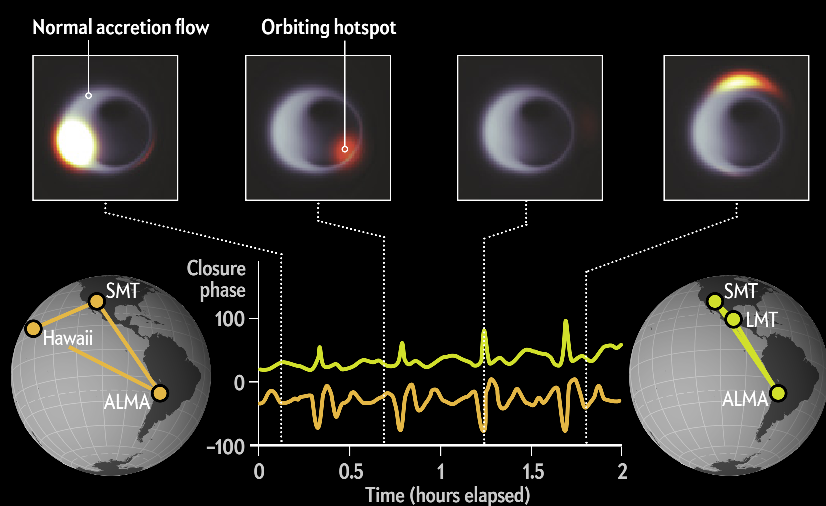
The Shape of the Shadow



A black hole casts a shadow on the emission from the hot matter surrounding it. The shape and size of the shadow depend, in principle, on how fast the black hole is spinning, on the amount that light rays are gravitationally bent in its vicinity, and on the orientation of the observer. Because of a lucky coincidence, all three effects conspire to make the shadow nearly circular for all black holes and observers (a). This coincidence, however, occurs only if Einstein's theory is correct and the no-hair theorem—which states that a black hole can be completely described by its mass, spin and charge—is satisfied. If observations reveal an elliptical shadow, as shown in images (b) and (c), then Einstein's theory will not pass this test.

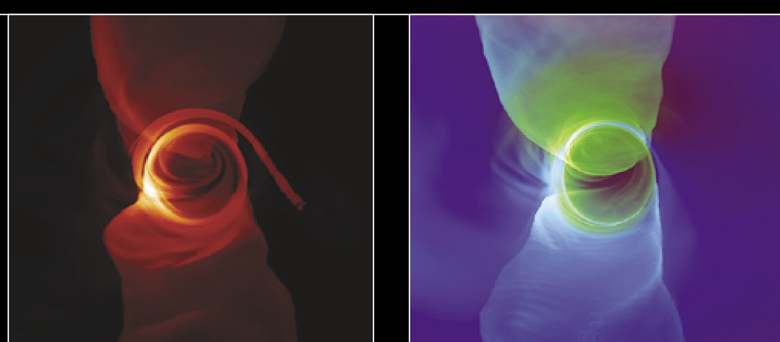
Tracking Closure Phase

Black holes sometimes flare up, and one explanation is that the normal steady accretion flow may be disrupted by "hot-spots," regions of increased temperature that orbit the black hole before dissipating. The EHT will use trios of telescopes to measure the difference in time of arrival of light emitted by hotspots; with these data, it is possible to triangulate the position of hot-spots. The simulation at the right shows such a signal (called closure phase) based on data from two different triangles. The orbit of the hotspot creates a "heartbeat" pattern—a time signature in closure phase. Measuring these signals will make it possible to map the spacetime of the black hole and test the predictions of Einstein's theory.



Simulating a Complex Reality

Scientists affiliated with the EHT are using supercomputers to perform elaborate numerical simulations of accreting black holes that exhibit the expected complexities of astrophysical objects. The right image depicts a black hole in a fairly quiet state of emission; on the left is a magnetically active region during a flare. With these simulations, scientists have developed algorithms that will allow them to extract the properties of black hole shadows from real-world observational data.



AVERY E. BRODERICK *University of Waterloo and Perimeter Institute for Theoretical Physics (a, b, c and top four images in Tracking Closure Phase)*; CHI-KWAN CHAN *University of Arizona (bottom two images)*; TERRA CARTA *(globes)*

deep physical reason for this, we still have not uncovered it—no matter how we alter the parameters in our models, the size and shape of the black hole shadow remain practically unchanged. This coincidence is excellent news if our goal is to test Einstein's theory because it happens only if the general theory of relativity holds up [*see box above*]. If Sagittarius A* has an event horizon, and if the size or the shape of its shadow deviates from our predictions, that would constitute a violation of the no-hair theorem—and, thus, of general relativity.

TRACING ORBITS AND MORE

EHT OBSERVATIONS will generate a great deal more data than are used to make images. The antennas will record the full polarization of the radiation emitted by the black hole, which will enable us to create maps of the magnetic fields near the event horizon. Such maps could help us understand the physics behind the powerful "jets" emanating from the centers of galaxies such as M87—beams of extraordinarily energetic matter traveling near the speed of light for up to thousands of light-years. Astrophysicists believe that magnetic fields near the event horizon of supermassive black holes power these jets; mapping the magnetic fields could help us test that hypothesis.

We can learn other things by watching the motion of matter around a black hole. The accretion flows around the black holes are expected to be highly turbulent and variable. Computer simulations often show the presence of localized, short-lived, magnetically active regions in them—"hotspots" similar to magnetic eruptions on the surface of the sun. These hotspots, which may explain the brightness variations that are often seen in Sagittarius A*, would circle the black hole at nearly the speed of light, along with the underlying accretion flow, completing full orbits in less than half an hour. In some cases, they become gravitationally lensed as they move behind the black hole and generate nearly complete Einstein rings—bright, gravitationally warped circles of light just like those the Hubble Space Telescope has detected from distant quasars. In other cases, they orbit around the black hole a few times before they lose their energy and dissipate.

Hotspots could complicate the process of making an image because the VLBI technique uses telescopes much like a time-lapse camera, leaving the virtual shutter open for the full duration of the observation and using the natural rotation of Earth to get as many different angles on the black hole as possible. If a bright spot in the accretion flow orbits the black hole, its appearance will be smeared, just as a photograph of a sprinter will be blurry if the camera shutter is left open too long.

Yet hotspots could also enable us to perform an entirely different test of general relativity. The EHT can trace the orbits of hotspots using a technique that goes by the fancy name of closure phase variability tracking. The method involves measuring the delays between the time of arrival of light from the hotspot at three telescopes and then using basic triangulation to infer the position of the hotspot in the sky. Orbiting hotspots will produce distinctive signatures in the raw data collected by the telescopes. And in much the same way that Einstein's equations predict the size and shape of the black hole shadow, they also disclose everything we need to know about the orbits that hotspots should trace. This hotspot model is somewhat schematic, and reality may be more complex. Nevertheless, at full sensitivity the EHT will be able to monitor structure in the accretion flow as it orbits the black hole, and that could provide yet another way of checking to see whether the predictions of general relativity hold up near the edge of a black hole.

EXTRAORDINARY EVIDENCE

WHAT HAPPENS if our observations appear to disagree with Einstein's theory? To use an expression popularized by Carl Sagan, extraordinary claims require extraordinary evidence. In the natural sciences, extraordinary evidence often means one or more verifications of any claim by independent methods. In the coming years, powerful optical and radio telescopes, as well as space-based gravitational-wave detectors, may provide such verification by monitoring the orbits of stars, neutron stars—tiny, incredibly dense objects produced by the gravitational collapse of massive stars—and other objects around supermassive black holes.

The Laser Interferometer Gravitational-wave Observatory (LIGO) detected gravitational waves from the coalescence of much smaller black holes than those found in the centers of galaxies, and as it will be accumulating more detections in the near future, it will help address whether these small black holes follow Einstein's predictions. The optical interferometer GRAVITY on the European Southern Observatory's Very Large Telescope (VLT) in Chile has tracked the orbits of stars in our galaxy that lie fairly close to Sagittarius A*'s event horizon—and found no evidence for unexpected phenomena at distances of about 1,000 times the radius of the black hole while continuing to push this limit closer to the event horizon itself. Once completed, the Square Kilometer Array (SKA), a radio interferometer under construction in South Africa and in Australia, will begin monitoring the orbits of rapidly spinning neutron stars, called pulsars, around the same black hole. Finally, the evolved Laser Interferometer Space Antenna (eLISA) will detect gravitational waves emitted as small compact objects orbit around supermassive black holes in nearby galaxies.

Because of the very strong gravitational fields of the black holes, the elliptical orbits of these objects will shift (precess) rapidly; this effect is so pronounced that the points of maximum distance from the black holes should trace a complete circle in only a few orbits. At the same time, the black holes will drag spacetime around with them, causing the orbital planes of objects within those spacetimes to precess as well. Measuring the rates of orbital precession for objects at different distances from a black hole will lead to a complete three-dimensional reconstruction of spacetime around a black hole, providing many tests of general relativity in the presence of extremely strong gravity.

Together all these instruments will help decide whether Einstein's general theory of relativity—in particular, its predictions about black holes—will survive intact for another century or be sacrificed on the altar of scientific progress. SA

Sheperd S. Doeleman is an astronomer at the Harvard-Smithsonian Center for Astrophysics, where he leads the team that develops instruments and algorithms and makes ultrahigh-resolution observations of black holes. He is director of the Event Horizon Telescope project.

Dimitrios Psaltis is a professor of astronomy and physics at the University of Arizona. He has pioneered the development of tests of Einstein's general theory of relativity in strong gravitational fields using observations of black holes and neutron stars in the electromagnetic spectrum.

MORE TO EXPLORE

Detecting Flaring Structures in Sagittarius A* with High-Frequency VLBI. Sheperd S. Doeleman et al. in *Astrophysical Journal,* Vol. 695, No. 1, pages 59–74; April 10, 2009.

Testing the No-Hair Theorem with Observations in the Electromagnetic Spectrum. II. Black Hole Images. Tim Johannsen and Dimitrios Psaltis in *Astrophysical Journal,* Vol. 718, No. 1, pages 446–454; July 20, 2010.

Jet-Launching Structure Resolved Near the Supermassive Black Hole in M87. Sheperd S. Doeleman et al. in *Science,* Vol. 338, pages 355–358; October 19, 2012.

The Power of Imaging: Constraining the Plasma Properties of GRMHD Simulations using EHT Observations of Sgr A*. Chi-Kwan Chan et al. in *Astrophysical Journal,* Vol. 799, No. 1, Article No. 1; January 20, 2015.

scientificamerican.com/magazine/sa

HEAD TRIP

Einstein's thought experiments left a long and somewhat mixed legacy of their own

By Sabine Hossenfelder

GEDANKENEXPERIMENT, GERMAN FOR "THOUGHT EXPERIMENT," WAS Albert Einstein's famous name for the imaginings that led to his greatest breakthroughs in physics. He traced his realization of light's finite speed—the core idea of special relativity—to his teenage daydreams about riding beams of light. General relativity, his monumental theory of gravitation, has its origins in his musings about going up and down in an elevator. In both cases, Einstein crafted new theories about the natural world by using his mind's eye to push beyond the limitations of laboratory measurements.

Einstein was neither the first nor the last theorist to do this, but his remarkable achievements were pivotal in establishing the *Gedankenexperiment* as a cornerstone of modern theoretical physics. Today physicists regularly use thought experiments to craft new theories and to seek out inconsistencies or novel effects within existing ones.

But the modern embrace of thought experiments raises some uncomfortable questions. In the search for a grand unified theory that would wed the small-scale world of quantum mechanics with Einstein's relativistic description of the universe at large, the most popular current ideas are bereft of observational support from actual experiments. Can thought alone sustain them? How far can we trust logical deduction? Where is the line between scientific intuition and fantasy? Einstein's legacy offers no certain answers: On one hand, his reliance on the power of thought was a spectacular success. On the other, many of his best-known thought experiments were based on data from real experimentation, such as the classic Michelson-Morley experiment that first measured the constancy of the speed of light. Moreover, Einstein's fixation on that which can be measured at times blinded him to deeper layers of reality—although even his mistakes in thought experiments contributed to later breakthroughs.

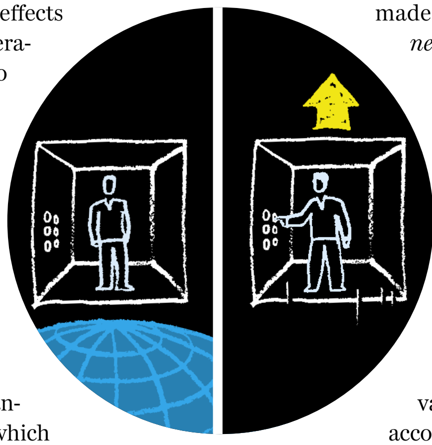
Here we will walk through some of Einstein's most iconic thought experiments, highlighting how they succeeded, where they failed and how they remain vital to questions now at the frontiers of theoretical physics.

THE WINDOWLESS ELEVATOR

IN HIS THOUGHT EXPERIMENTS, Einstein's genius was in realizing which aspects of experience were essential

Illustration by Sam Falconer

and which could be discarded. Consider his most famous one: the elevator thought experiment, which he began devising in 1907. Einstein argued that inside a windowless elevator, a person cannot tell whether the elevator is at rest in a gravitational field or is instead being hauled up with constant acceleration. He then conjectured that the laws of physics themselves must be identical in both situations. According to this "principle of equivalence," locally (in the elevator), the effects of gravitation are the same as that of acceleration in the absence of gravity. Converted into mathematical equations, this principle became the basis for general relativity. In other words, the elevator thought experiment motivated Einstein to make the daring intellectual leap that ultimately led to his greatest achievement, his geometric description of gravity.



SPOOKY ACTION

LATER IN HIS CAREER, Einstein fought hard against the tenets of quantum mechanics, particularly the uncertainty principle, which dictates that the more you know about one aspect of a fundamental particle, such as its position, the less you can know about another related aspect of that particle, such as its momentum—and vice versa. Einstein thought that the uncertainty principle was a sign that quantum theory was deeply flawed.

During a years-long exchange with Danish quantum theorist Niels Bohr, Einstein conceived of a series of thought experiments meant to demonstrate that it is possible to violate the uncertainty principle, but Bohr dissected every one of them. This exchange bolstered Bohr's conviction that quantum uncertainty was a fundamental aspect of nature. If not even the great Einstein could devise a way to precisely measure both the position and the momentum of a particle, then certainly there must be something to the uncertainty principle!

In 1935, along with his colleagues Boris Podolsky and Nathan Rosen, Einstein published what was meant to be his most potent critique of the uncertainty principle. Perhaps because Podolsky, not Einstein, drafted the actual text of the paper, this Einstein-Podolsky-Rosen (EPR) thought experiment was presented not as an easy-to-imagine scenario of boxes, clocks and light beams but as an abstract series of equations describing interactions between two generalized quantum systems.

The simplest version of the EPR experiment studies the paradoxical behavior of "entangled" particles—pairs of particles that share a common quantum state. It unfolds as follows: imagine an unstable particle with a spin of zero decaying into two daughter particles, which speed off in opposite directions. (Spin is a measure of a particle's angular momentum, but counterintuitively, it has little to do with a particle's rate of rotation.) Conservation laws dictate that the spins of those two daughter particles must add up to zero; one particle, then, could possess a spin value of "up," and the other could have a spin value of "down." The laws of quantum mechanics dictate that in the absence of measurement, neither of the particles possesses a definite spin until one of the two speeding entangled particles is measured. Once a measurement of one particle is made, the state of the other changes *instantaneously,* even if the particles are separated by vast distances!

Einstein believed this "spooky action at a distance" was nonsense. His own special theory of relativity held that nothing could travel faster than light, so there was no way for two particles to communicate with each other instantaneously from opposite sides of the universe. He suggested instead that the measurement outcomes must be determined prior to measurement by "hidden variables" that quantum mechanics failed to account for. Decades of discussion followed until 1964, when physicist John Bell developed a theorem quantifying exactly how the information shared between entangled particles differs from the information that Einstein postulated would be shared through hidden variables.

Since the 1970s lab experiments with entangled quantum systems have repeatedly confirmed that Einstein was wrong, that quantum particles indeed share mutual information that cannot be accounted for by hidden variables. Spooky action at a distance is real, but experiments have demonstrated that it cannot be used to transmit information faster than light, making it perfectly consistent with Einstein's special relativity. This counterintuitive truth remains one of the most mysterious conundrums in all of physics, and it was Einstein's stubborn, mistaken opposition that proved crucial to confirming it.

ALICE AND BOB

TODAY SOME OF THE MOST SIGNIFICANT thought experiments in physics explore how to reconcile Einstein's clockwork, relativistic universe with the fuzzy uncertainties inherent to quantum particles.

Consider, for instance, the widely discussed black hole information paradox. If you combine general relativity and quantum field theory, then you find that black holes evaporate, slowly radiating away their mass because of quantum effects. You also find that this process is not reversible: regardless of what formed the black hole, the evaporating black hole always produces the same featureless bath of radiation from which no information about its contents can be retrieved. But such a process is prohibited in quantum theory, which states that any occurrence can, in

IN BRIEF

One of Einstein's enduring contributions to physics was his use of *Gedankenexperiments,* or thought experiments, in which he used his mind's eye to explore the natural world.

His intuition about falling elevators, for example, led to his greatest achievement, the general theory of relativity.

