SPECIAL EDITION

SCIENTIFIC AMERICAN

WWW.SCIAM.COM

Display until May 31, 2003

wormholes and warp drive

stopping light cold

teleportation

where's the antimatter?

the edge of physics

extreme experiments

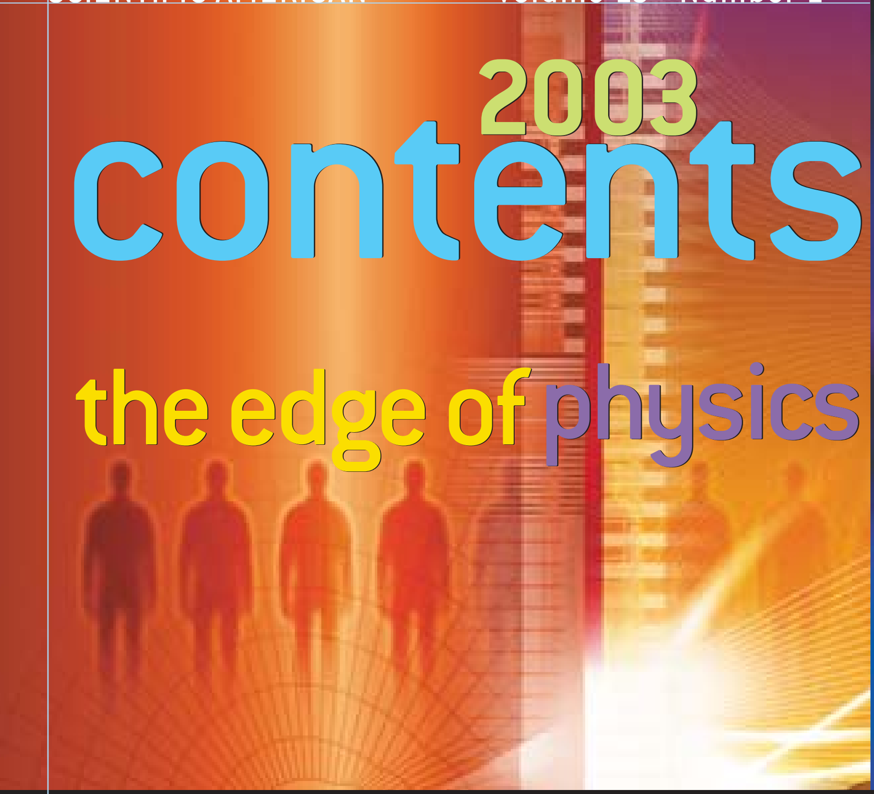
a theory of everything

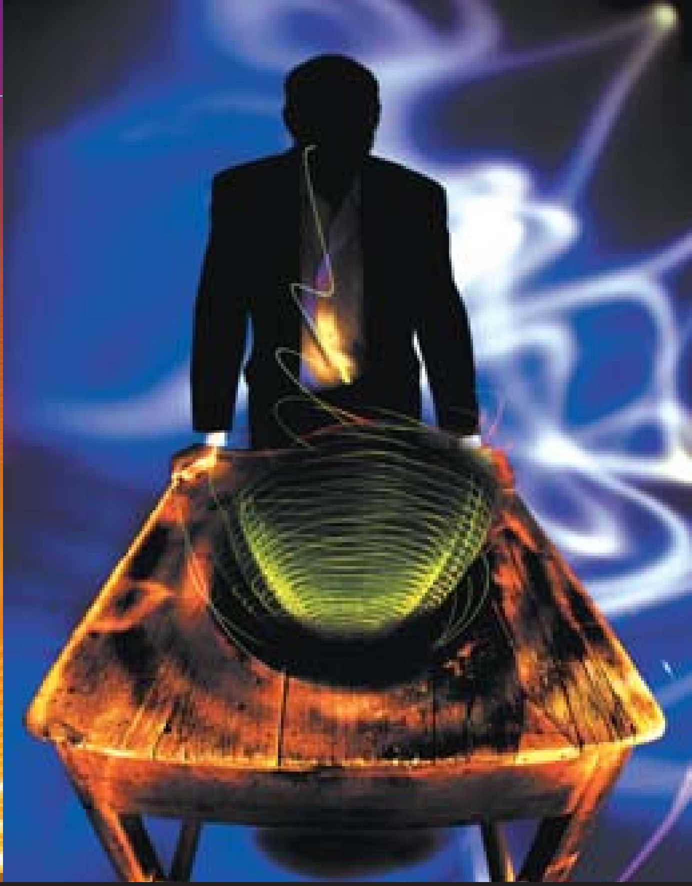
$5.95 U.S. $6.50 CAN.

SCIENTIFIC AMERICAN Volume 13 Number 1

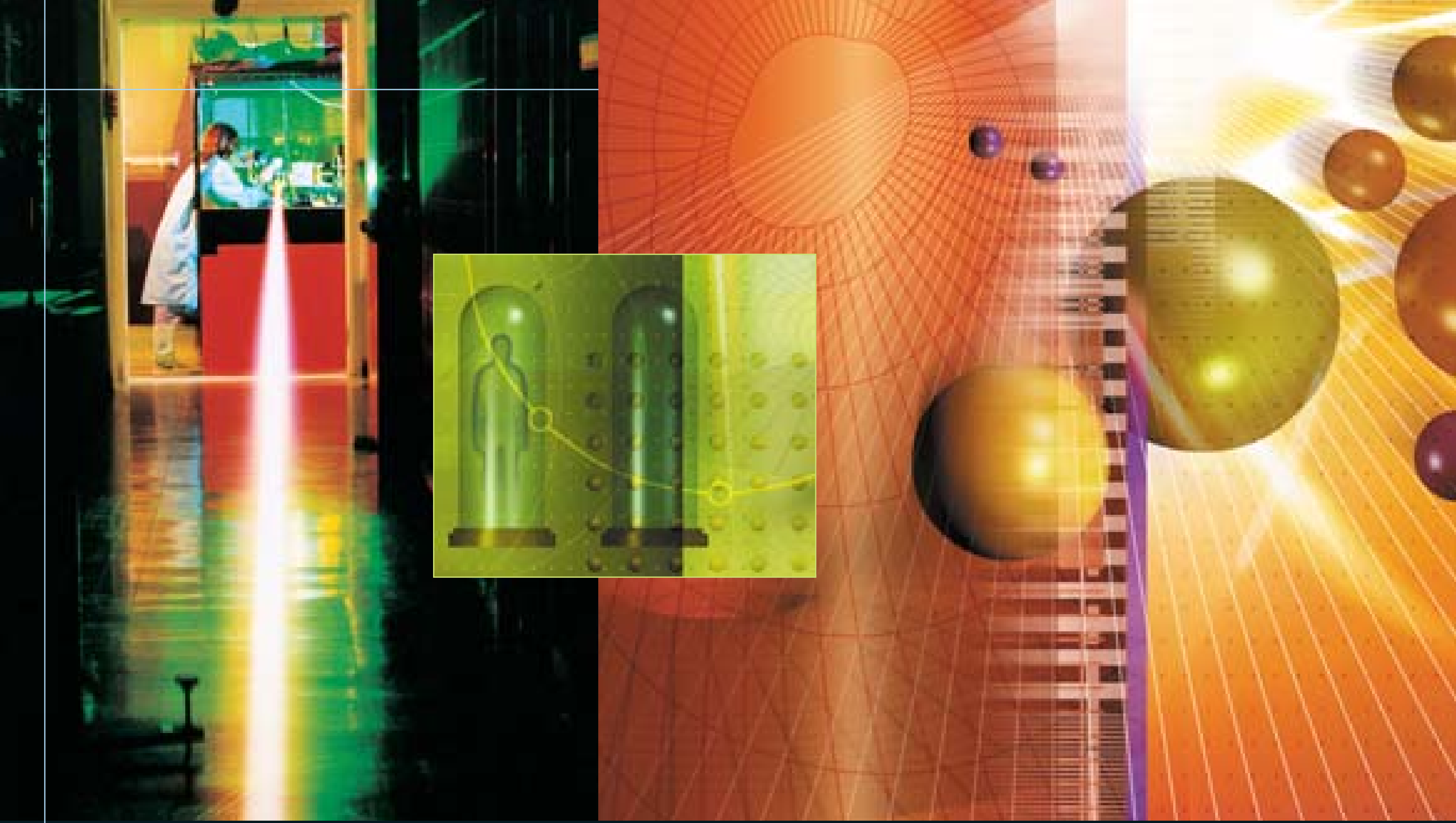
2003 contents

the edge of physics





Cover illustration by Tom Draper Design; Tom Draper Design (*left*); Chip Simons (*right*); page 2: William Pelletier/Photo Services, Inc., courtesy of the University of Michigan (*left*); Tom Draper Design (*inset* and *right*)



extremeexperiments

exoticspaces

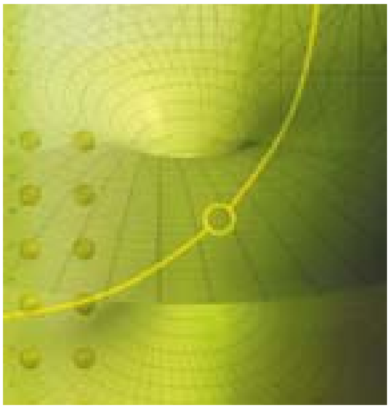
Scientific American Special (ISSN 1048-0943), Volume 13, Number 1, 2003, published by Scientific American, Inc., 415 Madison Avenue, New York, NY 10017-1111.

Canadian BN No. 127387652RT; QST No. Q1015332537. To purchase additional quantities: 1 to 9 copies: U.S. $5.95 each plus $2.00 per copy for postage and handling (outside U.S. $5.00 P&H). Send payment to Scientific American, Dept. PHY, 415 Madison Avenue, New York, NY 10017-1111. Inquiries: Fax 212-355-0408 or telephone 212-451-8890. Printed in U.S.A.

Postcards from the Edge

Anyone who understands science knows that it is often a messy, complex business that can't be conveniently packaged into neat "breakthroughs," despite what may appear in the daily headlines. Yet the striving of scientists to reach beyond the current limits of human learning is constant and unyielding, a persistent tap, tap, tapping away at the obscuring shield that lies at the edge of the unknown.

Physics, frequently called the most fundamental of sciences, quests vigorously to solve great puzzles at least as much as any other discipline. In recent years, researchers have made strides toward a Theory of Everything, one that could someday wrap together the classical physics inspired by Isaac Newton with the rules that govern events on quantum scales. Scientists have begun to forge a quantum theory of gravity, found ways to "beam" particles of light from one place to another, and even stopped light cold, the better to scrutinize its nature. They have learned that the laws of physics don't preclude an unusual form of energy—negative energy—that could be used in the construction of even more fantastic phenomena, such as shortcuts through space called wormholes and faster-than-light warp drives.



TOM DRAPER DESIGN

Clearly, much work remains. Giant experiments that are now under way or soon becoming active will let researchers probe an exotic new layer of reality, delve into the reasons behind the puzzling asymmetry between antimatter and matter in the universe, and detect "massive" neutrinos as the ghostly particles speed through the planet.

The latest developments in all these areas, and more, appear in this special edition from SCIENTIFIC AMERICAN. We invite you to explore these reports—postcards from those who are laboring in the field to push back the boundaries of knowledge, a little at a time.

John Rennie
Editor in Chief
Scientific American
editors@sciam.com

a unified physics by 2050?

By Steven Weinberg

The primary goal of physics is to understand the wonderful variety of nature in a unified way. The greatest advances of the past have been steps toward this goal: The unification of terrestrial and celestial mechanics by Isaac Newton in the 17th century. The theories of electricity and magnetism by James Clerk Maxwell in the 19th century. Spacetime geometry and the theory of gravitation by Albert Einstein from 1905 to 1916. And the unraveling of chemistry and atomic physics through the advent of quantum mechanics in the 1920s.

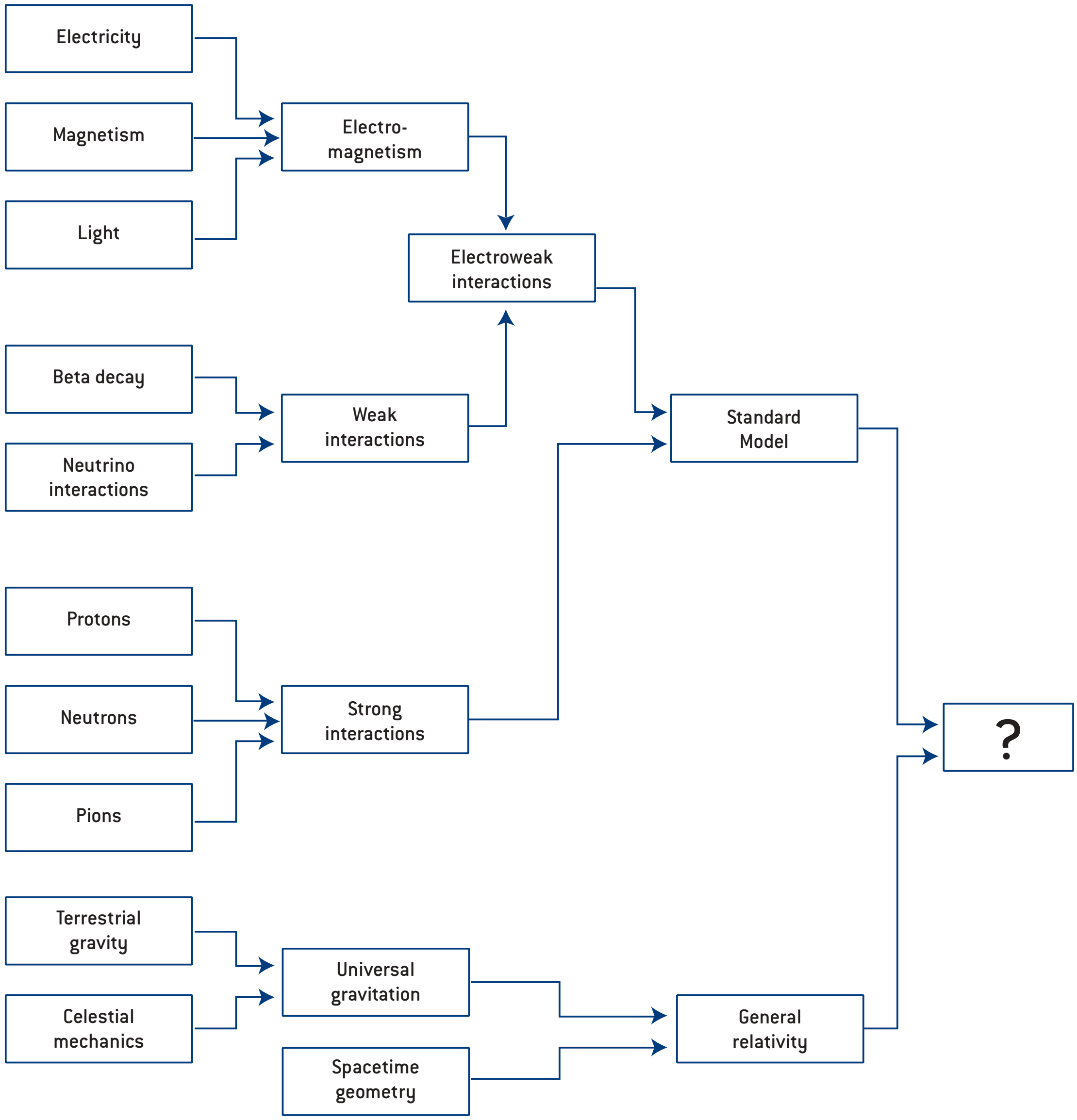
Einstein devoted the last 30 years of his life to an unsuccessful search for a "unified field theory," which would unite general relativity—his own theory of spacetime and gravitation—with Maxwell's theory of electromagnetism. Progress toward unification has been made more recently, but in a different direction. Our current theory of elementary particles and forces, known as the Standard Model of particle physics, has achieved a unification of electromagnetism with the weak interactions, the forces responsible for the change of neutrons and protons into each other in radioactive processes and in the stars. The Standard Model also gives a separate but similar description of the strong interactions, the forces that hold quarks together inside protons and neutrons and hold protons and neutrons together inside atomic nuclei.

We have ideas about how the theory of strong interactions can be unified with the theory of weak and electromagnetic interactions (often called Grand Unification), but this may only work if gravity is included, which presents grave difficulties. We suspect that the apparent differences among these forces have been brought about by events in

QUANTUM NATURE of space and time must be dealt with in a unified theory. At the shortest distance scales, space may be replaced by a continually reconnecting structure of strings and membranes—or by something stranger still.

 Updated from the December 1999 issue

Experiments at CERN and elsewhere should let us complete the Standard Model of particle physics, but a unified theory of all forces will probably require radically new ideas



UNIFICATION of disparate phenomena within one theory has long been a central theme of physics. The Standard Model of particle physics successfully describes three (electromagnetism, weak interactions and strong interactions) of the four known forces of nature but remains to be united definitively with general relativity, which governs the force of gravity and the nature of space and time.

the very early history of the big bang, but we cannot follow the details of cosmic history at those early times without a better theory of gravitation and the other forces. There is a chance the work of unification will be completed by 2050. But can we actually do it?

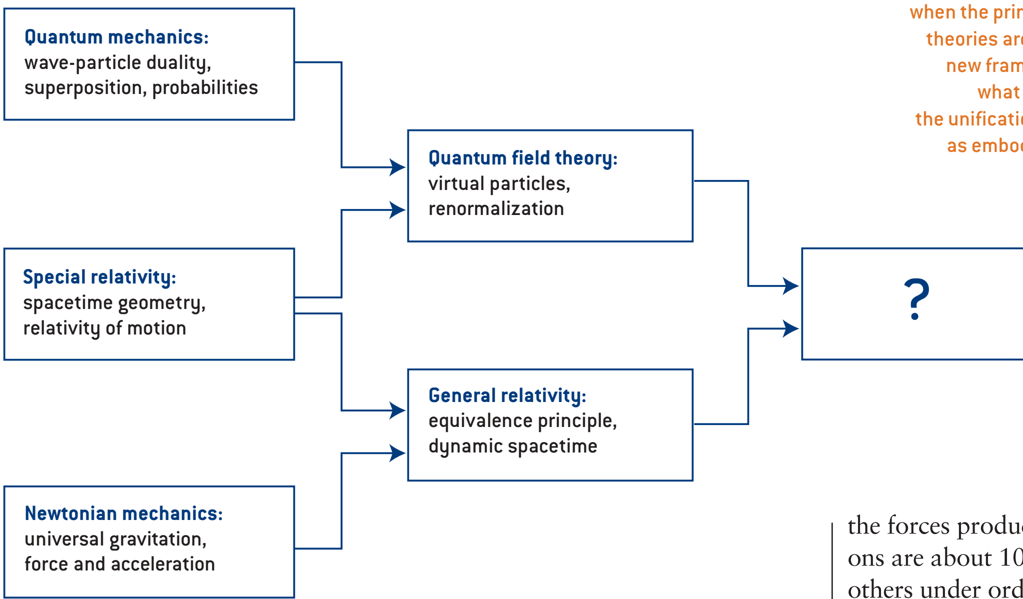
Quantum Fields

THE STANDARD MODEL of particle physics is a quantum field theory. Its basic ingredients are fields, among them the electric and magnetic fields of 19th-century electrodynamics. Little ripples in these fields carry energy and momentum from place to place, and quantum mechanics tells us that these ripples come in bundles, or quanta, that are recognized in the laboratory as elementary particles. For instance, the quantum of the electromagnetic field is a particle known as the photon.

The Standard Model includes a field for each type of elementary particle that has been observed in high-energy physics laboratories [*see top illustration on page 8*]. There are the lepton fields: their quanta include the familiar electrons, which make up the outer parts of ordinary atoms, similar heavier particles known as muons and tauons, and related electrically neutral particles known as neutrinos. There are fields for quarks of various types, some of which are bound together in the protons and neutrons that make up the nuclei of ordinary atoms. Forces between these particles are produced by the exchange of photons and similar elementary particles: the W^+, W^- and Z^0 transmit the weak force, and eight species of gluon produce the strong forces.

These particles exhibit a wide variety of masses that follow no recognizable pattern, with the electron 350,000 times as light as the heaviest quark, and neutrinos even lighter. The Standard Model has no mechanism that would account for any of these masses, unless we supplement it by adding additional fields, of a type known as scalar fields. "Scalar" means that these

ALFRED T. KAMAJIAN (*preceding page*); JOHNNY JOHNSON



MOST PROFOUND ADVANCES in fundamental physics tend to occur when the principles of different types of theories are reconciled within a single new framework. We do not yet know what guiding principle underlies the unification of quantum field theory, as embodied in the Standard Model, with general relativity.

fields do not carry a sense of direction, unlike the electric and magnetic fields and the other fields of the Standard Model. This opens up the possibility that these scalar fields can pervade all of space without contradicting one of the best established principles of physics, that space looks the same in all directions. (In contrast, if, for example, there were a significant magnetic field everywhere in space, we could then identify a preferred direction by using an ordinary compass.) The interaction of the other fields of the Standard Model with the all-pervasive scalar fields is believed to give the particles of the Standard Model their masses.

Beyond the Top

TO COMPLETE the Standard Model, we need to confirm the existence of these scalar fields and find out how many types there are. This is a matter of discovering new elementary particles, often called Higgs particles, that can be recognized as the quanta of these fields. We have every reason to expect that this task will be accomplished before 2020, when the accelerator called the Large Hadron Collider at CERN, the European laboratory for particle physics near Geneva, will have been operating for more than a decade.

The very least thing that will be discovered is a single electrically neutral scalar particle. It would be a disaster if this were all that were found by 2020, though, because that would leave us without a clue to the solution of a formidable puzzle called the hierarchy problem.

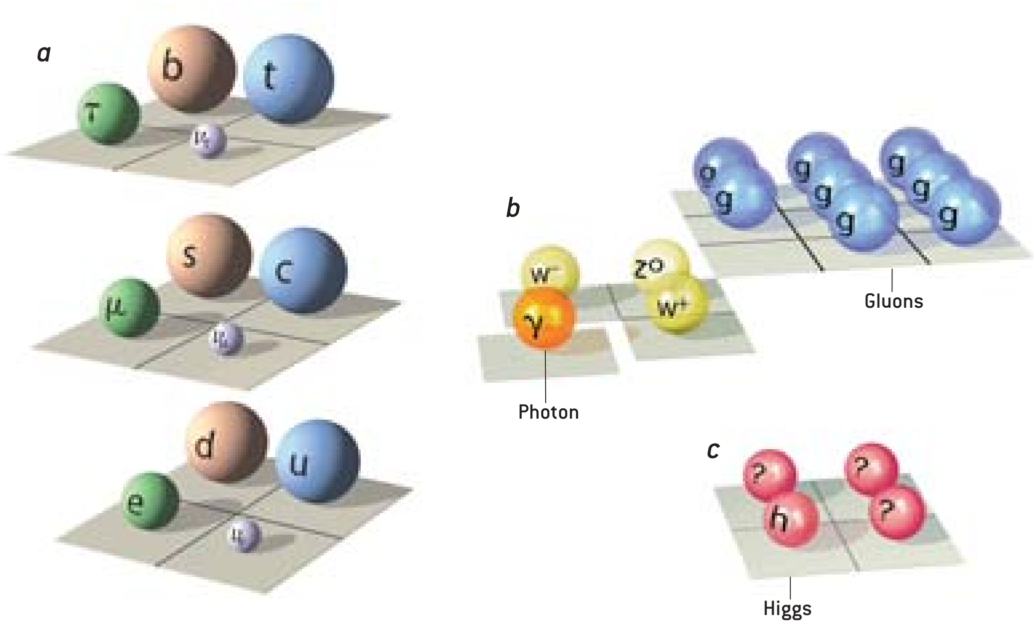
The heaviest known particle of the Standard Model is the top quark, with a mass equivalent to an energy of 175 giga-electron-volts (GeV). One GeV is a little more than the energy contained in a proton mass. [See "The Discovery of the Top Quark," by Tony M. Liss and Paul L. Tipton; SCIENTIFIC AMERICAN, September 1997.] The not yet discovered Higgs particles are expected to have similar masses, from 100 to several hundred GeV. But there is evidence of a much larger scale of masses that will appear in equations of the not yet formulated unified theory. The gluon, *W*, *Z* and photon fields of the Standard Model have interactions of rather different strengths with the other fields of this model; that is why the forces produced by exchange of gluons are about 100 times as strong as the others under ordinary conditions. Gravitation is vastly weaker: the gravitational force between the electron and proton in the hydrogen atom is about 10^{-39} the strength of the electric force.

But all these interaction strengths depend on the energy at which they are measured [*see top illustration on page 9*]. It is striking that when the interactions of the fields of the Standard Model are extrapolated, they all become equal to one another at an energy of a little more than 10^{16} GeV, and the force of gravitation has the same strength at an energy not much higher, around 10^{18} GeV. (Refinements to the theory of gravitation have been suggested that would even bring the strength of gravitation into equality with the other forces at about 10^{16} GeV.) We are used to some pretty big mass ratios in particle physics, like the 350,000 to 1 ratio of the top quark to the electron mass, but this is nothing compared with the enormous ratio of the fundamental unification energy scale of 10^{16} GeV (or perhaps 10^{18} GeV) to the energy scale of about 100 GeV that is

THE AUTHOR

STEVEN WEINBERG is head of the Theory Group at the University of Texas at Austin and a member of its physics and astronomy departments. His work in elementary particle physics has been honored with numerous prizes and awards, including the Nobel Prize for Physics in 1979 and the National Medal of Science in 1991. The third volume (*Supersymmetry*) of his treatise *The Quantum Theory of Fields* was published in 2000. The second volume (*Modern Applications*) was hailed by *Physics Today* as being "unmatched by any other book on quantum field theory for its depth, generality and definitive character."

STANDARD MODEL of particle physics describes each particle of matter and each force with a quantum field. The fundamental particles of matter are fermions; they come in three generations (*a*). Each generation of particles follows the same pattern of properties. The fundamental forces are caused by bosons (*b*), which are organized according to three closely related symmetries. In addition, one or more Higgs particles or fields (*c*) generate the masses of the other fields.



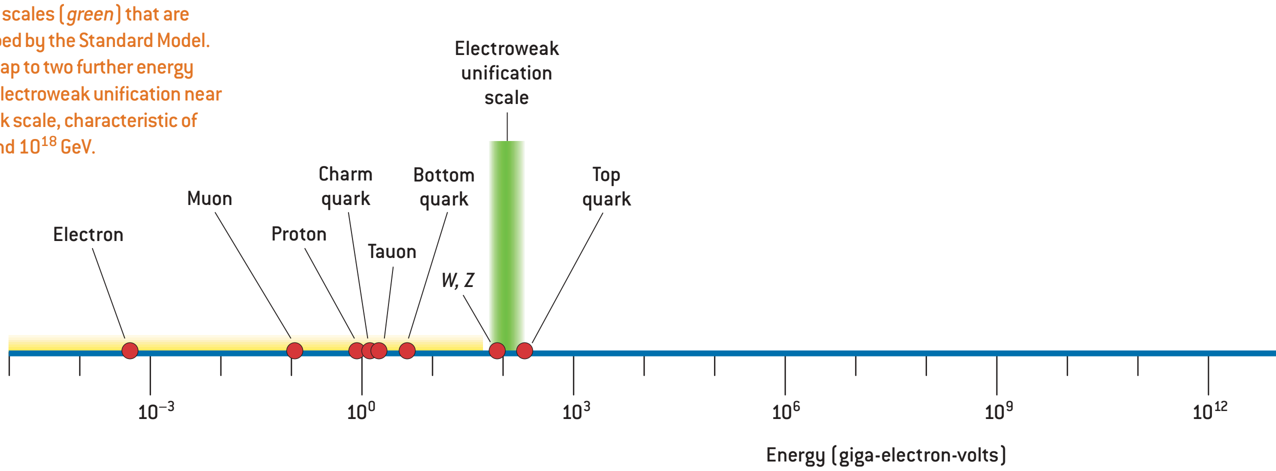
SLIM FILMS

typical of the Standard Model [*see illustration below*]. The crux of the hierarchy problem is to understand this huge ratio, this vast jump from one level to the next in the hierarchy of energy scales, and to do so not just by adjusting the constants in our theories to make the ratio come out right but as a natural consequence of fundamental principles.

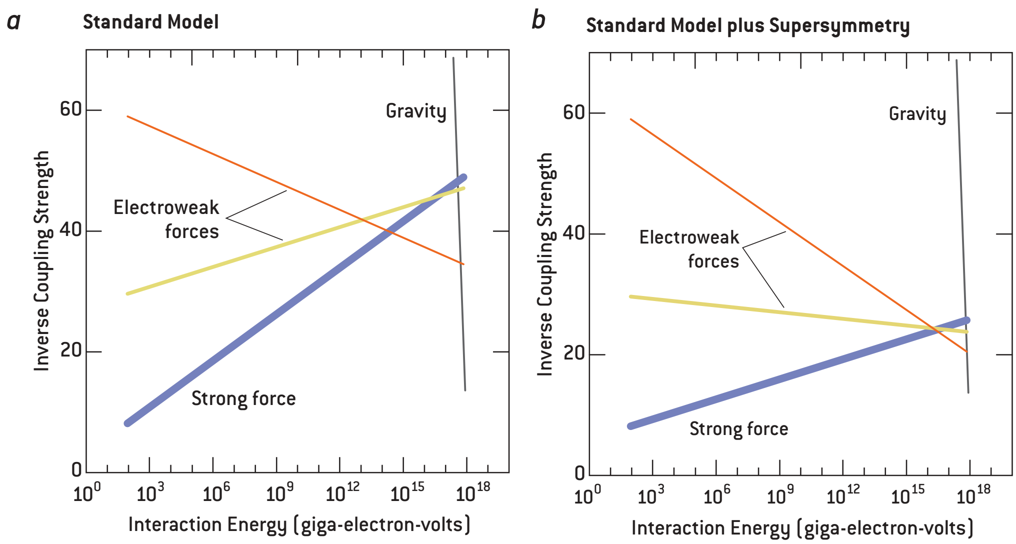
Theorists have proposed several interesting ideas for a natural solution to the hierarchy problem, incorporating a new symmetry principle known as supersymmetry (which also improves the accuracy with which the interaction strengths converge at 10^{16} GeV), or new strong forces known as technicolor, or both [*see illustration on page 10*]. All these theories contain additional forces that are unified with the strong, weak and electromagnetic forces at an energy of about 10^{16} GeV. The new forces become strong at some energy far below 10^{16} GeV, but we cannot observe them directly, because they do not act on the known particles of the Standard Model. Instead they act on other particles that are too massive to be created in our laboratories. These "very heavy" particles are nonetheless much lighter than 10^{16} GeV because they acquire their mass from the new forces, which are strong only far below 10^{16} GeV. In this picture, the known particles of the Standard Model would interact with the very heavy particles, and their masses would arise as a secondary effect of this relatively weak interaction. This mechanism would solve the hierarchy problem, making the known particles lighter than the very heavy particles, which are themselves much lighter than 10^{16} GeV.

All these ideas share another common feature: they require the existence of a zoo of new particles with masses not much larger than 1,000 GeV. If there is any truth to these ideas, then these particles should be discovered before 2020 at the Large Hadron Collider, and some of them may even show up before then at Fermilab or CERN, although it may take further decades and new accelerators to explore their properties fully. When these particles have been discovered and their

HIERARCHY PROBLEM is a measure of our ignorance. Experiments (*yellow band*) have probed up to an energy of about 200 GeV and have revealed an assortment of particle masses (*red*) and interaction energy scales (*green*) that are remarkably well described by the Standard Model. The puzzle is the vast gap to two further energy scales, that of strong-electroweak unification near 10^{16} GeV and the Planck scale, characteristic of quantum gravity, around 10^{18} GeV.



JOHNNY JOHNSON

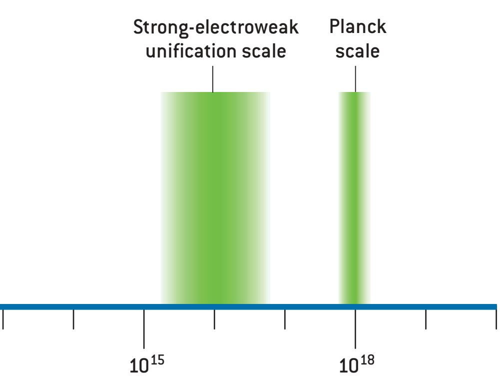


THEORETICAL EXTRAPOLATION shows that the three Standard Model forces (the strong force and the unified weak and electromagnetic forces) have roughly equal strength at very high energy (*a*), and the equality is improved by allowing for supersymmetry (*b*). Curve thickness indicates approximate uncertainty in the coupling strengths.