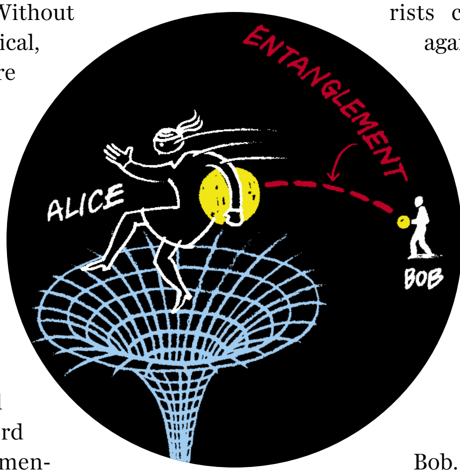
Today some of the most important questions in theoretical physics involve thought experiments about black holes.

Yet these thought experiments may be so far removed from empirical data as to be untestable.

Illustrations by Nigel Holmes

principle, be reversed in time. For instance, according to the laws of quantum mechanics, the leftovers of a burned book still contain all the information necessary to reassemble that book even though this information is not easily accessible. Not so for evaporating black holes. And so we arrive at a paradox, a logical inconsistency. A union of quantum mechanics and general relativity tells us that black holes must evaporate, but we conclude that the result is incompatible with quantum mechanics. We must be making some mistake—but where?

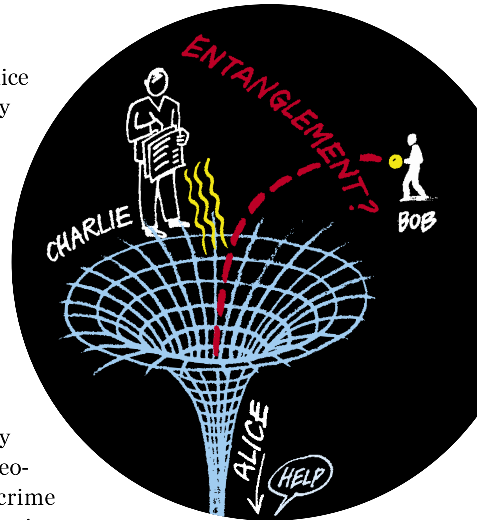
The thought experiments created to explore this paradox typically ask us to imagine a pair of observers, Bob and Alice, who share a pair of entangled particles—those spooky entities from the EPR experiment. Alice jumps into the black hole, carrying her particle with her, whereas Bob stays outside and far away with his. Without Alice, Bob's particle is just typical, with a spin that might measure up or down—the information that it once shared with its entangled partner is lost, along with Alice.



Bob and Alice play a central role in one of the most popular proposed solutions to the paradox, called black hole complementarity. Proposed in 1993 by Leonard Susskind, Lárus Thorlacius and John Uglum, all then at Stanford University, black hole complementarity rests on Einstein's golden rule for a *Gedankenexperiment:* a strict focus on that which can be measured. Susskind and his colleagues postulated that the information falling in with Alice must come out later with the evaporating black hole's radiation. This scenario would usually create another inconsistency because quantum mechanics allows only pair-wise entanglement with one partner at a time, a property called monogamy of entanglement. That is, if Bob's particle is entangled with Alice's, it cannot be entangled with anything else. But black hole complementarity requires that Bob's particle be entangled with Alice's *and* with the radiation the black hole later emits even though this violates monogamy. At first sight, then, black hole complementarity seems to exchange one inconsistency with another.

But like a perfect crime, if no one actually *witnesses* this inconsistency, perhaps it can subvert nature's otherwise strict laws. Black hole complementarity relies on the argument that it is physically impossible for any observer to see Alice and Bob's entangled particles breaking the rules.

To envision how this perfect quantum-mechanical crime could unfold, imagine a third observer, Charlie, hovering near the black hole, keeping an eye on Alice and Bob. He watches as Bob stays outside and as Alice falls in, measuring the black hole's emitted radiation all the while. In theory, information encoded in that radiation could tip off Charlie that Bob and Alice had violated the monogamy of their entanglement. To know for certain, however, Charlie would have to compare his observations not only with Bob's measurement but also with Alice's—inside the black hole. So he must hover at the horizon, measure the emitted radiation, then jump in to tell Alice what he has found. Amazingly enough, Susskind and Thorlacius showed that no matter how hard Charlie tries, it is impossible for him to enter the black hole and compare his information with Alice's before they are both torn apart by tidal forces. Their grisly fate suggests no violations of quantum mechanics can ever be measured by anybody around a black hole, and so theorists can commit this crime against nature with impunity.



Suffice it to say, not all theorists are convinced that this argument is valid. One criticism of black hole complementarity is that it might violate Einstein's equivalence principle—the one that grew out of his elevator thought experiment. Einstein's general relativity predicts that just as the elevator's passenger cannot distinguish between gravity and acceleration, an observer crossing a black hole's horizon should not notice anything unusual; there is no way an observer can tell that he or she has slipped past the point of no return.

Now let us return to the entanglement of Alice and Bob. If the radiation that Bob sees from far outside the hole contains all the information that we thought vanished with Alice behind the horizon, then this radiation must have been emitted with an extremely high energy; otherwise, it would not have escaped the strong gravitational pull near the horizon. This energy is high enough to vaporize any infalling observer before he or she slips past the black hole's horizon. In other words, black hole complementarity implies that black holes have a "firewall" just outside the horizon—and yet the firewall directly contradicts the predictions of Einstein's equivalence principle.

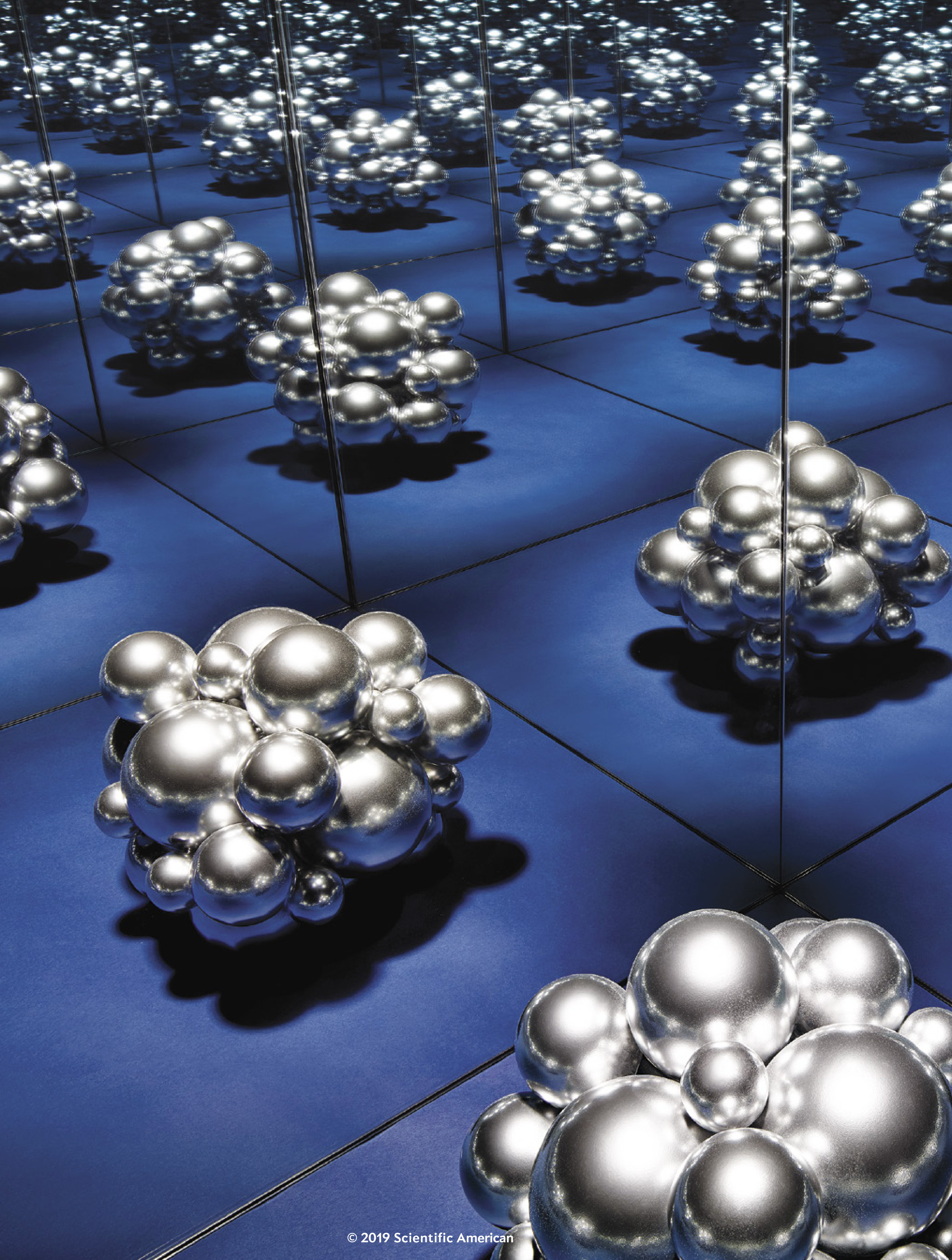
At this point, we have ventured deep into the realm of theory. Indeed, we might never know the solutions to these puzzles. But because those solutions could lead to an understanding of the quantum nature of space and time, these puzzles are, for better or worse, some of the most vibrant areas of research in theoretical physics. And it all goes back to Einstein's musings about falling elevators. SA

Sabine Hossenfelder is a theoretical physicist at the Frankfurt Institute for Advanced Studies in Germany, who researches physics beyond the Standard Model. She is author of the physics blog Backreaction and the book *Lost in Math: How Beauty Leads Physics Astray* (Basic Books, 2018).

MORE TO EXPLORE

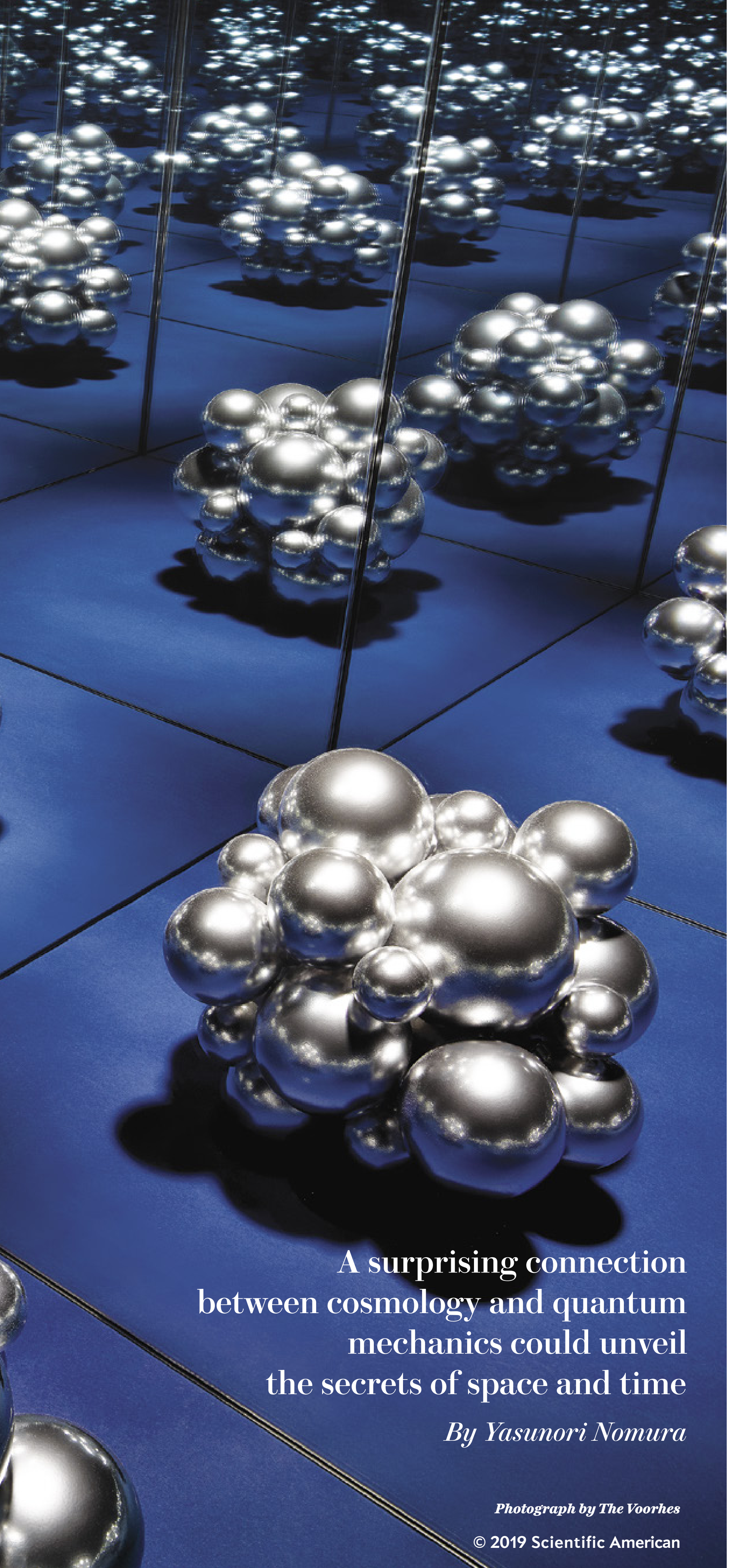
Einstein's Dice and Schrödinger's Cat: How Two Great Minds Battled Quantum Randomness to Create a Unified Theory of Physics. Paul Halpern. Basic Books, 2015.

scientificamerican.com/magazine/sa



EXTREME PHYSICS

THE QUANTUM MULTIVERSE



A surprising connection between cosmology and quantum mechanics could unveil the secrets of space and time

By Yasunori Nomura

Photograph by The Voorhes

MANY COSMOLOGISTS NOW ACCEPT THE EXTRAORDINARY IDEA THAT what seems to be the entire universe may actually be only a tiny part of a much larger structure called the multiverse. In this picture, multiple universes exist, and the rules we once assumed were basic laws of nature take different forms in each; for example, the types and properties of elementary particles may differ from one universe to another. The multiverse idea emerges from a theory that suggests the very early cosmos expanded exponentially. During this period of "inflation," some regions would have halted their rapid expansion sooner than others, forming what are called bubble universes, much like bubbles in boiling water. Our universe would be just one of these bubbles, and beyond it would lie infinitely more.

The idea that our entire universe is only a part of a much larger structure is, by itself, not as outlandish as it sounds. Throughout history scientists have learned many times over that the visible world is far from all there is. Yet the multiverse notion, with its unlimited number of bubble universes, does present a major theoretical problem: it seems to erase the ability of the theory to make predictions—a central requirement of any useful theory. In the words of Alan Guth of the Massachusetts Institute of Technology, one of the creators of inflation theory, "in an eternally inflating universe, anything that can happen will happen; in fact, it will happen an infinite number of times."

In a single universe where events occur a finite number of times, scientists can calculate the relative probability of one event occurring versus another by comparing the number of times these events happen. Yet in a multiverse where everything happens an infinite number of times, such counting is not possible, and nothing is more likely to occur than anything else. One can make any prediction one wants, and it is bound to come true in some universe, but that fact tells you nothing about what will go on in our specific world.

This apparent loss of predictive power has long troubled physicists. Some researchers, including me, have now realized that quantum theory—which, in contrast to the multiverse notion, is concerned with the very smallest particles in existence—may, ironically, point the way to a solution. Specifically, the cosmological picture of the eternally inflating multiverse may be mathematically equivalent to the "many worlds" interpretation of quantum mechanics, which attempts to explain how particles can seem to be in many places at once. As we will see, such a connection between the theories not only solves the prediction problem, it may also reveal surprising truths about space and time.

QUANTUM MANY WORLDS

I CAME TO THE IDEA of a correspondence between the two theories after I revisited the tenets of the many-worlds interpretation of quantum mechanics. This concept arose to make sense of some of the stranger aspects of quantum physics. In the quantum world—a nonintuitive place—cause and effect work differently than they do in the macro world, and the outcome of any process is always probabilistic. Whereas in our macroscopic experience, we can predict where a ball will land when it is thrown based on its starting point, speed and other factors, if that ball were a quantum particle, we could only ever say it has a certain chance of ending up here and another chance of ending up there. This probabilistic nature cannot be avoided by knowing more about the ball, the air currents or such details; it

IN BRIEF

The theory of cosmic inflation, which implies that the early cosmos expanded exponentially, suggests that we live not in a universe but a vast multiverse. **The problem with the multiverse idea,** however, is that all events that can occur will occur infinitely many times, ruining the theory's predictive ability. **Physicists realized** they can resolve the issue by viewing the multiverse as equivalent to a notion from quantum mechanics called the many-worlds interpretation, which suggests that our universe is one of many that coexist in "probability space" rather than in a single real space.

HUBBLE SPACE TELESCOPE'S Ultra Deep Field shows galaxies as far away as 13 billion light-years. Objects much farther out will forever be beyond reach because the expansion of space causes them to recede faster than the speed of light. This so-called cosmological horizon has important implications for the theory of the multiverse.

NASA, ESA, S. BECKWITH *STScI* AND HUBBLE DEEP FIELD TEAM

is an intrinsic property of the quantum realm. The same exact ball thrown under the same exact conditions will sometimes land at point A and other times at point B. This conclusion may seem strange, but the laws of quantum mechanics have been confirmed by innumerable experiments and truly describe how nature works at the scale of subatomic particles and forces.