JOHNNY JOHNSON

properties measured, we will be able to tell whether any of them would have survived the early moments of the big bang and could now furnish the "dark matter" in intergalactic space that is thought to make up most of the present mass of the universe. At any rate, it seems likely that by 2050 we will understand the reason for the enormous ratio of energy scales encountered in nature.

What then? There is virtually no chance that we will be able to do experiments involving processes at particle energies like 10^{16} GeV. With current technology the diameter of an accelerator is proportional to the energy given to the accelerated particles. To accelerate particles to an energy of 10^{16} GeV would require an accelerator a few light-years across. Even if someone found another way to concentrate macroscopic amounts of energy on a single particle, the rates of interesting processes at these energies would be too slow to yield useful information. But even though we cannot study processes at energies like 10^{16} GeV directly, there is a very good chance that these processes produce effects at accessible energies that can be recognized experimentally because they go beyond anything allowed by the Standard Model.



The Standard Model is a quantum field theory of a special kind, one that is "renormalizable." This term goes back to the 1940s, when physicists were learning how to use the first quantum field theories to calculate small shifts of atomic energy levels. They discovered that calculations using quantum field theory kept producing infinite quantities, which usually means that a theory is flawed or is being pushed beyond its limits of validity. In time, they found a way to deal with the infinite quantities by absorbing them into a redefinition, or "renormalization," of only a few physical constants, such as the charge and mass of the electron. (The minimum version of the Standard Model, with just one scalar particle, has 18 of these constants.) Theories in which this procedure worked were called renormalizable and had a simpler structure than nonrenormalizable theories.

Suppressed Interactions

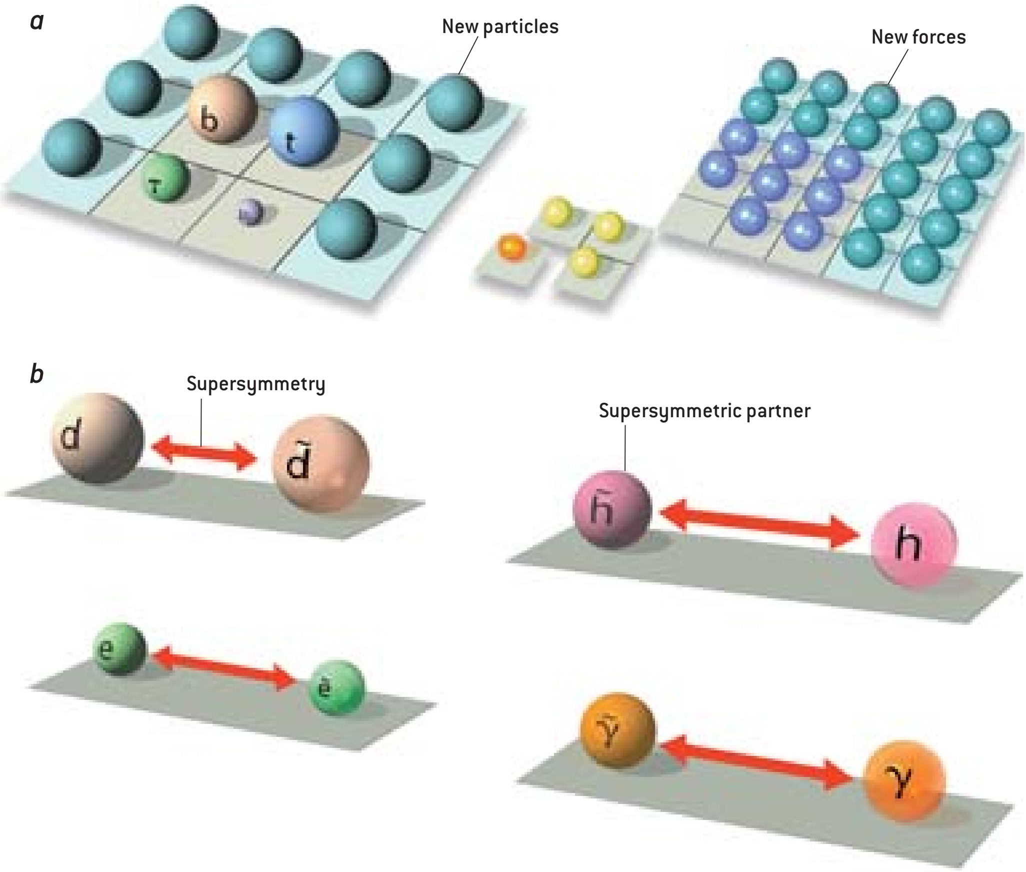
IT IS THIS SIMPLE, renormalizable structure of the Standard Model that has let us derive specific quantitative predictions for experimental results, predictions the success of which has confirmed the validity of the theory.

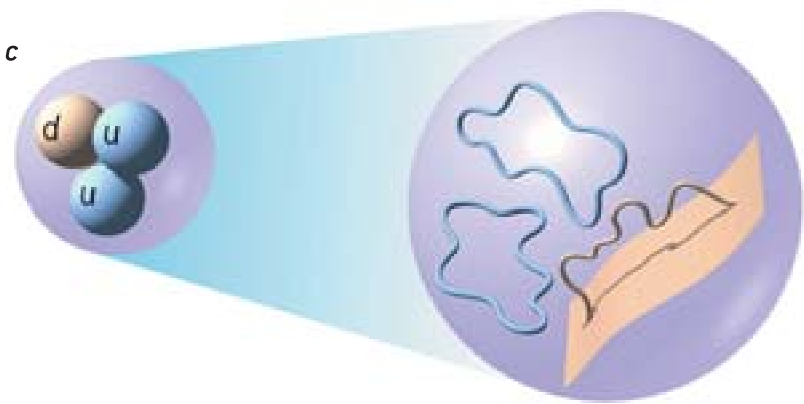
In particular, the principle of renormalizability, together with various symmetry principles of the Standard Model, rules out unobserved processes such as the decay of isolated protons and forbids the neutrinos from having masses. Physicists commonly used to believe that for a quantum field theory to have any validity, it had to be renormalizable. This requirement was a powerful guide to theorists in formulating the Standard Model. It was terribly disturbing that it seemed impossible, for fundamental reasons, to formulate a renormalizable quantum field theory of gravitation.

Today our perspective has changed. Particle physics theories look different depending on the energy of the processes and reactions being considered. Forces produced by exchange of a very massive particle will typically be extremely weak at energies that are low compared with that mass.

Other effects can be similarly suppressed, so that at low energies one has what is known as an effective field theory, in which these interactions are negligible. Theorists have realized that any fundamental quantum theory that is consistent with the special theory of relativity will look like a renormalizable quantum field theory at low energies. But although the infinities are still canceled, these effective theories do not have the

WHAT COMES NEXT? There are several possibilities for the unified physics that lies beyond the Standard Model. Technicolor models (*a*) introduce new interactions analogous to the "color" force that binds quarks. Accompanying the interactions are new generations of particles unlike the three known generations. Supersymmetry (*b*) relates fermions to bosons and adds the supersymmetric partners of each known particle to the model. M-theory and string theory (*c*) recast the entire model in terms of new entities such as tiny strings, loops and membranes that behave like particles at low energies.





SLIM FILMS

simple structure of theories that are renormalizable in the classic sense. Additional complicated interactions are present; instead of being completely excluded, they are merely highly suppressed below some characteristic energy scale.

Gravitation itself is just such a suppressed nonrenormalizable interaction. It is from its strength (or rather weakness) at low energies that we infer that its fundamental energy scale is roughly 10^{18} GeV. Another suppressed nonrenormalizable interaction would make the proton unstable, with a half-life in the range of 10^{31} to 10^{34} years, which might be too slow to be observed even by 2050 [see my article "The Decay of the Proton"; SCIENTIFIC AMERICAN, June 1981]. Yet another suppressed nonrenormalizable interaction would give the neutrinos tiny masses, about 10^{-11} GeV. There is now strong evidence being collected at giant detectors for neutrino masses, very likely of this order [see "Detecting Massive Neutrinos," on page 68].

Observations of this kind will yield valuable clues to the unified theory of all forces, but the discovery of this theory will probably not be possible without radically new ideas. Some promising ones are already in circulation. There are five theories of tiny one-dimensional entities known as strings, which in their different modes of vibration appear at low energy as various kinds of particles and apparently furnish perfectly finite theories of gravitation and other forces in 10 spacetime dimensions. Of course, we do not live in 10 dimensions, but it is plausible that six of these dimensions could be rolled up so tightly that they could not be observed in processes at energies below 10^{16} GeV per particle. Evidence has appeared in the past several years that these five string theories (and also a quantum field theory in 11 dimensions) are all versions of a single fundamental theory (sometimes called M-theory) that apply under different approximations [see "The

Perhaps when we understand how particles and forces behave at energies up to 10^{18} GeV we will find new mysteries, but I doubt it.

Theory Formerly Known as Strings," on page 12]. But no one knows how to write down the equations of this theory.

Outside of Spacetime

TWO GREAT OBSTACLES stand in the way of this task. One is that we do not know what physical principles govern the fundamental theory. In developing general relativity, Einstein was guided by a principle he had inferred from the known properties of gravitation, the principle of the equivalence of gravitational forces to inertial effects such as centrifugal force. The development of the Standard Model was guided by a principle called gauge symmetry, a generalization of the well-known property of electricity that it is only differences of voltages that matter, not voltages themselves.

But we have not discovered any fundamental principle that governs M-theory. The various approximations to this theory look like string or field theories in spacetimes of different dimensionalities, but it seems probable that the fundamental theory is not to be formulated in spacetime at all. Quantum field theory is powerfully constrained by principles concerning the nature of four-dimensional spacetime that are incorporated in the special theory of relativity. How can we get the ideas we need to formulate a truly fundamental theory, when this theory is meant to describe a realm where all intuitions derived from life in spacetime become inapplicable?

The other obstacle is that even if we were able to formulate a fundamental theory, we might not know how to use it to make predictions that could confirm its validity. Most of the successful predictions of the Standard Model have been based on a method of calculation known as perturbation theory. In quantum mechanics, the rates of physical processes are given by sums over all possible sequences of intermediate steps by which the process might occur. Using perturbation theory, one first considers just the simplest intermediate steps, then the next simplest, and so on. This works only if increasingly complicated intermediate steps make decreasingly large contributions to the rate, which is usually the case if the forces involved are sufficiently weak. Sometimes a theory with very strong forces is equivalent to another theory with very weak forces, which can be solved by the methods of perturbation theory. This seems to be true of certain pairs of the five string theories in 10 dimensions and the field theory in 11 dimensions mentioned earlier. Unfortunately, the forces of the fundamental theory are probably neither very strong nor very weak, ruling out any use of perturbation theory.

Recognizing the Answer

IT IS IMPOSSIBLE to say when these problems will be overcome. They may be solved in a preprint put out tomorrow by some young theorist. They may not be solved by 2050, or even 2150. But when they are solved, even though we cannot do experiments at 10^{16} GeV or look into higher dimensions, we will not have any trouble recognizing the truth of the fundamental unified theory. The test will be whether the theory successfully accounts for the measured values of the physical constants of the Standard Model, along with whatever other effects beyond the Standard Model may have been discovered by then.

It is possible that when we finally understand how particles and forces behave at energies up to 10^{18} GeV, we will just find new mysteries, with a final unification as far away as ever. But I doubt it. There are no hints of any fundamental energy scale beyond 10^{18} GeV, and string theory even suggests that higher energies have no meaning.

The discovery of a unified theory that describes nature at all energies will put us in a position to answer the deepest questions of cosmology: Did the expanding cloud of galaxies we call the big bang have a beginning at a definite time in the past? Is our big bang only one episode in a much larger universe in which big and little bangs have been going on eternally? If so, do what we call the constants—or even the laws—of nature vary from one bang to another?

This will not be the end of physics. It probably won't even help with some of the outstanding problems of today's physics, such as understanding turbulence and high-temperature superconductivity. But it will mark the end of a certain kind of physics: the search for a unified theory that entails all other facts of physical science. SA

MORE TO EXPLORE

Unified Theories of Elementary-Particle Interaction. Steven Weinberg in *Scientific American,* Vol. 231, No. 1, pages 50–59; July 1974.

Dreams of a Final Theory. Steven Weinberg. Pantheon Books, 1992.

Reflections on the Fate of Spacetime. Edward Witten in *Physics Today,* Vol. 49, No. 4, pages 24–30; April 1996.

Duality, Spacetime and Quantum Mechanics. Edward Witten in *Physics Today,* Vol. 50, No. 5, pages 28–33; May 1997.

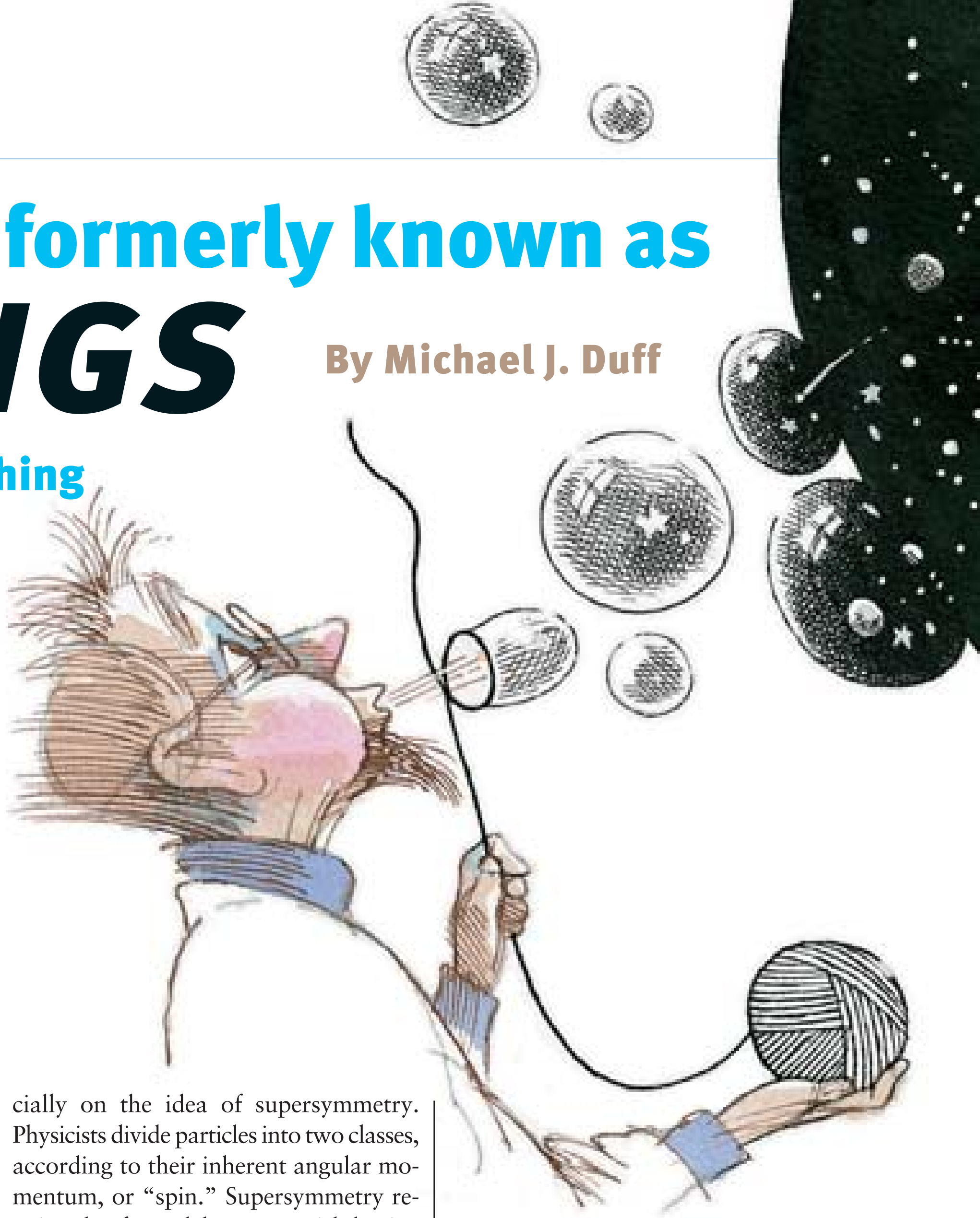
The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory. Brian Greene. W. W. Norton, 1999.

übertheory

the theory formerly known as STRINGS

By Michael J. Duff

The Theory of Everything is emerging as one in which not only strings but also membranes and black holes play a role



DUSAN PETRICIC

At a time when certain pundits claim that all the important discoveries have already been made, it is worth emphasizing that the two main pillars of 20th-century physics, quantum mechanics and Einstein's general theory of relativity, are mutually incompatible. General relativity fails to comply with the quantum rules that govern the behavior of elementary particles, while black holes are challenging the very foundations of quantum mechanics. Something big has to give.

Until recently, the best hope for a theory that would unite gravity with quantum mechanics and describe all physical phenomena was based on strings: one-dimensional objects whose modes of vibration represent the elementary particles. In 1995, however, strings were subsumed by M-theory. In the words of the guru of string theory, Edward Witten of the Institute for Advanced Study in Princeton, N.J., "M stands for magic, mystery or membrane, according to taste." New evidence in favor of this theory is appearing daily, representing the most exciting development since strings first swept onto the scene.

M-theory, like string theory, relies crucially on the idea of supersymmetry. Physicists divide particles into two classes, according to their inherent angular momentum, or "spin." Supersymmetry requires that for each known particle having integer spin—0, 1, 2 and so on, measured in quantum units—there is a particle with the same mass but half-integer spin ($1/2$, $3/2$, $5/2$ and so on), and vice versa.

Unfortunately, no such superpartner has yet been found. The symmetry, if it exists at all, must be broken, so that the postulated particles do not have the same mass as known ones but instead are too heavy to be seen in current accelerators. Even so, theorists believe in supersymmetry because it provides a framework within which the weak, electromagnetic and strong forces may be united with the most elusive force of all: gravity.

Supersymmetry transforms the coordinates of space and time such that the laws of physics are the same for all observers. Einstein's general theory of relativity derives from this condition, and so supersymmetry implies gravity. In fact, supersymmetry predicts "supergravity," in which a particle with a spin of 2—the graviton—transmits gravitational interactions and has as a partner a gravitino, with a spin of $3/2$.

Conventional gravity does not place any limits on the possible dimensions of spacetime: its equations can, in principle, be formulated in any dimension. Not so with supergravity, which places an upper limit of 11 on the dimensions of spacetime. The familiar universe, of course, has three dimensions of space: height, length and breadth; time is the fourth dimension of spacetime. But in the early 1920s Polish physicist Theodore Kaluza and Swedish physicist Oskar Klein suggested that spacetime may have a hidden fifth dimension. This extra dimension would not be

Updated from the February 1998 issue

LIFE, THE UNIVERSE AND EVERYTHING may arise from the interplay of strings, bubbles and sheets in higher dimensions of spacetime.

infinite, like the others; instead it would close in on itself, forming a circle. Around that circle could reside quantum waves, fitting neatly into a loop. Only integer numbers of waves could fit around the circle; each of these would correspond to a particle with a different energy. So the energies would be "quantized," or discrete.

An observer living in the other four dimensions, however, would see a set of particles with discrete charges, rather than energies. The quantum, or unit, of charge would depend on the circle's radius. In the real world as well, electrical charge is quantized, in units of *e,* the charge on the electron. To get the right value for *e,* the circle would have to be tiny, about 10^{-33} centimeter in radius.

The unseen dimension's small size explains why humans, or even atoms, are unaware of it. Even so, it would yield electromagnetism. And gravity, already present in the four-dimensional world, would be united with that force.

In 1978 Eugene Cremmer, Bernard Julia and Joel Scherk of the École Normale Supérieure in Paris realized that supergravity not only permits up to seven extra dimensions but is most elegant when existing in a spacetime of 11 dimensions (10 of space and one of time). The kind of real, four-dimensional world the theory ultimately predicts depends on how the extra dimensions are rolled up, à la Kaluza and Klein. The several curled dimensions could conceivably allow physicists to derive, in addition to electromagnetism, the strong and weak nuclear forces. For these reasons, many physicists began to look to supergravity in 11 dimensions for the unified theory.

In 1984, however, 11-dimensional supergravity was rudely knocked off its pedestal. An important feature of the real world is that nature distinguishes between right and left. Witten and others emphasized that such "handedness" cannot readily be derived by reducing spacetime from 11 dimensions down to four.

P-Branes

SUPERGRAVITY'S position was usurped by superstring theory in 10 dimensions. Five competing theories held sway, designated by their mathematical characteristics as the $E_8 \times E_8$ heterotic, the SO(32) heterotic, the SO(32) Type I, and the Type IIA and Type IIB strings. (The Type I is an "open" string consisting of just a segment; the others are "closed" strings that form loops.) The $E_8 \times E_8$ seemed—at least in principle—capable of explaining the elementary particles and forces, including their handedness. And strings seemed to provide a theory of gravity consistent with quantum effects. All these virtues enabled string theory to sweep physicists off their feet and supergravity into the doghouse.

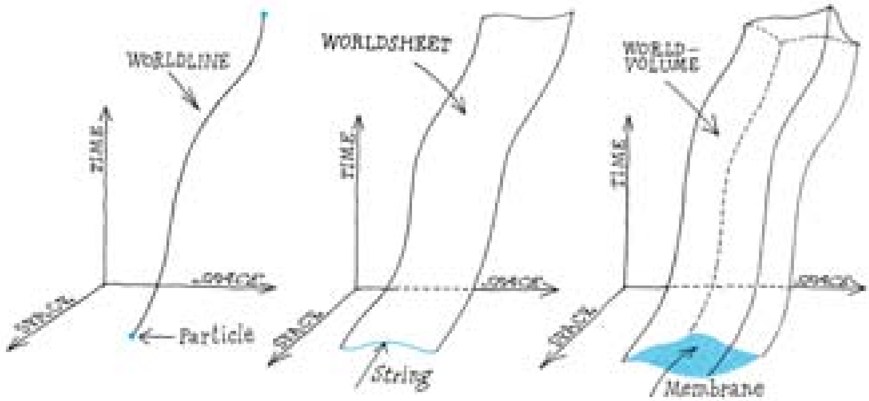
After the initial euphoria over strings, however, doubts began to creep in. First, important questions—especially how to confront the theory with experiment—seemed incapable of being answered by traditional methods of calculation. Second, why were there five different string theories? If one is looking for a unique Theory of Everything, surely this is an embarrassment of riches. Third, if supersymmetry permits 11 dimensions, why do superstrings stop at 10? Finally, if we are going to conceive of pointlike particles as strings, why not as membranes or more generally as p-dimensional objects, inevitably dubbed p-branes?

Consequently, while most theorists were tucking into super-spaghetti, a small group was developing an appetite for super-ravioli. A particle, which has zero dimensions, sweeps out a one-dimensional trace, or "worldline," as it evolves in spacetime [*see top illustration on next page*]. Similarly a string—having one dimension: length—sweeps out a two-dimensional "worldsheet," and a membrane—having two dimensions: length and breadth—sweeps out a three-dimensional "worldvolume." In general, a p-brane sweeps out a worldvolume of p + 1 dimensions.

As early as 1962, Paul A. M. Dirac

THE AUTHOR

MICHAEL J. DUFF conducts research on unified theories of elementary particles, quantum gravity, supergravity, superstrings, supermembranes and M-theory. He earned his Ph.D. in theoretical physics in 1972 at Imperial College, London, and joined the faculty there in 1980. He became a Distinguished Professor at Texas A&M University in 1992. Duff is now Oskar Klein Professor of Physics at the University of Michigan and director of the Michigan Center for Theoretical Physics.



TRAJECTORY of a particle in spacetime traces a worldline. Similarly, that of a string or a membrane sweeps out a worldsheet or worldvolume, respectively.

had constructed an imaginative model based on a membrane. He postulated that the electron, instead of resembling a point, was in reality a minute bubble, a membrane closed in on itself. Its oscillations, Dirac suggested, might generate particles such as the muon, a heavier version of the electron. Although his attempt failed, the equations that he postulated for the membrane are essentially the ones we use today.

Supersymmetry severely restricts the possible dimensions of a p-brane. In the spacetime of 11 dimensions floats a membrane, which may take the form of a bubble or a two-dimensional sheet. Paul S. Howe of King's College London, Takeo Inami of Kyoto University, Kellogg Stelle of Imperial College, London, and I were able to show that if one of the 11 dimensions is a circle, we can wrap the sheet around it once, pasting the edges together to form a tube. If the radius becomes sufficiently small, the rolled-up membrane ends up looking like a string in 10 dimensions; it yields precisely the Type IIA superstring.

Notwithstanding such results, the membrane enterprise was largely ignored by the string community. Fortunately, the situation was about to change because of progress in an apparently unrelated field.

In 1917 German mathematician Amalie Emmy Noether had shown that the mass, charge and other attributes of elementary particles are conserved because of symmetries of the laws of physics. For instance, conservation of electrical charge follows from a symmetry under a change of a particle's wave function.

Sometimes, however, attributes may be maintained because of deformations in fields. Such conservation laws are called topological. Thus, it may happen that a knot in a set of field lines, called a soliton, cannot be smoothed out. As a result, the soliton is prevented from dissipating and behaves much like a particle. A classic example is a magnetic monopole, which has not been found in nature but shows up as twisted configurations in some field theories.

In the traditional view, then, particles such as electrons and quarks (which carry Noether charges) are seen as fundamental, whereas particles such as magnetic monopoles (which carry topological charge) are derivative. In 1977, however, Claus Montonen, now at the Helsinki Institute of Physics in Finland, and David I. Olive, now at the University of Wales at Swansea, made a bold conjecture. Might there exist an alternative formulation of physics in which the roles of Noether charges (like electrical charge) and topological charges (like magnetic charge) are reversed? In such a "dual" picture, the magnetic monopoles would be the elementary objects, whereas the familiar particles—quarks, electrons and so on—would arise as solitons.

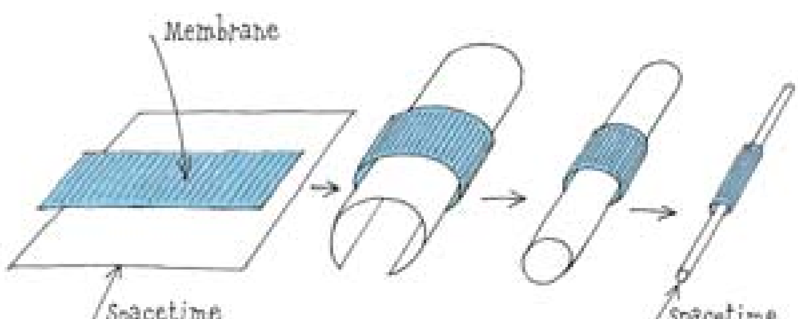
More precisely, a fundamental particle with charge e would be equivalent to a solitonic particle with charge $1/e$. Because its charge is a measure of how strongly a particle interacts, a monopole would interact weakly when the original particle interacts strongly (that is, when e is large), and vice versa.

The conjecture, if true, would lead to a profound mathematical simplification. In the theory of quarks, for instance, physicists can make hardly any calculations when the quarks interact strongly. But any monopoles in the theory must then interact weakly. One could imagine doing calculations with a dual theory based on monopoles and automatically getting all the answers for quarks, because the dual theory would yield the same final results.

Unfortunately, the idea presented a chicken-and-egg problem. Once proved, the Montonen-Olive conjecture could leap beyond conventional calculational techniques, but it would need to be proved by some other method in the first place.

As it turns out, p-branes can also be viewed as solitons. In 1990 Andrew Strominger of the Institute for Theoretical Physics in Santa Barbara, Calif., found that a 10-dimensional string can yield a soliton that is a five-brane. Reviving a conjecture of mine, Strominger suggested that a strongly interacting string is the dual equivalent of weakly interacting five-branes.

There were two major impediments to this duality. First, the duality proposed by Montonen and Olive—between electricity and magnetism in four dimensions—was still unproved, so duality between strings and five-branes in 10 dimensions was even more tenuous. Second, there were issues about how to find the quantum properties of five-branes and



SIMULTANEOUS SHRINKING of a membrane and a dimension of spacetime can result in a string. As the underlying space, shown here as a two-dimensional sheet, curls into a cylinder, the membrane wraps around it. The curled dimension becomes a circle so small that the two-dimensional space ends up looking one-dimensional, like a line. The tightly wrapped membrane then resembles a string.

DUSAN PETRICIC

EXTRA DIMENSION curled into a tube offers insights into the fabric of spacetime.

hence how to prove the new duality.

The first of these impediments was removed, however, when Ashoke Sen of the Tata Institute of Fundamental Research in Bombay, India, established that supersymmetric theories would require the existence of certain solitons with both electrical and magnetic charges. These objects had been predicted by the Montonen-Olive conjecture. This seemingly inconspicuous result converted many skeptics and unleashed a flood of papers. In particular, it inspired Nathan Seiberg of Rutgers University and Edward Witten to look for duality in more realistic (though still supersymmetric) versions of quark theories. They provided a wealth of information on quantum fields, of a kind unthinkable just a few years before.

Duality of Dualities

IN 1990 SEVERAL theorists generalized the idea of Montonen-Olive duality to four-dimensional superstrings, in whose realm the idea becomes even more natural. This duality, which was then speculative, goes by the name S-duality.

In fact, string theorists had already become used to a totally different kind of duality called T-duality. T-duality relates two kinds of particles that arise when a string loops around a compact dimension. One kind (call them “vibrating” particles) is analogous to those predicted by Kaluza and Klein and comes from vibrations of the loop of string [*see box on next page*]. Such particles are more energetic if the circle is small. In addition, the string can wind many times around the circle, like a rubber band on a wrist; its energy becomes higher the more times it wraps around and the larger the circle is. Moreover, each energy level represents a new particle (call them “winding” particles).

T-duality states that the winding particles for a circle of radius *R* are the same as the vibrating particles for a circle of radius $1/R$, and vice versa. To a physicist, the two sets of particles are indistinguishable: a fat, compact dimension may yield the same particles as a thin one.

This duality has a profound implication. For decades, physicists have been struggling to understand nature at the extremely small scales near the Planck length of 10^{-33} centimeter. We have always supposed that laws of nature break down at smaller distances. What T-duality suggests, however, is that at these scales, the universe looks just the same as it does at large scales. One may even imagine that if the universe were to shrink to less than the Planck length, it would transform into a dual universe that grows bigger as the original one collapses.

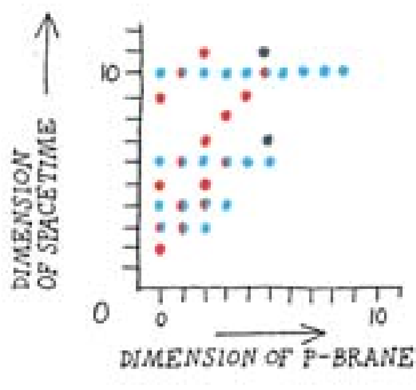
Duality between strings and five-branes was still conjectural, however, because of the problem of quantizing five-branes. Starting in 1991, a team at Texas A&M University, with Jianxin Lu, Ruben Minasian, Ramzi Khuri and myself, dealt with the problem by sidestepping it. If four of the 10 dimensions curl up and the five-brane wraps around these, the latter ends up as a one-dimensional object—a (solitonic) string in six-dimensional spacetime. In addition, a fundamental string in 10 dimensions remains fundamental even in six dimensions. So the concept of duality between strings and five-branes gave way to another conjecture, duality between a solitonic and a fundamental string.

The advantage is that we do know how to quantize a string. Hence, the predictions of string-string duality could be tested. One can show, for instance, that the strength with which the solitonic strings interact is given by the inverse of the fundamental string’s interaction strength, in agreement with the conjecture.

In 1994 Christopher M. Hull of Queen Mary and Westfield College at the University of London, along with Paul K. Townsend of the University of Cambridge, suggested that a weakly interacting heterotic string can even be the dual of a strongly interacting Type IIA string, if both are in six dimensions. The barriers between the different string theories were beginning to crumble.

It occurred to me that string-string duality has another unexpected payoff. If we reduce the six-dimensional spacetime to four dimensions by curling up two dimensions, the fundamental string and the solitonic string each acquire a T-duality. But here is the miracle: the T-duality of the solitonic string is just the S-duality of the fundamental string, and vice versa. This phenomenon—in which the interchange of charges in one picture is the inversion of length in the dual picture—is called the Duality of Dualities. It places the previously speculative S-duality on as firm a footing as the well-established T-duality. In addition, it predicts that the strength with which objects interact—their charges—is related to the size of the invisible dimensions. What is charge in one universe may be size in another.

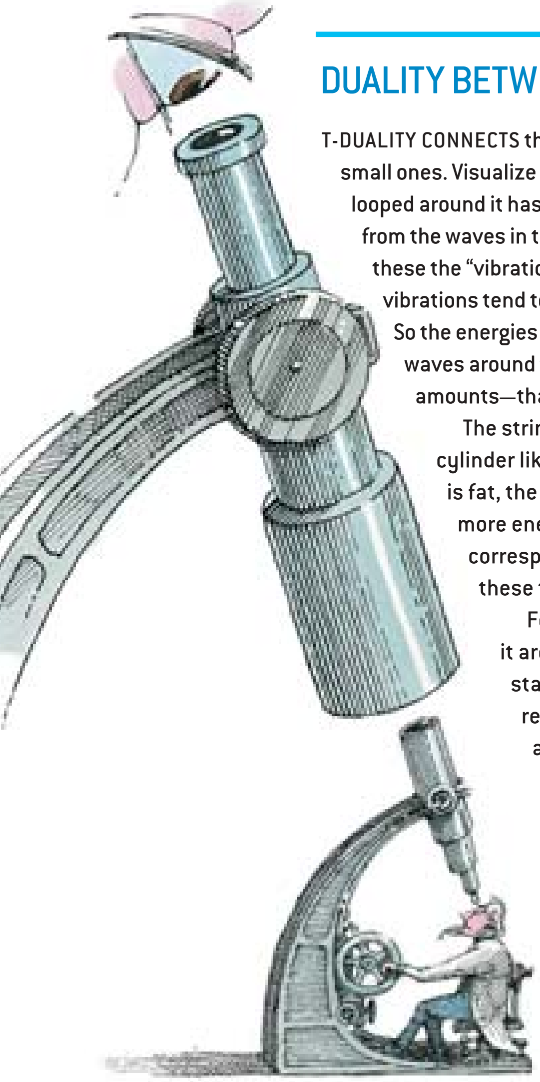
In a landmark talk at the University of Southern California in 1995, Witten



“BRANE” SCAN lists the membranes that arise in spacetimes of different dimensions. A p-brane of dimension 0 is a particle, that of dimension 1 is a string and that of dimension 2 is a sheet or bubble. Some branes have no spin (*red*), but Dirichlet-branes have spin of 1 (*blue*).

DUSAN PETRICIC

DUALITY BETWEEN LARGE AND SMALL

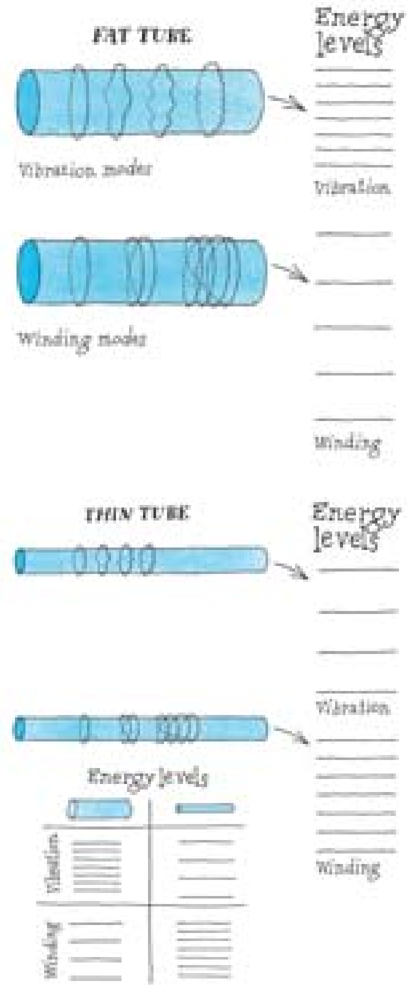


T-DUALITY CONNECTS the physics of large spacetimes with that of small ones. Visualize a curled spacetime as a cylinder. A string looped around it has two kinds of energy states. One set arises from the waves in the string that fit around the cylinder; call these the "vibration" modes. If the cylinder is fat, the vibrations tend to have long wavelengths and less energy. So the energies corresponding to different numbers of waves around the cylinder are separated by small amounts—that is, they are "closely spaced."

The string can, however, also loop around the cylinder like a stretched rubber band. If the cylinder is fat, the string needs to stretch more, requiring more energy. Thus, the energies of the states corresponding to different numbers of loops—call these the "winding" modes—are widely spaced.

For a thin cylinder, the waves fitting around it are small and have high energy; the vibration states are widely spaced. But the loops require less energy, so the winding modes are closely spaced.

To an outside observer, the physical origins of the vibration and winding states are not apparent. Both the thin and the fat tube yield the same energy levels, which physicists interpret as particles. As such, the minute scales of the thin spacetime may yield the same physics as the large scales of our universe. —*M.J.D.*

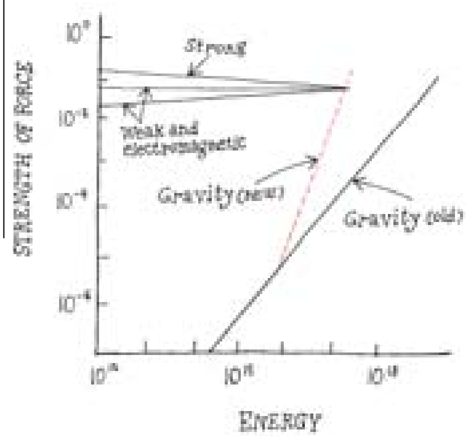


drew together all the work on T-duality, S-duality and string-string duality under the umbrella of M-theory in 11 dimensions. In the following months, literally hundreds of papers appeared on the Internet confirming that whatever M-theory may be, it certainly involves membranes in an important way.

Even the $E_8 \times E_8$ string, whose handedness was thought impossible to derive from 11 dimensions, acquired an origin in M-theory. Witten, along with Petr Horava of Princeton University, showed how to shrink the extra dimension of M-theory into a segment of a line. The resulting picture has two 10-dimensional universes (each at an end of the line) connected by a spacetime of 11 dimensions. Particles—and strings—exist only in the parallel universes at the ends, which can communicate with each other only via gravity. (One can speculate that all visible matter in our universe lies on one wall, whereas the "dark matter," believed to account for the invisible mass in the universe, resides in a parallel universe on the other wall.)

This scenario may have important consequences for confronting M-theory with experiment. For example, physicists know that the intrinsic strengths of all the forces change with the energy of the relevant particles. In supersymmetric theories, one finds that the strengths of the strong, weak and electromagnetic forces all converge at an energy E of 10^{16} giga-electron-volts. Further, the interaction strengths almost equal—but not quite—the value of the dimensionless number GE^2, where G is Newton's gravitational constant. This near miss, most likely not a coincidence, seems to call for an explanation; it has been a source of great frustration for physicists.

But in the bizarre spacetime envisioned by Horava and Witten, one can choose the size of the 11th dimension so that all four forces meet at this common scale. It is far less than the Planck energy of 10^{19} giga-electron-volts, at which gravity was formerly expected to become strong. (High



THREE FORCES CONVERGE to the same strength when particles are as energetic as 10^{16} giga-electron-volts. Until now, gravity was believed to miss this meeting point. But calculations including the 11th dimension of M-theory suggest that gravity may indeed converge.

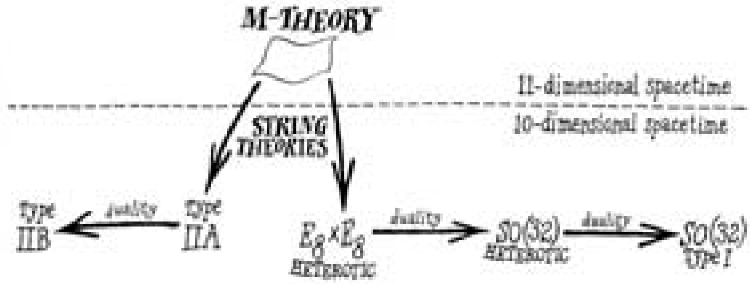
DUSAN PETRICIC

energy is connected to small distance via quantum mechanics. So Planck energy is simply Planck length expressed as energy.) Quantum-gravitational effects may thus be far closer in energy to everyday events than physicists previously believed, a result that would have all kinds of cosmological consequences. The Horava-Witten idea has prompted a variation on the Kaluza-Klein theme known as "braneworld," in which our universe is a three-brane in a higher-dimensional universe. The strong, weak and electromagnetic forces are confined to the brane, but gravity lives in the bulk. The extra dimension may be as a large as a millimeter.

In 1995 Joseph Polchinski of the Institute for Theoretical Physics realized that some p-branes resemble a surface discovered by 19th-century German mathematician Peter G. L. Dirichlet. On occasion these branes can be interpreted as black holes or, rather, black-branes—objects from which nothing, not even light, can escape. Open strings, for instance, may be regarded as closed strings, part of which are hidden behind the black-branes. Such breakthroughs have led to a new interpretation of black holes as intersecting black-branes wrapped around seven curled dimensions. As a result, there are strong hints that M-theory may even clear up the paradoxes of black holes raised by Stephen W. Hawking of the University of Cambridge.

In 1974 Hawking showed that black holes are not entirely black but may radiate energy. In that case, black holes must possess entropy, which measures the disorder by accounting for the number of quantum states available. Yet the microscopic origin of these states stayed a mystery. The technology of Dirichlet-branes has enabled Strominger and Cumrun Vafa of Harvard University to count the number of quantum states in black-branes. They find an entropy that agrees with Hawking's prediction, placing another feather in the cap of M-theory.

Black-branes also promise to solve one of the biggest problems of string theory: there seem to be billions of different ways of crunching 10 dimensions down to four. So there are many competing predictions of how the real world works—in



M-THEORY in 11 dimensions gives rise to the five string theories in 10 dimensions. When the extra dimension curls into a circle, M-theory yields the Type IIA superstring, further related by duality to the Type IIB string. If the extra dimension shrinks to a line segment, M-theory becomes the physically plausible $E_8 \times E_8$ heterotic string, connected to the SO(32) string theories by dualities.

DUSAN PETRICIC

other words, no prediction at all. It turns out, however, that the mass of a black-brane can vanish as a hole it wraps around shrinks. This feature miraculously affects the spacetime itself, allowing one spacetime with a certain number of internal holes to change to another with a different number of holes, violating the laws of classical topology.

If all the spacetimes are thus related, finding the right one becomes a more tractable problem. The string may ultimately choose the spacetime with, say, the lowest energy and inhabit it. Its undulations would then give rise to the elementary particles and forces as we know them—that is, the real world.

In an interesting offshoot of Dirichlet-branes, Juan Maldacena of the Institute for Advanced Study has posed a five-dimensional spacetime known as anti de Sitter space, a negatively curved, saddle-shaped spacetime. This world, including all its gravitational interactions, may be described by a nongravitational theory that resides on its four-dimensional boundary. This may shed light on the four-dimensional quark theories that govern the strong nuclear interactions. If this so-called holographic picture is correct, then the universe is like the wall of Plato's cave, and we are the shadows projected on it.

In another variation, Lisa Randall of Harvard and Raman Sundrum of Johns Hopkins University combine the brane-world and holographic ideas to suggest that our universe is a three-brane sitting on a five-dimensional anti de Sitter space. It has even been suggested that the big bang was simply the collision of two three-branes.

Thus, branes are no longer the ugly ducklings of string theory. They have taken center stage as the microscopic constituents of M-theory, as the higher-dimensional progenitors of black holes and as entire universes in their own right.

10 to 11: Not Too Late

DESPITE ALL THESE successes, physicists are glimpsing only small corners of M-theory; the big picture is still lacking. Physicists have long suspected that unifying gravity—the geometry of spacetime—with quantum physics will lead to spacetime's becoming similarly ill defined, at least until a new definition is discovered. Over the next few years we hope to discover what M-theory really is.

Witten is fond of imagining how physics might develop on a planet where discoveries such as general relativity, quantum mechanics and supersymmetry were made in a different order than on Earth. In a similar vein, I would like to suggest that on planets more logical than ours, 11 dimensions would have been the starting point from which 10-dimensional string theory was subsequently derived. Indeed, future terrestrial historians may judge the late 20th century as a time when theorists were like children playing on the seashore, diverting themselves with the smooth pebbles of superstrings while the great ocean of M-theory lay undiscovered before them. SA

MORE TO EXPLORE

Unity from Duality. Paul Townsend in *Physics World,* Vol. 8, No. 9, pages 1–6; September 1995.

Explaining Everything. Madhusree Mukerjee in *Scientific American,* Vol. 274, No. 1, pages 88–94; January 1996.

Duality, Spacetime and Quantum Mechanics. Edward Witten in *Physics Today,* Vol. 50, No. 5, pages 28–33; May 1997.

The Universe's Unseen Dimensions. Nima Arkani-Hamed, Savas Dimopoulos and Georgi Dvali in *Scientific American,* pages 62–69; August 2000.

BLACK HOLES and the INFORMATION PARADOX

What happens to the information in matter destroyed by a black hole? Searching for that answer, physicists are groping toward a quantum theory of gravity

By Leonard Susskind

SOMEWHERE in outer space, Professor Windbag's time capsule has been sabotaged by his rival, Professor Goulash. The capsule contains a mathematical formula vital to future generations. But Goulash's diabolical scheme to plant a bomb on board has succeeded. Bang! The formula is vaporized into a cloud of electrons, nucleons, photons and an occasional neutrino. Windbag is distraught. He has no record of the formula and cannot remember its derivation.

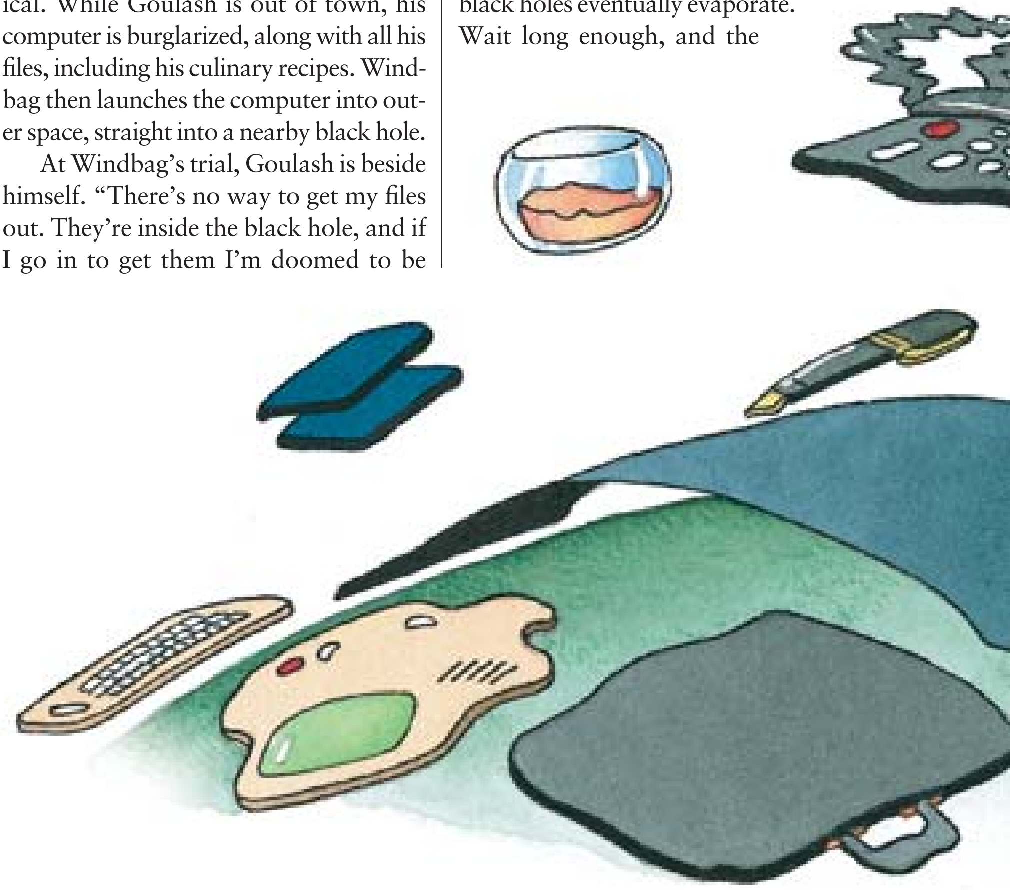
Later, in court, Windbag charges that Goulash has sinned irrevocably: "What that fool has done is irreversible. Off with his tenure!"

"Nonsense," says an unflustered Goulash. "Information can never be destroyed. It's just your laziness, Windbag. All you have to do is go and find each particle in the debris and reverse its motion. The laws of nature are time symmetric, so on reversing everything, your stupid formula will be reassembled. That proves, beyond a shadow of a doubt, that I could never have destroyed your precious information." Goulash wins the case.

Windbag's revenge is equally diabolical. While Goulash is out of town, his computer is burglarized, along with all his files, including his culinary recipes. Windbag then launches the computer into outer space, straight into a nearby black hole.

At Windbag's trial, Goulash is beside himself. "There's no way to get my files out. They're inside the black hole, and if I go in to get them I'm doomed to be crushed. You've truly destroyed information, and you'll pay."

"Objection, Your Honor!" Windbag jumps up. "Everyone knows that black holes eventually evaporate. Wait long enough, and the



BLACK HOLE'S SURFACE looks to Windbag (in the spaceship) like a spherical membrane, called the horizon. Windbag sees Goulash, who is falling into the black hole, being slowed down and flattened at the horizon; according to string theory, Goulash also seems to be spread all over it. Thus, Windbag, who represents the outside observer, sees the information contained in everything that falls into the black hole as stopping at the surface. But Goulash finds himself falling right through the horizon to the center of the black hole, where he becomes crushed.

Updated from the April 1997 issue

black hole will radiate away all its mass and turn into outgoing photons and other particles. True, it may take 10^{70} years, but it's the principle that counts. All Goulash has to do is reverse the paths of the debris, and his computer will come flying back out of the black hole."

"Not so!" cries Goulash. "This is different. My recipe was lost behind the black hole's boundary, its horizon. Once something crosses the horizon, it can never get back out without exceeding the speed of light, and nothing can do that. There is no way the evaporation products, which come from outside the horizon, can contain my recipes even in scrambled form. He's guilty, Your Honor."

Her Honor is confused. "We need some expert witnesses. Professor Hawking, what do you say?"

Stephen W. Hawking of the University of Cambridge comes to the stand. "Goulash is right. In most situations, information is scrambled and in a practical sense is lost. For example, if a new deck of cards is tossed in the air, the original order of the cards vanishes. But in principle, if we know the exact details of how the cards are thrown, the original order can be reconstructed. This is called microreversibility. But in my 1976 paper I showed that the principle of microreversibility, which has always held in classical and quantum physics, is violated by black holes. Because information cannot escape from behind the horizon, black holes are a fundamental new source of irreversibility in nature. Windbag really did destroy information."

Her Honor turns to Windbag: "What do you have to say to that?" Windbag calls on Professor Gerard 't Hooft of Utrecht University in the Netherlands.

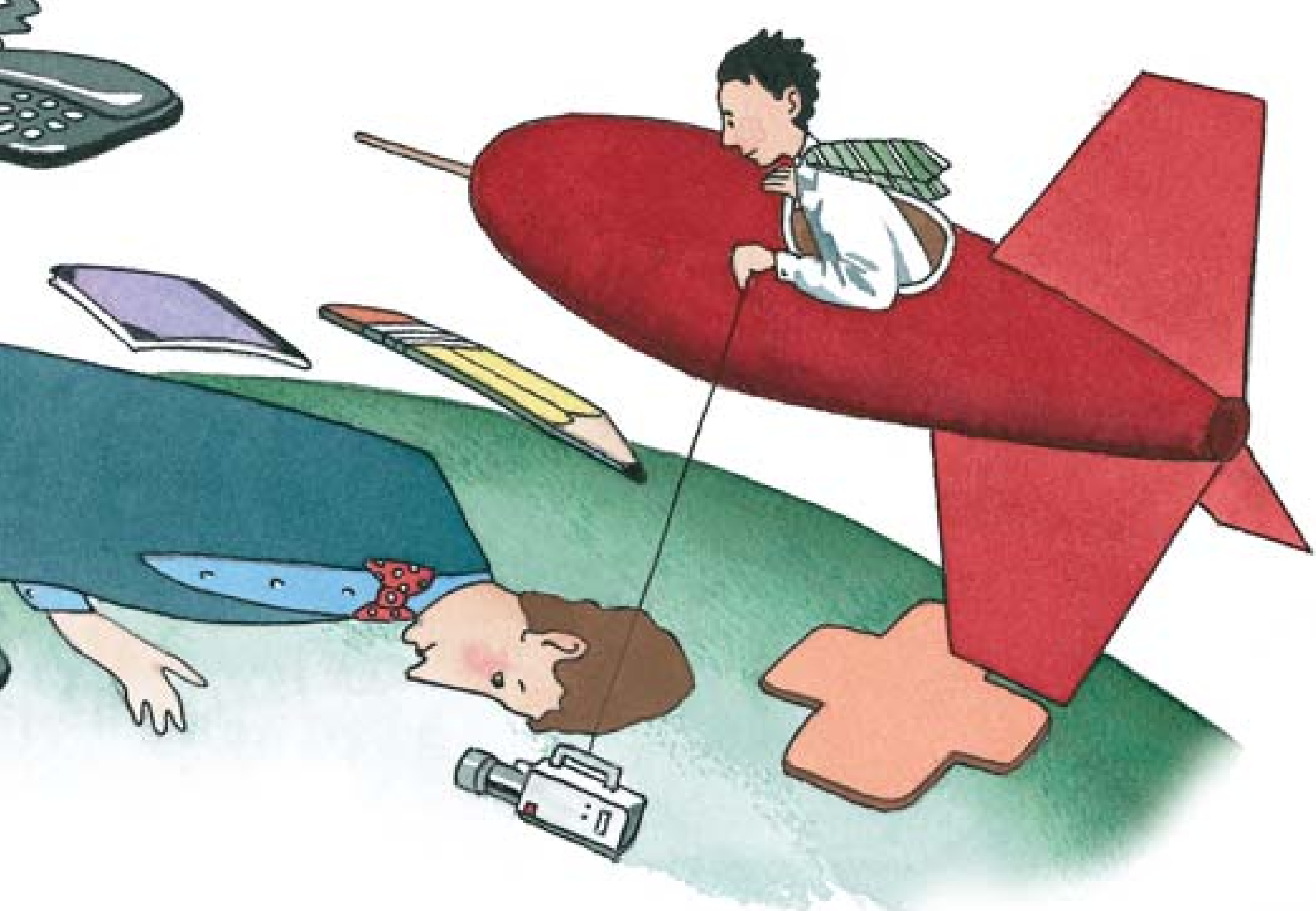
"Hawking is wrong," begins 't Hooft. "I believe black holes must not lead to violation of the usual laws of quantum mechanics. Otherwise the theory would be out of control. You cannot undermine microscopic reversibility without destroying energy conservation. If Hawking were right, the universe would heat up to a temperature of 10^{31} degrees in a tiny fraction of a second. Because this has not happened, there must be some way out."

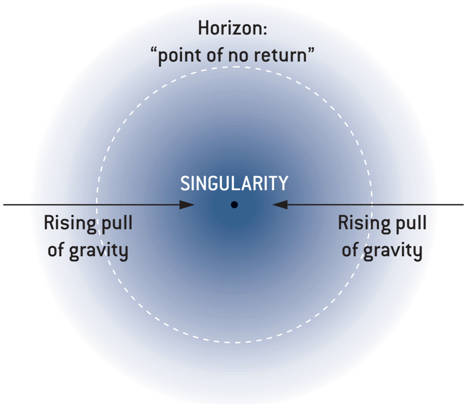
Twenty more famous theoretical physicists are called to the stand. All that becomes clear is that they cannot agree.

The Information Paradox

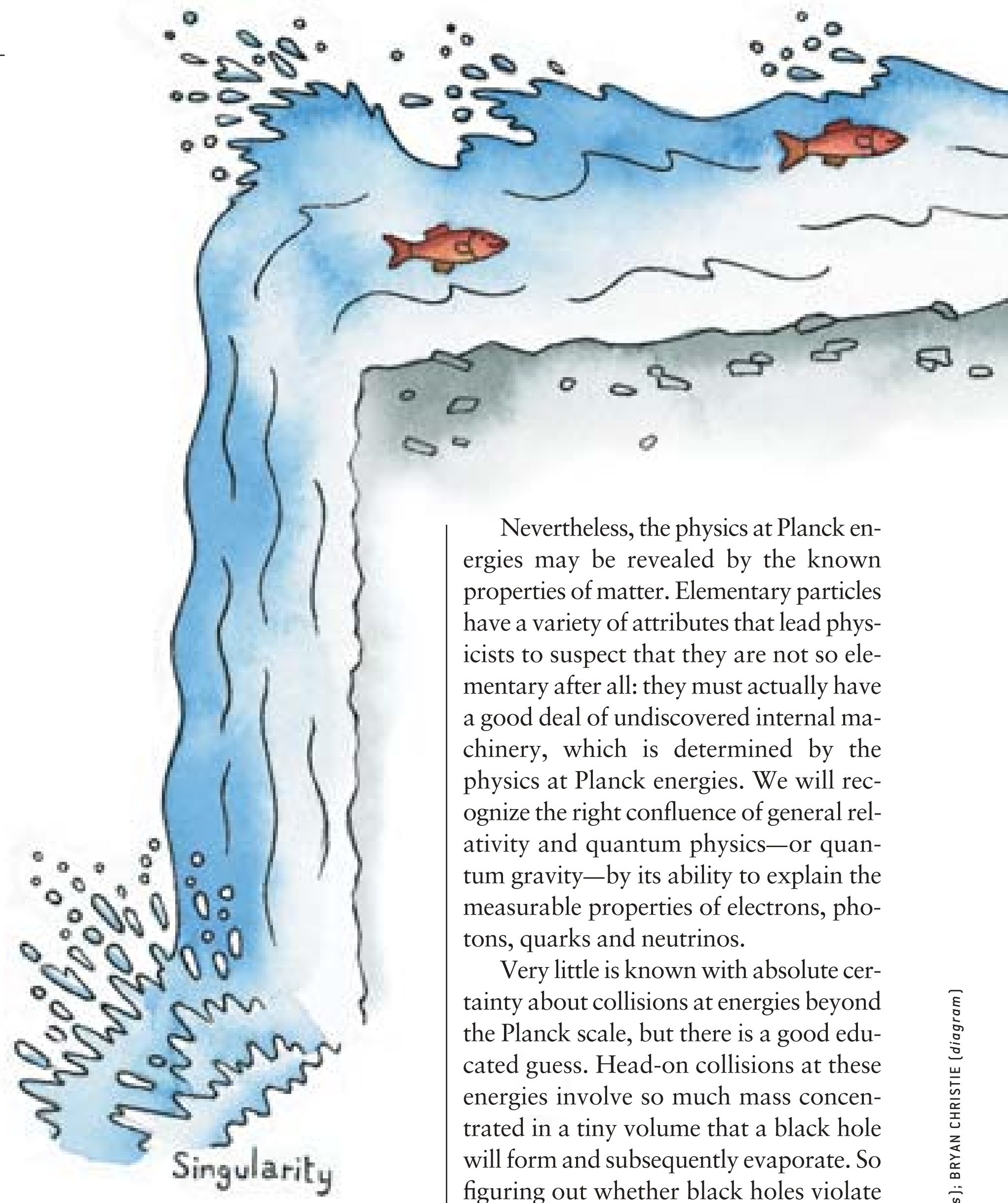
WINDBAG AND GOULASH are, of course, fictitious. Not so Hawking and 't Hooft, nor the controversy of what happens to information that falls into a black hole. Hawking's claim that a black hole consumes information has drawn attention to a potentially serious conflict between quantum mechanics and the general theory of relativity. The problem is known as the information paradox.

When something falls into a black hole, one cannot expect it ever to come flying back out. The information coded in the properties of its constituent atoms is, according to Hawking, impossible to retrieve. Albert Einstein once rejected quan-





INVISIBLE HORIZON is represented in this analogy as a point of no return in a river. To the left of it, water flows faster than a "lightfish" can swim. If a lightfish happens to drift beyond this line, it can never get back upstream; it is doomed to be crushed in the falls. But the fish notices nothing special at the line. Likewise, a light ray or person who is inside the horizon of a black hole can never get out; the object inevitably falls into the singularity at the center but without noticing anything special about the horizon.



tum mechanics with the protest: "God does not play dice." But Hawking states that "God not only plays dice, He sometimes throws the dice where they cannot be seen"—into a black hole.

The problem, 't Hooft points out, is that if the information is truly lost, quantum mechanics breaks down. Despite its famed indeterminacy, quantum mechanics controls the behavior of particles in a very specific way: it is reversible. When one particle interacts with another, it may be absorbed or reflected or may even break up into other particles. But one can always reconstruct the initial configurations of the particles from the final products.

If this rule is broken by black holes, energy may be created or destroyed, threatening one of the most essential underpinnings of physics. The conservation of energy is ensured by the mathematical structure of quantum mechanics, which also guarantees reversibility; losing one means losing the other. As Thomas Banks, Michael Peskin and I showed in 1980 at Stanford University, information loss in a black hole leads to enormous amounts of energy being generated. For such reasons, 't Hooft and I believe the information that falls into a black hole must somehow become available to the outside world.

Some physicists feel the question of what happens in a black hole is academic or even theological. But at stake are the future rules of physics. Processes inside a black hole are merely extreme examples of interactions between elementary particles. At the energies that particles can acquire in today's largest accelerators (about 10^{12} electron volts), the gravitational attraction between them is negligible. But if the particles have a "Planck energy" of about 10^{28} electron volts, so much energy—and therefore mass—becomes concentrated in a tiny volume that gravitational forces outweigh all others. The resulting collisions involve quantum mechanics and the general theory of relativity in equal measure.

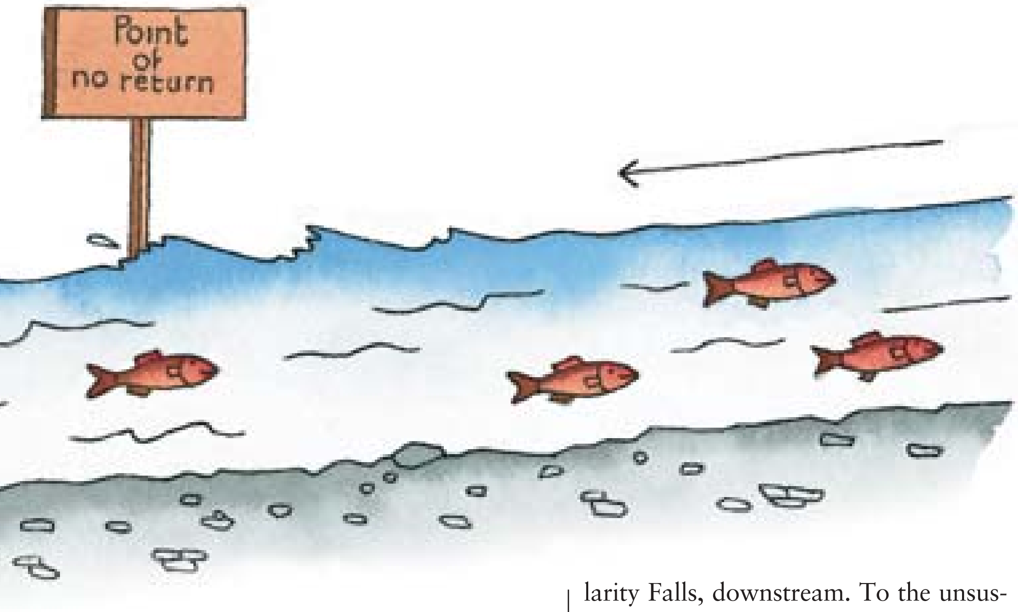
It is to Planckian accelerators that we would nominally look for guidance in building future theories of physics. Alas, Shmuel Nussinov of Tel Aviv University concludes that such an accelerator would have to be at least as big as the entire known universe.

Nevertheless, the physics at Planck energies may be revealed by the known properties of matter. Elementary particles have a variety of attributes that lead physicists to suspect that they are not so elementary after all: they must actually have a good deal of undiscovered internal machinery, which is determined by the physics at Planck energies. We will recognize the right confluence of general relativity and quantum physics—or quantum gravity—by its ability to explain the measurable properties of electrons, photons, quarks and neutrinos.

Very little is known with absolute certainty about collisions at energies beyond the Planck scale, but there is a good educated guess. Head-on collisions at these energies involve so much mass concentrated in a tiny volume that a black hole will form and subsequently evaporate. So figuring out whether black holes violate the rules of quantum mechanics or not is essential to unraveling the ultimate structure of particles.