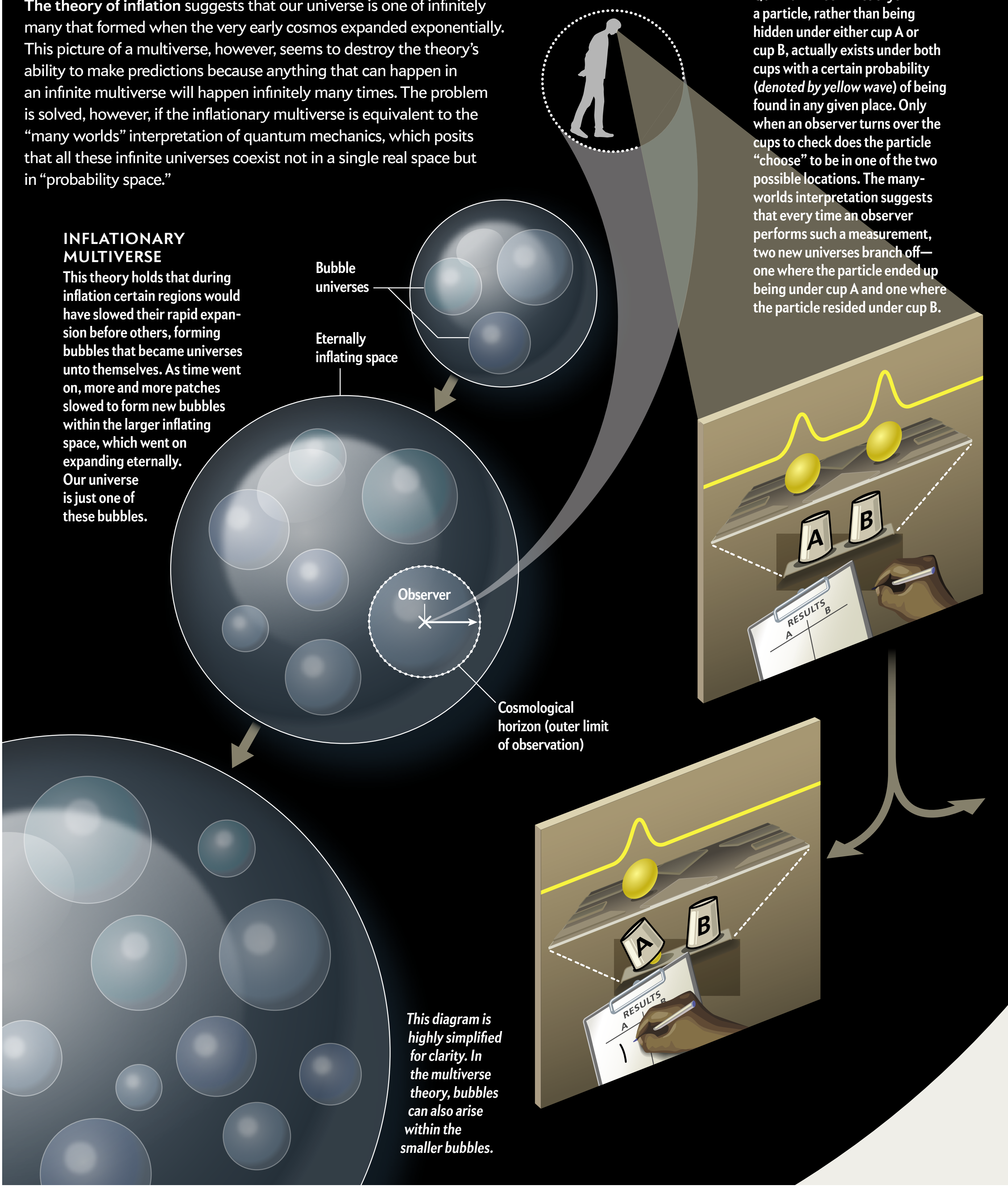
In the quantum world, we say that after the ball is thrown, but before we look for its landing spot, it is in a so-called superposition state of outcomes A and B—that is, it is neither at point A nor point B but located in a probabilistic haze of *both* points A and B (and many other locations as well). Once we look, however, and find the ball in a certain place—say, point A—then anyone else who examines the ball will also confirm that it sits at A. In other words, before any quantum system is measured, its outcome is uncertain, but afterward all subsequent measurements will find the same result as the first.

In the conventional understanding of quantum mechanics, called the Copenhagen interpretation, scientists explain this shift by saying that the first measurement changed the state of the system from a superposition state to the state A. But although the Copenhagen interpretation does predict the outcomes of laboratory experiments, it leads to serious difficulties at the conceptual level. What does the "measurement" really mean, and why does it change the state of the system from a superposition of possibilities to a single certainty? Does the change of state occur when a dog or even a fly observes the system? What about when a molecule in the air interacts with the system, which we expect to be occurring all the time yet which we do not usually treat as a measurement that can interfere with the outcome? Or is there some special physical significance in a human consciously learning the state of the system?

In 1957 Hugh Everett, then a graduate student at Princeton University, developed the many-worlds interpretation of quantum mechanics that beautifully addresses this issue—although at the time many received it with ridicule, and the idea is still less favored than the Copenhagen interpretation. Everett's key insight was that the state of a quantum system reflects the state of the *whole* universe around it, so that we must include the observer in a complete description of the measurement. In other words, we cannot consider the ball, the wind and the hand that throws it in isolation—we must also include in the fundamental description the person who comes along to inspect its landing spot, as well as everything else in the cosmos at that time. In this picture, the quantum state after the measurement is still a superposition—not just a superposition of two landing spots but of two entire worlds! In the first world, the observer finds that the state of the system has changed to A, and therefore any observer in this particular world will obtain result A in all subsequent measurements. But when the measurement was made, another universe split off from the first in

Inflation Meets Many Worlds

The theory of inflation suggests that our universe is one of infinitely many that formed when the very early cosmos expanded exponentially. This picture of a multiverse, however, seems to destroy the theory's ability to make predictions because anything that can happen in an infinite multiverse will happen infinitely many times. The problem is solved, however, if the inflationary multiverse is equivalent to the "many worlds" interpretation of quantum mechanics, which posits that all these infinite universes coexist not in a single real space but in "probability space."



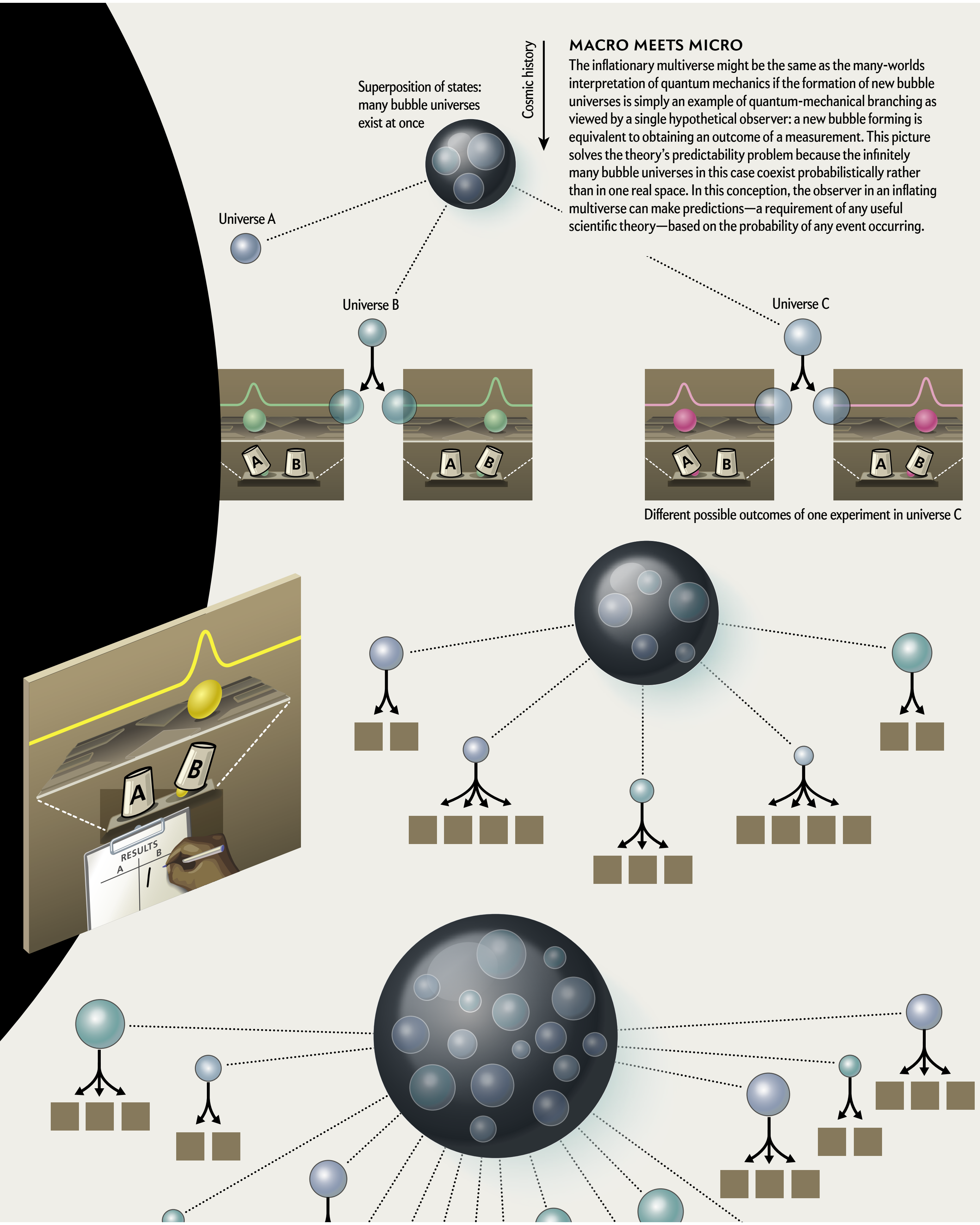
INFLATIONARY MULTIVERSE

This theory holds that during inflation certain regions would have slowed their rapid expansion before others, forming bubbles that became universes unto themselves. As time went on, more and more patches slowed to form new bubbles within the larger inflating space, which went on expanding eternally. Our universe is just one of these bubbles.

MANY WORLDS

Quantum mechanics says that a particle, rather than being hidden under either cup A or cup B, actually exists under both cups with a certain probability (*denoted by yellow wave*) of being found in any given place. Only when an observer turns over the cups to check does the particle "choose" to be in one of the two possible locations. The many-worlds interpretation suggests that every time an observer performs such a measurement, two new universes branch off—one where the particle ended up being under cup A and one where the particle resided under cup B.

Illustration by Jen Christiansen



MACRO MEETS MICRO

The inflationary multiverse might be the same as the many-worlds interpretation of quantum mechanics if the formation of new bubble universes is simply an example of quantum-mechanical branching as viewed by a single hypothetical observer: a new bubble forming is equivalent to obtaining an outcome of a measurement. This picture solves the theory's predictability problem because the infinitely many bubble universes in this case coexist probabilistically rather than in one real space. In this conception, the observer in an inflating multiverse can make predictions—a requirement of any useful scientific theory—based on the probability of any event occurring.

which the observer finds, and keeps finding, that the ball landed at point B. This feature explains why the observer—let us say it is a man—thinks that his measurement changes the state of the system; what actually happens is that when he makes a measurement (interacts with the system), he himself divides into two different people who live in two different parallel worlds corresponding to two separate outcomes, A and B.

According to this picture, humans making measurements have no special significance. The state of the entire world continuously branches into many possible parallel worlds that coexist as a superposition. A human observer, being a part of nature, cannot escape from this cycle—the observer keeps splitting into many observers living in many possible parallel worlds, and all are equally "real." An obvious but important implication of this picture is that everything in nature obeys the laws of quantum mechanics, whether small or large.

What does this interpretation of quantum mechanics have to do with the multiverse discussed earlier, which seems to exist in a continuous real space rather than as parallel realities? In 2011 I argued that the eternally inflating multiverse and quantum-mechanical many worlds à la Everett are the same concept in a specific sense. In this understanding, the infinitely large space associated with eternal inflation is a kind of "illusion"—the many bubble universes of inflation do not all exist in a single real space but represent the possible different branches on the probabilistic tree. Around the same time that I made this proposal, Raphael Bousso of the University of California, Berkeley, and Leonard Susskind of Stanford University put forth a similar idea. If true, the many-worlds interpretation of the multiverse would mean that the laws of quantum mechanics do not operate solely in the microscopic realm—they also play a crucial role in determining the global structure of the multiverse even at the largest distance scales.

I AND OTHER PHYSICISTS ARE ALSO PURSUING THE QUANTUM MULTIVERSE IDEA FURTHER. HOW CAN WE DETERMINE THE QUANTUM STATE OF THE ENTIRE MULTIVERSE? WHAT IS TIME, AND HOW DOES IT EMERGE?

BLACK HOLE QUANDARY

TO BETTER EXPLAIN how the many-worlds interpretation of quantum mechanics could describe the inflationary multiverse, I must digress briefly to talk about black holes. Black holes are extreme warps in spacetime whose powerful gravity prevents objects that fall into them from escaping. As such, they provide an ideal testing ground for physics involving strong quantum and gravitational effects. A particular thought experiment about these entities reveals where the traditional way of thinking about the multiverse goes off track, thereby making prediction impossible.

Suppose we drop a book into a black hole and observe from the outside what happens. Whereas the book itself can never escape the black hole, theory predicts that the information in the book will not be lost. After the book has been shredded by the black hole's gravity and after the black hole itself has gradually evaporated by emitting faint radiation (a phenomenon known as Hawking radiation, discovered by physicist Stephen Hawking of the University of Cambridge), outside observers can reconstruct all the information contained in the initial book by closely examining the radiation released. Even before the black hole has completely evaporated, the book's information starts to slowly leak out via each piece of Hawking radiation.

Yet a puzzling thing occurs if we think about the same situation from the viewpoint of someone who is falling into the black hole along with the book. In this case, the book seems to simply pass through the boundary of the black hole and stay inside. Thus, to this inside observer, the information in the book is also contained within the black hole forever. On the other hand, we have just argued that from a distant observer's point of view, the information will be *outside*. Which is correct? You might think that the information is simply duplicated: one copy inside and the other outside. Such a solution, however, is impossible. In quantum mechanics, the so-called no-cloning theorem prohibits faithful, full copying of information. Therefore, it seems that the two pictures seen by the two observers cannot both be true.

Physicists Gerard 't Hooft of Utrecht University in the Netherlands, Susskind and their collaborators have proposed the following solution: the two pictures can both be valid but not at the same time. If you are a distant observer, then the information is outside. You need not describe the interior of the black hole, because you can never access it even in principle; in fact, to avoid cloning information, you must think of the interior spacetime as nonexistent. On the other hand, if you are an observer falling into the hole, then the interior is all you have, and it contains the book and its information. This view, however, is possible only at the cost of ignoring the Hawking radiation being emitted from the black hole—but such a conceit is allowed because you yourself have crossed the black hole boundary and accordingly are trapped inside, cut off from the radiation emitted from the boundary. There is no inconsistency in either of these two viewpoints; only if you artificially "patch" the two, which you can never physically do, given that you cannot be both a distant and a falling observer at the same time, does the apparent inconsistency of information cloning occur.

COSMOLOGICAL HORIZONS

THIS BLACK HOLE CONUNDRUM may seem unrelated to the issue of how the many-worlds notion of quantum mechanics and the multiverse can be connected, but it turns out that the boundary of a black hole is similar in important ways to the so-called cosmological horizon—the boundary of the spacetime region within which we can receive signals from deep space. The horizon exists because space is expanding exponentially, and objects farther than this cutoff are receding faster than the speed of light, so any message from them can never reach us. The situation, therefore, is akin to a black hole viewed by a distant observer. Also, as in the case of the black hole, quantum mechanics requires an observer inside the horizon to view spacetime on the other side of the boundary—in this case, the exterior of the cosmological horizon—as nonexistent. If we consider such spacetime in addition to the information that can be retrieved from the horizon later (analogous to Hawking radiation in the black hole case), then we are overcounting the information. This problem implies that any description of the quantum state of the universe should include only the region within (and on) the horizon—in particular, there can be no infinite space in any single, consistent description of the cosmos.

If a quantum state reflects only the region within the horizon, then where is the multiverse, which we thought existed in an eternally inflating infinite space? The answer is that the creation of bubble universes is probabilistic, like any other process in quantum mechanics. Just as a quantum measurement could spawn many different results distinguished by their probability of occurring, inflation could produce many different universes, each with a different probability of coming into being. In other words, the quantum state representing eternally inflating space is a superposition of worlds—or branches—representing different universes, with each of these branches including only the region within its own horizon.

Because each of these universes is finite, we avoid the problem of predictability that was raised by the prospect of an infinitely large space that encompasses all possible outcomes. The multiple universes in this case do not all exist simultaneously in real space—they coexist only in "probability space," that is, as possible outcomes of observations made by people living inside each world. Thus, each universe—each possible outcome—retains a specific probability of coming into being.

This picture unifies the eternally inflating multiverse of cosmology and Everett's many worlds. Cosmic history then unfolds like this: the multiverse starts from some initial state and evolves into a superposition of many bubble universes. As time passes, the states representing each of these bubbles further branch into more superpositions of states representing the various possible outcomes of "experiments" performed within those universes (these need not be scientific experiments—they can be any physical processes). Eventually the state representing the whole multiverse will thus contain an enormous number of branches, each of which represents a possible world that may arise from the initial state. Quantum-mechanical probabilities therefore determine outcomes in cosmology and in microscopic processes. The multiverse and quantum many worlds are really the same thing; they simply refer to the same phenomenon—superposition—occurring at vastly different scales.

In this new picture, our world is only one of all possible worlds that are allowed by the fundamental principles of quantum physics and that exist simultaneously in probability space.