A black hole is born when so much mass or energy gathers in a small volume that gravitational forces overwhelm all others and everything collapses under its own weight. The material squeezes into an unimaginably small region called a singularity, the density inside of which is essentially infinite. Surrounding the singularity is an imaginary surface called the horizon. For a black hole with the mass of a galaxy, the horizon is 10^{11} kilometers

YAN NASCIMBENE (*illustrations on these and preceding pages*); BRYAN CHRISTIE (*diagram*)



from the center—as far as the outermost reaches of the solar system are from the sun. For a black hole of solar mass, the horizon is roughly a kilometer away; for a black hole with the mass of a small mountain, the horizon is 10^{-13} centimeter away, roughly the size of a proton.

The horizon separates space into two regions that we can think of as the interior and exterior of the black hole. Suppose that Goulash, who is scouting for his computer near the black hole, shoots a particle away from the center. If he is not too close and the particle has a high velocity, then it may overcome the gravitational pull of the black hole and fly away. It will be most likely to escape if it is shot with the maximum velocity—that of light. If, however, Goulash is too close to the singularity, the gravitational force will be so great that even a light ray will be sucked in. The horizon is the place with the (virtual) warning sign: POINT OF NO RETURN. No particle or signal of any kind can cross it from the inside to the outside.

At the Horizon

AN ANALOGY inspired by William G. Unruh of the University of British Columbia, a pioneer in black hole quantum mechanics, helps to explain the relevance of the horizon. Imagine a river that gets swifter downstream. Among the fish that live in it, the fastest swimmers are the "lightfish." But at some point, the river flows at the fish's maximum speed; clearly, any lightfish that drifts behind this point can never get back up. It is doomed to be crushed on the rocks below Singularity Falls, downstream. To the unsuspecting lightfish, though, passing the point of no return is a nonevent. No currents or shock waves warn it of the crossing.

What happens to Goulash, who in a careless moment gets too close to the black hole's horizon? Like the freely drifting fish, he senses nothing special: no great forces, no jerks or flashing lights. His pulse and breathing rate remain normal. To him the horizon is just like any other place.

But Windbag, watching Goulash from a spaceship safely outside the horizon, sees Goulash acting in a bizarre way. Windbag has lowered to the horizon a cable equipped with a camcorder and other probes. As Goulash falls toward the black hole, his speed increases until it approaches that of light. Einstein found that if two persons are moving fast relative to each other, each sees the other's clock slow down; in addition, a clock that is near a massive object will run slowly compared with one in empty space. Windbag sees an oddly lethargic Goulash. As he falls, the latter shakes his fist at Windbag, but Windbag sees Goulash's motions slow to a halt. Although Goulash falls through the horizon, Windbag never quite sees him get there.

In fact, not only does Goulash seem to slow down, but his body looks as if it is being squashed into a thin layer. Einstein also showed that if two persons move fast with respect to each other, each will see the other as being flattened in the direction of motion. More strangely, Windbag should also see all the material that ever fell into the black hole, including the original matter that made it up—and Goulash's computer—similarly flattened and frozen at the horizon. With respect to an outside observer, all of that matter suffers a relativistic time dilation. To Windbag, the black hole consists of an immense junkyard of flattened matter at its horizon. But Goulash sees nothing unusual until much later, when he reaches the singularity, there to be crushed by ferocious forces.

Black hole theorists have discovered over the years that from the outside, the properties of a black hole can be described in terms of a mathematical membrane above the horizon. This layer has many physical qualities, such as electrical conductivity and viscosity. Perhaps the most surprising of its properties was postulated in the early 1970s by Hawking, Unruh and Jacob D. Bekenstein of the Hebrew University of Jerusalem. They found that as a consequence of quantum mechanics, a black hole—in particular, its horizon—behaves as though it contains heat. The horizon is a layer of hot material of some kind.

The temperature of the horizon depends on where it is measured. Suppose one of the probes that Windbag has attached to his cable is a thermometer. Far from the horizon he finds that the temperature is inversely proportional to the black hole's mass. For a black hole of solar mass, this "Hawking temperature" is about 10^{-8} degree—far colder than intergalactic space. As Windbag's thermometer approaches the horizon, however, it registers higher. At a distance of a centimeter, it measures about a thousandth of a degree; at a nuclear diameter, it records 10 billion degrees. The temperature ultimately becomes so high that no imaginable thermometer could measure it.

THE AUTHOR

LEONARD SUSSKIND is one of the early inventors of string theory. He holds a Ph.D. from Cornell University and has been a professor at Stanford University since 1978. He has made many contributions to elementary particle physics, quantum field theory, cosmology and, most recently, the theory of black holes. His current studies in gravitation have led him to suggest that information can be compressed into one lower dimension, a concept he calls the holographic universe.

Hot objects also possess an intrinsic disorder called entropy, which is related to the amount of information a system can hold. Think of a crystal lattice with *N* sites; each site can house one atom or none at all. Thus, every site holds one "bit" of information, corresponding to whether an atom is there or not; the total lattice has *N* such bits and can contain *N* units of information. Because there are two choices for each site and *N* ways of combining these choices, the total system can be in any one of 2^N states (each of which corresponds to a different pattern of atoms). The entropy (or disorder) is defined as the logarithm of the number of possible states. It is roughly equal to *N*—the same number that quantifies the capacity of the system for holding information.

Bekenstein found that the entropy of a black hole is proportional to the area of its horizon. The precise formula, derived by Hawking, predicts an entropy of 3.2×10^{64} per square centimeter of horizon area. Whatever physical system carries the bits of information at the horizon must be extremely small and densely distributed: their linear dimensions have to be $1/10^{20}$ the size of a proton's. They must also be quite special for Goulash to miss them completely as he passes through.

The discovery of the thermodynamic properties of black holes led Hawking to a very interesting conclusion. Like other hot bodies, a black hole must radiate energy and particles into the surrounding space. The radiation comes from the horizon and does not violate the rule that nothing can escape from within. But it causes the black hole to lose energy and mass. In time an isolated black hole radiates away all its mass and vanishes.

All of the above, though peculiar, has been known to relativists for some decades. The true controversies arise when, following Hawking, we seek the fate of the information that fell into the black hole during and after its formation. In particular, can it be carried away by the evaporation products—albeit in a very scrambled form—or is it lost forever behind the horizon?

Goulash, who followed his computer into the black hole, would insist that its contents passed behind the horizon, where they were lost to the outside world; this in a nutshell is Hawking's argument. The opposing point of view might be described by Windbag: "I saw the computer fall toward the horizon, but I never saw it fall through. The temperature and radiation grew so intense I lost track of it. I believe the computer was vaporized; later its energy and mass came back out in the form of thermal radiation. The consistency of quantum mechanics requires that this evaporating energy also carried away all the information in the computer." This is the position that 't Hooft and I take.

Black Hole Complementarity

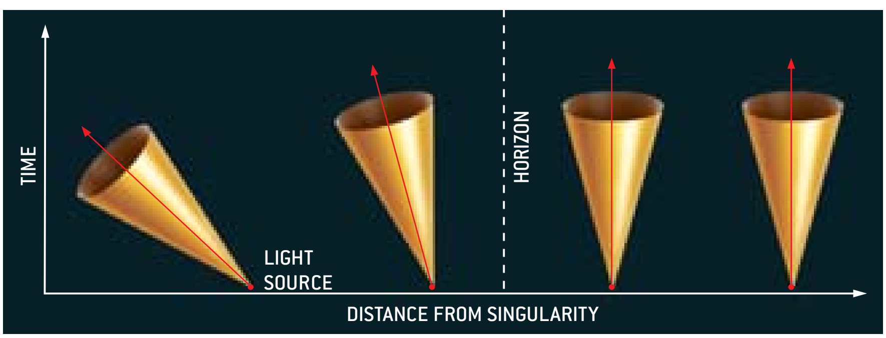
IS IT POSSIBLE that Goulash and Windbag are in a sense both correct? Can it be that Windbag's observations are indeed consistent with the hypothesis that Goulash and his computer are thermalized and radiated back into space before ever reaching the horizon, even though Goulash discovers nothing unusual until long after, when he encounters the singularity? The idea that these are not contradictory but complementary scenarios was first put forward as the principle of black hole complementarity by Lárus Thorlacius, John Uglum and me at Stanford. Very similar ideas are also found in 't Hooft's work. Black hole complementarity is a new principle of relativity. In the special theory of relativity, we find that although different observers disagree about the lengths of time and space intervals, events take place at definite spacetime locations. Black hole complementarity does away with even that.

Suppose that Windbag, whose cable is also equipped with a powerful microscope, watches an atom fall toward the horizon. At first he sees the atom as a nucleus surrounded by a blur of negative charge. But as the atom gets closer to the black hole, its internal motions seem to slow down and the electrons become visible. A little later the electrons freeze, and the protons and neutrons start to show up. Later yet, the quarks making up these particles are revealed. (Goulash, who falls with the atom, sees no changes.)

Quite a few physicists believe elementary particles are made of even smaller constituents. Although there is no definitive theory for this machinery, one candidate stands out: string theory. In this theory, an elementary particle does not resemble a point; rather it is like a tiny rubber band that can vibrate in many modes. The fundamental mode has the lowest frequency; then there are higher harmonics, which can be superimposed on top of one another. There are an infinite number of such modes, each of which corresponds to a different elementary particle.

Here another analogy helps. One cannot see the wings of a hovering hummingbird, because its wings flutter too fast. But in a photograph taken with a fast shutter speed, one can see the wings—so the bird looks bigger. If a hummer falls into the black hole, Windbag will see its wings take form as the bird approaches the horizon and the vibrations appear to slow down; it seems to grow. Now suppose that the wings have feathers that flap even faster. Soon these, too, would come into view, adding further to the apparent size of the bird. Windbag sees the hummer enlarge continuously. But Goulash, who falls with the bird, sees no such strange growth.

Like the hummingbird's wings, the string's oscillations are usually too rapid



LIGHT CONES describe the path of light rays emanating from a point. Outside the horizon the cones point upward—that is, forward in time. But inside, the cones tip so that light falls into the black hole's center.

BRYAN CHRISTIE

to detect. A string is a minute object, $1/10^{20}$ the size of a proton. But as it falls into a black hole, its vibrations slow down and more of them become visible. Mathematical studies done at Stanford by Thorlacius, Amanda W. Peet, Arthur Mezhlumian and me have demonstrated the behavior of a string as its higher modes freeze out. The string spreads and grows, just as if it were being bombarded by particles and radiation in a very hot environment. In a relatively short time the string and all the information it carries are smeared over the entire horizon.

This picture applies to all the material that ever fell into the black hole—because according to string theory, everything is ultimately made of strings. Each elementary string spreads and overlaps all the others until a dense tangle covers the horizon. Each minute segment of string, measuring 10^{-33} centimeter across, functions as a bit. Thus, strings provide a means for the black hole's surface to hold the immense amount of information that fell in during its birth and thereafter.

CASCADE OF VIBRATIONS on a string slows down and becomes visible if the string falls into a black hole. Strings are small enough to encode all the information that ever fell into a black hole, thereby offering a way out of the information paradox.

String Theory

IT SEEMS, THEN, that the horizon is made of all the substance in the black hole, resolved into a giant tangle of strings. The information, as far as an outside observer is concerned, never actually fell into the black hole; it stopped at the horizon and was later radiated back out. String theory offers a concrete realization of black hole complementarity and therefore a way out of the information paradox. To outside observers—that is, us—information is never lost. Most important, it appears that the bits at the horizon are minute segments of strings.

Tracing the evolution of a black hole from beginning to end is far beyond the current techniques available to string theorists. But some exciting new results are giving quantitative flesh to these ghostly ideas. Mathematically, the most tractable black holes are the "extremal" black holes. Whereas black holes that have no electrical charge evaporate until all their mass is radiated away, black holes with electrical or (in theory) magnetic charge cannot do that; their evaporation ceases when the gravitational attraction equals the electrostatic or magnetostatic repulsion of whatever is inside the black hole. The remaining stable object is called an extremal black hole.

Ashoke Sen of the Tata Institute of Fundamental Research (TIFR) in Mumbai, India, showed in 1995 that for certain extremal black holes with electrical charge, the number of bits predicted by string theory exactly accounts for the entropy as measured by the area of the horizon. This agreement was the first powerful evidence that black holes are consistent with quantum-mechanical strings.

Sen's black holes were, however, microscopic. More recently, Andrew Strominger of the University of California at Santa Barbara, Cumrun Vafa of Harvard University and, slightly later, Curtis G. Callan and Juan Maldacena of Princeton University extended this analysis to black holes with both electrical and magnetic charge. These new black holes could be large enough to allow Goulash to fall through unharmed. Again, the theorists find complete consistency.

Two groups have done an even more exciting new calculation of Hawking radiation: Sumit R. Das of TIFR, with Samir Mathur of the Massachusetts Institute of Technology; and Avinash Dhar, Gautam Mandal and Spenta R. Wadia, also at TIFR. The researchers studied the process by which an extremal black hole with some excess energy or mass radiates off this flab. String theory fully accounted for the Hawking radiation that was produced. Just as quantum mechanics describes the radiation of an atom by showing how an electron jumps from a high-energy "excited" state to a low-energy "ground" state, quantum strings seem to account for the spectrum of radiation from an excited black hole. The information paradox is well on its way to being resolved. Windbag will be right.

The principle of black hole complementarity has received spectacular mathematical confirmation by Maldacena and others. Following the introduction by 't Hooft and myself of a so-called holographic principle, Maldacena discovered a powerful "holographic" equivalence between quantum gravity in a dimension called anti de Sitter space and a conventional quantum system. He gives a compelling argument that information in black holes in this space is never lost behind the horizon. As a result of Maldacena's work, physicists have made black hole complementarity one of the working assumptions of modern string theory.

Quantum mechanics, I believe, will in all likelihood turn out to be consistent with the theory of gravitation; these two great streams of physics are merging into a quantum theory of gravity based on string theory. The information paradox has played an extraordinary role in this ongoing revolution in physics. And although Goulash would never admit it, Windbag will probably turn out to be right: his recipe for *matelote d'anguilles* is not forever lost to the world. SA

BRYAN CHRISTIE

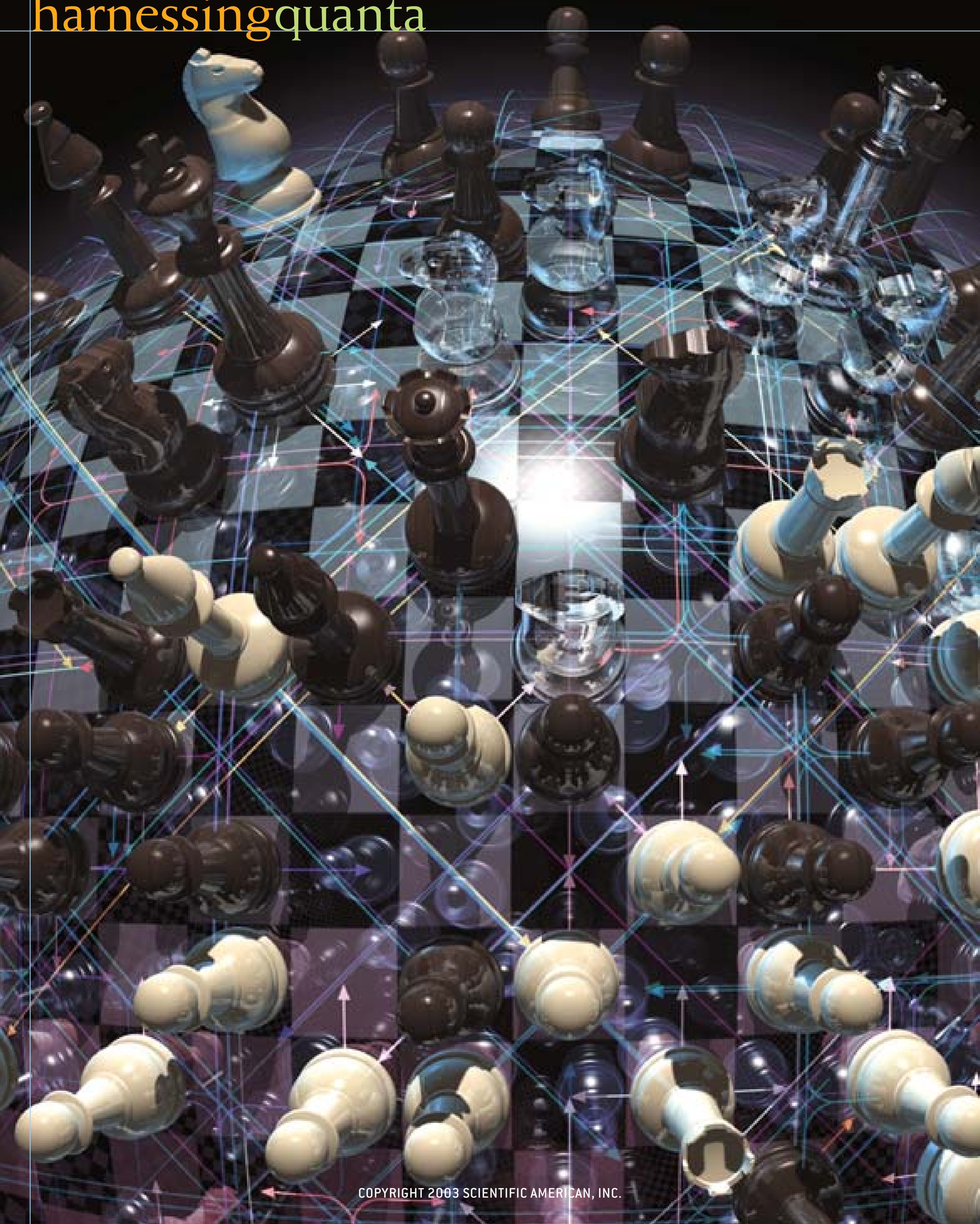
MORE TO EXPLORE

Black Holes and Time Warps: Einstein's Outrageous Legacy. Kip S. Thorne. W. W. Norton, 1994.

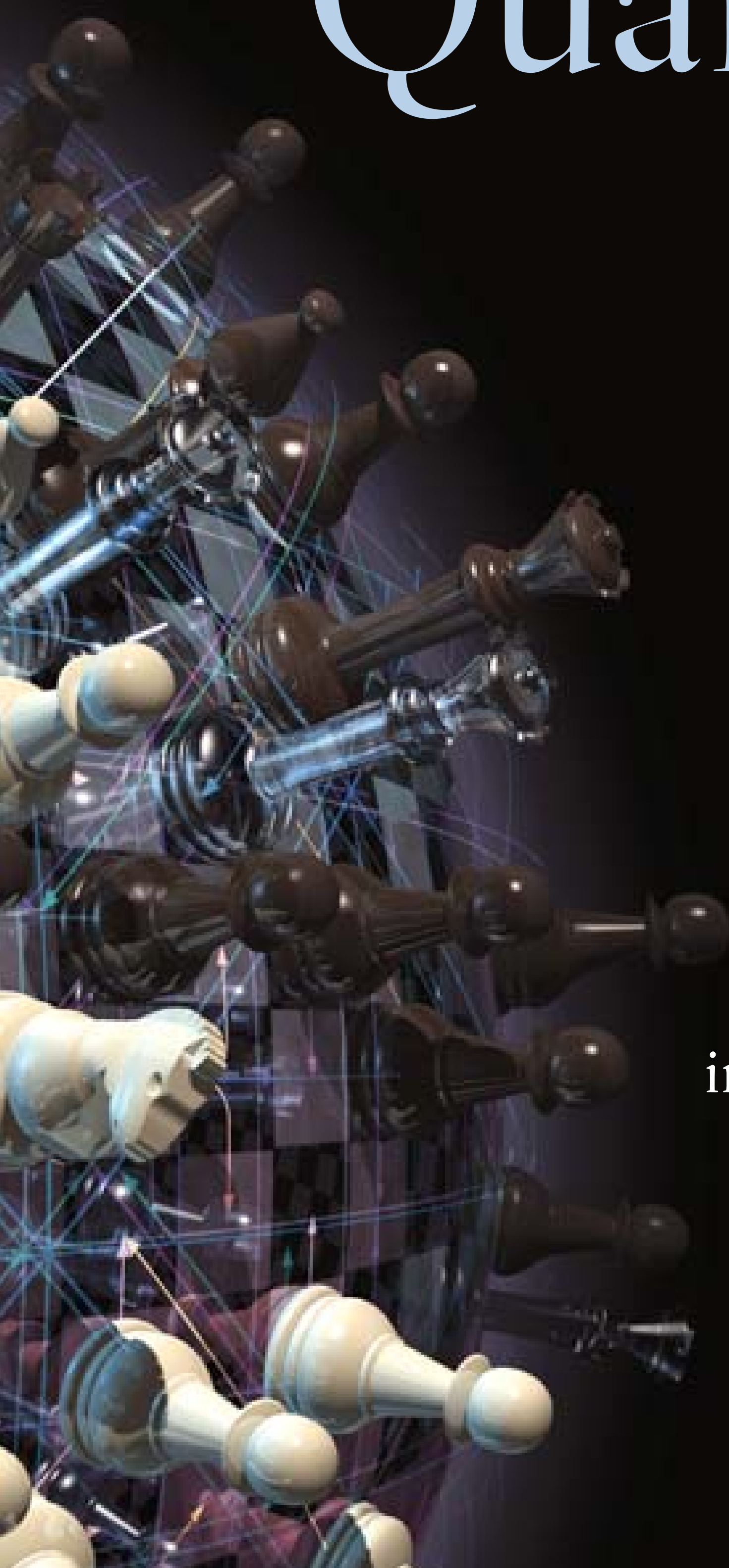
The Illustrated A Brief History of Time. Stephen W. Hawking. Bantam Books, 1996.

Trends in Theoretical Physics: Explaining Everything. Madhusree Mukerjee in *Scientific American*, Vol. 274, No. 1, pages 88–94; January 1996.

harnessingquanta



Simple Rules for a Complex Quantum World



An exciting new fundamental discipline of research combines information science and quantum mechanics

By Michael A. Nielsen

Over the past few decades, scientists have learned that simple

rules can give rise to very rich behavior. A good example is chess. Imagine you're an experienced chess player introduced to someone claiming to know the game. You play a few times and realize that although this person knows the rules of chess, he has no idea how to play well. He makes absurd moves, sacrificing his queen for a pawn and losing a rook for no reason at all. He does not truly *understand* chess: he is ignorant of the high-level principles and heuristics familiar to any knowledgeable player. These principles are collective or emergent properties of chess, features not immediately evident from the rules but arising from interactions among the pieces on the chessboard.

Scientists' current understanding of quantum mechanics is like that of a slow-learning student of chess. We've known the rules for more than 70 years, and we have a few clever moves that work in some special situations, but we're only gradually learning the high-level principles that are needed to play a skillful overall game.

The discovery of these principles is the goal of quantum information science, a fundamental field that is opening up in response to a new way of comprehending the world. Many articles about quantum information science focus on technological applications: research groups "teleport" quantum states from one location to another. Other physicists use quantum states to create cryptographic keys that are absolutely secure from eavesdropping. Information scientists devise algorithms for hypothetical quantum-mechanical computers, much faster than the best known algorithms for conventional, or classical, computers.

These technologies are fascinating, but they obscure the fact that they are a by-product of investigations into deep new scientific questions. Applications such as quantum teleportation play a role similar to the steam engines and other machines that spurred the development of thermodynamics in the 18th and 19th centuries. Thermodynamics was motivated by profound, basic questions about how energy, heat and temperature are related, the transformations among these quantities in physical processes, and the key role of entropy.

Similarly, quantum information scientists are fathoming the relation between classical and quantum units of information, the novel ways that quantum information can be processed, and the pivotal importance of a quantum feature called entanglement, which entails peculiar connections between different objects.

Popular accounts often present entanglement as an all-or-nothing property in which quantum particles are either entangled or not. Quantum information science has revealed that entanglement is a quantifiable physical resource, like energy, that enables information-processing tasks: some systems have a little entanglement; others have a lot. The more entanglement available, the better suited a system is to quantum information processing.

Furthermore, scientists have begun to develop powerful quantitative laws of entanglement (analogous to the laws of thermodynamics governing energy), which provide a set of high-level principles for understanding the behavior of entanglement and describing how we can use it to do information processing.

Quantum information science is new enough that researchers are still coming to grips with its very nature, and they disagree about which questions lie at its heart. From my point of view, the central goal of quantum information science is to develop general principles, like the laws of entanglement, that will enable us to understand complexity in quantum systems.

Overview/*Quantum Information*

- Information is not purely mathematical. Instead it always has a physical embodiment. In traditional information science the embodiment follows classical, or nonquantum, physics. The burgeoning field of quantum information science puts information in a quantum context.
- The basic resource of classical information is the bit, which is always either a 0 or a 1. Quantum information comes in quantum bits, or qubits (pronounced "cue-bits"). Qubits can exist in superpositions, which simultaneously involve 0 and 1, and groups of qubits can be "entangled," which gives them counterintuitive correlations.
- Quantum computers processing qubits, particularly entangled qubits, can outperform classical computers. Entanglement behaves like a resource, similar to energy, that can be used to do quantum information processing.
- The goal of quantum information science is to understand the general high-level principles that govern complex quantum systems such as quantum computers. These principles relate to the laws of quantum mechanics in the way that heuristics for skillful play at chess relate to the game's basic rules.

Complexity and Quanta

NUMEROUS STUDIES in complexity concentrate on systems, such as the weather or piles of sand, that are described by classical physics rather than quantum physics. That focus is natural because complex systems are usually macroscop-

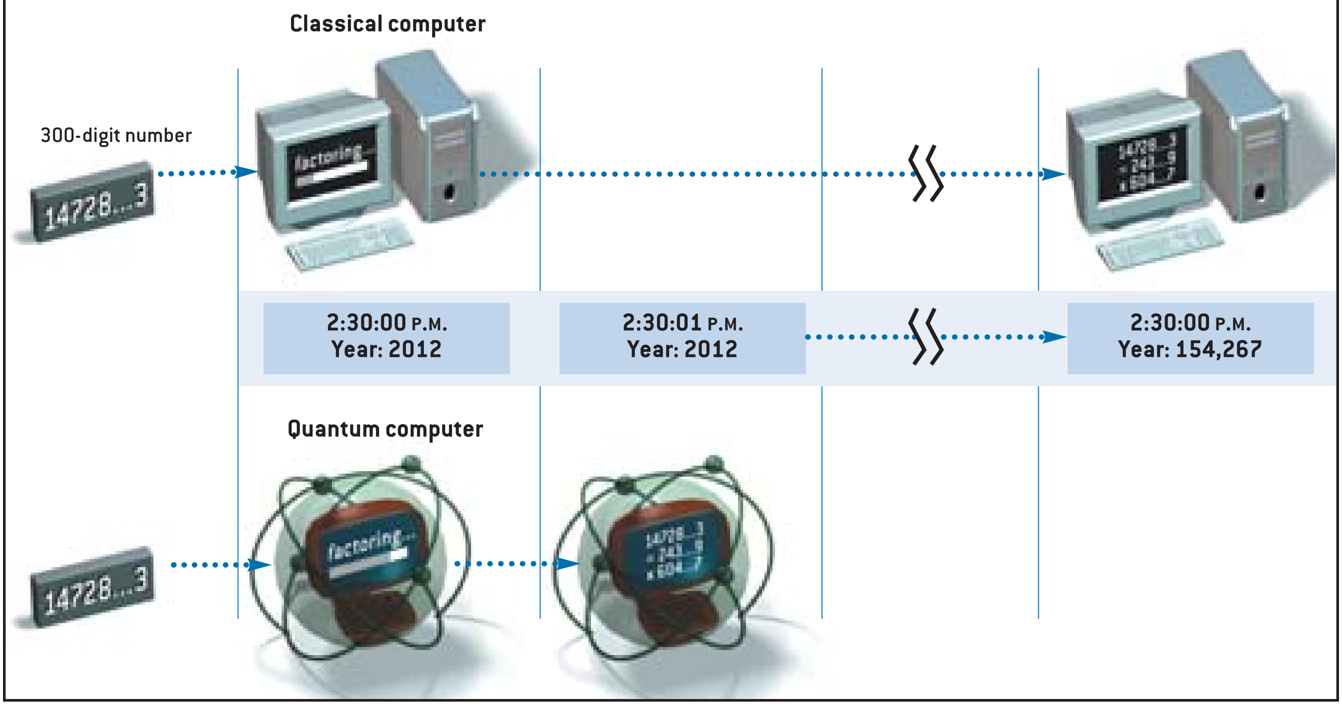
SLIM FILMS (*preceding pages*)

THE FUNDAMENTAL QUESTION

MUCH OF INFORMATION SCIENCE, both classical and quantum, can be summed up by analyzing variants of a basic question:

"What quantity of an information resource is needed to perform a specific information-processing task?"

For example: "How many computational steps are needed to find the prime factors of a 300-digit number?" The best classical algorithm known would take about 5×10^{24} steps, or about 150,000 years at terahertz speed. By taking advantage of innumerable quantum states, a quantum factoring algorithm would take only 5×10^{10} steps, or less than a second at terahertz speed.



ic, containing many constituent parts, and most systems lose their quantum nature as their size is increased. This quantum-to-classical transition occurs because large quantum systems generally interact strongly with their environment, causing a process of decoherence, which destroys the system's quantum properties [see "100 Years of Quantum Mysteries," by Max Tegmark and John A. Wheeler; SCIENTIFIC AMERICAN, February 2001].

As an example of decoherence, think of Erwin Schrödinger's famous cat inside a box. In principle, the cat ends up in a weird quantum state, somewhere between dead and alive; it makes no sense to describe it as either one or the other. In a real experiment, however, the cat interacts with the box by exchange of light, heat and sound, and the box similarly interacts with the rest of the world. In nanoseconds, these processes destroy the delicate quantum states inside the box and replace them with states describable, to a good approximation, by the laws of classical physics. The cat inside really is either alive or dead, not in some mysterious nonclassical state that combines the two.

The key to seeing truly quantum behavior in a complex system is to isolate the system extremely well from the rest of the world, preventing decoherence and preserving fragile quantum states. This isolation is relatively easy to achieve with small systems, such as atoms suspended in a magnetic trap in a vacuum, but is much more difficult with the larger ones in which complex behavior may be found. Accidental laboratory discoveries of remarkable phenomena such as superconductivity and the quantum Hall effect are examples in which physicists have achieved large, well-isolated quantum systems. These phenomena demonstrate that the simple rules of quantum mechanics can give rise to emergent principles governing complex behaviors.

Resources and Tasks

WE ATTEMPT TO understand the high-level principles that govern in those rare instances when the quantum and the complex meet by abstracting, adapting and extending tools from classical information theory. In 2001 Benjamin W. Schumacher of Kenyon College proposed that the essential elements of information science, both classical and quantum, can be summarized as a three-step procedure:

1. Identify a *physical resource*. A familiar classical example is a string of bits. Although bits are often thought of as abstract entities—0's and 1's—all information is inevitably encoded in real physical objects, and thus a string of bits should be regarded as a physical resource.

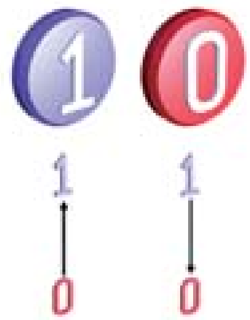
2. Identify an *information-processing task* that can be performed using the physical resource of step 1. A classical example is the two-part task of compressing the output from an information source (for example, the text in a book) into a bit string and then decompressing it—that is, recovering the original information from the compressed bit string.

3. Identify a *criterion for successful completion* of the task of step 2. In our example, the criterion could be that the

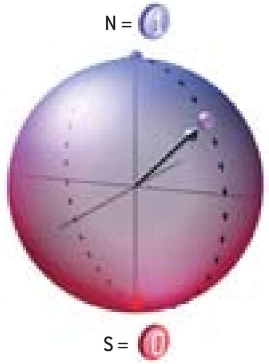
BRYAN CHRISTIE DESIGN

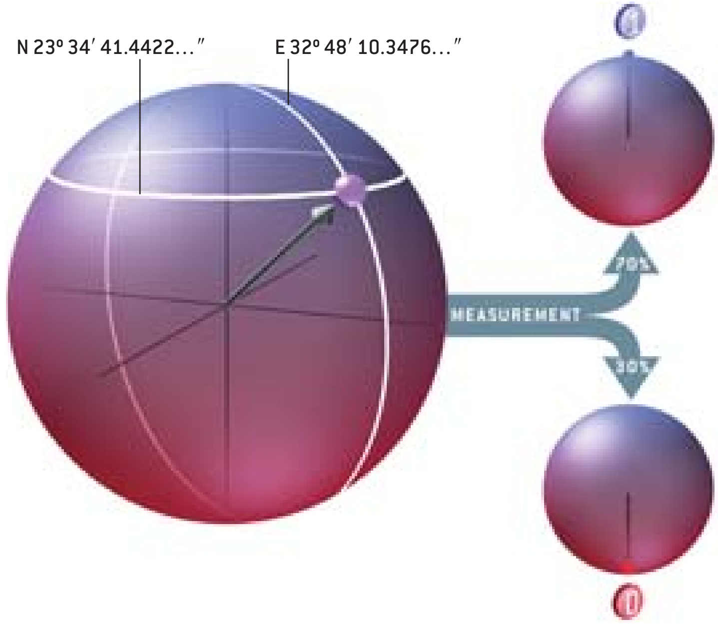
QUBITS EXPLAINED

A BIT can have one of two states: 0 or 1. A bit can be represented by a transistor switch set to "off" or "on" or abstractly by an arrow pointing up or down.