THE REALM BEYOND

TO KNOW IF THIS IDEA is correct, we would want to test it experimentally. But is that feasible? It turns out that discovery of one particular phenomenon would lend support to the new thinking. The multiverse could lead to a small amount of negative spatial curvature in our universe—in other words, objects would travel through space not along straight lines as in a flat cosmos but along curves, even in the absence of gravity. Such curvature could happen because, even though the bubble universes are finite as seen from the perspective of the entire multiverse, observers inside a bubble would perceive their universe to be infinitely large, which would make space seem negatively curved (an example of negative curvature is the surface of a saddle, whereas the surface of a sphere is positively curved). If we were inside one such bubble, space should likewise appear to us to be bent.

Evidence so far indicates that the cosmos is flat, but experiments studying how distant light bends as it travels through the cosmos are likely to improve measures of the curvature of our universe by about two orders of magnitude in the next few decades. If these experiments find any amount of negative curvature, they will support the multiverse concept because, although such curvature is technically possible in a single universe, it is implausible there. Specifically, a discovery supports the quantum multiverse picture described here because it can naturally lead to curvature large enough to be detected, whereas the traditional inflationary picture of the multiverse tends to produce negative curvature many orders of magnitude smaller than we can hope to measure.

Interestingly, the discovery of positive curvature would falsify the multiverse notion presented here because inflation theory suggests that bubble universes could produce only negative curvature. On the other hand, if we are lucky, we may even see dramatic signs of a multiverse—such as a remnant from a "collision" of bubble universes in the sky, which may be formed in a single branch in the quantum multiverse. Scientists are, however, far from certain if we will ever detect such signals.

I and other physicists are also pursuing the quantum multiverse idea further on a theoretical level. We can ask fundamental questions such as: How can we determine the quantum state of the entire multiverse? What is time, and how does it emerge? The quantum multiverse picture does not immediately answer these questions, but it does provide a framework to address them. Lately, for instance, I have found that constraints imposed by the mathematical requirement that our theory must include rigorously defined probabilities may enable us to determine the unique quantum state of the entire multiverse. These constraints also suggest that the overall quantum state stays constant even though a physical observer, who is a part of the multiverse state, will see that new bubbles constantly form. This implies that our sense of the universe changing over time and, indeed, the concept of time itself may be an illusion. Time, according to this notion, is an "emergent concept" that arises from a more fundamental reality and seems to exist only within local branches of the multiverse.

Many of the ideas I have discussed are still quite speculative, but it is thrilling that physicists can talk about such big and deep questions based on theoretical progress. Who knows where these explorations will finally lead us? It seems clear, though, that we live in an exciting era in which our scientific explorations reach beyond what we thought to be the entire physical world—our universe—into a potentially limitless realm. SA

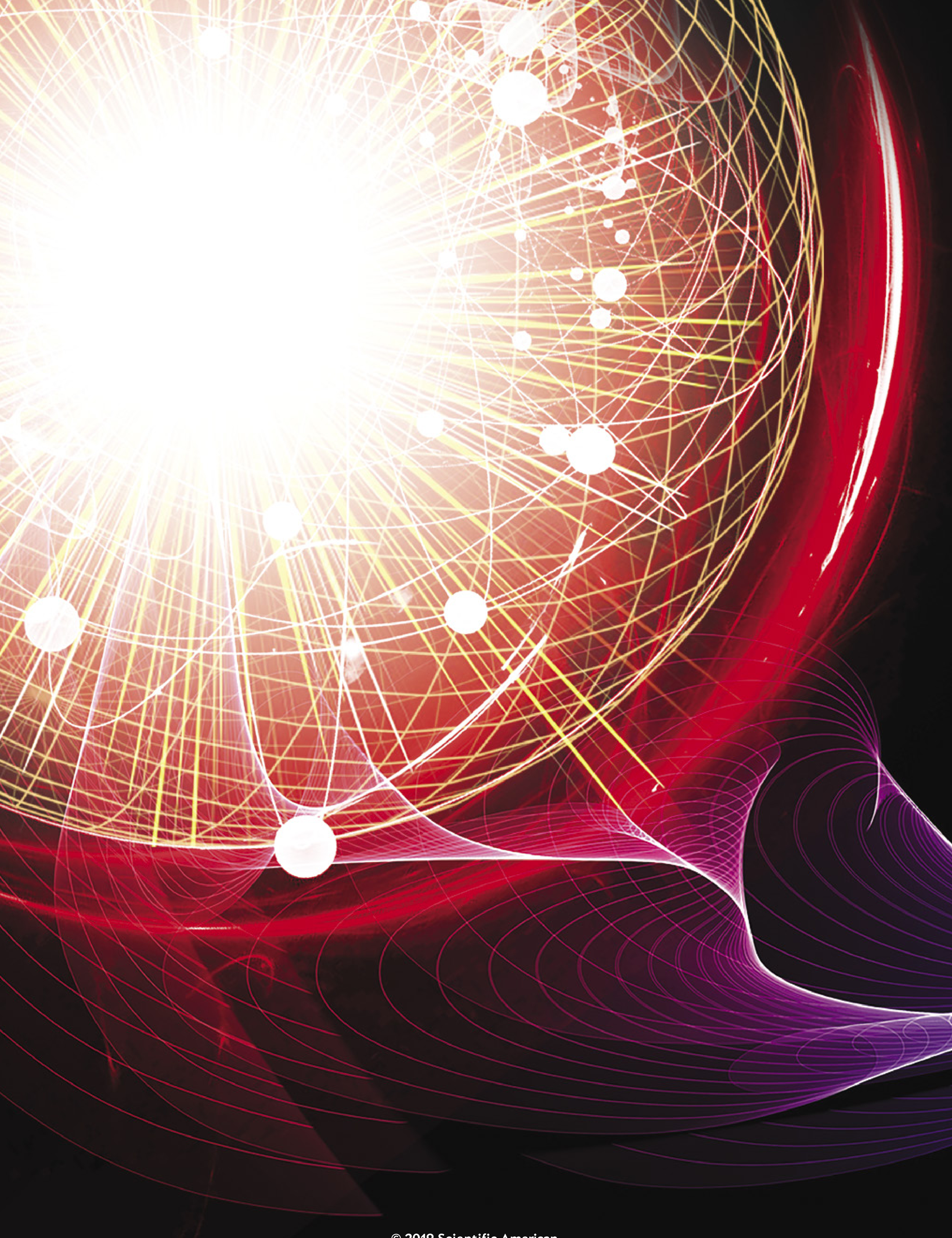
Yasunori Nomura is a professor of physics and director of the Berkeley Center for Theoretical Physics at the University of California, Berkeley. He is also a senior faculty scientist at Lawrence Berkeley National Laboratory and a principal investigator at the University of Tokyo's Kavli Institute for the Physics and Mathematics of the Universe.

MORE TO EXPLORE

Physical Theories, Eternal Inflation, and the Quantum Universe. Yasunori Nomura in *Journal of High Energy Physics,* Vol. 2011, No. 11, Article No. 063; November 2011. Preprint available at **https://arxiv.org/abs/1104.2324**

Multiverse Interpretation of Quantum Mechanics. Raphael Bousso and Leonard Susskind in *Physical Review D,* Vol. 85, No. 4, Article No. 045007. Published online February 6, 2012. Preprint available at **https://arxiv.org/abs/1105.3796**

scientificamerican.com/magazine/sa

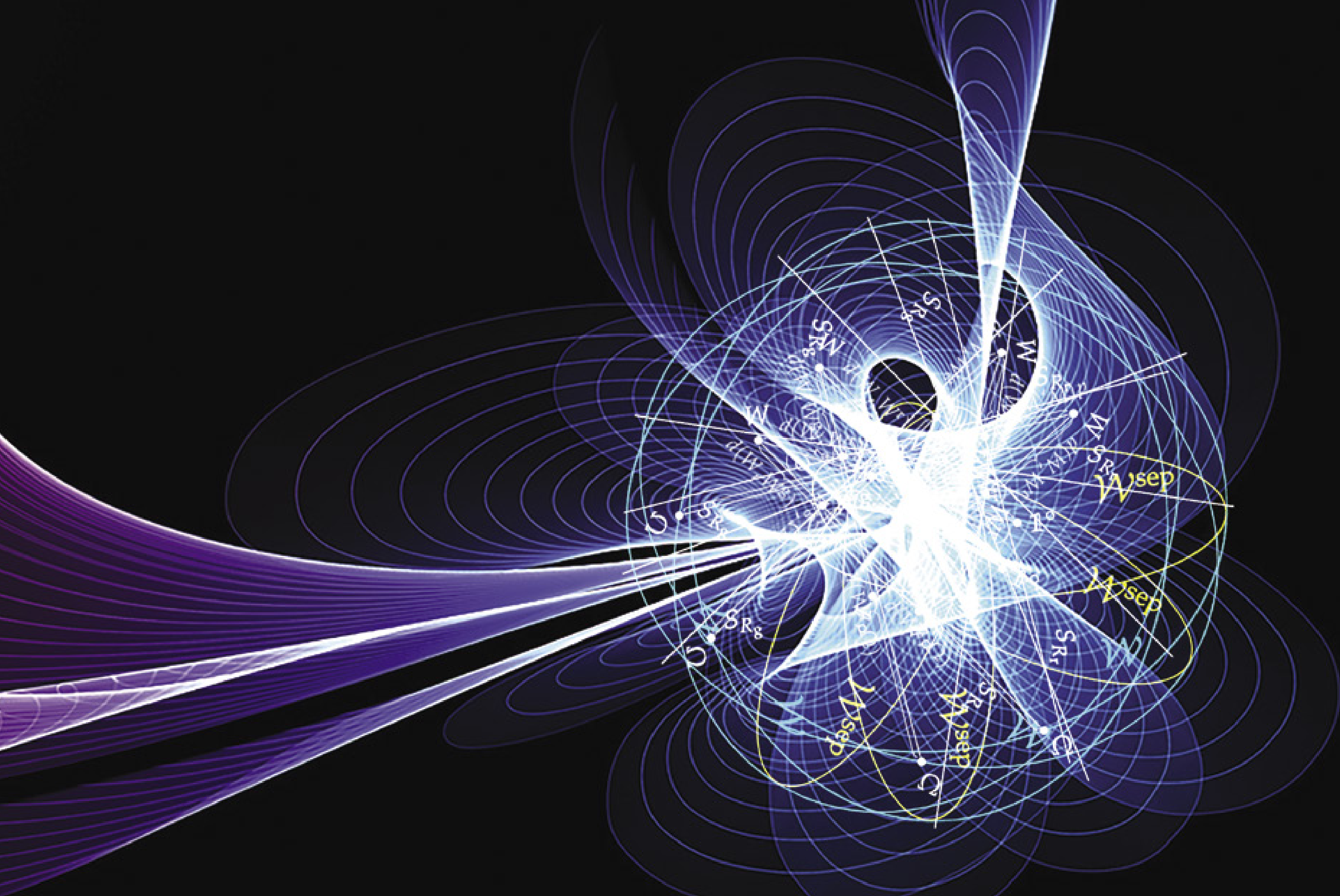


SPOOKY ACTION

Recent experiments quash the hope that the unsettling phenomenon of quantum entanglement can be explained away

By Ronald Hanson and Krister Shalm

Illustration by Kenn Brown and Chris Wren, Mondolithic Studios



Not all revolutions start big. In the case of quantum mechanics, a quiet one began in 1964, when physicist John Bell published an equation. This equation, in the form of a mathematical inequality, proposed a test to address deep philosophical questions that troubled many of the early founders of quantum mechanics.

The issue was whether particles separated by vast distances could retain a connection so that measurements performed on one affect the other. According to classical physics, this should be impossible. But under quantum theory, it happens all the time. Through his equation, Bell proposed a way to determine whether the universe could actually be that strange.

Over the past half a century his simple equation has profoundly changed the way we think about quantum theory. Today many of the quantum technologies that physicists are inventing owe their beginnings to Bell's test. Yet it was not until 2015, more than 50 years after Bell proposed his inequality, that scientists were able to verify the predictions of Bell's theorem in the most complete manner possible. These experiments close a quest that has spanned generations and mark the start of a new era in developing quantum technologies.

HIDDEN VARIABLES

TO UNDERSTAND BELL'S EQUATION, we must go back to the roots of quantum mechanics. This set of rules describes the behavior of light and matter at the smallest scales. Atoms, electrons, photons and other subatomic particles act differently from things we experience in our everyday lives. One of the major deviations is that these small particles exist in uncertain states. Take an electron's spin, for example. If an electron whose spin is sideways passes through a magnetic field oriented up and down, half the time it will veer upward and half the time it will veer downward, but the outcome is truly random. Compare this to a coin flip. We might think that flipping a coin is equally random, but if we knew precisely the mass of the coin, how much force was used to flip it and all the details about the air currents hitting it, we would be able to predict exactly how the coin would land. Electron spin is different, however. Even if we had perfect knowledge about all the properties of the electron and its spin before it passes through the magnetic field, quantum fuzziness prevents us from knowing which way it will go (we can, however, calculate the *probability* of it going up or down). When scientists actually measure a quantum system, though, all these possibilities cease to exist somehow, and a single outcome is decided—the electron ends up having a spin that is oriented either up or down.

When physicists formulated quantum theory in the early 20th century, some of its founding members, such as Albert Einstein and Erwin Schrödinger, felt uncomfortable with the fuzziness of quantum states. Perhaps, they thought, nature is not really fuzzy, and a theory that goes beyond quantum mechanics could exactly predict the behavior of particles. Then it would be possible to foresee the outcome of a measurement of the spin of an electron in the same way it is possible to know exactly how a coin will land if you have enough information.

Schrödinger introduced the idea of entanglement (*Verschränkung* in German) to describe quantum fuzziness spread across two or more particles. According to quantum theory, properties of particles can be entangled such that their joint value is precisely known, but the individual values remain completely uncertain. An analogy would be two dice that, when rolled, would each yield a random result but together always add up to 7. Schrödinger used the idea of entanglement in a famous thought experiment in which the fuzziness of the state of an atom becomes entangled with a cat being dead or alive. Surely any cat is either dead or alive and not in an absurd limbo in between, Schrödinger reasoned, and therefore we should question the notion that atoms can be fuzzy at all.

IN BRIEF

In 1964 physicist John Bell discovered that the phenomenon of quantum entanglement—where two particles can retain a "spooky" connection even when far apart—leads to a mathematical conflict with our intuitive picture of nature.

Since Bell's proposal, experimenters have staged many versions of his test. Most results seemed in agreement with the existence of entanglement. Yet each of these experiments has contained loopholes that make it possible for "hidden variables" to act behind the scenes, producing results that masquerade as entanglement.

Finally, in 2015, several groups conducted the first loophole-free Bell tests, ruling out any local hidden variable explanation.

Einstein, with his collaborators Boris Podolsky and Nathan Rosen (known together as EPR), took the argument a step further by analyzing two entangled electrons that are far apart. Imagine that the spins of the particles are entangled such that when measured along the same orientation, opposite values will always result. For instance, if scientists measure one electron spin and find it to be pointing up, the other will point down. Such correlations are certainly surprising when the electrons are far enough apart that it is impossible for them to communicate at the speed of light before their individual spins are measured. How does the second particle know that the first one was up? Einstein famously called this synchronization "spooky actions at a distance."

The EPR analysis of this case, published in 1935 in a now classic paper, started from two very reasonable assumptions. First, if scientists can predict a measurement outcome with certainty, there must be some property in nature that corresponds to this outcome. Einstein named these properties "elements of reality." For example, if we know that an electron's spin is up, we can predict with certainty that if it travels through an appropriate magnetic field it will always be deflected upward. In this situation, the electron's spin would be an element of reality because it is well defined and not fuzzy. Second, an event in one place cannot instantaneously affect a faraway event; influences cannot travel faster than the speed of light.

Taking these assumptions, let us analyze two entangled electrons held at distant places by two people, Alice and Bob. Suppose Alice measures her electron spin along the z direction. Because of the perfect anticorrelation, she immediately knows what the outcome will be if Bob measures his electron spin along z as well. According to EPR, the z component of Bob's electron spin would thus be an element of reality. Similarly, if Alice decides to measure the spin along the x direction, she would know with certainty the outcome of a measurement on Bob's electron spin along x. In this case, the x component of Bob's electron spin would be an element of reality. But because Alice and Bob are far apart, Alice's decision to measure along the z direction or the x direction cannot influence what happens at Bob's. Therefore, to account for the perfect anticorrelations predicted by quantum theory, the value of Bob's electron spin must be perfectly predictable both along the z direction and the x direction. This appears to contradict quantum theory, which states, through the so-called Heisenberg uncertainty principle, that the spin can have a well-defined value along a single direction only and must be fuzzy along the others.

This conflict led EPR to conclude that quantum theory is incomplete. They suggested that it might be possible to resolve the contradiction by supplementing the theory with extra variables. In other words, there might be a deeper theory that goes beyond quantum mechanics in which the electrons possess extra properties that describe how they will behave when jointly measured. These extra variables might be hidden from us, but if we had access to them, we could predict exactly what would happen to the electrons. The apparent fuzziness of quantum particles is a result of our ignorance. Physicists call any such successor to quantum mechanics that contains these hidden variables a "local hidden variable theory." The "local" here refers to the hidden signals not being able to travel faster than the speed of light.

BELL'S TWIST

EINSTEIN DID NOT QUESTION the predictions of quantum mechanics itself; rather he believed there was a deeper truth in the form of hidden variables that govern reality. After the 1935 EPR paper, interest in these foundational issues in quantum mechanics died down. The possibility of hidden variables was seen as a philosophical question without any practical value—the predictions of theories with and without hidden variables appeared to be identical. But that changed in 1964, when Bell startlingly showed that in certain circumstances hidden variable theories and quantum mechanics predict different things. This revelation meant it is possible to test experimentally whether local hidden variable theories—and thus Einstein's hoped-for deeper truth of nature—can really exist.