A QUBIT, the quantum version of a bit, has many more possible states. The states can be represented by an arrow pointing to a location on a sphere. The north pole is equivalent to 1, the south pole to 0. The other locations are quantum superpositions of 0 and 1.





A QUBIT MIGHT SEEM TO CONTAIN an infinite amount of information because its coordinates can encode an infinite sequence of digits. But the information in a qubit must be extracted by a measurement. When the qubit is measured, quantum mechanics requires that the result is always an ordinary bit—a 0 or a 1. The probability of each outcome depends on the qubit's "latitude."

output from the decompression stage perfectly matches the input to the compression stage.

The fundamental question of information science is then "What is the minimal quantity of the physical resource (1) we need to perform the information-processing task (2) in compliance with the success criterion (3)?" Although this question does not quite capture all of information science, it provides a powerful lens through which to view much research in the field [*see box on preceding page*].

The data-compression example corresponds to a basic question of classical information science—namely, what is the minimum number of bits needed to store the information produced by some source? This problem was solved by Claude E. Shannon in his famous 1948 papers founding information theory. In so doing, Shannon quantified the information content produced by an information source, defining it to be the minimum number of bits needed to reliably store the output of the source. His mathematical expression for the information content is now known as the Shannon entropy.

Shannon's entropy arises as the answer to a simple, fundamental question about classical information processing. It is perhaps not surprising, then, that studying the properties of the Shannon entropy has proved fruitful in analyzing processes far more complex than data compression. For example, it plays a central role in calculating how much information can be transmitted reliably through a noisy communications channel and even in understanding phenomena such as gambling and the behavior of the stock market. A general theme in information science is that questions about elementary processes lead to unifying concepts that stimulate insight into more complex processes.

In quantum information science, all three elements of Schumacher's list take on new richness. What novel physical resources are available in quantum mechanics? What information-processing tasks can we hope to perform? What are appropriate criteria for success? The resources now include superposition states, like the idealized alive and dead cat of Schrödinger. The processes can involve manipulations of entanglement (mysterious quantum correlations) between widely separated objects. The criteria of success become more subtle than in the classical case, because to extract the result of a quantum information-processing task we must observe, or measure, the system—which almost inevitably changes it, destroying the special superposition states that are unique to quantum physics.

Qubits

QUANTUM INFORMATION science begins by generalizing the fundamental resource of classical information—bits—to quantum bits, or qubits. Just as bits are ideal objects abstracted from the principles of classical physics, qubits are ideal quantum objects abstracted from the principles of quantum mechanics. Bits can be represented by magnetic regions on disks, voltages in circuitry, or graphite marks made by a pencil on paper. The functioning of these classical physical states as bits does not depend on the de-

BRYAN CHRISTIE DESIGN

tails of how they are realized. Similarly, the properties of a qubit are independent of its specific physical representation as the spin of an atomic nucleus, say, or the polarization of a photon of light.

A bit is described by its state, 0 or 1. Likewise, a qubit is described by its quantum state. Two possible quantum states for a qubit correspond to the 0 and 1 of a classical bit. In quantum mechanics, however, *any* object that has two different states necessarily has a range of other possible states, called superpositions, which entail both states to varying degrees. The allowed states of a qubit are precisely all those states that must be available, in principle, to a classical bit that is transplanted into a quantum world. Qubit states correspond to points on the surface of a sphere, with the 0 and 1 being the south and north poles [*see box on opposite page*]. The continuum of states between 0 and 1 fosters many of the extraordinary properties of quantum information.

How much classical information can we store in a qubit? One line of reasoning suggests the amount is infinite: To specify a quantum state we need to specify the latitude and longitude of the corresponding point on the sphere, and in principle each may be given to arbitrary precision. These numbers can encode a long string of bits. For example, 011101101… could be encoded as a state with latitude 01 degrees, 11 minutes and 01.101… seconds.

This reasoning, though plausible, is incorrect. One can encode an infinite amount of classical information in a single qubit, but one can never retrieve that information from the qubit. The simplest attempt to read the qubit's state, a standard direct measurement of it, will give a result of either 0 or 1, south pole or north pole, with the probability of each outcome determined by the latitude of the original state. You could have chosen a different measurement, perhaps using the "Melbourne–Azores Islands" axis instead of north-south, but again only one bit of information would have been extracted, albeit one governed by probabilities with a different dependence on the state's latitude and longitude. Whichever measurement you choose erases all the information in the qubit except for the single bit that the measurement uncovers.

The principles of quantum mechanics prevent us from ever extracting more than a single bit of information, no matter how cleverly we encode the qubit or how ingeniously we measure it afterward. This surprising result was proved in 1973 by Alexander S. Holevo of the Steklov Mathematical Institute in Moscow, following a 1964 conjecture by J. P. Gordon of AT&T Bell Laboratories. It is as though the qubit contains hidden information that we can manipulate but not access directly. A better viewpoint, however, is to regard this hidden information as being a unit of quantum information rather than an infinite number of inaccessible classical bits.

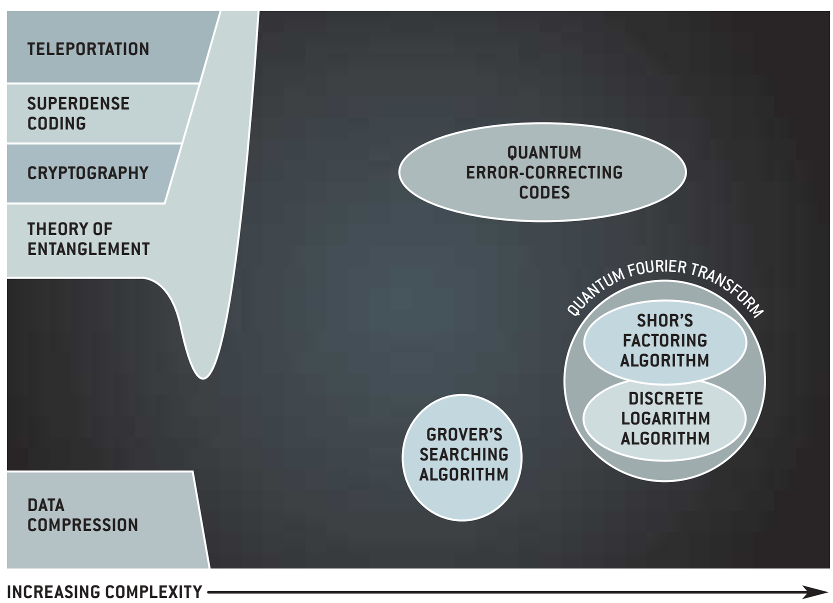
Notice how this example follows Schumacher's paradigm for information science. Gordon and Holevo asked how many qubits (the physical resource) are required to store a given amount of classical information (the task) in such a way that the information can be reliably recovered (the criterion for success). Furthermore, to answer this question, they introduced a mathematical concept, now known as the Holevo chi (represented by the Greek letter χ), that has since been used to simplify the analysis of more complex phenomena, similar to the simplifications enabled by Shannon's entropy. For example, Michal Horodecki of the University of Gdansk in Poland has shown that the Holevo chi can be used to analyze the problem of compressing quantum states produced by a quantum information source, which is analogous to the classical data compression considered by Shannon.

Entangled States

SINGLE QUBITS are interesting, but more fascinating behavior arises when several qubits are brought together. A key feature of quantum information science is the understanding that groups of two or

HERE THERE BE QUANTUM TYGERS

QUANTUM INFORMATION SCIENTISTS are still mapping out the broad topography of their nascent field. Some simpler processes, such as teleportation and quantum cryptography, are well understood. In contrast, complex phenomena such as quantum error correction and Peter W. Shor's factorization algorithm are surrounded by large tracts of terra incognita. One effort to bridge the gaps between the simple and the complex is work on a comprehensive theory of entanglement, analogous to the theory of energy embodied in thermodynamics.



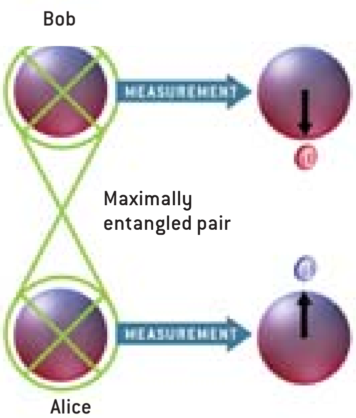
BRYAN CHRISTIE DESIGN

DISENTANGLING ENTANGLEMENT

IF DICE COULD BE "entangled" in the manner of quantum particles, each entangled pair would give the same outcome, even if rolled light-years apart or at very different times.

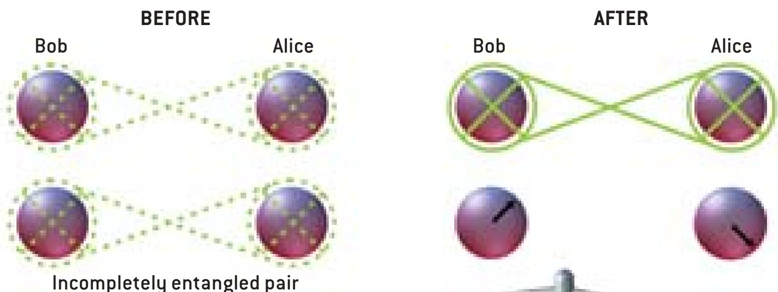
The Standard E-Bit

WHEN TWO QUBITS are entangled, they no longer have individual quantum states. Instead a relation between the qubits is defined. For example, in one type of maximally entangled pair, the qubits give opposite results when measured. If one gives 0, the other returns 1, and vice versa. A maximally entangled pair carries one "e-bit" of entanglement.



Weighing Entanglement

INCOMPLETELY ENTANGLED PAIRS carry less than one e-bit. If Alice and Bob share two partially entangled pairs, they can try to "distill" the entanglement onto a single pair. If distillation produces a maximally entangled pair, then Alice and Bob know their pairs originally carried a total of at least one e-bit of entanglement.

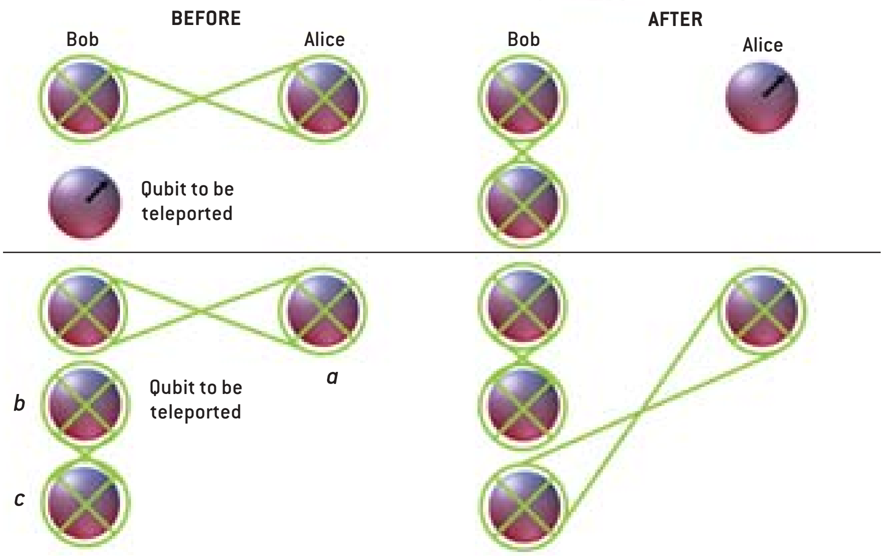


By using distillation (and the inverse process, entanglement dilution), one constructs a virtual set of scales for weighing the entanglement of various states against the standard e-bit.

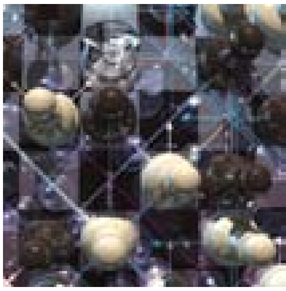
Quantum Teleportation

IF ALICE AND BOB share one e-bit, they can teleport one qubit. The shared e-bit is "used up," in that they no longer share it after teleporting.

If Bob teleports a member (*b*) of an entangled pair to Alice, that particle's entanglement with its original partner (*c*) is transferred to Alice's particle (*a*). Alice and Bob cannot use teleportation, however, to increase their stock of shared e-bits.



BRYAN CHRISTIE DESIGN



Entangled quantum systems behave in ways impossible in any classical world.

more quantum objects can have states that are entangled. These entangled states have properties fundamentally unlike anything in classical physics and are coming to be thought of as an essentially new type of physical resource that can be used to perform interesting tasks.

Schrödinger was so impressed by entanglement that in a seminal 1935 paper (the same year that he introduced his cat to the world) he called it "not *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought." The members of an entangled collection of objects do not have their own individual quantum states. Only the group as a whole has a well-defined state [*see box on opposite page*]. This phenomenon is much more peculiar than a superposition state of a single particle. Such a particle does have a well-defined quantum state even though that state may superpose different classical states.

Entangled objects behave as if they were connected with one another no matter how far apart they are—distance does not attenuate entanglement in the slightest. If something is entangled with other objects, a measurement of it simultaneously provides information about its partners. It is easy to be misled into thinking that one could use entanglement to send signals faster than the speed of light, in violation of Einstein's special relativity, but the probabilistic nature of quantum mechanics stymies such efforts.

Despite its strangeness, for a long time entanglement was regarded as a curiosity and was mostly ignored by physicists. This changed in the 1960s, when John S. Bell of CERN, the European laboratory for particle physics near Geneva, predicted that entangled quantum states allow crucial experimental tests that distinguish between quantum mechanics and classical physics. Bell predicted, and experimenters have confirmed, that entangled quantum systems exhibit behavior that is impossible in a classical world—impossible even if one could change the laws of physics to try to emulate the quantum predictions within a classical framework of any sort! Entanglement represents such an essentially novel feature of our world that even experts find it very difficult to think about. Although one can use the mathematics of quantum theory to reason about entanglement, as soon as one falls back on analogies, there is a great danger that the classical basis of our analogies will mislead us.

In the early 1990s the idea that entanglement falls wholly outside the scope of classical physics prompted researchers to ask whether entanglement might be useful as a resource for solving information-processing problems in new ways. The answer was yes. The flood of examples began in 1991, when Artur K. Ekert of the University of Cambridge showed how to use entanglement to distribute cryptographic keys impervious to eavesdropping. In 1992 Charles H. Bennett of IBM and Stephen Wiesner of Tel Aviv University showed that entanglement can assist the sending of classical information from one location to another (a process called superdense coding, in which two bits are transferred on a particle that seems to have room to carry only one). In 1993 an international team of six collaborators explained how to teleport a quantum state from one location to another using entanglement. An explosion of further applications followed.

Weighing Entanglement

AS WITH INDIVIDUAL qubits, which can be represented by many different physical objects, entanglement also has properties independent of its physical representation. For practical purposes, it may be more convenient to work with one system or another, but in principle it does not matter. For example, one could perform quantum cryptography with an entangled photon pair or an entangled pair of atomic nuclei or even a photon and a nucleus entangled together.

Representation independence suggests a thought-provoking analogy between entanglement and energy. Energy obeys the laws of thermodynamics regardless of whether it is chemical energy, nuclear energy or any other form. Could a general theory of entanglement be developed along similar lines to the laws of thermodynamics?

This hope was greatly bolstered in the late 1990s, when researchers showed that different forms of entanglement are qualitatively equivalent—the entanglement of one state could be transferred to another, similar to energy flowing from, say, a battery charger to a battery. Building on these qualitative relations, investigators have begun introducing quantitative measures of entanglement. These developments are ongoing, and researchers have not yet agreed as to the best way of quantifying entanglement. The most successful scheme thus far is based on the notion of a standard unit of entanglement, akin to a standard unit of mass or energy [*see box on opposite page*].

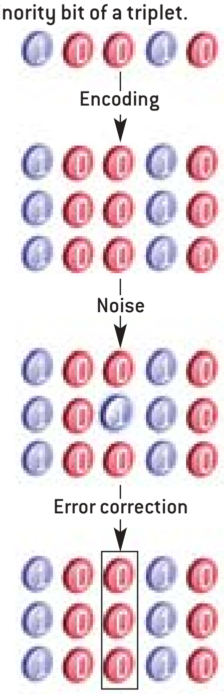
This approach works analogously to measuring masses by using a balance. The mass of an object is defined by how many copies of the standard mass are needed to balance it on a set of scales. Quantum information scientists have developed a theoretical "entanglement balance" to compare the entanglement in two different states. The amount of entanglement in a state is defined by seeing how many copies of some fixed standard unit of entanglement are needed to balance it. Notice that this method of quantifying entanglement is another example of the fundamental question of information science. We have identified a physical resource (copies of our entangled state) and a task with a criterion for success. We define our measure of entanglement by asking how much of our physical resource we need to do our task successfully.

The quantitative measures of entan-

DEALING WITH ERRORS

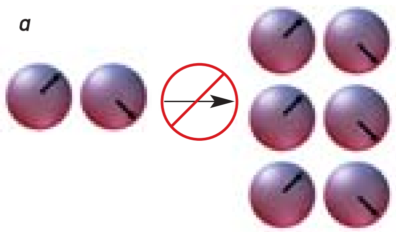
Classical Repetition Code

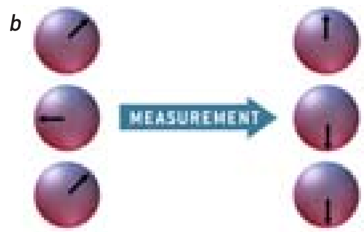
THIS SIMPLE CLASSICAL scheme for reducing errors encodes each bit as a triplet of identical bits. If noise flips one bit, the error can be corrected by fixing the minority bit of a triplet.



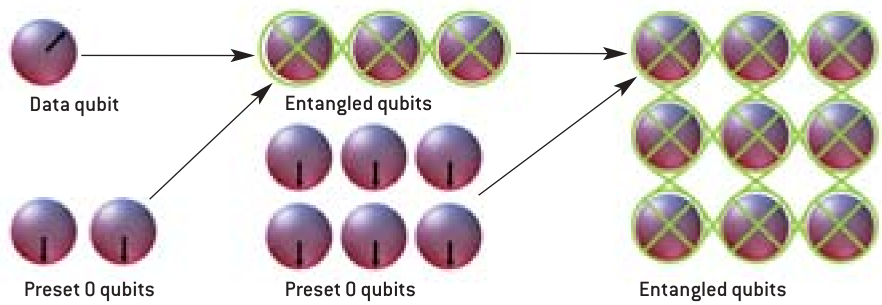
Error Correction for Qubits

THE REPETITION STRATEGY IS IMPOSSIBLE for qubits for two reasons. First, qubits in unknown states cannot be perfectly cloned (*a*). Even if duplicates are produced (for example, by running multiple copies of the computation), a simple measurement will not reveal errors (*b*).





ONE QUANTUM ERROR-CORRECTING CODE works by entangling each data qubit with two preset 0 qubits. These three qubits are in turn entangled with six others. Joint measurements on pairs of qubits will reveal whether one of these nine qubits suffers an error and, if so, how to correct it without disrupting the qubits' individual states.



glement developed by following this program are proving enormously useful as unifying concepts in the description of a wide range of phenomena. Entanglement measures improve how researchers can analyze tasks such as quantum teleportation and algorithms on quantum-mechanical computers. The analogy with energy helps again: to understand processes such as chemical reactions or the operation of an engine, we study the flow of energy between different parts of the system and determine how the energy must be constrained at various locations and times. In a similar way, we can analyze the flow of entanglement from one subsystem to another required to perform a quantum information-processing task and so obtain constraints on the resources needed to perform the task.

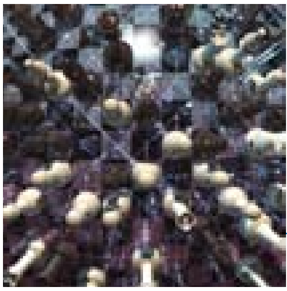
The development of the theory of entanglement is an example of a bottom-up approach—starting from simple questions about balancing entanglement, we gradually gain insight into more complex phenomena. In contrast, in a few cases, people have divined extremely complex phenomena through a great leap of insight, allowing quantum information science to proceed from the top down. The most celebrated example is an algorithm for quickly finding the prime factors of a composite integer on a quantum computer, formulated in 1994 by Peter W. Shor of AT&T Bell Labs. On a classical computer, the best algorithms known take exponentially more resources to factor larger numbers. A 500-digit number needs 100 million times as many computational steps as a 250-digit number. The cost of Shor's algorithm rises only polynomially—a 500-digit number takes only eight times as many steps as a 250-digit number.

Shor's algorithm is a further example of the basic paradigm (how much computational time is needed to find the factors of an *n*-bit integer?), but the algorithm appears isolated from most other results of quantum information science [*see box on page* 29]. At first glance, it looks like merely a clever programming trick with little fundamental significance. That appearance is deceptive; researchers have shown that Shor's algorithm can be interpreted as an instance of a procedure for determining the energy levels of a quantum system, a process that is more obviously fundamental. As time goes on and we fill in more of the map, it should become easier to grasp the principles underlying Shor's and other quantum algorithms and, one hopes, to develop new algorithms.

THE AUTHOR

MICHAEL A. NIELSEN is associate professor in the department of physics at the University of Queensland in Brisbane, Australia. Born in Brisbane, he received his Ph.D. in physics as a Fulbright Scholar at the University of New Mexico in 1998. He is the author, with Isaac L. Chuang of the Massachusetts Institute of Technology, of the first comprehensive graduate-level textbook on quantum information science, *Quantum Computation and Quantum Information.*

BRYAN CHRISTIE DESIGN



Quantum error correction might improve the precision of the world's best clocks.

One final application, quantum error correction, provides the best evidence to date that quantum information science is a useful framework for studying the world. Quantum states are delicate, easily destroyed by stray interactions, or noise, so schemes to counteract these disturbances are essential.

Classical computation and communications have a well-developed assortment of error-correcting codes to protect information against the depredations of noise. A simple example is the repetition code [*see box on opposite page*]. This scheme represents the bit 0 as a string of three bits, 000, and the bit 1 as a string of three bits, 111. If the noise is relatively weak, it may sometimes flip one of the bits in a triplet, changing, for instance, 000 to 010, but it will flip two bits in a triplet far less often. Whenever we encounter 010 (or 100 or 001), we can be almost certain the correct value is 000, or 0. More complex generalizations of this idea provide very good error-correcting codes to protect classical information.

Quantum Error Correction

INITIALLY IT APPEARED to be impossible to develop codes for quantum error correction because quantum mechanics forbids us from learning with certainty the unknown state of a quantum object—the obstacle, again, of trying to extract more than one bit from a qubit. The simple classical triplet code therefore fails because one cannot examine each copy of a qubit and see that one copy must be discarded without ruining each and every copy in the process. Worse still, making the copies in the first place is nontrivial: quantum mechanics forbids taking an unknown qubit and reliably making a duplicate, a result known as the no-cloning theorem.

The situation looked bleak in the mid-1990s, when prominent physicists such as the late Rolf Landauer of IBM wrote skeptical articles pointing out that quantum error correction would be necessary for quantum computation but that the standard classical techniques could not be used in the quantum world. The field owes a great debt to Landauer's skepticism for pointing out problems of this type that had to be overcome [see "Riding the Back of Electrons," by Gary Stix; Profile, SCIENTIFIC AMERICAN, September 1998].

Happily, clever ideas developed independently by Shor and Andrew M. Steane of the University of Oxford in 1995 showed how to do quantum error correction without ever learning the states of the qubits or needing to clone them. As with the triplet code, each value is represented by a set of qubits. These qubits are passed through a circuit (the quantum analogue of logic gates) that will successfully fix an error in any one of the qubits without actually "reading" what all the individual states are. It is as if one ran the triplet 010 through a circuit that could spot that the middle bit was different and flip it, all without determining the identity of any of the three bits.

Quantum error-correcting codes are a triumph of science. Something that brilliant people thought could not be done—protecting quantum states against the effects of noise—was accomplished using a combination of concepts from information science and basic quantum mechanics. These techniques have now received preliminary confirmation in experiments conducted at Los Alamos National Laboratory, IBM and the Massachusetts Institute of Technology, and more extensive experiments are planned.

Quantum error correction has also stimulated many exciting new ideas. For example, the world's best clocks are currently limited by quantum-mechanical noise; researchers are asking whether the precision of those clocks can be improved by using quantum error correction. Another idea, proposed by Alexei Kitaev of the California Institute of Technology, is that some physical systems might possess a type of natural noise tolerance. Those systems would in effect use quantum error correction without human intervention and might show extraordinary inherent resilience against decoherence.

We have explored how quantum information science progresses from fundamental questions to build up an understanding of more complex systems. What does the future hold? By following Schumacher's program, we will surely obtain novel insights into the information-processing capabilities of the universe. Perhaps the methods of quantum information science will even yield insights into systems not traditionally thought of as information-processing systems. For instance, condensed matter exhibits complex phenomena such as high-temperature superconductivity and the fractional quantum Hall effect. Quantum properties such as entanglement are involved, but their role is currently unclear. By applying what we have learned from quantum information science, we may greatly enhance our skills in the ongoing chess match with the complex quantum universe. SA

MORE TO EXPLORE

Quantum Theory and Measurement. Edited by John A. Wheeler and Wojciech H. Zurek. Contains reprints of landmark papers, including a translation of Erwin Schrödinger's 1935 "cat paradox" paper. Princeton University Press, 1983.

The Fabric of Reality. David Deutsch. Penguin Books, 1998.

The Bit and the Pendulum. Tom Siegfried. John Wiley & Sons, 2000.

Quantum Computation and Quantum Information. Michael A. Nielsen and Isaac L. Chuang. Cambridge University Press, 2000.

The Center for Quantum Computation's Web site: **www.qubit.org**

John Preskill's lecture notes are available at **www.theory.caltech.edu/people/preskill/ph229/**

See **www.sciam.com** for *Scientific American* articles related to quantum information science.

harnessingquanta

QUANTUM TELEPORTATION

The science-fiction dream of "beaming" objects from place to place is now a reality—at least for particles of light

By Anton Zeilinger

The scene is a familiar one from science fiction and TV: an intrepid band of explorers enters a special chamber; lights pulse, sound effects warble, and our heroes shimmer out of existence to reappear on the surface of a faraway planet. This is the dream of teleportation—the ability to travel from place to place without having to pass through the tedious intervening miles accompanied by a physical vehicle and airline-food rations. Although the teleportation of large objects or humans still remains a fantasy, quantum teleportation has become a laboratory reality for photons, the individual particles of light.

Quantum teleportation exploits some of the most basic (and peculiar) features of quantum mechanics, a branch of physics invented in the first quarter of the 20th century to explain processes that occur at the level of individual atoms. From the beginning, theorists realized that quantum physics led to a plethora of new phenomena, some of which defy common sense. Technological progress in the final quarter of the 20th century enabled researchers to conduct many experiments that not only have demonstrated fundamental, sometimes bizarre aspects of quantum mechanics but, as in the case of quantum teleportation, have applied them to achieve previously inconceivable feats.

In science-fiction stories, teleportation often permits travel that is instantaneous, violating the speed limit set down by Albert Einstein, who concluded from his theory of relativity that nothing can travel faster than light. Teleportation is also less cumbersome than the more ordinary means of space travel. It is said that Gene Roddenberry, the creator of *Star Trek*, conceived of the "transporter beam" as a way to save the expense of simulating landings and takeoffs on strange planets.

The procedure for teleportation in science fiction varies from story to story but generally goes as follows: A device scans the original object to extract all the information needed to describe it. A transmitter sends the information to a receiving station, where it is used to obtain an exact replica of the original. In some cases, the material that made up the original is also transported to the receiving station, perhaps as energy of some kind; in other cases, the replica is made of atoms and molecules that were already present at the receiving station.

Quantum mechanics seems to make such a teleportation scheme impossible in principle. Heisenberg's uncertainty principle rules that one cannot know both the precise position of an object and its momentum at the same time. Thus, one cannot perform a perfect scan of the object to be teleported; the location or velocity of every atom and electron would be subject to errors. Heisenberg's uncertainty principle also applies to other pairs of quantities, making it impossible to measure the exact, total quantum state of any object with certainty. Yet such measurements would be necessary to obtain all the information needed to describe the original exactly. (In *Star Trek* the "Heisenberg Compensator" somehow miraculously overcomes that difficulty.)

A team of physicists overturned this conventional wisdom in 1993, when they discovered a theoretical way to use quantum mechanics itself for teleportation. The team—Charles H. Bennett of IBM; Gilles Brassard, Claude Crépeau and Richard Josza of the University of Montreal; Asher Peres of Technion–Israel Institute of Technology; and William K. Wootters of Williams College—found that a peculiar but fundamental feature of quantum mechanics, entanglement, can be used to circumvent the limitations imposed by Heisenberg's uncertainty principle without violating it.

Entanglement

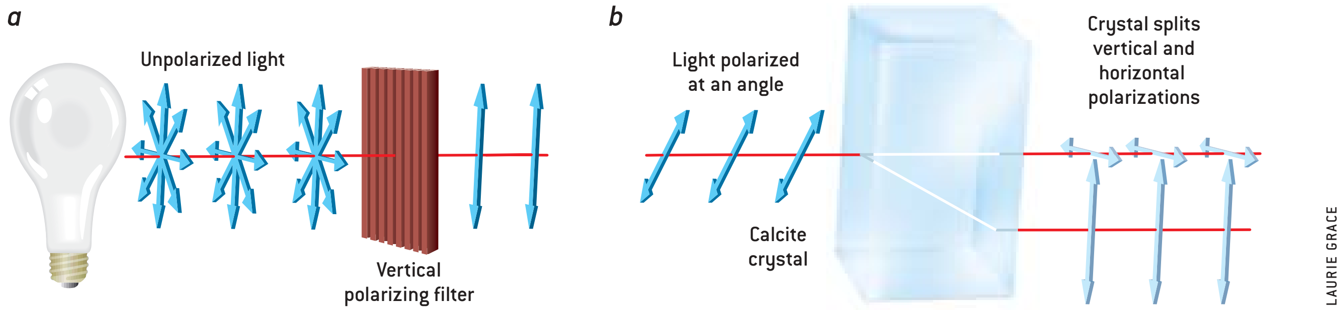
IT IS THE YEAR 2100. A friend who likes to dabble in physics and party tricks has brought you a collection of pairs of dice. He lets you roll them once, one pair at a time. You handle the first pair gingerly, remembering the fiasco with the micro black hole

TRAVELERS ARRIVE at Grand Central Station's teleport terminal. Although teleporting large objects, let alone living beings, will never be practical, teleportation of elementary quantum states has been demonstrated.

SPACE CHANNEL/PHILIP SAUNDERS

Updated from the April 2000 issue

< Teleport Exit
< Baggage
< Hover Cabs
< Skyway
< Rentals
QUANTUM TRANSPORT ANALYZER
ENTROPIC DECAY
ELECTRON LOCK
DECOHERENCE
PLANCK FACTOR
NEW YORK
QUANTUM
TRANSIT
AUTHORIT



UNPOLARIZED LIGHT consists of photons polarized in all directions (*a*). In polarized light the photons' electric-field oscillations (*arrows*) are all aligned. A calcite crystal (*b*) splits a light beam, sending photons that are polarized parallel with its axis into one beam and those that are perpendicular into the other. Intermediate angles go into a quantum superposition of both beams. Each such photon can be detected in one beam or the other, with probability depending on the angle. Because probabilities are involved, we cannot measure the polarization of a single photon with certainty.

last Christmas. Finally, you roll the two dice and get double 3. You roll the next pair. Double 6. The next: double 1. They always match.

The dice in this fable are behaving as if they were quantum-entangled particles. Each die on its own is random and fair, but its entangled partner somehow always gives the correct matching outcome. Such behavior has been demonstrated and intensively studied with real entangled particles. In typical experiments, pairs of atoms, ions or photons stand in for the dice, and properties such as polarization stand in for their different faces.

Consider the case of two photons whose polarizations are entangled to be random but identical. Beams of light and even individual photons consist of oscillations of electromagnetic fields, and polarization refers to the alignment of the electric field oscillations [*see illustration above*]. Suppose that Alice has one of the entangled photons and Bob has its partner. When Alice measures her photon to see if it is horizontally or vertically polarized, each outcome has a 50 percent chance. Bob's photon has the same probabilities, but the entanglement ensures that he will get exactly the same result as Alice. As soon as Alice gets the result "horizontal," say, she knows that Bob's photon will also be horizontally polarized. Before Alice's measurement the two photons do not have individual polarizations; the entangled state specifies only that a measurement will find that the two polarizations are equal.

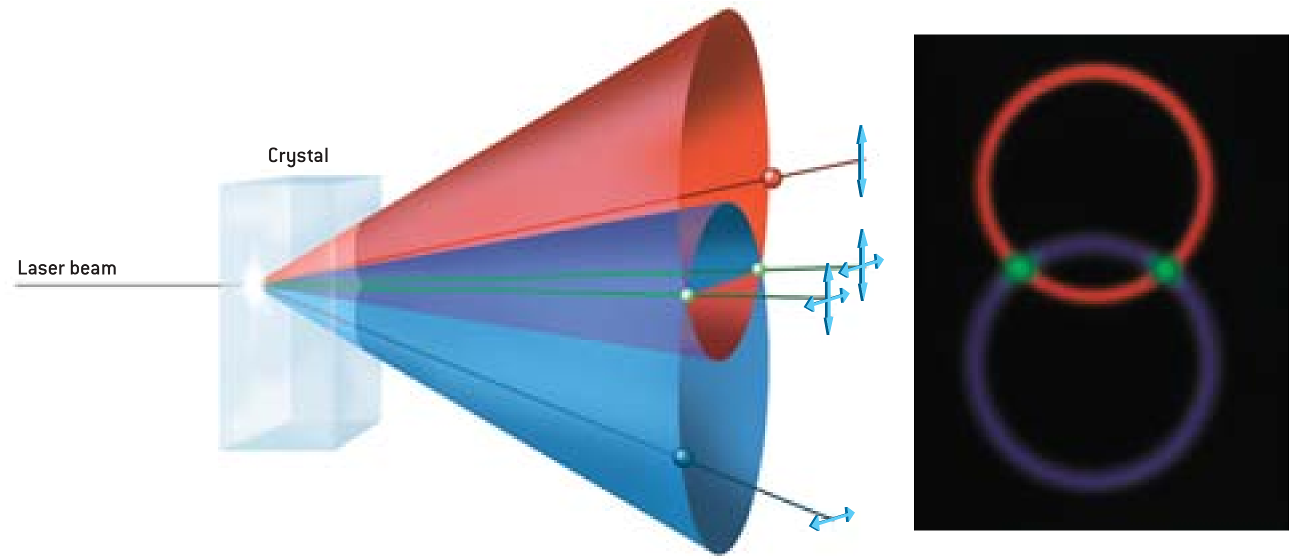
An amazing aspect of this process is that it doesn't matter if Alice and Bob are far away from each other; the process works so long as their photons' entanglement has been preserved. Even if Alice is on Alpha Centauri and Bob on Earth, their results will agree when they compare them. In every case, it is as if Bob's photon is magically influenced by Alice's distant measurement, and vice versa.

You might wonder if we can explain the entanglement by imagining that each particle carries within it some recorded instructions. Perhaps when we entangle the two particles, we synchronize some hidden mechanism within them that determines what results they will give when they are measured. This would explain away the mysterious effect of Alice's measurement on Bob's particle. In the 1960s, however, Irish physicist John Bell proved a theorem that in certain situations any such "hidden variables" explanation of quantum entanglement would have to produce results different from those predicted by standard quantum mechanics. Experiments have confirmed the predictions of quantum mechanics to a very high accuracy.

Austrian physicist Erwin Schrödinger, one of the co-inventors of quantum mechanics, called entanglement "the essen-

PREPARING FOR QUANTUM TELEPORTATION ...

QUANTUM TELEPORTATION OF A PERSON (impossible in practice but a good example to aid the imagination) would begin with the person inside a measurement chamber (*left*) alongside an equal mass of auxiliary material (*green*). The auxiliary matter has previously been quantum-entangled with its counterpart, which is at the faraway receiving station (*right*).



LAURIE GRACE (*left*); P. G. KWIAT AND M. RECK *Institute for Experimental Physics, University of Vienna* (*right*)

ENTANGLED PHOTON PAIRS are created when a laser beam passes through a crystal such as beta barium borate. The crystal occasionally converts a single ultraviolet photon into two photons of lower energy, one polarized vertically (*on red cone*), one polarized horizontally (*on blue cone*). If the photons happen to travel along the cone intersections (*green*), neither photon has a definite polarization, but their relative polarizations are complementary; they are then entangled. Colorized image (*at right*) is a photograph of down-converted light. Colors do not represent the color of the light.

tial feature" of quantum physics. Entanglement is often called the EPR effect and the particles EPR pairs, after Einstein, Boris Podolsky and Nathan Rosen, who in 1935 analyzed the effects of entanglement acting across large distances. Einstein talked of it as "spooky action at a distance." If one tried to explain the results in terms of signals traveling between the photons, the signals would have to travel faster than the speed of light. Naturally, many people have wondered if this effect could be used to transmit information faster than the speed of light.

Unfortunately, the quantum rules make that impossible. Each local measurement on a photon, considered in isolation, produces a completely random result and so can carry no information from the distant location. It tells you nothing more than what the distant measurement result probabilities would be, depending on what was measured there. Nevertheless, we can put entanglement to work in an ingenious way to achieve quantum teleportation.

Teleporting Photons

ALICE AND BOB anticipate that they will want to teleport a photon in the future. In preparation, they share an entangled auxiliary pair of photons, Alice taking photon A and Bob photon B. Instead of measuring them, they each store their photon without disturbing the delicate entangled state [*see top illustration on next page*].

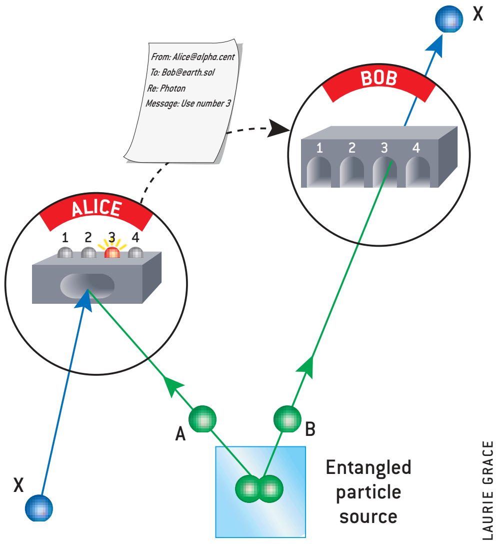
In due course, Alice has a third photon—call it photon X—that she wants to teleport to Bob. She does not know what photon X's state is, but she wants Bob to have a photon with that same polarization state. She cannot simply measure the photon's polarization and send Bob the result. In general, her measurement result would not be identical to the photon's original state. This is Heisenberg's uncertainty principle at work.

... A QUANTUM MEASUREMENT ...

JOINT MEASUREMENT carried out on the auxiliary matter and the person (*left*) changes them to a random quantum state and produces a vast amount of random (but significant) data—two bits per elementary state. By "spooky action at a distance," the measurement also instantly alters the quantum state of the faraway counterpart matter (*right*).

MORE >>>

DAVID FIERSTEIN



LAURIE GRACE

IDEAL QUANTUM TELEPORTATION relies on Alice, the sender, and Bob, the receiver, sharing a pair of entangled particles A and B. Alice has a particle that is in an unknown quantum state X. Alice performs a Bell-state measurement on particles A and X, producing one of four possible outcomes. She tells Bob about the result by ordinary means. Depending on Alice's result, Bob leaves his particle unaltered (1) or rotates it (2, 3, 4). Either way it ends up a replica of particle X.

Instead, to teleport photon X, Alice measures it jointly with photon A, without determining their individual polarizations. She might find, for instance, that their polarizations are "perpendicular" to each other (she still does not know the absolute polarization of either one, however). Technically, the joint measurement entangles photon A and photon X and is called a Bell-state measurement. Alice's measurement produces a subtle effect: it changes Bob's photon to correlate with a combination of her measurement result and the state that photon X originally had. In fact, Bob's photon now carries her photon X's state, either exactly or modified in a simple way.

To complete the teleportation, Alice must send a message to Bob—one that travels by conventional means, such as a telephone call or a note on a scrap of paper. After he receives this message, if necessary Bob can transform his photon B, with the end result that it becomes an exact replica of the original photon X. Which transformation Bob must apply depends on the outcome of Alice's measurement. There are four possibilities, corresponding to four quantum relations between her photons A and X. A typical transformation that Bob must apply to his photon is to alter its polarization by 90 degrees, which he can do by sending it through a crystal with the appropriate optical properties.

Which of the four possible results Alice obtains is completely random and independent of photon X's original state. Bob therefore does not know how to process his photon until he learns the result of Alice's measurement. One can say that Bob's photon instantaneously contains all the information from Alice's original, transported there by quantum mechanics. Yet to know how to read that information, Bob must wait for the classical information, consisting of two bits that can travel no faster than the speed of light.

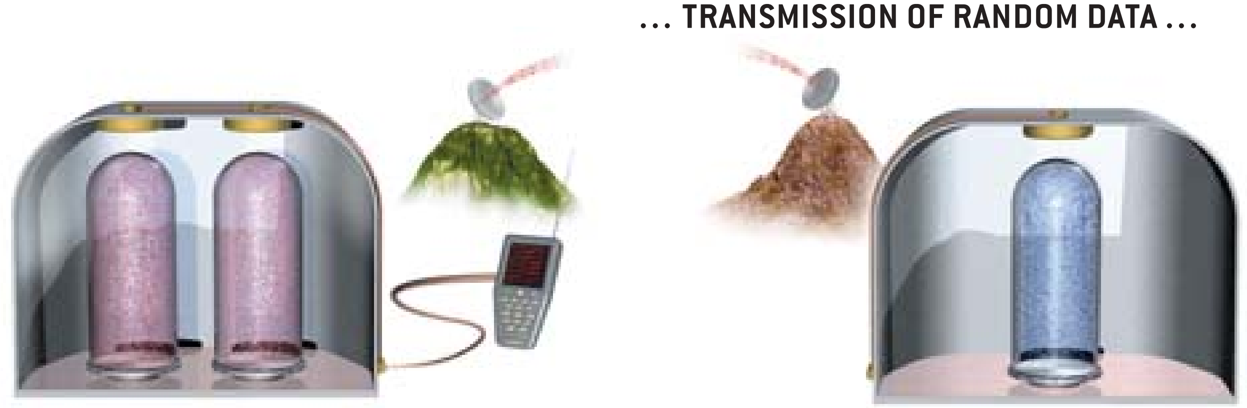
Skeptics might complain that the only thing teleported is the photon's polarization state or, more generally, its quantum state, not the photon "itself." But because a photon's quantum state is its defining characteristic, teleporting its state is completely equivalent to teleporting the particle [*see box on page 41*].

Note that quantum teleportation does not result in two copies of photon X. Classical information can be copied any number of times, but perfect copying of quantum information is impossible, a result known as the no-cloning theorem, which was proved in 1982 by Wootters and Wojciech H. Zurek of Los Alamos National Laboratory. (If we could clone a quantum state, we could use the clones to violate Heisenberg's principle.) Alice's measurement actually entangles her photon A with photon X, and photon X loses all memory, one might say, of its original state. As a member of an entangled pair, it has no individual polarization state. Thus, the original state of photon X disappears from Alice's domain.

Circumventing Heisenberg

FURTHERMORE, photon X's state has been transferred to Bob with neither Alice nor Bob learning anything about what the state is. Alice's measurement result, being random, tells them nothing about the state. This is how the process circumvents Heisenberg's principle, which stops us from determining the complete quan-

... TRANSMISSION OF RANDOM DATA ...

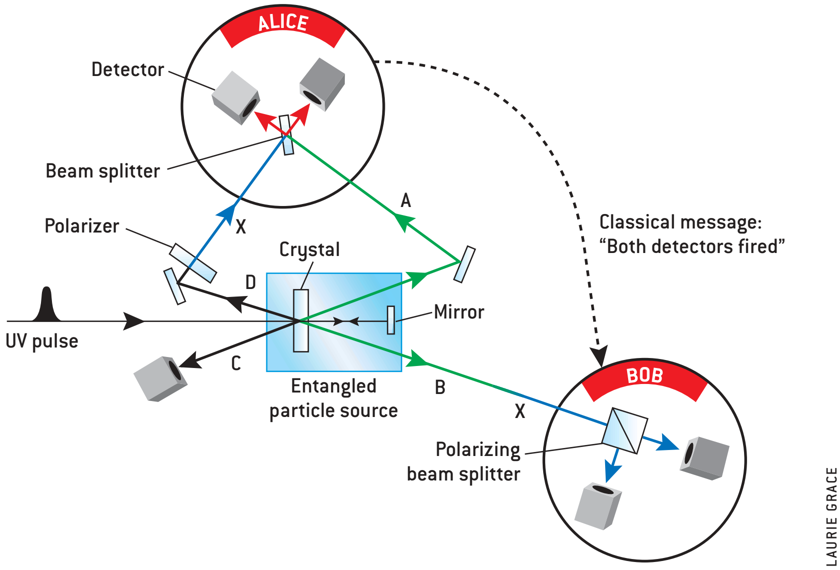


MEASUREMENT DATA must be sent to the distant receiving station by conventional means. This process is limited by the speed of light, making it impossible to teleport the person faster than the speed of light.

tum state of a particle but does not preclude teleporting the complete state so long as we do not try to see what the state is!

Also, the teleported quantum information does not travel materially from Alice to Bob. All that travels materially is the message about Alice's measurement result, which tells Bob how to process his photon but carries no information about photon X's state itself.

In one out of four cases, Alice is lucky with her measurement, and Bob's photon immediately becomes an identical replica of Alice's original. It might seem as if information has traveled instantly from Alice to Bob, beating Einstein's speed limit. Yet this strange feature cannot be used to send information, because Bob has no way of knowing that his photon is already an identical replica. Only when he learns the result of Alice's Bell-state measurement, transmitted to him via classical means, can he exploit the information in the teleported quantum state. If he tries to guess in which cases teleportation was instantly successful, he will be wrong 75 percent of the time, and he will not know which guesses are correct. If he uses the photons based on such guesses, the results will be the same as they would had he taken a beam of photons with random polarizations. Thus, Einstein's relativity prevails; even the spooky instantaneous action at a distance of quantum mechanics fails to send usable information faster than the speed of light.



INNSBRUCK EXPERIMENT begins with a short pulse of ultraviolet laser light. Traveling left to right through a crystal, this pulse produces the entangled pair of photons A and B, which travel to Alice and Bob. Reflected back through the crystal, the pulse creates two more photons, C and D. A polarizer prepares photon D in a specific state, X. Photon C is detected, confirming that photon X has been sent to Alice. Alice combines photons A and X with a beam splitter. If she detects one photon in each detector (as occurs at most 25 percent of the time), she notifies Bob, who uses a polarizing beam splitter to verify that his photon has acquired X's polarization, thus demonstrating teleportation.

It would seem that the theoretical proposal described above laid out a clear blueprint for building a teleporter; on the contrary, it presented a great experimental challenge. Producing entangled pairs of photons has become routine in physics experiments in the past decade, but carrying out a Bell-state measurement on two independent photons had never been done before.

Building a Teleporter

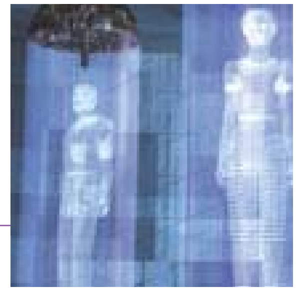
A POWERFUL WAY to produce entangled pairs of photons is spontaneous parametric down-conversion: a single photon passing through a special crystal sometimes generates two new photons that are entangled so that they will show opposite polarization when measured.

A much more difficult problem is to entangle two independent photons that already exist, as must occur during the operation of a Bell-state analyzer. This means that the two photons (A and X) somehow have to lose their private features. In 1997 my group (Dik Bouwmeester, Jian-Wei Pan, Klaus Mattle, Manfred Eibl and Harald Weinfurter),

... RECONSTRUCTION OF THE TRAVELER

RECEIVER RE-CREATES THE TRAVELER, exact down to the quantum state of every atom and molecule, by adjusting the counterpart matter's state according to the random measurement data sent from the scanning station.

Teleportation could transfer quantum information between quantum processors in a quantum computer.



then at the University of Innsbruck, applied a solution to this problem in our teleportation experiment [*see top illustration on preceding page*].

In our experiment, a brief pulse of ultraviolet light from a laser passes through a crystal and creates the entangled photons A and B. One travels to Alice, and the other goes to Bob. A mirror reflects the ultraviolet pulse back through the crystal again, where it may create another pair of photons, C and D. (These will also be entangled, but we don't use their entanglement.) Photon C goes to a detector, which alerts us that its partner, D, is available to be teleported. Photon D passes through a polarizer, which we can orient in any conceivable way. The resulting polarized photon is our photon X, the one to be teleported, which travels on to Alice. Once it passes through the polarizer, X is an independent photon, no longer entangled. And although *we* know its polarization because of how we set the polarizer, Alice does not. We reuse the same ultraviolet pulse in this way to ensure that Alice has photons A and X at the same time.

Now we arrive at the problem of performing the Bell-state measurement. To do this, Alice combines her two photons (A and X) using a semireflecting mirror, a device that reflects half the incident light. An individual photon has a 50–50 chance of passing through or being reflected. In quantum terms, the photon goes into a superposition of these two possibilities.

Now suppose that two photons strike the mirror from opposite sides, with their paths aligned so that the reflected path of one photon lies along the transmitted path of the other, and vice versa. A detector waits at the end of each path. Ordinarily the two photons would be reflected independently, and there would be a 50 percent chance of them arriving in separate detectors. If the photons are indistinguishable and arrive at the mirror at the same instant, however, quantum interference takes place: some possibilities cancel out and do not occur, whereas others reinforce and occur more often. When the photons interfere, they have only a 25 percent likelihood of ending up in separate detectors. Furthermore, that outcome corresponds to detecting one of the four possible Bell states of the two photons—the case that we called "lucky" earlier. The other 75 percent of the time the two photons both end up in one detector, which corresponds to the other three Bell states but does not discriminate among them.

When Alice simultaneously detects one photon in each detector, Bob's photon instantly becomes a replica of Alice's original photon X. We verified that this teleportation occurred by showing that Bob's photon had the polarization that we imposed on photon X. Our experiment was not perfect, but the correct polarization was detected 80 percent of the time (random photons would achieve 50 percent). We demonstrated the procedure with a variety of polarizations: vertical, horizontal, linear at 45 degrees and even a nonlinear, circular polarization.

The most difficult aspect of our Bell-state analyzer is making photons A and X indistinguishable. Even the timing of when the photons arrive could be used to identify which photon is which, so it is important to "erase" the time information carried by the particles. In our experiment, we used a clever trick first suggested by Marek Zukowski of the University of Gdansk in Poland: we send the photons through very narrow bandwidth wavelength filters. This process makes the wavelength of the photons extremely precise, and by Heisenberg's uncertainty relation it smears out the photons in time.

A mind-boggling case arises when the teleported photon is itself entangled with another and thus does not have its own polarization. In 1998 my Innsbruck group demonstrated this scenario by giving Alice photon D without polarizing it, so that it was still entangled with photon C. We showed that when the teleportation succeeded, Bob's photon B ended up entangled with C. The *entanglement* with C had been transmitted from D to B.

My current group at the University of Vienna was able to perform teleportation of entanglement in such high fidelity that the nonlocal correlation between photons B and C violated a Bell inequality. The quality was high enough to make quantum repeaters possible, necessary to connect quantum computers over large distances. Shortly thereafter we overcame a limitation of our initial experiment. Earlier Bob had to actually detect and so destroy his photon X to make sure that teleportation succeeded. Our experiment provided a freely propagating beam of teleported qubits emerging from Bob's side, thus showing that this step is not essential. This is important in a case where the qubits will be used again in some way.

Piggyback States

OUR EXPERIMENT clearly demonstrated teleportation, but it had a low rate of success. Because we could identify just one Bell state, we could teleport Alice's photon only 25 percent of the time—the occasions when that state occurred. No complete Bell-state analyzer exists for independent photons or for any two independently created quantum particles, so at present there is no experimentally proven way to improve our scheme's efficiency to 100 percent.

In 1994 a way to circumvent this

THE AUTHOR

ANTON ZEILINGER (anton.zeilinger@quantum.at) is at the Institute for Experimental Physics at the University of Vienna, having teleported there in 1999 from the University of Innsbruck. He considers himself very fortunate to have the privilege of working on exactly the mysteries and paradoxes of quantum mechanics that drew him into physics nearly 40 years ago. In his little free time, Zeilinger interacts with classical music and with jazz and loves to ski.

problem was proposed by Sandu Popescu, then at the University of Cambridge. He suggested that the state to be teleported could be a quantum state riding piggyback on Alice's auxiliary photon A. Francesco De Martini's group at the University of Rome I "La Sapienza" successfully demonstrated this scheme in 1997. The auxiliary pair of photons was entangled according to the photons' locations: photon A was split, as by a beam splitter, and sent to two different parts of Alice's apparatus, with the two alternatives linked by entanglement to a similar splitting of Bob's photon B. The state to be teleported was also carried by Alice's photon A—its polarization state. With both roles played by one photon, detecting all four possible Bell states becomes a standard single-particle measurement: detect Alice's photon in one of two possible locations with one of two possible polarizations. The drawback of the scheme is that Alice cannot teleport a separate unknown photon X. Doing that would require her to somehow transfer its state onto her photon A, which is essentially a teleportation procedure by itself.

Polarization of a photon, the feature employed by the Innsbruck and Rome experiments, is a discrete quantity, in that any polarization state can be expressed as a superposition of just two discrete states, such as vertical and horizontal polarization. The electromagnetic field associated with light also has continuous features that amount to superpositions of an infinite number of basic states. For example, a light beam can be "squeezed," meaning that one of its properties is made extremely precise, or noise-free, at the expense of greater randomness in another property (à la Heisenberg). In 1998 Jeffrey Kimble's group at the California Institute of Technology teleported such a squeezed state from one beam of light to another, thus demonstrating teleportation of a continuous feature. In 2002 a group at the Australian National University in Canberra led by Ping Koy Lam realized such a teleportation with unprecedented high fidelity.

Remarkable as all these experiments are, they are a far cry from quantum teleportation of large objects. There are two essential problems: First, one needs an entangled pair of such large objects. Second, the object to be teleported and the entangled pairs must be sufficiently isolated from the environment. If enough information leaks to or from the environment through stray interactions, the objects' quantum states degrade, a process called decoherence. It is hard to imagine how we could achieve such extreme isolation for an object, let alone a living creature that breathes air and radiates heat. But who knows how fast development might go in the future?

Certainly we could use existing technology to teleport elementary states, like those of the photons in our experiment, across distances of a few kilometers and maybe even up to satellites. The technology to teleport states of individual atoms is at hand today: the group led by Serge Haroche at the École Normale Supérieure in Paris has demonstrated entanglement of atoms. The entanglement and teleportation of molecules may reasonably be expected within the next decade.

What happens beyond that is anybody's guess. In 2002 Eugene Polzik's group at the University of Århus in Denmark demonstrated entanglement of the spin of two ensembles, each containing about 10^{12} atoms. This experiment opens up the possibility of teleporting systems containing large numbers of atoms.

An important application of teleportation might be in quantum computation,

SKEPTICS CORNER

ANSWERS TO COMMON TELEPORTATION QUESTIONS

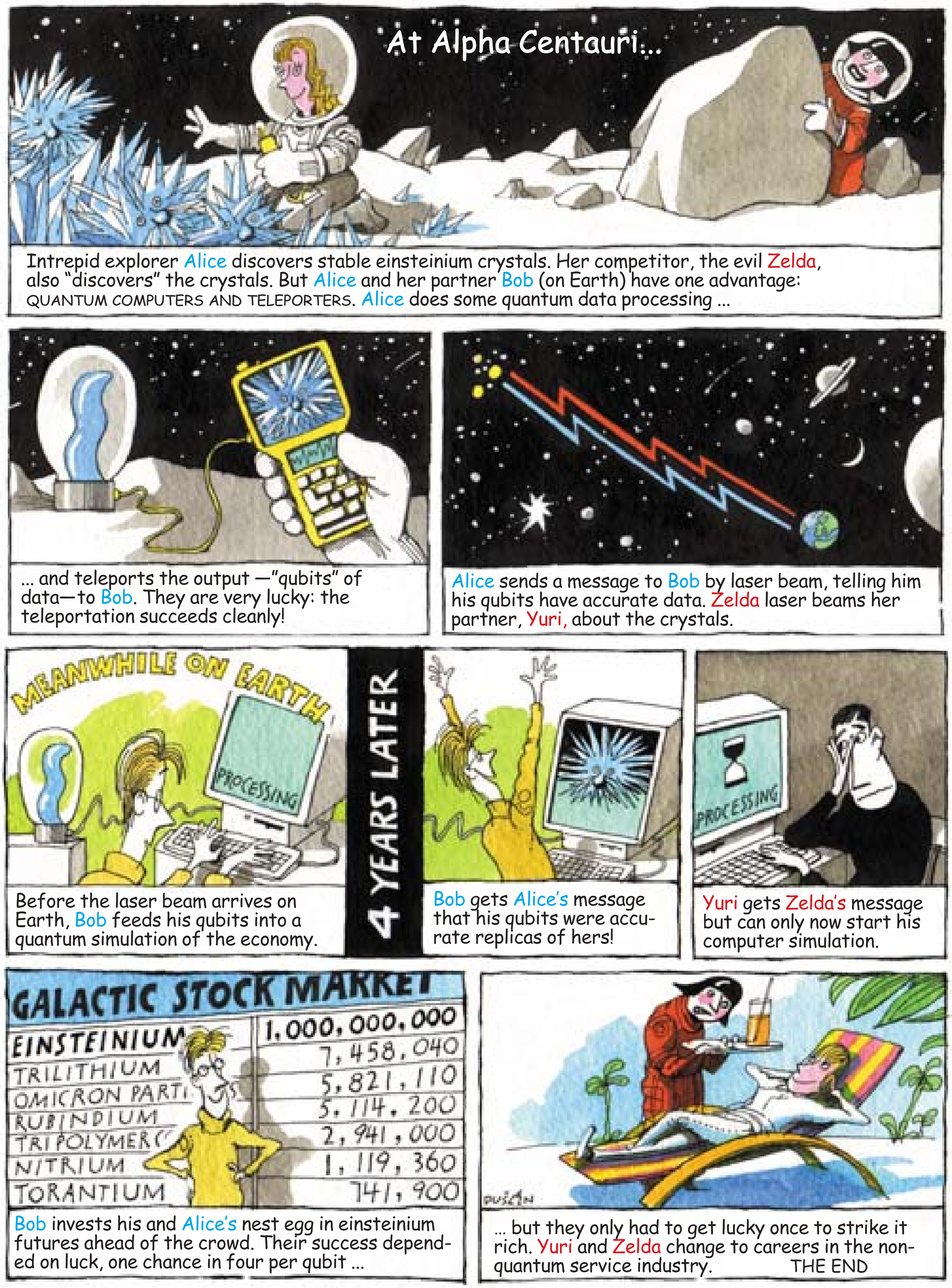
Isn't it an exaggeration to call this teleportation? After all, it is only a quantum state that is teleported, not an actual object. What do we mean by identity? How do we know that an object—say, the car we find in our garage in the morning—is the same one we saw a while ago? When it has all the right features and properties. Quantum physics reinforces this point: particles of the same type in the same quantum state are indistinguishable even in principle. If one could carefully swap all the iron atoms in the car with those from a lump of ore and reproduce the atoms' states exactly, the end result would be identical, at the deepest level, to the original car. Identity cannot mean more than this: being the same in all properties.

Isn't it more like "quantum faxing"? Faxing produces a copy that is easy to distinguish from the original. Moreover, because of the quantum no-cloning theorem, in quantum teleportation the original must be destroyed.

Can we hope to teleport a complicated object? There are severe obstacles. First, the object has to be in a pure quantum state, and such states are very fragile. Experiments with atoms and larger objects must be done in a vacuum to avoid collisions with gas molecules. Even a tiny lump of matter would be disturbed merely by thermal radiation from the walls of the apparatus. This is why we do not routinely see quantum effects in our everyday world. Another problem is the Bell-state measurement. What would it mean to do a Bell-state measurement of a virus consisting of, say, 10^7 atoms? How would we extract the 10^8 or more bits of information that such a measurement would generate? For an object of just a few grams the numbers become impossible: more than 10^{24} bits of data.

Would teleporting a person require quantum accuracy? Being in the same quantum state does not seem necessary for being the same person. We change our states all the time and remain the same people—at least as far as we can tell! Conversely, identical twins or biological clones are not "the same people," because they have different memories. Does Heisenberg uncertainty prevent us from replicating a person precisely enough for her to think she was the same as the original? Who knows. It is intriguing, however, that the quantum no-cloning theorem prohibits us from making a perfect replica of a person. *—A.Z.*

THE QUANTUM ADVENTURES OF ALICE & BOB



DUSAN PETRICIC

where the ordinary notion of bits (0's and 1's) is generalized to quantum bits, or qubits, which can exist as superpositions and entanglements of 0's and 1's. Teleportation could be used to transfer quantum information between quantum processors. Quantum teleporters can also be used to build a quantum computer [*see box at right*]. The cartoon on the opposite page illustrates an intriguing situation in which a combination of teleportation and quantum computation could occasionally yield an advantage, as if one had received the teleported information instantly instead of having to wait for it to arrive by normal means.

Quantum mechanics is probably one of the most profound theories ever discovered. The problems that it poses for our everyday intuition about the world led Einstein to criticize it very strongly. He insisted that physics should be an attempt to grasp a reality that exists independently of its observation. Yet he realized that we run into deep problems when we try to assign such an independent physical reality to the individual members of an entangled pair. His great counterpart, Danish physicist Niels Bohr, insisted that one has to take into account the whole system—in the case of an entangled pair, the arrangement of both particles together, no matter how far they may be separated from each other. Einstein's desideratum, the independent real state of each particle, has no meaning for an entangled quantum system.

Quantum teleportation is a direct descendant of the scenarios debated by Einstein and Bohr. We run into all kinds of problems if we ask ourselves what the properties of the individual particles *really* are when they are entangled. We have to analyze carefully what it means to "have" a polarization. We cannot escape the conclusion that all we can talk about are certain experimental results obtained by measurements. In our polarization measurement, a click of the detector lets us construct a picture in our mind in which the photon actually "had" a certain polarization. Yet we must always remember that this is just a made-up story. It is valid only if we talk about that specific experiment, and we should be cautious when using it in other situations.

QUANTUM COMPUTERS

PERHAPS THE MOST IMPORTANT, yet still hypothetical, application of quantum teleportation outside of physics research is in quantum computation. A conventional digital computer works with bits, which take definite values of 0 or 1, but a quantum computer uses quantum bits, or qubits. Qubits can be in quantum superpositions of 0 and 1 just as a photon can be in a superposition of horizontal and vertical polarization. Indeed, in sending a single photon, the basic quantum teleporter transmits a single qubit of quantum information.

Superpositions of numbers may seem strange, but as the late Rolf Landauer of IBM put it, "When we were little kids learning to count on our very sticky classical fingers, we gained the wrong intuition. We thought that information was classical. We thought that we could hold up three fingers, then four. We didn't realize that there could be a superposition of both."

A quantum computer can work on a superposition of many different inputs at once. It could run an algorithm simultaneously on one million inputs, using only as many qubits as a conventional computer would need bits to run the algorithm once on a single input. Theorists have proved that the algorithms running on quantum computers can solve certain problems faster (in fewer computational steps) than any known algorithm running on a classical computer. The problems include finding items in a database and factoring large numbers, which is of great interest for breaking secret codes.

So far only the most rudimentary elements of quantum computers have been built: logic gates that can process one or two qubits. The realization of even a small-scale quantum computer is still far away. A key problem is transferring quantum data reliably between different logic gates or processors, whether within a single quantum computer or across quantum networks. Quantum teleportation is one solution.