Bell analyzed the EPR thought experiment but with one twist: he let Alice and Bob measure their electron spins along any possible direction. In the traditional experiment, Alice and Bob must measure along the same direction and therefore find that their results are 100 percent correlated—if Alice measures up, then Bob always measures down. But if Alice and Bob are sometimes measuring along different axes, sometimes their outcomes are not synchronized, and that is where the differences between quantum theory and hidden variable theories come in. Bell showed that for certain sets of directions, the correlations between the outcomes of Alice and Bob's measurements would be stronger according to quantum theory than according to any local hidden variable theory—a difference known as Bell's inequality. These differences arise because the hidden variables cannot influence one another faster than the speed of light and therefore are limited in how they can coordinate their efforts. In contrast, quantum mechanics allows the two electrons' spins to exist jointly in a single entangled fuzzy state that can stretch over vast distances. Entanglement causes quantum theory to predict correlations that are up to 40 percent stronger.

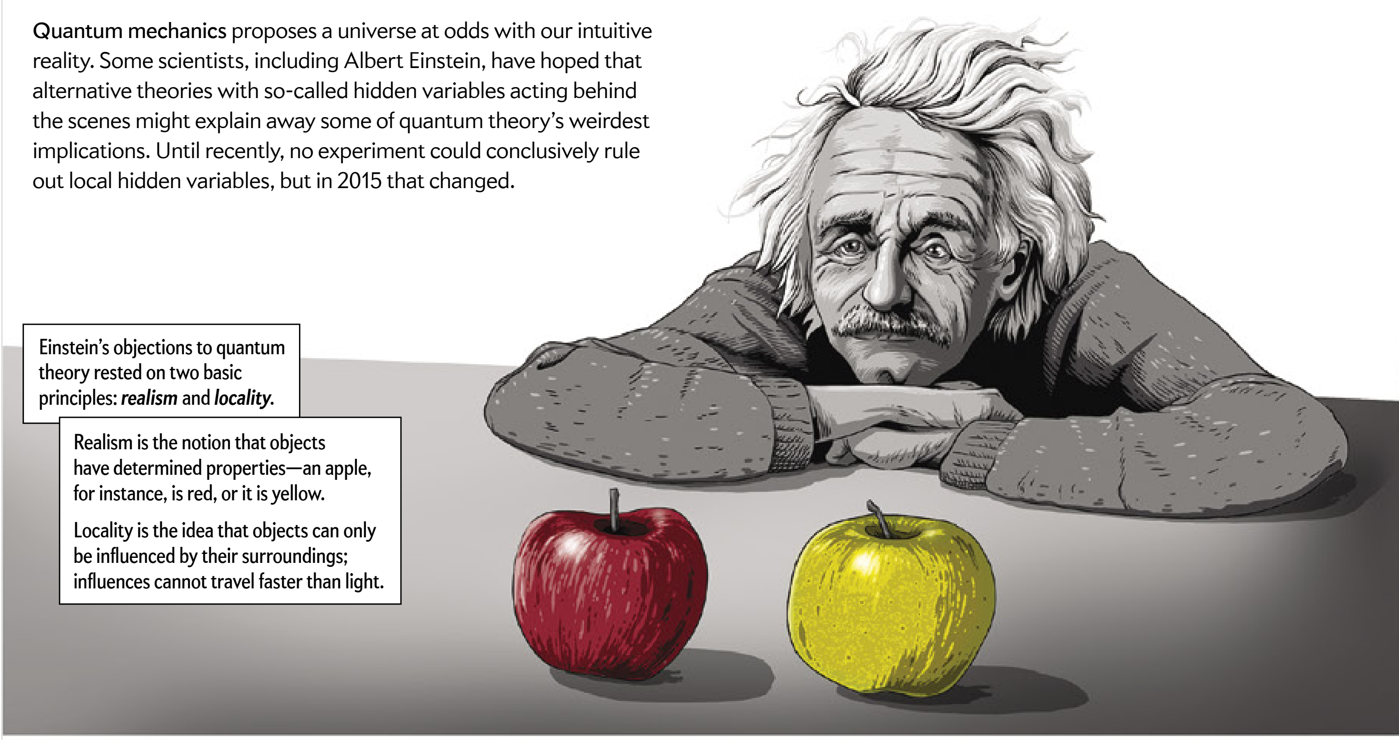
Bell's theorem completely changed physicists' thinking. It showed a mathematical conflict between Einstein's view and quantum theory and outlined a powerful way for experimentally testing the two. Because Bell's theorem is a mathematical inequality that limits how high correlations can be under any local hidden variable theory, experimental data that exceed these bounds—in other words, that "violate" Bell's inequality—will show that local hidden variable theories cannot describe nature.

Soon after Bell's publication, physicists John Clauser, Michael Horne, the late Abner Shimony and Richard Holt (known as CHSH) found similar inequalities that were easier to test in experiments. Researchers performed the first trials in the late 1960s, and since then experiments have come closer and closer to the ideal of Bell's proposed setup. The experiments have found correlations that violate Bell's inequality and seemingly cannot be explained by local hidden variable theories. Until 2015, though, all experiments necessarily relied on one or more additional assumptions because of imperfections in the setups. These assumptions provide loopholes that local hidden variable theories could in principle use to pass the test.

In virtually all such experiments in the 20th century, scientists generated entangled photons at a source and sent them to measurement stations (standing in for Alice and Bob). The Alice and Bob stations each measured their respective photon along one of two orientations, noting its polarization—the

Closing All the Loopholes

Quantum mechanics proposes a universe at odds with our intuitive reality. Some scientists, including Albert Einstein, have hoped that alternative theories with so-called hidden variables acting behind the scenes might explain away some of quantum theory's weirdest implications. Until recently, no experiment could conclusively rule out local hidden variables, but in 2015 that changed.



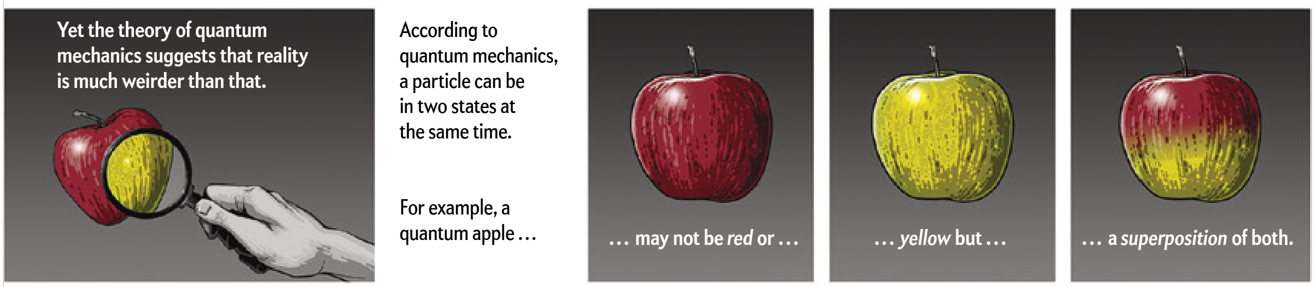
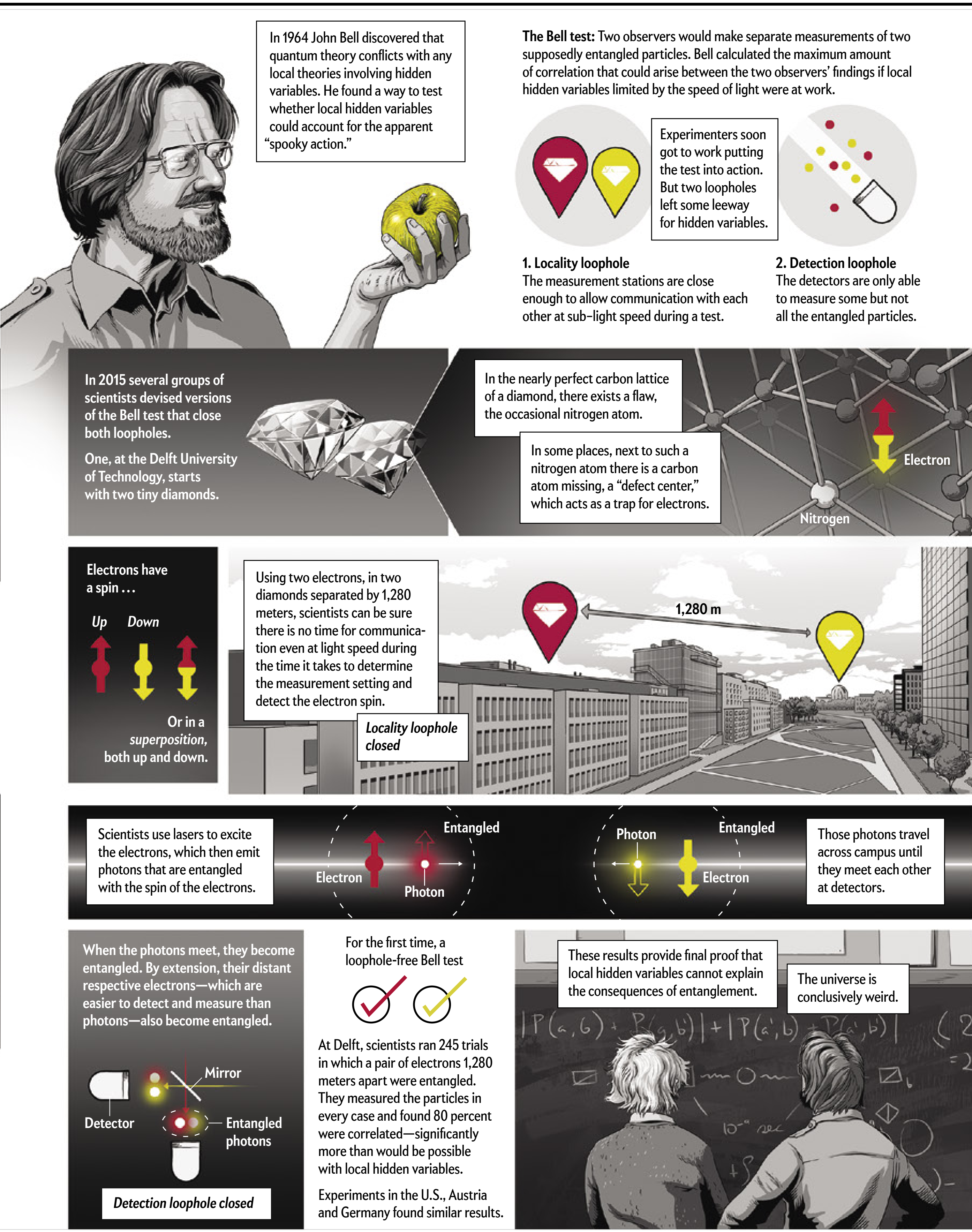


Illustration by Matthew Twombly



SOURCE: MODIFIED FROM *TU DELFT—A LOOPHOLE-FREE BELL TEST.* TEXT BY MICHEL VAN BAAL AND GRAPHICS BY SCIXEL. TU DELFT, 2015 direction in which the photon's electric field oscillates (polarization can be thought of as the spin of a photon). The scientists then calculated the average correlations between the two stations' outcomes and plugged those into Bell's equation to check whether the results violated the inequality.

TESTS WITH CAVEATS

THE FIRST SERIES OF EXPERIMENTS used fixed measurement directions. In these cases, there is ample time for hidden variables (using knowledge of the measurement directions on either side) to influence the outcomes. That is, hidden signals could tell Bob which direction Alice used to measure her photon without traveling faster than light. This so-called locality loophole means that a hidden variable theory could match the quantum correlations. In 1982 French physicist Alain Aspect and co-workers performed a test where the photons were sent to opposite ends of a large room and their polarization measured. While these entangled photons were in flight, the polarization angle of the measurement device changed periodically. In the late 1990s Anton Zeilinger, now at the University of Vienna, and his colleagues further improved this strategy by using truly random (as opposed to periodic) polarization-measurement directions. In addition, these measurement directions were determined very shortly before the measurements took place, so hidden signals would have had to travel faster than light to affect this experiment. The locality loophole was firmly closed.

These experiments had one drawback, however: photons are hard to work with. Most of the time the tests got no answer at all, simply because the photons were not created in the first place or were lost along the way. The experimenters were forced to assume that the trials that worked were representative of the full trial set (the "fair sampling assumption"). If this assumption was dropped, the results would not violate Bell's inequality. It is possible that something different was happening in the trials where photons were lost, and if their data were included, the results would not be in conflict with local hidden variable theories. Scientists were able to close this so-called detection loophole in the past couple of decades by giving up photons and using matter, such as trapped ions, atoms, superconducting circuits and nuclei in diamond atoms, which can all be entangled and measured with high efficiency. The problem is that in these cases the particles were all located extremely close to one another, leaving the locality loophole open. Thus, although these Bell tests were ingenious, they could all, at least in principle, be explained by a local hidden variable theory. A Bell test with all the loopholes closed simultaneously became one of the grandest challenges in quantum science.

Thanks to rapid progress in scientists' ability to control and measure quantum systems, it became possible in 2015, 80 years after the EPR paper and 51 years after Bell's equation, to carry out a Bell test in the ideal setting, often referred to as a loophole-free Bell test. In fact, within a short span of time, four different groups found results that violated Bell's inequality with all loopholes closed—providing ironclad evidence against local hidden variable theories.

CLOSING THE LOOPHOLES

ONE OF US (Hanson) and his collaborators performed the first experiment to close all loopholes at the Delft University of Technology in the Netherlands using a setup [*see box on preceding page*] that closely resembles the original EPR concept. We entangled the spins of two electrons contained inside a diamond, in a space called a defect center, where a carbon atom should have been but was missing. The two entangled electrons were in different laboratories across campus, and to make sure no communication was possible between them, we used a fast random-number generator to pick the direction of measurement. This measurement was finished and locally recorded on a hard drive before any information from the measurement on the other side could have arrived at light speed. A hidden signal telling one measuring station which direction the other had used would not have had time to travel between the labs, so the locality loophole was firmly closed.

These strict timing conditions required us to separate the two electrons by more than a kilometer, about two orders of magnitude farther apart than the previous world record for entangled matter systems. We achieved this separation by using a technique called entanglement swapping, in which we first entangle each electron with a photon. We then send the photons to meet halfway between the two labs on a semitransparent mirror where we have placed detectors on either side. If we detect the photons on different sides of the mirror, then the spins of the electrons entangled with each photon become entangled themselves. In other words, the entanglement between the electrons and the photons is transferred to the two electrons. This process is prone to failure—photons can be lost between the diamonds and the mirror, just as in the earlier photon-based experiments. But we start a Bell trial only if both photons are detected; thus, we deal with photon loss beforehand. In this way, we close the detection loophole because we do not exclude the findings of any Bell test trials from our final results. Although the photon loss related to the large separation in our case does not limit the quality of the entanglement, it does severely restrict the rate at which we can conduct Bell trials—just a few per hour.

After running the experiment nonstop for several weeks in June 2015, we found Bell's inequality was violated by as much as 20 percent, in full agreement with the predictions of quantum theory. The probability that such results could have arisen in any local hidden variable model—even allowing the devices to have maliciously conspired using all available data—was 0.039. A second experimental run conducted in December 2015 found a similar violation of Bell's inequalities.

In the same year, three other groups performed loophole-free Bell tests. In September physicists at the National Institute of Standards and Technology (NIST) and their colleagues, led by one of us (Shalm), used entangled photons, and in the same month Zeilinger's group did so as well. Not too long after, Harald Weinfurter of Ludwig Maximilian University of Munich and his team used rubidium atoms separated by 400 meters in a scheme similar to that of the Hanson group (the results were published in 2017).

Both the NIST and Vienna teams entangled the polarization state of two photons by using intense lasers to excite a special crystalline material. Very rarely, about one in a billion of the laser photons entering the crystal underwent a transformation and split into a pair of daughter photons whose polarization states were entangled. With powerful enough lasers, it was possible to generate tens of thousands of entangled photon pairs per second. We then sent these photons to distant stations (separated by 184 meters in the NIST experiment and 60 meters in the Vienna experiment) where we measured the polarization states. While the photons were in flight toward the measurement stations, our system decided which direction to measure their polarization in such a way that it would be impossible for any hidden variables to influence the results. The locality loophole is therefore closed. The most challenging aspect of using photons is preventing them from being lost, as we must detect more than two thirds of the photons we create in our setup to avoid the detection loophole. Most conventional single-photon detectors operate at around 60 percent efficiency—a nonstarter for this test. But at NIST we developed special single-photon detectors, made of cold superconducting materials, capable of observing more than 90 percent of the photons that reach it. Thus, we closed the detection loophole as well.

Repeating these polarization measurements on many different entangled photon pairs more than 100,000 times per second, we were able to quickly accumulate statistics on the correlations between the photon polarization states. The correlations observed in both experiments were much stronger than those predicted by hidden variable theories. In fact, the probability that the NIST results could have arisen by chance is on the order of one in a billion (even less likely than winning the Powerball lottery), and the chances are even smaller for the Vienna experiment. Today our NIST group regularly uses an improved version of our setup to violate Bell's inequalities to a similar degree in less than a minute, and future improvements will speed this up by two orders of magnitude.

HARNESSING ENTANGLEMENT

THESE EXPERIMENTS FORCE US TO CONCLUDE that any local hidden variable model, such as those Einstein advocated, is incompatible with nature. The correlations between particles we have observed defy our intuition, showing that spooky action does indeed take place.