Daniel Gottesman of Microsoft and Isaac L. Chuang of IBM proved that a general-purpose quantum computer can be built out of three basic components: entangled particles, quantum teleporters and gates that operate on a single qubit at a time. This result provides a systematic way to construct two-qubit gates. In general, building a two-qubit gate for independent qubits provides the same experimental challenge as realizing a Bell-state analyzer for independent systems, and either one, once realized, can be used to build the other one. *—A.Z.*

Indeed, following Bohr, I would argue that we can understand quantum mechanics if we realize that science does not describe how nature *is* but rather articulates what we can *say* about nature. Expressed in modern language, this means that quantum mechanics is a science of knowledge, of information. This is where the current value of fundamental experiments such as teleportation lies: in helping us to reach a deeper understanding of our mysterious quantum world. SA

MORE TO EXPLORE

Experimental Quantum Teleportation. D. Bouwmeester, J. W. Pan, K. Mattle, M. Eibl, H. Weinfurter and A. Zeilinger in *Nature,* Vol. 390, pages 575–579; December 11, 1997.

Quantum Information. Special issue of *Physics World,* Vol. 11, No. 3; March 1998.

Quantum Theory: Weird and Wonderful. A. J. Leggett in *Physics World,* Vol. 12, No. 12, pages 73–77; December 1999.

Entanglement: The Greatest Mystery in Physics. Amir D. Aczel. Four Walls Eight Windows, New York, 2002.

More about quantum teleportation is available at **www.quantum.at**

LAURIE GRACE

Slowing a beam of light to a halt may pave the way for new optical communications technology, tabletop black holes and quantum computers

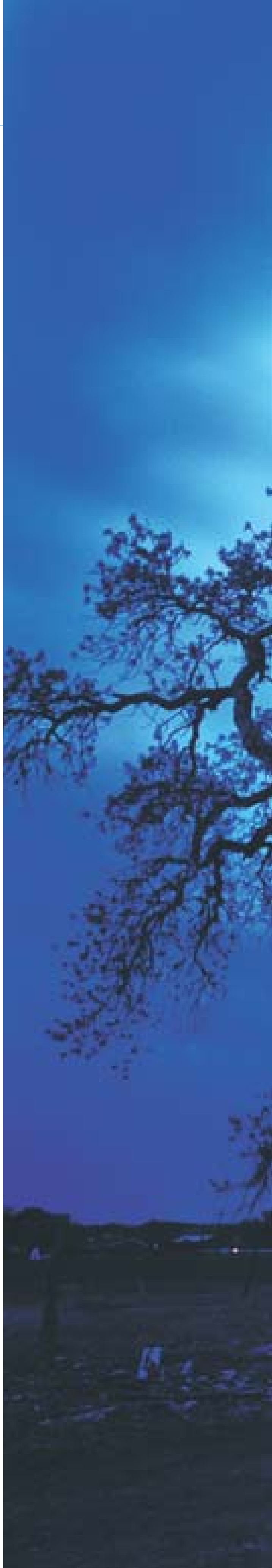
Frozen LIGHT

EVERYONE KNOWS OF THE SPEED OF LIGHT AS one of the unshakable properties of the universe. It's not surprising, then, that experiments to radically alter light's speed require some serious equipment and hard work. Running such an experiment requires first a careful tune-up and optimization of the setup and then a long period of painstaking data-gathering to get a consistent set of measurements. At the Rowland Institute for Science in Cambridge, Mass., our original ultraslow-light experiments took place in stints lasting 27 hours nonstop. Instead of breaking for meals, we learned to balance a slice of pizza in one hand, leaving the other clean to flip mirrors in and out on the optics table during 38 seconds of total darkness at a crucial stage of each run.

Our goal was to drastically slow down light, which travels through empty space at the universe's ultimate speed limit of nearly 300,000 kilometers a second. We saw the first sign of light pulses slowing down in March 1998. As happens so often in experimental physics—because it can take so many hours to get all the components working together for the first time—this occurred in the wee hours of the morning, at 4 A.M. By July we were down to airplane speed. At that time I had to go to the Niels Bohr Institute in Copenhagen to teach a class. I remember sitting in the plane marveling that I was traveling "faster than light"—that I could beat one of our slow pulses to Denmark by a full hour.

Needless to say, I was restless during the week in Copenhagen and eager to get back to Cambridge to continue the light-slowing experiments. In the next month we reached 60 kilometers an hour and decided that it was time to publish. The real payoff for the hard work, prior to those results, was sitting in the lab in the middle of the night and observing the slow-light pulses, knowing that we were the first in history to

BY LENE VESTERGAARD HAU



 Updated from the July 2001 issue

FREEZING OF LIGHT begins with a process in which a carefully tuned laser beam renders an opaque material transparent to a second laser beam.

Compressing a kilometer-long laser pulse to one millionth of a meter sets off quantum shock waves in sodium atoms near absolute zero.

see light go so slowly that you could outpace it on a bicycle.

In the summer of 2000 we brought pulses of light to a complete halt within tiny gas clouds cooled to near absolute zero. We could briefly keep the pulses on ice, so to speak, and then send them back on their way.

As well as being of great intrinsic interest, slowing and freezing light have a number of applications. At sufficiently low temperatures the ultracold clouds of atoms used in our slow-light experiments form Bose-Einstein condensates, remarkable systems in which all the atoms gather in a single quantum state and act in synchrony. Further uses could involve sending a light pulse through a condensate as slowly as a sound wave, which we expect would cause a wave of atoms to "surf" on the light pulse, setting off oscillations of the entire condensate.

The slow and frozen light work also opens up new possibilities for optical communications and data storage and for quantum-information processing—that is, for quantum computers, which would utilize quantum phenomena to outperform conventional computers. The freezing-light system essentially converts between motionless forms of quantum information and photons flying around at the usual speed of light.

Getting Atoms into a State

MANY ORDINARY MATERIALS slow down light. Water, for instance, slows light to about 75 percent of its velocity in a vacuum. But that type of speed reduction, associated with a material's refractive index, is limited. Diamond, which has one of the highest refractive indices of a transparent material, slows light by a factor of only 2.4. Reducing light's speed by factors of tens of millions requires new effects that depend on quantum mechanics. My group produces the conditions for these effects in a cigar-shaped cloud of sodium atoms—typically 0.2 millimeter long and 0.05 millimeter in diameter—trapped in a magnetic field and cooled to within a millionth of a degree of absolute zero.

Sodium belongs to the family of alkali atoms, which have a single outermost, or valence, electron. The valence electron produces almost all the action: Different excited states of a sodium atom correspond to that electron's being promoted to larger orbits around the nucleus, with higher energies than its usual lowest energy state, or ground state. These states determine how the atom interacts with light—which frequencies it will absorb strongly and so on. In addition, both the valence electron and the atom's nucleus are magnets, in effect acting like tiny compass needles. The electron's magnetism is associated with its intrinsic angular momentum, or spin, a little like the association of the earth's rotational axis with magnetic north but with exact alignment. The precise energies of an atom's excited states depend on how the spins of the nucleus and the valence electron are aligned.

Although an atom can assume many states, we use just three to slow light. When we finish preparing and cooling the atom cloud, every atom is internally in state 1, its ground state: the valence electron is in its lowest orbit, and its spin is exactly opposite, or anti-aligned, with the nuclear spin. Also, the total magnetism of each atom is anti-aligned with the magnetic field that we use to hold the cloud in place. State 2 is very similar, only with the electron and nuclear spins aligned, which raises the atom's energy a little. State 3 has about 300,000 times more energy than state 2 (with state 1 as the reference level) and is produced by boosting the valence electron up to a larger orbit. Atoms relaxing from state 3 down to state 1 or 2 generate the characteristic yellow glow of sodium streetlights.

The pulse of laser light that we wish to slow (the "probe" pulse) is tuned to the energy difference between states 1 and 3. If we sent a pulse of that light into the cloud without doing any other preparation, the atoms would completely absorb the pulse and jump from state 1 to state 3. After a brief time, the excited atoms would relax by reemitting light, but at random and in all directions. The cloud would glow bright yellow, but all information about the original light pulse would be obliterated.

To prevent this absorption, we use electromagnetically induced transparency, a phenomenon first observed in 1990 by Stephen E. Harris's group at Stanford University. In electromagnetically induced transparency, a laser beam with a carefully chosen frequency shines on the cloud and changes it from being as opaque as a wall to being as clear as glass for light of another specific frequency.

The transparency-inducing laser beam, or coupling beam,

Overview/*Stopping Light*

- Nothing travels faster than light in a vacuum, but even light is slowed down in many media. Scientists have manipulated clouds of atoms with lasers so that pulses of light travel through the clouds at one 20-millionth of their normal speed—slower than highway traffic.
- A similar technique completely halts the pulses, turning them into a quantum imprint on the atoms. Later, another laser beam converts the frozen pulse back into a moving light pulse with all the properties of the original.
- The process of slowing and stopping light has many research and technological applications.

CHIP SIMONS (*preceding pages*)

is tuned to the energy difference between states 2 and 3. The atoms, in state 1, cannot absorb this beam. As the light of the probe laser pulse, tuned to state 3, arrives, the two beams shift the atoms to a quantum superposition of states 1 and 2, meaning that each atom is in both states at once. State 1 alone would absorb the probe light, and state 2 would absorb the coupling beam, each by moving atoms to state 3, which would then emit light at random. Together, however, the two processes cancel out, like evenly matched competitors in a tug of war—an effect called quantum interference. The superposition state is called a dark state because the atoms in essence cannot see the laser beams (they remain "in the dark"). The atoms appear transparent to the probe beam because they cannot absorb it in the dark state. Which superposition is dark—what ratio of states 1 and 2 is needed—varies according to the ratio of light in the coupling and probe beams at each location. But once the system starts in a dark state (in this case, 100 percent coupling beam and 100 percent state 1), it adjusts to remain dark even when the probe beam lights up.

A similar cancellation process makes the refractive index exactly one—like empty space—for probe light tuned precisely to state 3. At very slightly different frequencies, however, the cancellation is less exact and the refractive index changes. A short pulse of light "sniffs out" this variation in the index because a pulse actually contains a small range of frequencies. Each of these frequency components sees a different refractive index and therefore travels at a different velocity. This velocity, that of a continuous beam of one pure frequency, is the phase velocity. The pulse of light is located where all these components are precisely in sync (or, more technically, in phase). In an ordinary medium such as air or water, all the components move at practically the same velocity, and the place where they are in sync—the location of the pulse—also travels at that speed. When the components move with the range of velocities that occurs in the transparent atoms, the place where they are in sync gets shifted progressively farther back; in other words, the pulse is slowed. The velocity of the pulse is called the group velocity, because the pulse consists of a group of beams of different frequencies.

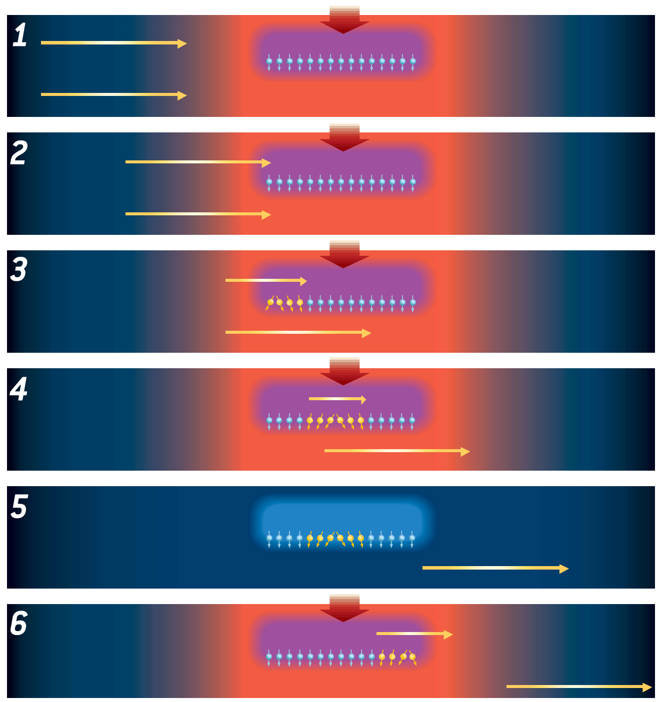
This process differs in a number of important respects from the usual slowing of light by a medium with a refractive index greater than one: the group velocity is slowed, not the phase velocity; a steep variation of the refractive index with frequency, not a large value of the index itself, causes the slowing; and the coupling laser beam has to be on the entire time.

Ultracold Atoms for Freezing Light

THE MORE RAPIDLY the refractive index changes with frequency, the slower the pulse travels. How rapidly the index can change is limited by the Doppler effect: the incessant thermal motion of the atoms in the gas smears out each atomic state across a small range of energies. The Doppler effect is like the change in tone of a siren moving toward or away from you. Imagine the cacophony you would hear if many police cars were racing toward and away from you at various speeds.

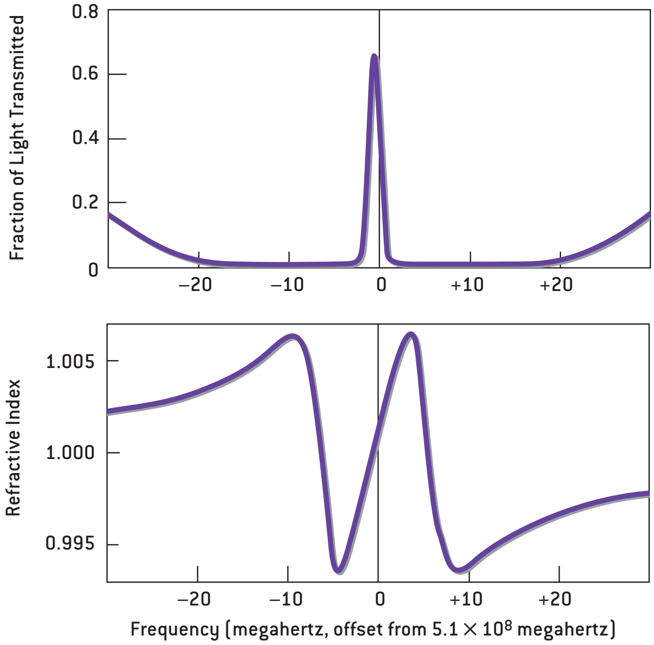
My research group uses extremely cold atoms (which move slowly) to minimize this Doppler spreading. Consequently, the energy states are sharply defined, and the variation of refractive index can be made very steep. Slow light in hot gases has since been obtained by Marlan O. Scully's group at Texas A&M University and Dmitry Budker's group at the University of California at Berkeley. The use of hot atoms removes the need to produce ultracold atoms, but it puts severe constraints on, for example, the geometry of the setup; the probe and coupling beams must propagate in exactly the same direction.

We chill our sodium atoms with a combination of laser beams, magnetic fields and radio waves. The atoms first emerge from a hot source as an intense beam, traveling about 2,600 kilometers an hour. A laser beam hits the atoms head-on and in a millisecond slows them to 160 kilometers an hour—a deceleration of 70,000 times gravity produced by a laser beam that wouldn't burn your finger. Further laser cooling in an optical molasses—six beams bathing the atoms from all sides—chills the atoms to 50 millionths of a degree above absolute zero. In a few seconds we accumulate 10 billion atoms in the molasses. Next we turn off the laser beams, plunging the lab into total darkness, and turn on electromagnets, whose combined field holds the atom cloud like a trap. For 38 seconds we



WAYLAYING LIGHT: Before the light pulse (*yellow*) reaches the cloud of cold atoms (*purple*) that will freeze it, all the atoms' spins (*small arrows*) are aligned and a coupling laser beam (*red*) renders the cloud transparent to the pulse (*1, 2*). The cloud greatly slows and compresses the pulse (*3*), and the atoms' states change; an imprint of the pulse is created in the cloud, and it accompanies the slow light. When the pulse is fully inside the cloud (*4*), the coupling beam is turned off (*5*), halting the imprint and the light; at zero velocity the light vanishes but the imprint stays. Later the coupling beam is turned on (*6*), regenerating the light pulse and setting the imprint and the light back in motion.

SAMUEL VELASCO



OPTICAL PROPERTIES **induced in a cloud of atoms by a carefully tuned laser beam are the key to the light-slowing process. A coupling laser beam passing through the cloud makes it transparent to light of a precise frequency (*top*) and causes an associated sharp variation of its refractive index (*bottom*). The transparency allows properly tuned light to pass through the cloud without being absorbed, and the steeper the change in the refractive index, the slower the light travels.**

LAURIE GRACE

cool the atoms through evaporation, kicking out the hotter atoms and leaving the cooler ones behind. Specially tuned radio waves help to speed the hot atoms on their way. This whole process—from hot atom beam to cold, trapped atoms—takes place inside a vacuum chamber pumped out to 10^{-14} (10 quadrillionths) of atmospheric pressure.

When we cool the cloud to about 500 billionths of a degree, it forms a Bose-Einstein condensate, a very odd state of matter in which the several million atoms left after the evaporative cooling behave in a completely synchronized fashion [see "The Coolest Gas in the Universe," by Graham P. Collins; Scientific American, December 2000]. These ultracold atom clouds, freely suspended in the middle of a vacuum chamber by a magnetic field, are the coldest places in the universe. And yet the rest of the experimental setup, within one centimeter of the cloud, is at room temperature. Vacuum-sealed windows on the chamber let us see the atoms directly by eye during laser cooling: a cold atom cloud in the optical molasses looks like a little bright sun, five millimeters in diameter. Such easy optical access allows us to massage the atoms with laser beams and make them do exactly what we want.

When our cigar of cold atoms is in place, we illuminate it from the side with the coupling laser. Then we launch a probe pulse along the axis of the cigar. To measure the speed of the light pulse, we do the most direct measurement imaginable: we sit behind the atom cloud with a light detector and wait for the pulse to come out, to see how long it takes. Immediately after the pulse has gone through, we measure the length of the cloud with yet another laser beam, shone from below to project the cloud's shadow onto a camera. That length divided by the delay of the pulse gives us the velocity. The delays are in the range of microseconds to milliseconds; this might sound short, but it is equivalent to light taking a detour through kilometers of optical fiber wound in a coil—and our clouds are only 0.1 to 0.2 millimeter long.

When we slow a light pulse down by a factor of 20 million, more happens than just a change of speed. At the start our pulse of light is a kilometer long, racing through the air at nearly 300,000 kilometers a second. (Of course, our laboratory's length is much less than a kilometer, but if we could place our laser that far away, its pulses would be that long in the air.) The pulse's leading edge crosses the glass window into the vacuum chamber and enters our levitating speck of sodium atoms. Inside this tenuous cloud the light travels at 60 kilometers an hour. A cyclist on a racing bike could overtake such sluggish light.

Through the Gas, Darkly

WITH THE FRONT of the light pulse traveling so slowly and its tail still going full tilt through the air, the pulse piles into the gas like a concertina. Its length is compressed by a factor of 20 million to a mere twentieth of a millimeter. Even though the atom cloud is small, the pulse compresses so much that it fits completely inside—important for stopping light. One might expect the light's intensity to increase greatly because the same amount of energy is crammed into a smaller space. This amplification does not happen, however; instead the electromagnetic wave remains at the same intensity. To put it another way, in free space the pulse contains 50,000 photons, but the slow pulse contains $\frac{1}{400}$ of a photon (the factor of 20 million again). What has happened to all the other photons and their energy? Some of that energy goes into the sodium atoms, but most of it is transferred to the coupling laser beam. We have monitored the intensity of the coupling laser to observe this energy transfer directly.

These transfers of energy also change the states of the sodium atoms where the pulse is passing by. At the front of the pulse the atoms are changed from their original state 1 to a superposition of states 1 and 2, the dark state. This state has the largest proportion of state 2 at the central, most intense part of the pulse. As the rear of the slow pulse leaves a region of atoms, the atoms revert to state 1. The spatial pattern of atomic dark states in the cloud mimics the shape of the compressed slow-light pulse and accompanies it through the gas as an imprint. When this imprint and the light pulse reach the end of the gas cloud, the light pulse sucks energy back out of the atoms and the coupling beam

THE AUTHOR

LENE VESTERGAARD HAU is a MacArthur Fellow and Gordon McKay Professor of Applied Physics and professor of physics at Harvard University. She received her Ph.D. in theoretical solid state physics from the University of Århus in Denmark. The author wishes to thank the team of Zachary Dutton, Christopher Slowe, Chien Liu, Cyrus H. Behroozi, Brian Busch and Michael Budde, as well as Stephen E. Harris of Stanford University, for an extremely fruitful collaboration.

to dash away through the air at its customary 300,000 kilometers a second, restored to its original kilometer of length.

The velocity of the slow light pulse depends on several parameters. Some of these are fixed once we choose our atom species and which excited states to use, but two of the variables are under our control: the density of the atom cloud and the intensity of the coupling laser beam. Increasing the cloud's density decreases the light's speed, but we can push that only so far, in part because very dense clouds leak atoms out of the magnetic trap too rapidly. The pulse speed is also reduced if the coupling laser beam is weaker. Of course, if the coupling laser is too feeble, the cloud will not be transparent and will absorb the pulse. Nevertheless, we can still achieve the ultimate in slowing without losing the pulse to absorption if we turn off the coupling laser beam while the compressed slowed pulse is contained in the middle of the gas.

In response, the light pulse comes to a grinding halt and turns off. But the information that was in the light is not lost. Those data were already imprinted on the atoms' states, and when the pulse halts, that imprint is simply frozen in place, somewhat like a sound recorded on a magnetic tape. The stopping process does not compress the pattern of states, because the imprinted atoms in unison force the light pulse to turn off, unlike the earlier stage in which the pulse gradually entered the gas.

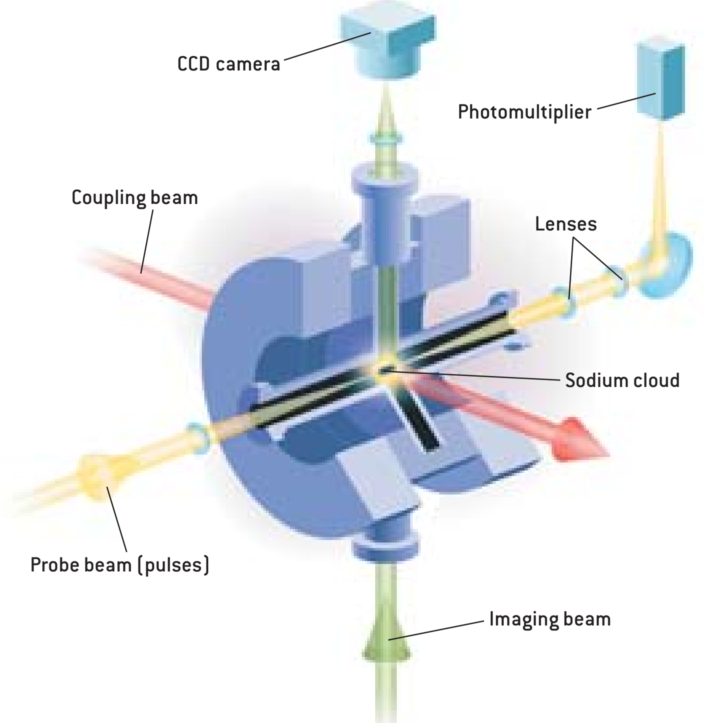
The frozen pattern imprinted on the atoms contains all the information about the original light pulse. We effectively have a hologram of the pulse written in the atoms of the gas. This hologram is read by turning the coupling laser back on. Like magic, the pulse reappears and sets off in slow motion again, along with the imprint in the atoms' states, as if nothing had interrupted it.

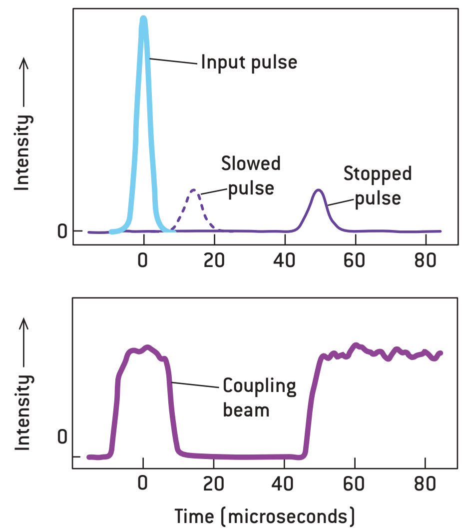
We can store the light for several milliseconds, long enough for a pulse to travel hundreds of kilometers in air. The pulse does degrade the longer it is stored, because even though the gas

A BENCHTOP GUIDE TO STOPPING LIGHT

EXPERIMENTAL SETUP

Three laser beams and an ultracold cloud of sodium atoms (*size exaggerated*) in a high vacuum lie at the heart of the slow-light experiment. The coupling beam interacts with the cloud, making it transparent but molasseslike to a pulse of the probe beam. A photomultiplier tube measures the pulse's time of arrival to better-than-microsecond precision. The imaging beam then measures the length of the cloud by projecting its shadow onto a camera. Not shown are the system that delivers and cools a new ultracold cloud for each pulse, electromagnets whose combined field holds the atoms in place, and additional details of the optics.





WHAT STOPPED LIGHT LOOKS LIKE

The precise times of detection of light pulses reveal the slowing and stopping of light. With no atom cloud present, the input pulse is detected at time "zero" (*top*). Slowing of the pulse by a cloud is revealed by the pulse's delay (*dotted curve*). To stop a pulse, the coupling beam (*bottom*) is turned off while the slowed pulse is inside the cloud. The time that the pulse is stopped—about 40 microseconds—adds to its delay. The slowed pulse loses intensity because the cloud is not perfectly transparent, but the pulse is stopped and revived with 100 percent efficiency.

GRAPHS BY LAURIE GRACE; EXPERIMENTAL SETUP BY SAMUEL VELASCO

CREATING QUANTUM SHOCK WAVES

IT TAKES A HURRICANE

FROZEN LIGHT can be used to examine the nature of superfluids, which flow without friction when close to absolute zero. To probe a Bose-Einstein condensate—a sodium superfluid—a light pulse was stopped at a roadblock within the condensate (*at the large black hemisphere in micrographs*), compressing a kilometer-long pulse to 0.001 millimeter. Within the compressed light pulse region, the sodium atoms were converted from their ground state to a slightly energized state, which leaves a narrow void in the condensate. This action causes density "dimples" that propagate out toward the boundaries, at the speed of sound in the condensate, which almost immediately forms quantum shock waves (moving to the right, at time 0 milliseconds). The fronts of the shock waves curve up (at 0.5 millisecond), creating quantized vortices with zero atom density at their centers (two white voids at 2.5 milliseconds)—the superfluid analogue of the eye of a hurricane. At 5 milliseconds there is a long sausage-shaped region (*black*) with high atom density, which relaxes over time and represents the presence of a large collective excitation of the condensate.

atoms are very cold they still move a bit, causing the pattern of dark states to diffuse slowly. In addition, collisions between atoms can disrupt the dark superposition states. After some milliseconds, the revived output pulse will begin to be weaker than the original. Yet these limits show that cold atoms allow for long storage times of immensely compressed optical information. And storage times are maximized with condensates in which atom collisions tend not to destroy the dark states.

We can also play some tricks. If the coupling beam is turned back on at a higher intensity, the output pulse will be brighter but shorter than the one we sent in. Turning the coupling beam on and off quickly several times regenerates the stored pulse in several pieces. Such manipulations demonstrate the degree of control that we have over stored pulses and may be extremely useful in future applications.

Since our initial observation in 2000 of stopped light, it has been obtained in a hot gas by Ronald L. Walsworth and Mikhail D. Lukin of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Mass., and in a cooled, doped solid by Philip R. Hemmer, then at the Air Force Research Laboratory in Hanscom, Mass.

Quantum Shock Waves

SLOW AND STOPPED LIGHT open up many interesting experiments. In our latest trial we turned everything around: rather than using cold atoms and condensates to slow light, we used slow light pulses to probe the odd properties of Bose-Einstein condensates.

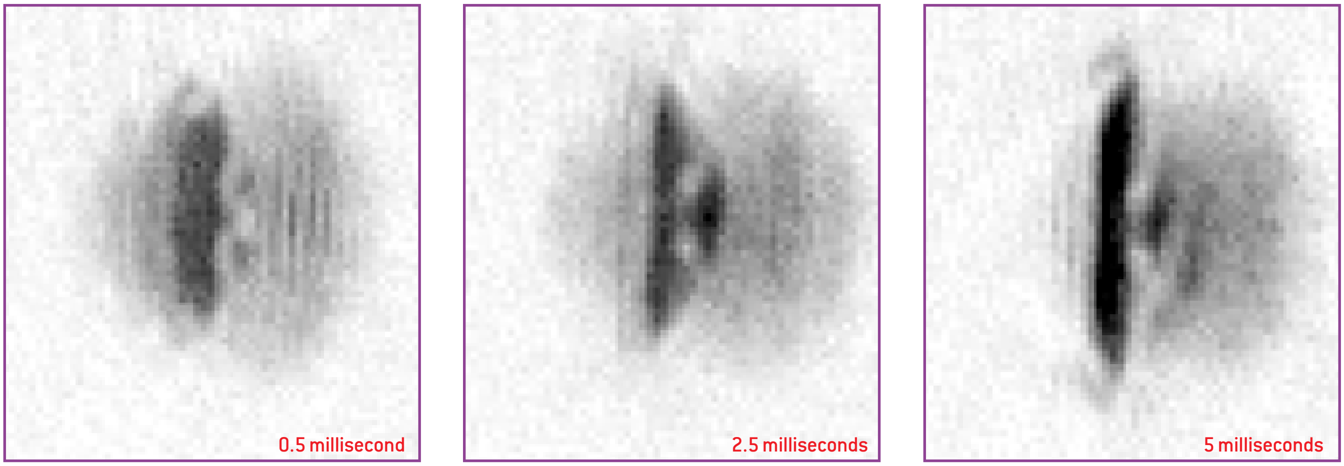
A condensate is a superfluid—it flows through a tube without friction, just as electric current flows through a superconductor without resistance. Helium turns superfluid when it is cooled to very low temperatures. Once a superfluid gets going, it flows forever with no need for further energy input. Superfluids are described by a characteristic length scale called the healing length, the minimum distance over which a superfluid can adjust (heal itself) to an external perturbation. In our condensates, the healing length is about 0.001 millimeter. So if we can compress our light pulses to that scale, we should be able to cause dramatic excitations of the superfluid, helping us to learn about the superfluid state.

We have succeeded in compressing a light pulse this drastically by creating a "roadblock" that forces the pulse to stop at a particular point in space, where it becomes totally compressed and localized. This is possible because when the probe light pulse is propagating at right angles to the coupling laser, we can spatially manipulate the intensity of the coupling laser along the propagation direction of the incoming light pulse. The speed of the pulse is controlled by the coupling laser intensity: the lower the intensity, the lower the speed.

To create the roadblock, we illuminate only the front half of the condensate with the coupling laser. When the light pulse enters the atom cloud, it slows to bicycle speed as before. Then, as the pulse runs into the roadblock region in the middle of the condensate, the pulse really slows down and compresses. Within the light pulse region at the roadblock, the atoms are almost entirely in state 2 (the dark state) because the coupling intensity is very low. Outside, the atoms are all in state 1. So we have a way of creating very localized defects in the condensate. We have taken direct images of this process in our laboratory.

As described earlier, for an atom in state 2, the spin of the valence electron is aligned opposite to the electron spin of an atom in state 1. This has severe consequences, because the electromagnet is used to trap atoms and hold them in place; atoms in state 2 will get kicked out of the magnet, as if a magnetic north pole was facing another north pole. So the localized defect of atoms in state 2 will zip out of the magnet in less than 0.5 millisecond. That process is by itself interesting and gives rise to what is called a pulsed atom laser.

Let's concentrate, however, on what is left in the trap. We end up with a condensate of state 1 atoms, with a hole punched in the middle. As you might imagine, condensates do not like holes punched in their middles, and the response we observed is fascinating. Two density "dimples" are created. They start propagating out toward the condensate boundaries, at the speed of sound in the condensate, about 1 to 10 millimeters a second. During this process, the back edge of the dimples steep-



ens because there is a dramatic variation of the atom density and thus the sound speed across the dimple structures. The back part of each dimple, where the density is high, will catch up to the central part, where the density is the lowest. Such a steepening of the back edge is what in a normal fluid, like water, would lead to shock wave formation. But because a condensate is a superfluid, the steepening forms the superfluid analogue of that phenomenon—a quantum shock wave, a quantized rotation pattern, a quantized vortex shaped like a smoke ring.

Our lab has photographed this process [*see illustration above*]. The central parts of the quantized vortices appear as white dots. In these places there is no atom density and the light just goes right through the condensate, as in the eye of a hurricane.

This process shows how the superfluid nature of the condensate will break down during the formation of these dramatic excitations. The vortices even seem to have a life of their own: they move around and sometimes bounce off one another like billiard balls. At other times they collide and seem to explode into a bunch of sound waves. Because a slow light pulse always creates vortices in pairs of opposite circulation, they are very much like particle and antiparticle pairs that can annihilate. We are excited about probing this rich system further.