Our results also hint at the remarkable power contained in entanglement that we may be able to put to use. A near-term application where loophole-free Bell tests can be useful is in generating randomness. Random numbers are a critical resource in many cryptographic and security techniques. If you can predict the next number a random-number generator will produce, you can hack many financial and communications systems. A good source of randomness that cannot be predicted is therefore of vital importance. Two of the most common ways to generate randomness are through mathematical algorithms or using physical processes. With mathematical algorithms, if you know the conditions used as a "seed," you can often predict the output perfectly. With physical processes, a detailed understanding of the underlying physics of the system is required. Miss even a single detail, and a hacker can exploit or control the randomness. The history of cryptography is littered with examples of both types of random-number generators being broken.

Quantum mechanics has handed us a gift, though. It is possible to "extract" the randomness inherent in quantum processes to produce true randomness. The correlations measured in a loophole-free Bell test can be distilled into a certifiably random string. Remarkably, it is possible to hand part of the experimental apparatus (the generation of the entangled particles) to a potential hacker to control. Even in this extreme case, it is possible to produce numbers that are as random as nature allows. In early 2018 our team at NIST was able to use our loophole-free Bell setup to extract 1,024 truly random bits from 10 minutes of experimental data. These bits were certified to be random to better than a part in one trillion. In contrast, it would take a conventional random-number generator several hundred thousand years to acquire enough data to directly measure the quality of their randomness to this level. We are working now to incorporate our random-number generator into a public randomness beacon. This tool could act as a time-stamped source of random numbers that is broadcast over the Internet at fixed intervals and can be used in security applications by anyone who needs it.

On a more general level, the techniques developed in loophole-free Bell experiments may enable fundamentally new types of communications networks. Such networks, often referred to as a quantum Internet, can perform tasks that are out of reach of classical information networks. A quantum Internet could enable secure communication, clock synchronization, quantum sensor networks, as well as secure access to remote quantum computers in the cloud. Another goal is "device-independent cryptography," in which (in close analogy to the randomness beacon) users can validate the secrecy of a shared key through a violation of Bell's inequalities.

The backbone of a future quantum Internet will be formed by entanglement links precisely like the setups used to test Bell's inequalities with diamond defect centers, trapped atoms and photons. In 2017 our team at Delft demonstrated a method to boost the quality of remote entangled spins, and in 2018 we improved the entangling rates by three orders of magnitude. Based on this progress, researchers are working toward a first rudimentary version of a quantum Internet that is scheduled to be realized among a few cities in the Netherlands in 2020.

Eight decades ago when quantum theory was being written, skeptics chafed at its apparent contradiction with the centuries of physical intuition that had been developed; now four experiments have dealt the final blow to that intuition. At the same time, these results have opened the door to exploit nature in ways that Einstein and Bell could not have foreseen. The quiet revolution that John Bell kicked off is now in full swing. SA

Ronald Hanson is a physicist at the Delft University of Technology and scientific director of its QuTech research center, a collaboration with the Netherlands Organization for Applied Scientific Research (TNO), focused on quantum computing and quantum Internet technology.

Krister Shalm is a physicist at the National Institute of Standards and Technology and the University of Colorado Boulder, where he develops tools to test foundational issues in quantum mechanics.

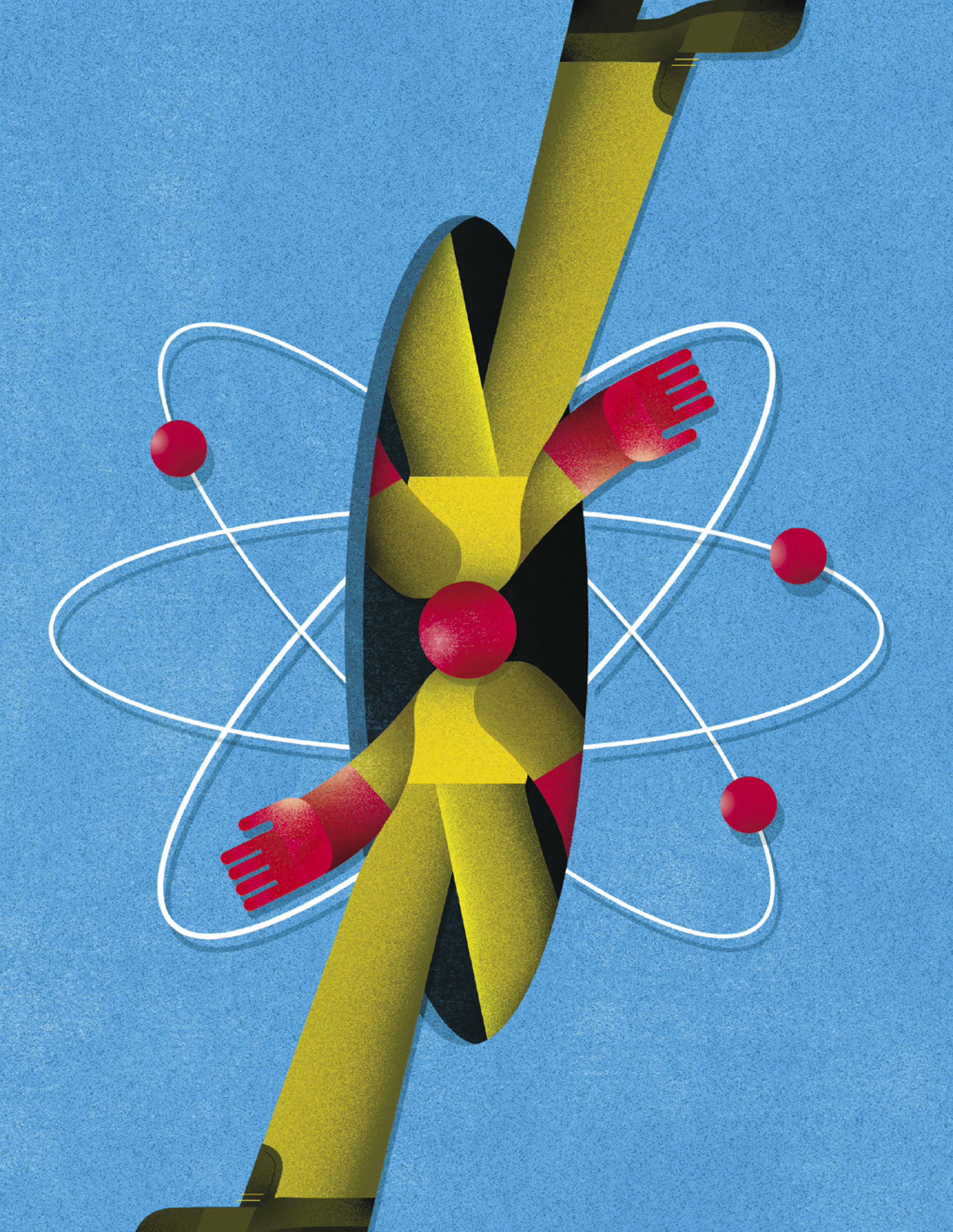
MORE TO EXPLORE

Loophole-Free Bell Inequality Violation Using Electron Spins Separated by 1.3 Kilometres. B. Hensen et al. in *Nature,* Vol. 526; pages 682–686; October 29, 2015.

Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons. Marissa Giustina et al. in *Physical Review Letters,* Vol. 115, Article No. 250401. Published online December 16, 2015.

Strong Loophole-Free Test of Local Realism. Lynden K. Shalm et al. in *Physical Review Letters,* Vol. 115, Article No. 250402. Published online December 16, 2015.

scientificamerican.com/magazine/sa



CROSSING THE QUANTUM DIVIDE

The universe according to quantum mechanics is strange and probabilistic, but our everyday reality seems nailed down. New experiments aim to probe where—and why—one realm passes into the other

By Tim Folger

Illustration by Maria Corte

MOST OF SIMON GRÖBLACHER'S HANDIWORK IS INVISIBLE TO THE NAKED EYE. One of the mechanical devices he fashioned in his laboratory at the Delft University of Technology in the Netherlands is just a few millionths of a meter long—not much bigger than a bacterium—and 250 nanometers thick—about a thousandth of the thickness of a sheet of paper. Gröblacher no doubt could continue to shrink his designs, but he has a different goal: he wants to scale things *up,* not down. "What we're trying to do is make things that are really, really big," he says as he brings up images of hardware on his computer. Keep in mind that for Gröblacher, an experimental physicist, "really, really big" means something just barely visible without a microscope, "a millimeter by a millimeter in size."

IN BRIEF

The microscopic and macroscopic worlds do not blend seamlessly: the probabilistic nature of quantum mechanics reigns over the first, whereas the second observes more logical "classical" rules.

Physicists have long been stymied over the question of where one realm ends and the other begins, but upcoming experiments offer hope of testing different theories.

One possibility, called continuous spontaneous localization, suggests that quantum probabilities randomly collapse into classical certainties. If true, these collapses would also create a sea of background vibrations in the universe that experiments could detect.

By working on that less than humongous scale, Gröblacher hopes to address an extraordinary question: Can a single macroscopic object be in two places at once? Could something the size of a pinhead, say, exist both here and there at the same time? That seemingly impossible condition is actually the norm for atoms, photons and all other particles. According to the surreal laws of quantum theory, reality at its most basic level defies our commonsense assumptions: Particles do not have fixed positions, energies or any other definite properties—at least while no one is looking. They exist in numerous states simultaneously.

But for reasons physicists do not understand, the reality we see is different. Our world—even the parts we cannot observe directly—appears to be distinctly *un*quantum. Really big things—meaning anything from a virus on up—always manifest in one place and one place only; there is just one Gröblacher talking to one jet-lagged, scribbling journalist in his Delft lab. And therein lies a mystery: Why, if everything is built on a quantum blur of matter and energy, do we not experience quantum weirdness ourselves? Where does the quantum world end and the so-called classical world of Newtonian physics begin? Is there a rift in reality, a scale beyond which quantum effects simply cease? Or does quantum mechanics reign everywhere, and we are somehow blind to it?

"We know the microworld is quantum, and we know in one way or another, we are classical—whatever that means," says Angelo Bassi, a theoretical physicist at the University of Trieste in Italy. "We are ignorant about the true nature of matter in between the micro and the macro." That no-man's-land has baffled physicists since the birth of quantum theory a century ago. But in recent years Gröblacher and other physicists have started running exquisitely sensitive tabletop experiments that may one day reveal how objects make the startling transition from quantum to quotidian. Whether those efforts will resolve the mysteries of quantum theory or deepen them, no one can yet say. But in probing the wild and woolly quantum borderlands, researchers stand a chance of discovering a whole new realm of physics.

THE MEASUREMENT PROBLEM

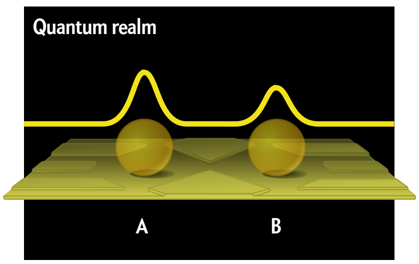
FOR ALL ITS PARADOXES, quantum mechanics is the most powerful and exacting scientific theory ever devised. The theory's predictions match experiment with ridiculous precision—to better than parts-per-trillion accuracy in some cases. By revolutionizing our understanding of atomic structure, it transformed every facet of science, from biology to astrophysics. Without quantum theory, there would be no electronics industry, no cell phones, no Google. Yet the theory has one glaring shortcoming, says Stephen L. Adler, a theoretical physicist at the Institute for Advanced Study in Princeton, N.J.: "In quantum mechanics, things don't happen."

Separate Realms

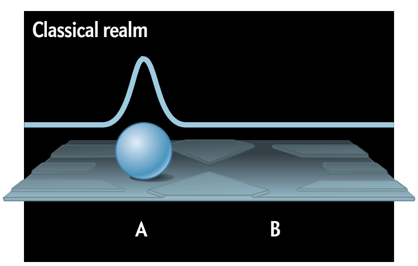
Quantum mechanics produces some bizarre effects in the microscopic world, but we do not see these phenomena in our macroscopic, "classical" reality. Why is that? Scientists have never understood why and how the universe crosses over between these realms, but several theories, as depicted here, offer possible explanations.

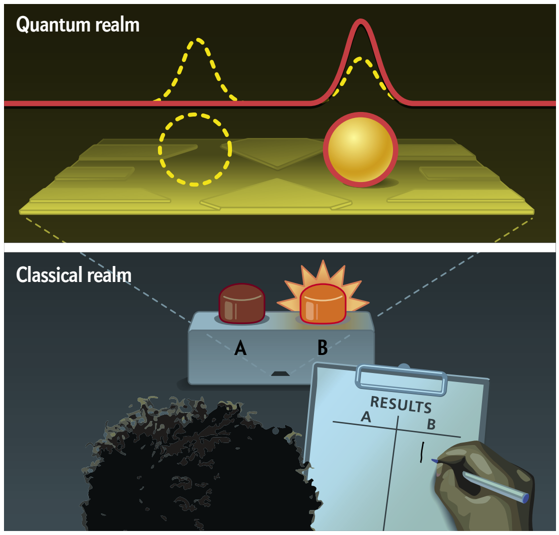
Quantum vs. Classical

According to quantum mechanics, particles do not exist in definite states—here or there, having this energy or that—but rather take on all possible states and positions. The theory describes particles with equations called wave functions, which are combinations, or "superpositions," of multiple waves. The amplitude of each peak in a wave function denotes the probability of a particle being found in any specific circumstances—for instance, at point A or B, as shown.



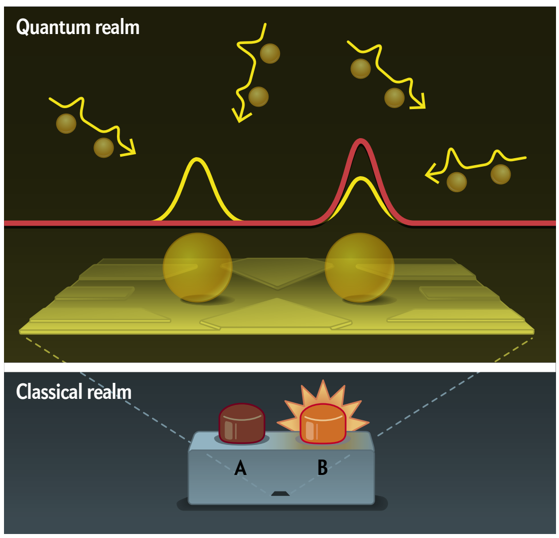
Strangely, when scientists make a measurement of a particle, this act appears to reduce all the quantum possibilities to one, seemingly chosen at random. The experiment will find the particle at point A, for example, and the particle enters the classical realm, ceasing to be in a superposition.





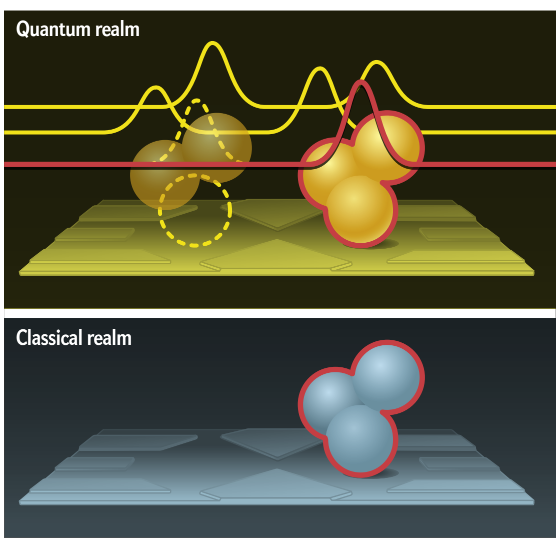
Collapse at Measurement

One theory for how the universe crosses over from quantum to classical is that the act of measurement intervenes. Particles can linger in quantum superpositions (*dotted yellow lines*) as long as no one looks too closely, but once humans make a measurement, the particle is forced to "choose" a specific state (*solid red lines*). How this happens, and why human measurement should take on such a significance in physics, remains mystifying.



Decoherence

Another theory posits that a particle's environment is to blame for moving it from the quantum world to the classical. As long as a particle is undisturbed by any outside influence, so the thinking goes, it can remain in superposition. But when the wave functions of other particles or objects nearby meet with its own, they interfere, causing the particle's many quantum possibilities to "collapse" into a single classical reality.



Continuous Spontaneous Localization

Another possibility is that the collapse of the wave function to a single possibility is a random event, not caused by human or environmental interference. The chances of any one particle collapsing at any given time are extremely small, but in macroscopic objects containing multitudes of atoms, the collapse of at least one is inevitable, which then causes the entire structure to collapse.

Illustrations by Jen Christiansen

Adler's cryptic comment refers to what the basic equations of quantum theory say—or do not say—about the nature of reality. Known as wave functions, the equations assign probabilities to an object's chances of being found in various states. Unlike Newtonian physics, where apples, planets and everything else always have well-defined properties, quantum physics is inherently probabilistic. In a sense, particles that are described by wave functions cannot even be said to fully exist; they have no fixed locations, speeds or energies—they have only probabilities.

But everything changes when scientists take a measurement. Then real, tangible properties arise, as if conjured up by merely attempting to observe them. Not only does the theory not say *why* measurements bring about this transformation—it does not tell us why one of those many possibilities manifests instead of others. Quantum mechanics describes what *might* happen as the outcome of a measurement but not what *will* happen. In other words, the theory provides no mechanism for the transition from the probable to the actual.

To "make things happen" in quantum mechanics, one of the theory's legendary founders argued for an almost metaphysical hack. In the late 1920s Werner Heisenberg formulated and spread the notion that the very act of measurement makes the wave function of a particle "collapse"—the many potential outcomes instantaneously reduce to a single observed result. The only flaw with the idea is that there is nothing in the equations of quantum theory that says a collapse occurs or offers a physical process to explain it. Heisenberg's "solution" essentially introduced a new mystery into physics: What exactly happens when a wave function collapses? That quantum conundrum is now known as the measurement problem.

Physicists may have gotten used to the collapse idea over the past 90 years, but they have never really liked it. The notion that a human action—measurement—plays a central role in our most fundamental theory of how the universe works does not sit well with anyone partial to the concept of an objective reality.

"Fundamentally, I have an ideal of what a physical theory should be," says Nobel laureate physicist Steven Weinberg of the University of Texas at Austin. "It should be something that doesn't refer in any specific way to human beings. It should be something from which everything else—including anything you can say systematically about chemistry, or biology, or human affairs—can be derived. It shouldn't have human beings at the beginning in the laws of nature. And yet I don't see any way of formulating quantum mechanics without an interpretive postulate that refers to what happens when people choose to measure one thing or another thing."

CHOOSE YOUR INTERPRETATION

ONE SLEIGHT-OF-HAND WAY OUT of the measurement problem is to assume that collapse simply does not happen. In the early 1970s the late H. Dieter Zeh, then at the University of Heidelberg in Germany, proposed a process that yields the *appearance* of collapse while preserving the full quantum multiplicity of the wave function. In the real world, Zeh argued, the wave function of any particular object becomes hopelessly enmeshed with that of everything else in its environment, making it impossible to keep track of all the countless quantum interactions going on around us. In quantum parlance, the wave functions become "entangled"—a special kind of correlation that preserves connectedness even over huge distances. An observer can only ever hope to look at a single small part of that vast entangled system, so any particular measurement captures just a sliver of the quantum world.

Zeh called this process "decoherence," and it has become the go-to explanation among physicists for why we do not witness quantum phenomena on a macroscopic level. It describes how an intact wave function—which comprises all the possible physical states a particle might have—decoheres as it mingles with the wave functions of other quantum systems around it. If the decoherence model is right, we ourselves live among the strands of that entangled quantum web but see only part of it.

Not all physicists believe that decoherence settles the measurement problem. For one thing, it still fails to explain why we see one strand of the quantum web and not others. "You still have to invoke the collapse postulate, which takes an entangled state and says that one of those possible states has to be selected, and that is usually done by fiat," says Miles P. Blencowe, a theoretical physicist at Dartmouth College. For Blencowe and others, the process does not capture the way we experience things. "I believe we have this one world that is evolving," he says. "How do you go from an entangled state to this perception of the world as always finding one unique path into the future? Many quantum mechanicians would feel that there needs to be a collapse to restore this oneness about the world as it evolves rather than this web of entanglement that keeps enlarging." Adler's assessment of decoherence is more blunt: "It doesn't supply a mechanism [for collapse] at all. It doesn't solve the problem, period."

Six decades ago a doctoral candidate at Princeton University proposed an even more radical solution to the collapse problem. In his 1957 Ph.D. thesis, Hugh Everett argued that the wave function neither collapses nor decoheres. Rather all its components are physically real, parts of an endlessly branching panoply of universes. Everett's "many worlds" interpretation, as it is called, has become popular among cosmologists, who have other reasons to think we might inhabit a multiverse. But no one has ever managed to experimentally distinguish the many-worlds idea from standard quantum theory.

The same holds for other interpretations of quantum mechanics. French physicist Louis de Broglie, one of the founders of quantum theory, sought to eliminate the need for collapse by introducing the notion of "pilot waves" that guide the paths of electrons and all other particles. In de Broglie's version of quantum theory, which American physicist David Bohm further developed in the 1950s, there is no mysterious collapse; measurements simply show the interactions of pilot waves and their associated particles. But again, no one has yet found experimental evidence that distinguishes de Broglie and Bohm's pilot-wave view of reality from Everett's many worlds or any of the other dozen or so different takes on quantum mechanics. In the end, quantum partisans choose their favorite description of reality based on aesthetics. "I still come back to the fact that we have this one world that is evolving," Blencowe says. "For that, one really needs some sort of collapse, which is more than just a rule for the results of experiments but is some actual process."

TESTING COLLAPSE

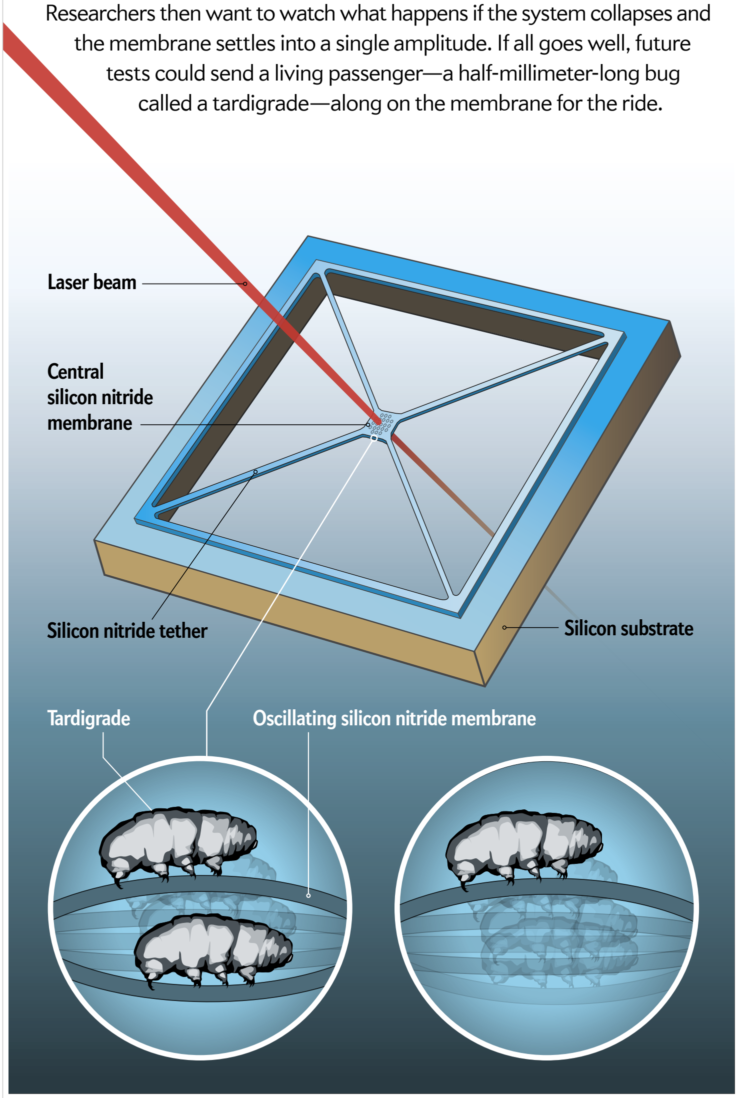
THE CITY OF DELFT might qualify as an entangled quantum system. Its placid canals and medieval brick buildings overlap in space and time with cars, bicyclists, cell-phone shops and students staggering home from all-night parties along the same narrow streets painter Johannes Vermeer once walked. Gröblacher's lab lies about two kilometers south of the old town center and what feels like several hundred years into the future. On a warm spring morning, he shows a visitor one of the "really, really big" things he and his colleagues have built: a millimeter-size membrane tethered to a silicon chip, just barely visible to the naked eye.

Seen up close (or blown up on a poster in the hallway outside Gröblacher's office), the membrane resembles a minuscule trampoline. It is made of silicon nitride, a durable ceramic material that was used for engine bearings in the space shuttles, and holds a highly reflective mirror at its center. A single jolt from a component on the chip can set the membrane vibrating for minutes at a time. Such membranes are "really good oscillators," Gröblacher says. "To put that in perspective, it would be like pushing someone on a swing, and the person would go back and forth, with one single push, for 10 years." Despite its Lilliputian dimensions, the membrane is extraordinarily robust. "We really put a lot of stress in it—six gigapascals" says Richard Norte, one of Gröblacher's collaborators. "It's about 10,000 times the stress you'd have in a bicycle tire, in something that's only about eight times thicker than the width of DNA."

Those robust qualities make the membrane an ideal place to study quantum phenomena—it vibrates reliably at room temperature without breaking down.

Tabletop Test

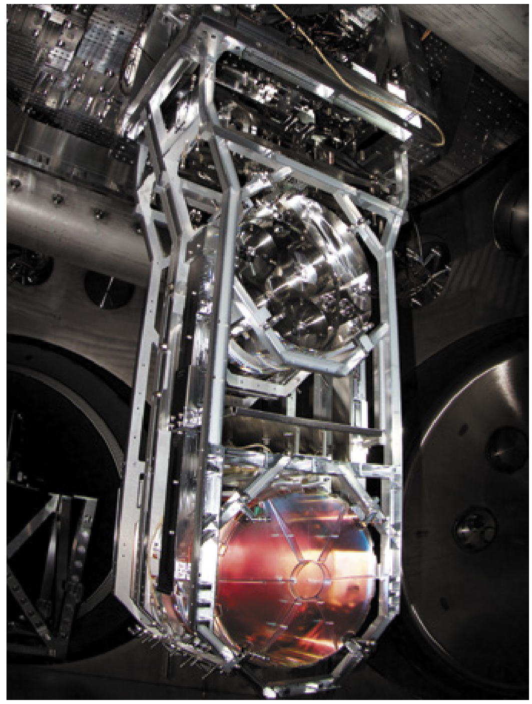
Physicists want to see if macroscopic objects can behave in quantum ways. One planned experiment will feature a millimeter-size membrane that looks like a tiny trampoline. Attached to a silicon chip, the membrane can be jolted into long-lasting vibrations. Ultimately scientists plan to use a laser to push the membrane into a quantum superposition. In the experiment, the membrane could be vibrating at two different amplitudes at once. Researchers then want to watch what happens if the system collapses and the membrane settles into a single amplitude. If all goes well, future tests could send a living passenger—a half-millimeter-long bug called a tardigrade—along on the membrane for the ride.



Gröblacher and Norte plan to eventually use a laser to nudge the membrane into a superposition—a quantum state where the membrane could be simultaneously oscillating at two different amplitudes. The membrane's ability to wiggle for minutes on end should, in principle, allow such quantum states to persist long enough to see what happens when—or if—the membrane collapses into a single classical state.

"That is exactly what you need to create some sort of quantumness," Gröblacher says. "You don't want to have it interact with its environment, because that induces decoherence—supposedly. So you want a really well-isolated system, get it in a quantum state, then switch on your own decoherence, something you control—a laser. We're still not at the point where we can actually create a superposition of the oscillations of the system. But in a few years that's what we're aiming for."

And Gröblacher and his colleagues do not plan to stop there. The researchers hope to ultimately place a living creature on the membrane and then put the membrane, and any passengers on it, into a quantum superposition. Leading candidates for that mission into quantum space are eight-legged microorganisms called tardigrades, also known as water bears. "They're amazing creatures," Gröblacher says. "You can cool them down—they're still alive; you can heat them up—they're still alive; put them in a vacuum—they're still alive." He admits this step is a bit of a ways off. "It's not crazy. It's nice as a long-term goal, but first we have to get our devices into superposition, then we can think about putting in a living organism."



MIRRORS at LIGO showed no evidence of having been nudged by quantum jiggles predicted by CSL theory.

LIGO LABORATORY *California Institute of Technology and Massachusetts Institute of Technology*

CONTINUOUS SPONTANEOUS LOCALIZATION

WITH OR WITHOUT TARDIGRADES, such an experiment would allow physicists to test whether nature somehow censors quantum effects above a certain size scale. Some physicists have proposed that collapse might be an actual physical phenomenon, with measurable effects. One idea—known as continuous spontaneous localization, or CSL—is that wave function collapse is simply a random event occurring constantly in the microscopic world. According to CSL, the chance that any one particle will collapse is extremely rare—it might happen once in hundreds of millions of years—but for large aggregates of particles, collapse becomes a certainty.

"A single proton has to wait about 10^{16} seconds to see a collapse, so it happens only a few times over the age of the universe," Bassi says. But the huge number of particles in any macroscopic object makes collapse inevitable. "If you take a table, which contains roughly Avogadro's number of particles—10^{24}—the collapse occurs almost immediately." If CSL is real, measurement and observation have no role in collapse. In any measurement, a given particle and the devices recording it become part of an immense quantum array that very rapidly collapses. Although it seems as if the particle went from a superposition to an actual position during a measurement, this transformation happened as soon as the particle interacted with the devices, before the measurement occurred.

If collapse turns out to be a real physical phenomenon, the practical consequences could be significant. For one thing, it might limit the nascent technology of quantum computers. "Ideally, you would like to make bigger and bigger quantum computers," Bassi says. "But you would not be able to run quantum algorithms, because the collapse would kill everything." For decades most physicists have regarded collapse as an essentially untestable aspect of quantum theory. But CSL and other collapse models have changed that. The CSL model, for example, predicts that the action of collapse imparts a slight jiggle to particles, creating an omnipresent background vibration that might be detectable in experiments. "The collapse [in CSL] is something universal for micro and macro systems," Bassi says. "Every time there is a collapse, you move the particle a little." He and other physicists have searched for such evidence in surprising places. They have combed through the calibration data for the Laser Interferometer Gravitational-wave Observatory (LIGO), an instrument capable of registering motions 10,000 times as small as the width of a proton.

In February 2016 LIGO reported detecting a gravitational wave for the first time. The wave—a ripple in spacetime caused by two distant colliding black holes—stretched and squeezed the space between two mirrors at the experiment's twin sites in Washington State and Louisiana. This passing wave shifted the positions of LIGO's mirrors by just four-thousandths the diameter of a proton, in perfect agreement with predictions by Einstein's general theory of relativity. But Bassi and his colleagues found no evidence in LIGO's data for any additional motion caused by the kind of quantum nudges predicted by CSL. The result did not surprise them. If quantum collapse is an actual physical phenomenon, it is an extraordinarily weak one. The question was: How weak? Now they have put extremely precise bounds on the effect. "If you apply the model to the mirror at LIGO, the mirror should move more than expected, but the mirror doesn't move much. Therefore, the collapse noise can't be too strong," Bassi says.

Physicists have also hunted for signs of collapse in experiments designed to look for dark matter—hypothetical particles thought to account for up to 85 percent of the matter in the universe. One such experiment, sheltered in the Spanish Pyrenees, uses germanium detectors to search for signs of dark matter particles zipping through and generating a flash of x-rays. A collapsing wave function should likewise create a flash, but experimenters have seen no such emissions.

These types of experiments have tightened the constraints on collapse models considerably but not fatally. In September 2017 Andrea Vinante, a physicist at the University of Southampton in England, along with Bassi and three colleagues, reported the discovery of tentative evidence in support of the CSL model. Vinante's team constructed a miniature cantilever (a horizontal beam fixed at one end), just half a millimeter long and two microns thick and tipped with a small magnet. The researchers carefully shielded the setup from any external vibrations and cooled the cantilever to 40 thousandths of a kelvin above absolute zero to eliminate any possibility of thermally induced movements.

Under those conditions the cantilever should have vibrated ever so slightly because of thermal motion of its particles. But the actual wobble was greater than this predictable motion. The experiment's motion detector—an extremely sensitive instrument called a superconducting quantum-interference device, or SQUID—found that the cantilever and its magnet vibrated like a diving board, bending up and down by a few trillionths of a meter. Twelve years ago Adler calculated that collapsing wave functions might produce vibrations of approximately that size.

"We could see some unexplained noise," says Vinante, describing his experimental results. "It's something that is consistent with what we expect from collapse models, but it could be from an effect we have not understood completely." He and his colleagues are working on upgrades to improve the experiment's sensitivity by at least a factor of 10 and perhaps a factor of 100. "We should be able to either confirm that there is something anomalous or rule out that what we observed was anything interesting." Vinante says it might take another year or two before they have new data. Given the century-long track record of quantum theory's dominance, the odds of discovering a deviation are slim.

But what if one of these experiments does pan out and confirms the phenomenon of quantum collapse? Would that mean an end to the mysteries and paradoxes of the theory? "If collapse really existed, it would divide the world into different scales," says Igor Pikovski, a theoretical physicist at the Harvard-Smithsonian Center for Astrophysics. "Above a certain scale quantum mechanics would cease to be the correct theory. But below that scale everything we know about quantum mechanics would still hold. So the same philosophical questions and interpretations that bug us would still hold for the lower scale. You'd still have many worlds for electrons or atoms—but not for the moon! So it doesn't solve some of the problems—I think it makes it more strange."

Models such as CSL are just preliminary efforts to unify those two realms. Although they are not full-fledged theories yet, they may eventually help physicists develop a more comprehensive model of reality than quantum mechanics now provides. "My own belief is that you need some modification of quantum mechanics," Adler says. "I don't see why that is a problem. Newtonian mechanics was believed to be exact for 200 years, and it's not. Most theories have a domain in which they work, and then there's a domain beyond which they don't work and where a broader theory is needed."

But for now, at least, quantum mechanics largely seems to withstand every test. "No, we're not facing any crisis. That's the problem!" Weinberg says. "In the past, we made progress when existing theories ran into difficulties. There's nothing like that with quantum mechanics. It's not in conflict with observation at all. It's a problem of failing to satisfy the reactionary philosophical preconceptions of people like me."

Yet for all the weirdness of quantum mechanics, most scientists are happy to leave it be. They carry on using the theory to operate their atom smashers and dark matter detectors and rarely stop to ponder what quantum mechanics says—or does not say—about the fundamental nature of reality. "I think most physicists have what seems to me a very healthy attitude," Weinberg says, "to go on using it, to try to push forward the frontiers of our knowledge and leave the philosophical questions for a future generation." More than a few, though, are not willing to wait that long. "Some people will tell you quantum mechanics has taught us that the world is strange, so we have to accept it," Bassi says. "I would say no. If something is strange, then we have to understand better." SA

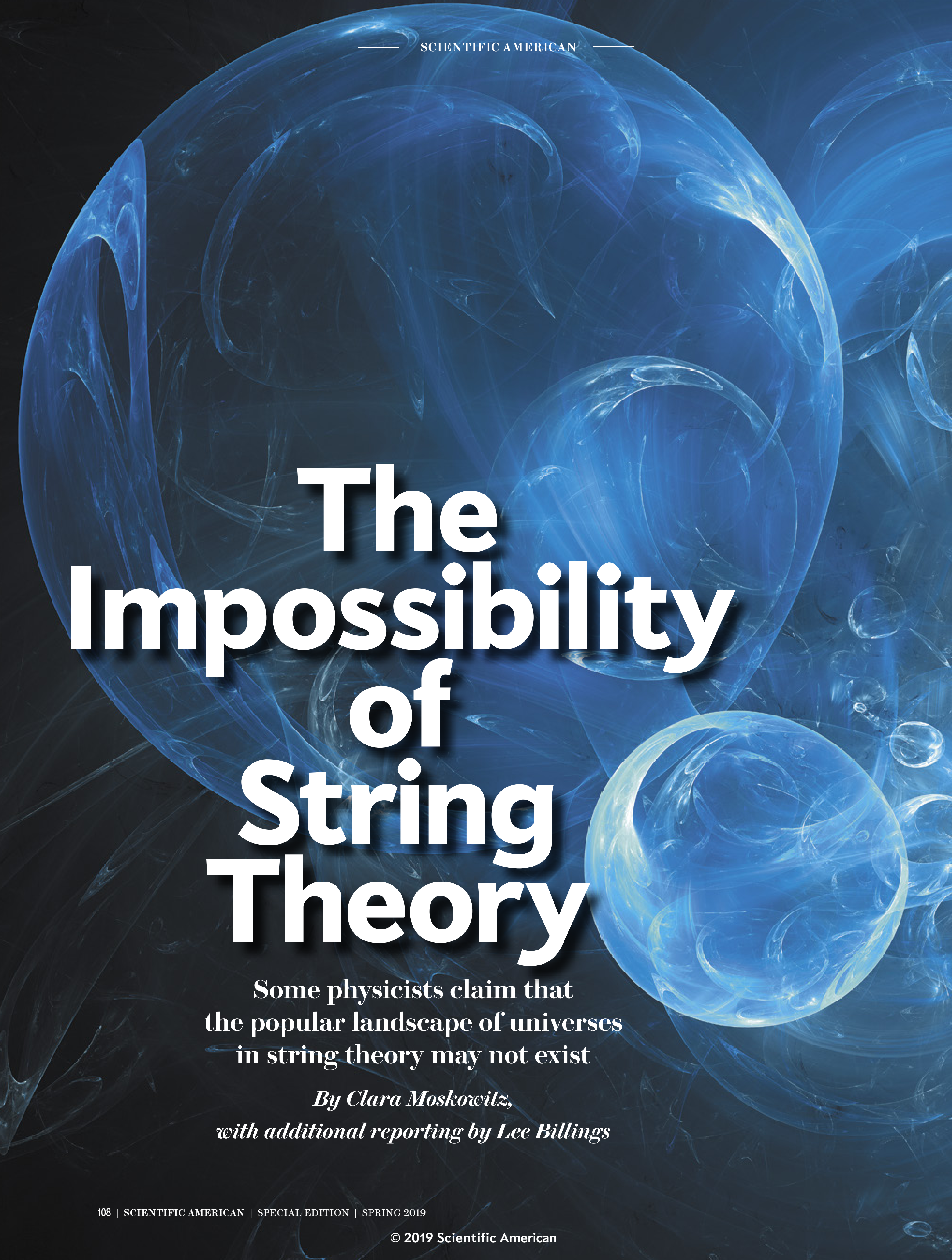
Tim Folger is a freelance journalist who writes for *National Geographic*, *Discover* and other national publications. He is also the series editor for *The Best American Science and Nature Writing*, an annual anthology published by Houghton Mifflin Harcourt.

MORE TO EXPLORE

Mechanical Resonators for Quantum Optomechanics Experiments at Room Temperature. R. A. Norte et al. in *Physical Review Letters*, Vol. 116, No. 14, Article No. 147202; April 8, 2016. Preprint available at https://arxiv.org/abs/1511.06235

Improved Noninterferometric Test of Collapse Models Using Ultracold Cantilevers. A. Vinante et al. in *Physical Review Letters*, Vol. 119, No. 11, Article No. 110401; September 15, 2017. Preprint available at https://arxiv.org/abs/1611.09776

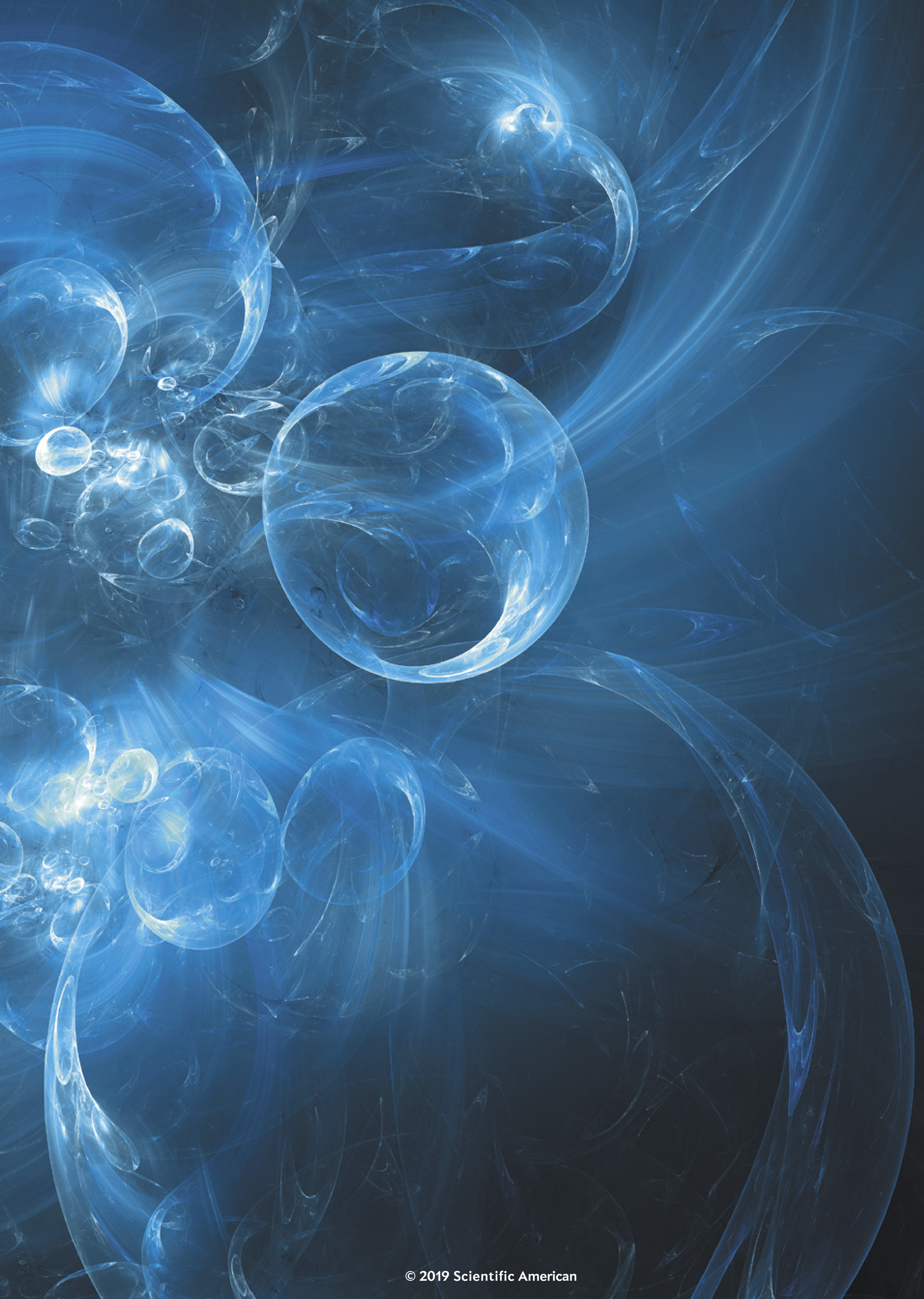
scientificamerican.com/magazine/sa



The Impossibility of String Theory

Some physicists claim that the popular landscape of universes in string theory may not exist

By Clara Moskowitz,
with additional reporting by Lee Billings



The problem with string theory, according to some physicists, is that it makes too many universes. It predicts not one but some 10,500 versions of spacetime, each with its own laws of physics. But with so many universes on the table, how can the theory explain why ours has the features it does?

Now some theorists suggest most—if not all—of those universes are actually forbidden, at least if we want them to have stable dark energy, the supposed force accelerating the expansion of the cosmos. To some, eliminating so many possible universes is not a drawback but a major step forward for string theory, offering new hope of making testable predictions. But others say the multiverse is here to stay, and the proposed problem with all those universes is not a problem at all.

The debate was a hot topic in the summer of 2018 in Japan, where string theorists convened for the conference Strings 2018. "This is really something new and it's led to a controversy within the field," says Ulf Danielsson, a physicist at Uppsala University in Sweden. The conversation centers on a pair of papers posted on the preprint server arXiv in June 2018 taking aim at the so-called landscape of string theory—the incomprehensible number of potential universes that result from the many different solutions to string theory's equations that produce the ingredients of our own cosmos, including dark energy. But the vast majority of the solutions found so far are mathematically inconsistent, the papers contend, putting them not in the landscape but in the so-called swampland of universes that cannot actually exist. Scientists have known many solutions must fall in this swampland for years, but the idea that most, or maybe all, of the landscape solutions might live there would be a major change. In fact, it may be theoretically impossible to find a valid solution to string theory that includes stable dark energy, says Cumrun Vafa, a Harvard University physicist who led the work on the two papers.

LOST IN THE MULTIVERSE

STRING THEORY is an attempt to describe the whole universe under a single "theory of everything" by adding extra dimensions of spacetime and thinking of particles as minuscule vibrating loops. Many string theorists contend it is still the most promising direction for pursuing Albert Einstein's dream of uniting his general theory of relativity with the conflicting microscopic world of quantum mechanics. Yet the notion of a string theory landscape that predicts not just one universe but many has put some physicists off. "If it's really the landscape, in my view it's death for the theory because it loses all predictive value," says Princeton University physicist Paul Steinhardt, who collaborated on one of the recent papers. "Literally anything is possible." To Steinhardt and others, the newfound problems with dark energy offer string theory a way out. "This picture with a big multiverse could be mathematically wrong," Danielsson says. "Paradoxically this makes things much more interesting because that means string theory is much more predictive than we thought it was."

Some string theorists such as Savdeep Sethi of the University of Chicago welcome the reevaluation that is happening now. "I think this is exciting," he says. "I've been a skeptic of the landscape for a long time. I'm really happy to see the paradigm shift away from this belief that we have this proven set of solutions." But not everyone buys the argument that the landscape actually belongs in the swampland—especially the research team that established one of the earliest versions of the landscape in the first place back in 2003, which goes by the acronym KKLT after the scientists' last names. "I think it's very healthy to make these conjectures and check what other things could be going on, but I don't see either theoretical or experimental reasons to take such a conjecture very seriously," says KKLT member Shamit Kachru of Stanford University. And Eva Silverstein, a Stanford physicist who also helped build the early landscape models, likewise doubts Vafa and his colleagues' argument. "I think the ingredients KKLT use and the way they put them together is perfectly valid," she says. Juan Maldacena, a theorist at the Institute for Advanced Study, says he also still supports the idea of string theory universes with stable dark energy.

And many theorists are perfectly happy with the string theory multiverse. "It is true that if this landscape picture is correct, the bit of the universe we're in compared to the multiverse will be like our solar system within the universe," Kachru says. And that is a good thing, he adds. Johannes Kepler originally sought a fundamental reason for why Earth lies the distance it does from the sun. But now we know the sun is just one of billions of stars in the galaxy, each with its own planets, and the Earth-sun distance is simply a random number rather than a result of some

PRECEDING PAGES: GETTY IMAGES

deep mathematical principle. Likewise, if the universe is one of trillions within the multiverse, the particular parameters of our cosmos are similarly random. The fact that these numbers seem perfectly fine-tuned to create a habitable universe is a selection effect—humans will of course find themselves in one of the rare corners of the multiverse where it is possible for them to have evolved.

THE ACCELERATING UNIVERSE

IF IT IS TRUE that string theory cannot accommodate stable dark energy, that may be a reason to doubt string theory. But to Vafa, it is a reason to doubt dark energy—that is, dark energy in its most popular form, called a cosmological constant. The idea originated in 1917 with Einstein and was revived in 1998 when astronomers discovered that not only is spacetime expanding—the rate of that expansion is picking up. The cosmological constant would be a form of energy in the vacuum of space that never changes and counteracts the inward pull of gravity. But it is not the only possible explanation for the accelerating universe. An alternative is "quintessence," a field pervading spacetime that can evolve. "Regardless of whether one can realize a stable dark energy in string theory or not, it turns out that the idea of having dark energy changing over time is actually more natural in string theory," Vafa says. "If this is the case, then one can measure this sliding of dark energy by astrophysical observations currently taking place."

So far all astrophysical evidence supports the cosmological constant idea, but there is some wiggle room in the measurements. Upcoming experiments such as Europe's Euclid space telescope, NASA's Wide-Field Infrared Survey Telescope (WFIRST) and Chile's Simons Observatory being built in the desert will look for signs that dark energy was stronger or weaker in the past than the present. "The interesting thing is that we're already at a sensitivity level to begin to put pressure on [the cosmological constant theory]." Steinhardt says. "We don't have to wait for new technology to be in the game. We're in the game now." And even skeptics of Vafa's proposal support the idea of considering alternatives to the cosmological constant. "I actually agree that [a changing dark energy field] is a simplifying method for constructing accelerated expansion," Silverstein says. "But I don't think there's any justification for making observational predictions about the dark energy at this point."

Quintessence is not the only other option. In the wake of Vafa's papers, Danielsson and his colleagues proposed another way of fitting dark energy into string theory. In their vision, our universe is the three-dimensional surface of a bubble expanding within a larger-dimensional space. "The physics within this surface can mimic the physics of a cosmological constant," Danielsson says. "This is a different way of realizing dark energy compared to what we've been thinking so far."

A BEAUTIFUL THEORY

ULTIMATELY THE DEBATE going on in string theory centers on a deep question: What is the point of physics? Should a good theory be able to explain the particular characteristics of the universe around us or is that asking too much? And when a theory conflicts with the way we think our universe works, do we abandon the theory or the things we think we know?

String theory is incredibly appealing to many scientists because it is "beautiful"—its equations are satisfying and its proposed explanations elegant. But so far it lacks any experimental evidence supporting it—and even worse, any reasonable prospects for gathering such evidence. Yet even the suggestion that string theory may not be able to accommodate the kind of dark energy we see in the cosmos around us does not dissuade some. "String theory is so rich and beautiful and so correct in almost all the things that it's taught us that it's hard to believe that the mistake is in string theory and not in us," Sethi says. But perhaps chasing after beauty is not a good way to find the right theory of the universe. "Mathematics is full of amazing and beautiful things, and most of them do not describe the world," physicist Sabine Hossenfelder of the Frankfurt Institute for Advanced Studies wrote in her recent book, *Lost in Math: How Beauty Leads Physics Astray* (Basic Books, 2018).

Despite the divergence of opinions, physicists are a friendly bunch and are united by their common goal of understanding the universe. Kachru, one of the founders of the landscape idea, worked with Vafa, the landscape's critic, as his undergraduate adviser—and the two are still friends. "He asked me once if I'd bet my life these [landscape solutions] exist," Kachru says. "My answer was, 'I wouldn't bet my life, but I'd bet his!'"

Clara Moskowitz is *Scientific American*'s senior editor specializing in space and physics. She has a bachelor's degree in astronomy and physics from Wesleyan University and a graduate degree in science journalism from the University of California, Santa Cruz.

Lee Billings is an associate editor for *Scientific American*. He covers space and physics.

MORE TO EXPLORE

On the Cosmological Implications of the String Swampland. Prateek Agrawal, Georges Obied, Paul J. Steinhard and Cumrun Vafa in *Physics Letters B*, Vol. 784, pages 271–276; September 10, 2018.

Emergent de Sitter Cosmology from Decaying Anti–de Sitter Space. Souvik Banerjee, Ulf Danielsson, Giuseppe Dibitetto, Suvendu Giri and Marjorie Schillo in *Physical Review Letters*, Vol. 121, Article 261301; December 27, 2018.

De Sitter Space and the Swampland. Georges Obied, Hirosi Ooguri, Lev Spodyneiko and Cumrun Vafa (in press). Preprint available at arXiv:1806.08362v3

scientificamerican.com/magazine/sa

END NOTE

Eyes into Deep Space

Shown is one of the four auxiliary telescopes that make up the ESO Very Large Telescope (VLT) Interferometer in the Atacama Desert of northern Chile. The light beams collected by each telescope are combined in an intricate system of underground mirrors and tunnels. The resulting composite images have a resolution 25 times finer than any single telescope. The lasers improve the optics of the telescopes by reducing the effects of atmospheric distortion. In the past few years alone, the VLT has captured images of planetary nebulas, granulation patterns on the surface of a red giant star and the glittering constellations of stellar nurseries. In 2018 scientists using the MUSE spectrograph on the VLT reported a colossal fountain of molecular gas powered by a black hole some one billion light-years away.

ENRICO SACCHETTI