Black Holes and Computers

CONDENSATES CAN BE PRODUCED in a vortex state wherein the gas rotates, reminiscent of water going down the drain. Ulf Leonhardt of the University of St. Andrews in Scotland has suggested that a pulse of slow light traveling through a vortex would find itself dragged along with the gas—very similar to a phenomenon believed to occur near black holes. With slow light, we can perhaps study this and some other black hole phenomena in the laboratory.

Slow light also enables a new kind of nonlinear optics that occurs, in particular, when one laser beam alters the properties of another beam. Nonlinear optics is a huge field of research, both of fundamental interest and with applications from imaging to telecommunications. Extremely intense beams are usually needed to achieve nonlinear optical effects, but with slow light the corresponding phenomena can be produced with a very small number of photons. Such effects could be useful for creating ultrasensitive optical switches.

Another application for slow and stopped light could be quantum computers, in which the usual 1's and 0's are replaced with quantum superpositions of 1's and 0's called qubits. Such computers, if they can be built, could solve certain problems that would take an ordinary computer a very long time. Two broad categories of qubits exist: those that stay in one place and interact with one another readily (such as quantum states of atoms) and those that travel rapidly from place to place but are difficult to make interact in the ways needed in a quantum computer (photons). The slow-light system, by transforming flying photons into stationary dark state patterns and back, provides a robust way to convert between these types of qubits, a process that could be essential for building large-scale quantum computers. We can imagine imprinting two pulses in the same atom cloud, allowing the atoms to interact and then reading out the result by generating new output light pulses.

Even if frozen light doesn't prove to be the most convenient and versatile component for building a quantum computer, it has opened up more than enough research applications to keep us—and other groups—busy for many more all-night sessions in the years to come. SA

MORE TO EXPLORE

Electromagnetically Induced Transparency. Stephen E. Harris in *Physics Today,* Vol. 50, No. 7, pages 36–42; July 1997.

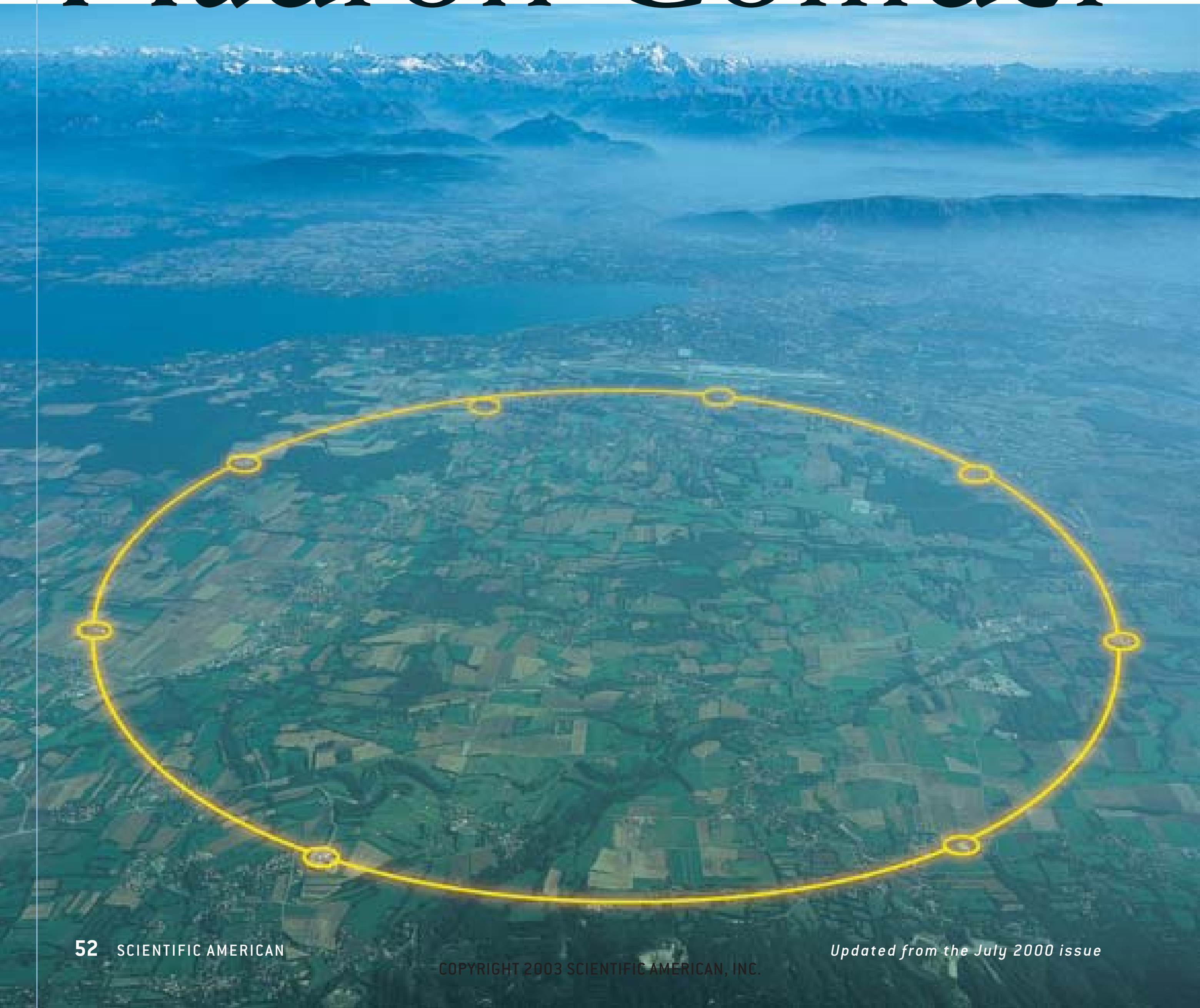
Light Speed Reduction to 17 Metres per Second in an Ultracold Atomic Gas. Lene Vestergaard Hau, S. E. Harris, Zachary Dutton and Cyrus H. Behroozi in *Nature,* Vol. 397, pages 594–598; February 18, 1999.

Observation of Coherent Optical Information Storage in an Atomic Medium Using Halted Light Pulses. Chien Liu, Zachary Dutton, Cyrus H. Behroozi and Lene Vestergaard Hau in *Nature,* Vol. 409, pages 490–493; January 25, 2001.

Observation of Quantum Shock Waves Created with Ultra-Compressed Slow Light Pulses in a Bose-Einstein Condensate. Zachary Dutton, Michael Budde, Christopher Slowe and Lene Vestergaard Hau in *Science,* Vol. 293, pages 663–668; July 27, 2001.

COURTESY OF LENE VESTERGAARD HAU

the Large Hadron Collider



Updated from the July 2000 issue

The Large Hadron Collider will be a particle accelerator of unprecedented energy and complexity, a global collaboration to uncover an exotic new layer of reality

By Chris Llewellyn Smith

When two protons traveling at 99.999999 percent of the speed of light collide head-on, the ensuing subatomic explosion provides nature with 14 trillion electron volts (TeV) of energy to play with. This energy, equal to 14,000 times that stored in the mass of a proton at rest, is shared among the smaller particles that make up each proton: quarks and the gluons that bind them together. In most collisions the energy is squandered when the individual quarks and gluons strike only glancing blows, setting off a tangential spray of familiar particles that physicists have long since catalogued and analyzed. On occasion, however, two of the quarks will themselves collide head-on with an energy as high as 2 TeV or more. Physicists are sure that nature has new tricks up her sleeve that must be revealed in those collisions—perhaps an exotic particle known as the Higgs boson, perhaps evidence of a miraculous effect called supersymmetry, or perhaps something unexpected that will turn theoretical particle physics on its head.

The last time that such violent collisions of quarks occurred in large numbers was billions of years ago, during the first picosecond of the big bang. They will start occurring again in 2007, in a circular tunnel under the Franco-Swiss countryside near Geneva. That's when thousands of scientists and engineers from dozens of countries expect to finish building the giant detectors for the Large Hadron Collider (LHC) and start experiments. This vast and technologically challenging project, coordinated by CERN (the European laboratory for particle physics), which is taking the major responsibility for constructing the accelerator, is already well under way.

The LHC will have about seven times the energy of the Tevatron collider based at Fermi National Accelerator Laboratory in Batavia, Ill., which discovered the long-sought "top" quark in experiments spanning from 1992 to 1995. The LHC will achieve its unprecedented energies despite being built within the confines of an existing 27-kilometer tunnel. That tunnel housed CERN's Large Electron-Positron Collider (LEP), which operated from 1989 to 2000 and was used to carry out precision tests of particle physics theory at about 1 percent of the LHC's energy. By using LEP's tunnel, the LHC avoids the problems and vast expense of siting and building a new, larger tunnel and constructing four smaller "injector" accelerators and supporting facilities. But bending the trajectories of the 7-TeV proton beams around the old tunnel's curves will require magnetic fields stronger than those any accelerator has used before. Those fields will be produced by 1,232 15-meter-long magnets installed around 85 percent of the tunnel's circumference. The magnets will be powered by superconducting cables carrying currents of 12,000 amps cooled by superfluid helium to –271 degrees Celsius, two degrees above the absolute zero of temperature.

STRADDLING THE FRANCO-SWISS BORDER, the location of the 27-kilometer tunnel that will house the Large Hadron Collider (LHC) 100 meters below the ground is indicated in gold. Smaller circles mark the positions of caverns that house detectors or ancillary equipment.

CERN

SUPERCONDUCTING MAGNET test string is laid out in the assembly hall; 1,232 large magnets will bend the trajectory of the two proton beams to follow the curve of the accelerator's tunnel.

LAURENT GUIRAUD *CERN*

To carry out productive physics experiments, one needs more than just high-energy protons. What counts is the energy of collisions between the protons' constituent quarks and gluons, which share a proton's energy in a fluctuating manner. The LHC will collide beams of protons of unprecedented intensity to increase the number of rare collisions between quarks and gluons carrying unusually large fractions of their parent protons' energy. The LHC's intensity, or luminosity, will be 100 times as great as that of previous colliders and 10 times that of the canceled Superconducting Super Collider (SSC). The SSC would have been a direct competitor to the LHC, colliding 20-TeV proton beams in an 87-kilometer-circumference tunnel around Waxahachie, Tex. The LHC's higher intensity will mostly compensate for the lower beam energy, but it will make the experiments much harder. Furthermore, such large intensities can provoke problems, such as chaos in the beam orbits, that must be overcome to keep the beams stable and well focused.

At four locations around the LHC's ring, a billion collisions will occur each second, each one producing about 100 secondary particles. Enormous detectors—the largest roughly the height of a six-story building—packed with thousands of sophisticated components will track all this debris. Elaborate computer algorithms will have to sift through this avalanche of data in real time to decide which cases (perhaps 10 to 100 per second) appear worthy of being recorded for full analysis later, off-line.

Unanswered Questions

AS WE STUDY nature with higher-energy probes, we are delving into the structure of matter at ever smaller scales. Experiments at existing accelerators have explored down to one billionth of one billionth of a meter (10^{-18} meter). The LHC's projectiles will penetrate even deeper into the heart of matter, down to 10^{-19} meter. This alone would be enough to whet scientific appetites, but pulses are really set racing by compelling arguments that the answers to major questions must lie in this new domain.

In the past 35 years, particle physicists have established a relatively compact picture—the Standard Model—that successfully describes the structure of matter down to 10^{-18} meter. The Standard Model [*see box on page 57*] succinctly characterizes all the known constituents of matter and three of the four forces that control their behavior. The constituents of matter are six particles called leptons and six called quarks. One of the forces, known as the strong force, acts on quarks, binding them together to form hundreds of particles known as hadrons. The proton and the neutron are hadrons, and a residual effect of the strong force binds them together to form atomic nuclei. The other two forces are electromagnetism and the weak force, which operates only at very short range but is responsible for radioactive beta decay and is essential for the sun's fuel cycle. The Standard Model elegantly accounts for these two forces as a "unified" electroweak force, which relates their properties despite their appearing very different.

More than 20 physicists have won Nobel Prizes for work that has contributed to the Standard Model, from the theory of quantum electrodynamics (the 1965 prize) to the discovery of the neutrino and the tau particle (1995) and the theoretical work of Gerardus 't Hooft and Martinus J. G. Veltman while at the University of Utrecht (1999). Nevertheless, although it is a great scientific achievement, confirmed by a plethora of detailed experiments, the Standard Model has a number of serious flaws.

First, it does not consistently include Albert Einstein's theory of the properties of spacetime and its interaction with matter. This theory, general relativity, provides a beautiful, experimentally very well verified description of the fourth force, gravity. The difficulty is that unlike general relativity, the Standard Model is a fully quantum-mechanical theory, and its predictions must therefore break down at very small scales (very far from the domain in which it has been tested). The absence of a quantum-mechanical description of gravity renders the Standard Model logically incomplete.

Second, although it successfully describes a huge range of data with simple underlying equations, the Standard Model contains many apparently arbitrary features. It is too byzantine to be the full story. For example, it does not indicate why there are six quarks and six leptons instead of, say, four. Nor does it explain why there are equal numbers of leptons and quarks—is this just a coincidence? On paper, we can construct theories that explain why there are deep connections between quarks and leptons, but we do not know if any of these theories is correct.

Third, the Standard Model has an unfinished, untested element. This is not some minor detail but a central component: a mechanism to generate the observed masses of the particles. Particle

THE AUTHOR

CHRIS LLEWELLYN SMITH is a senior research fellow in the physics department of the University of Oxford. He served as president and provost of University College London from 1999 to 2002. He spent periods at the Lebedev Institute in Moscow, at CERN near Geneva and at the Stanford Linear Accelerator Center in California before returning to Oxford for most of his career as a theoretical physicist. His 1994–98 term as director general of CERN saw the approval of the Large Hadron Collider project and the negotiation of new relations with nonmember countries, including Japan and the U.S., that will participate in the project.

masses are profoundly important—altering the mass of the electron, for example, would change all of chemistry, and the masses of neutrinos affect the expansion of the universe. (A neutrino's mass is at most a few millionths of an electron's mass, but recent experiments show that it is not zero. The scientists who led two pioneering experiments that made this discovery were awarded a share of the 2002 Nobel Prize for Physics.)

Higgs Mechanism

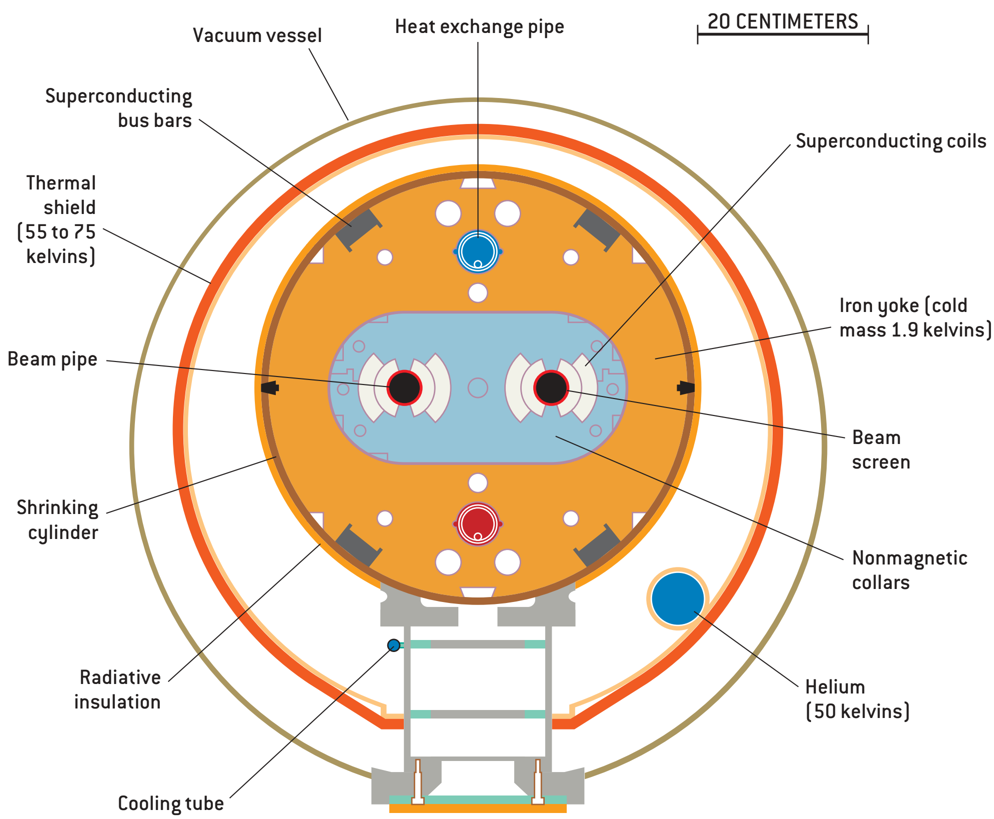
PHYSICISTS BELIEVE that particle masses are generated by interactions with a field that permeates the entire universe; the stronger a particle interacts with the field, the more massive it is [*see illustration on page 59*]. The nature of this field, however, remains unknown. It could be a new elementary field, called the Higgs field after British physicist Peter Higgs. Alternatively, it may be a composite object, made of new particles ("techniquarks") tightly bound together by a new force ("technicolor"). Even if it is an elementary field, there are many variations on the Higgs theme: How many Higgs fields are there, and what are their detailed properties?

Nevertheless, we know with virtually mathematical certainty that *whatever* mechanism is responsible, it must produce new phenomena in the LHC's energy range, such as observable Higgs particles (which would be a manifestation of ripples in the underlying field) or techniparticles. The principal design goal of the LHC is therefore to discover these phenomena and pin down the nature of the mass-generating mechanism.

The LHC experiments will also be sensitive to other new phenomena that could confirm one or another of the speculative theories that extend or complete the Standard Model. For example, it is widely thought that the more complete theory must incorporate a "super" symmetry. Supersymmetry would greatly increase the web of relations among the elementary particles and forces. Furthermore, so-called local supersymmetry automatically includes gravity; conversely, the only known theory (string theory) that could successfully combine general relativity and quantum mechanics requires supersymmetry. If supersymmetry is correct, physicists have very good reason to believe that the LHC can find the new particles that it predicts.

These new phenomena may be discovered before the LHC comes into operation by experiments at Fermilab's Tevatron, which started colliding beams of protons and antiprotons again in 2001 after a major upgrade. These experiments could find new phenomena beyond the range already explored by LEP. But even if they do "scoop" the LHC, they will reveal only the tip of a new iceberg, and the LHC will be where physicists make comprehensive studies of the new processes.

If the Tevatron does not observe these new phenomena, then the LHC will pick up the chase. The exploratory power of the LHC overlaps that of LEP and the Tevatron, leaving no gaps in which new physics could hide. Moreover, high-precision measurements made in the past decade at LEP, the Stanford Linear Accelerator Center and Fermilab have



ACCELERATOR MAGNET is shown in cross section. The superconducting coils carry 12,000 amps of current and must be kept cooled to below two kelvins. Each beam pipe carries one of the two countermoving proton beams. Other magnets focus the beams and bend them to cross at collision points within the detectors.

SLIM FILMS

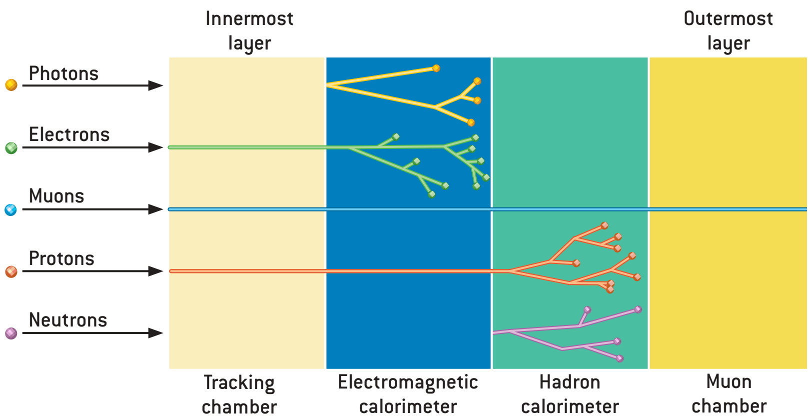
A TOROIDAL LHC APPARATUS (ATLAS) detector (*bottom*) uses a novel toroidal magnet system. Protons collide in the center, producing a spray of particles. The concentric layers of ATLAS detect different species of particles, some precisely tracking the particle trajectories, others ("calorimeters") measuring the energy carried. The simplified diagram (*below, left*) illustrates how such layers work. The toroidal magnets curve the tracks of charged particles, allowing their momenta to be measured. The image (*below, right*) shows simulated data of a collision in which a Higgs particle decays into four muons (*yellow tracks*).

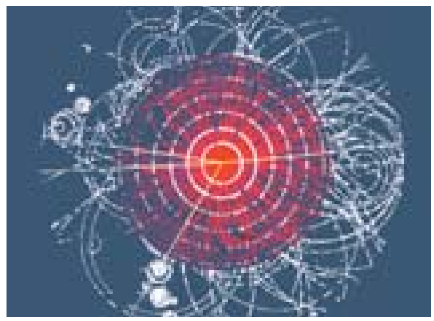
essentially eliminated worries that the Higgs boson might be out of reach of the LHC's energy range. It is now clear that either the Higgs boson or other new physics associated with the generation of mass will be found at the LHC.

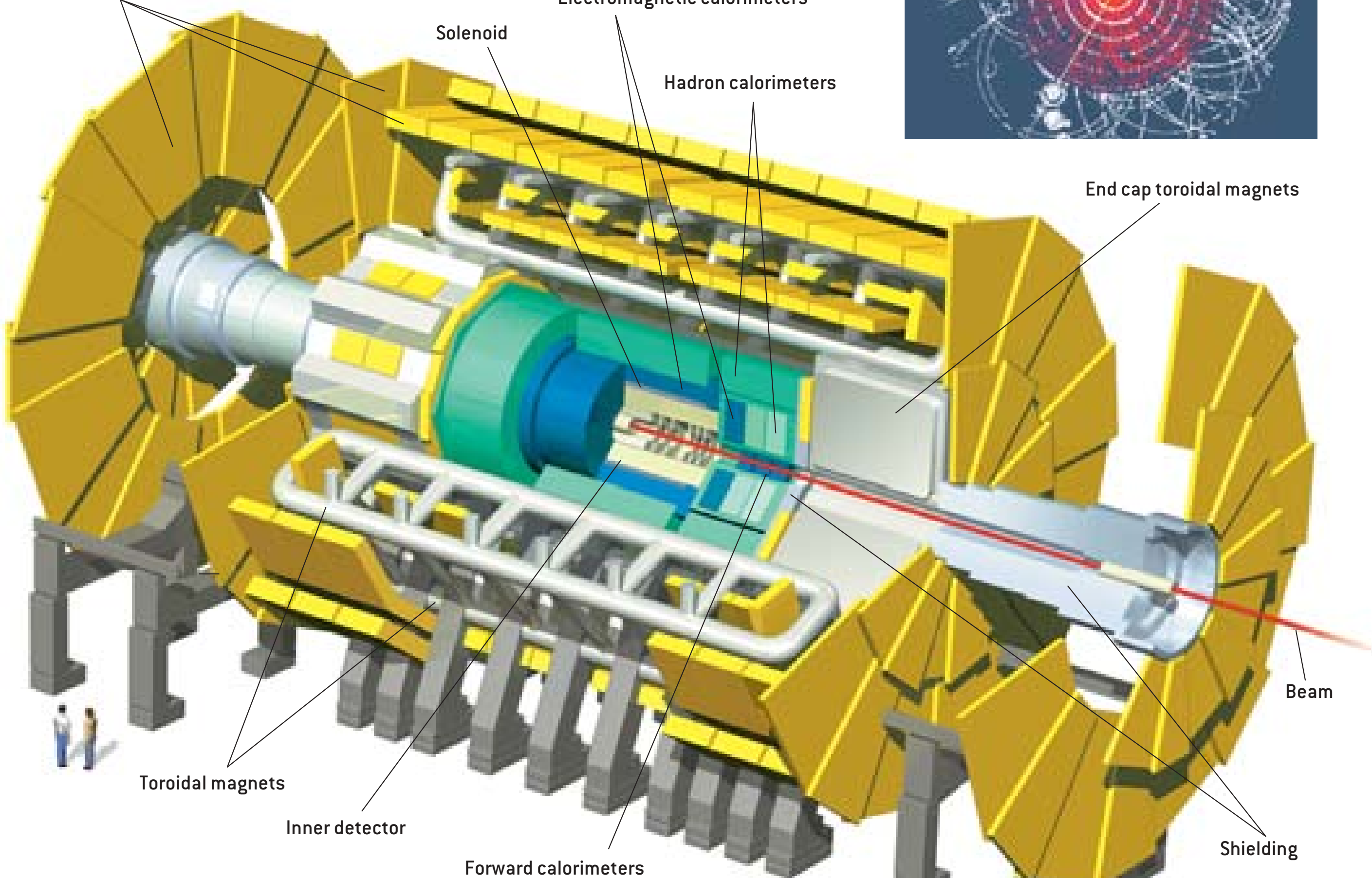
Emulating the Big Bang

TO ADDRESS THIS kind of physics requires re-creating conditions that existed just a trillionth of a second after the big bang, a task that will push modern technologies to their limits and beyond. To hold the 7-TeV proton beams on course, magnets must sustain a magnetic field of 8.3 tesla, almost 100,000 times the earth's magnetic field and the highest ever used in an accelerator. They will rely on superconductivity: large currents flowing without resistance through thin superconducting wires, resulting in compact magnets that can generate magnetic-field strengths unobtainable with conventional magnets made with copper wires [*see illustration on preceding page*]. To maintain the superconductivity under operating conditions—with 12,000 amps of current—the magnets' cores must be held at –271 degrees C around 22.4 kilometers of the tunnel. Cryogenics on this scale has never before been attempted.

In December 1994 a full prototype section of the LHC was operated for 24







SLIM FILMS (*bottom* and *top left*); CERN (*top right*) hours, demonstrating that the key technical choices for the magnets are correct. Since then, tests on prototypes have simulated about 10 years of running the LHC. Magnets that surpass the design criteria are now being produced in industry and delivered to CERN for final testing and subsequent installation.

With the 1993 demise of the planned 40-TeV SSC, the 14-TeV LHC became the only accelerator project in the world that can support a diverse research program at the high-energy frontier. The LHC's intense beams present those designing the experiments with remarkable challenges of data acquisition. The beams will consist of proton bunches strung like beads on a chain, 25 billionths of a second apart. At each collision point, pairs of these bunches will sweep through each other 40 million times a second, each time producing about 20 proton-proton collisions. Collisions will happen so often that particles from one collision will still be flying through the detectors when the next one occurs!

Of these 800 million collisions a second, only about one in a billion will involve a head-on quark collision. To keep up with this furious pace, information from the detector will go into electronic pipelines that are long enough to hold the data from a few thousand collisions. This will give "downstream" electronics enough time to decide whether a collision is interesting and should be recorded before the data reach the end of the pipeline and are lost. LHC detectors will have tens of millions of readout channels. Matching up all the pipelined signals that originate from the same proton-proton collision will be a mind-boggling task.

When Quarks Collide

PARTICLE DETECTORS are the physicists' electronic eyes, diligently watching each collision for signs of interesting events. LHC will have four particle detectors. Two will be giants, each built like a Russian *matryoshka* doll, with modules fitting snugly inside modules and a beam collision point at the center. Each module, packed with state-of-the-art technology, is custom-built to perform specific observations before the particles fly out to the next layer. These general-purpose detectors, ATLAS and CMS, standing up to 22 meters high, will look for Higgs particles and supersymmetry and will be on the alert for the unexpected, recording as much as possible of the collision debris. Two smaller detectors, ALICE and LHCb, will concentrate on different specific areas of physics.

Both ATLAS and CMS are optimized to detect energetic muons, electrons and photons, whose presence could signal the production of new particles, including Higgs bosons. Yet they follow very different strategies. Years of computer simulations of their performance have shown that they are capable of detecting whatever new phenomena nature may exhibit. ATLAS (*a t*oroidal *LHC a*pparatus) is based on an enormous toroidal magnet equipped with detectors designed to identify muons in air [*see illustration on opposite page*]. CMS (*c*ompact *m*uon *s*olenoid) follows the more traditional approach of using chambers inside the return yoke of a very powerful solenoidal magnet to detect muons [*see illustration on next page*].

Part of the CMS detector will consist of crystals that glow, or scintillate, when electrons and photons enter them. Such crystals are extremely difficult to make, and CMS benefits from the experience gained from a recent CERN experiment, L3, which also used crystals. (The L3 detector was one of four that operated from

SLIM FILMS

THE STANDARD MODEL

THE STANDARD MODEL of particle physics encompasses our knowledge of the fundamental particles. It contains particles of matter and particles that transmit forces. For example, the electromagnetic force between a proton and an electron is generated by photons (particles of light) being passed back and forth between them.

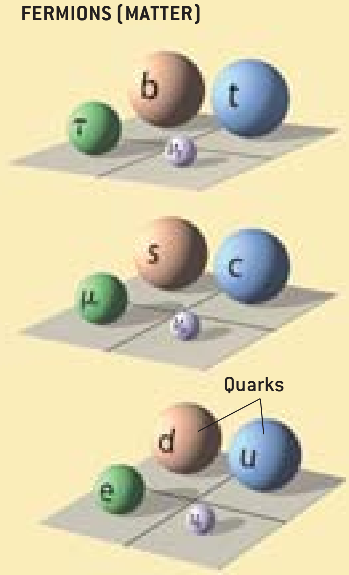
The matter particles come in three families of four, each family differing only by mass. All the matter around us is made of particles from the lightest family. These are "up" quarks, "down" quarks, electrons and electron-neutrinos. The other two families of matter particles exist only ephemerally after being created in high-energy collisions (neutrinos, however, are long-lived).

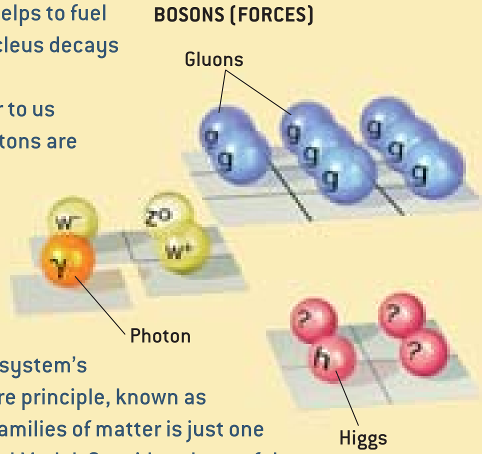
The quarks are stuck together by the strong force, carried by gluons, to form hadrons, which include the protons and neutrons that make atomic nuclei. Electrons, attracted to these nuclei by the electromagnetic force, orbit nuclei to form atoms and molecules. The weak interaction, carried by the *W* and *Z* particles, helps to fuel the sun and is responsible when an atomic nucleus decays and emits an electron and a neutrino.

Gravity, the weakest force, is most familiar to us because it acts on mass. Particles called gravitons are assumed to carry gravity, but they have not been detected, because the force is so weak. Also, gravitons are not yet properly incorporated into the Standard Model.

The entire system of matter and forces (except gravity) is encapsulated in a few simple equations derived from a function (the system's "Lagrangian") that is organized around one core principle, known as local gauge symmetry. Why nature has three families of matter is just one of many questions unanswered by the Standard Model. Considered one of the great intellectual triumphs of 20th-century science, the Standard Model can only be a stepping-stone to a more complete description of nature's forces.

—Graham P. Collins, staff writer





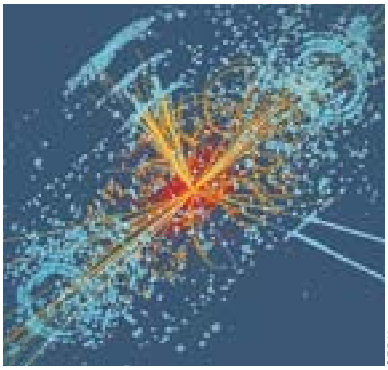
1989 to 2000 at the LEP collider, performing precision studies of the weak force that told us that exactly three types of zero- or low-mass neutrino exist.) Before L3, such crystals had been made only in small quantities, but L3 needed 11,000 of them. Crystals of the type developed for L3 have been widely used in medical imaging devices. CMS needs more than seven times as many crystals made of a more robust material. In due course the superior CMS crystals are likely to have an even bigger effect on the medical field.

ALICE (*a* *l*arge *i*on *c*ollider *e*xperiment) is a more specialized experiment that will come into its own when the LHC collides nuclei of lead with the colossal energy of 1,150 TeV. That energy is expected to "melt" the more than 400 protons and neutrons in the colliding nuclei, releasing their quarks and gluons to form a globule of quark-gluon plasma (QGP), which dominated the universe about 10 microseconds after the big bang. ALICE is based around the magnet of the L3 experiment, with new detectors optimized for QGP studies.

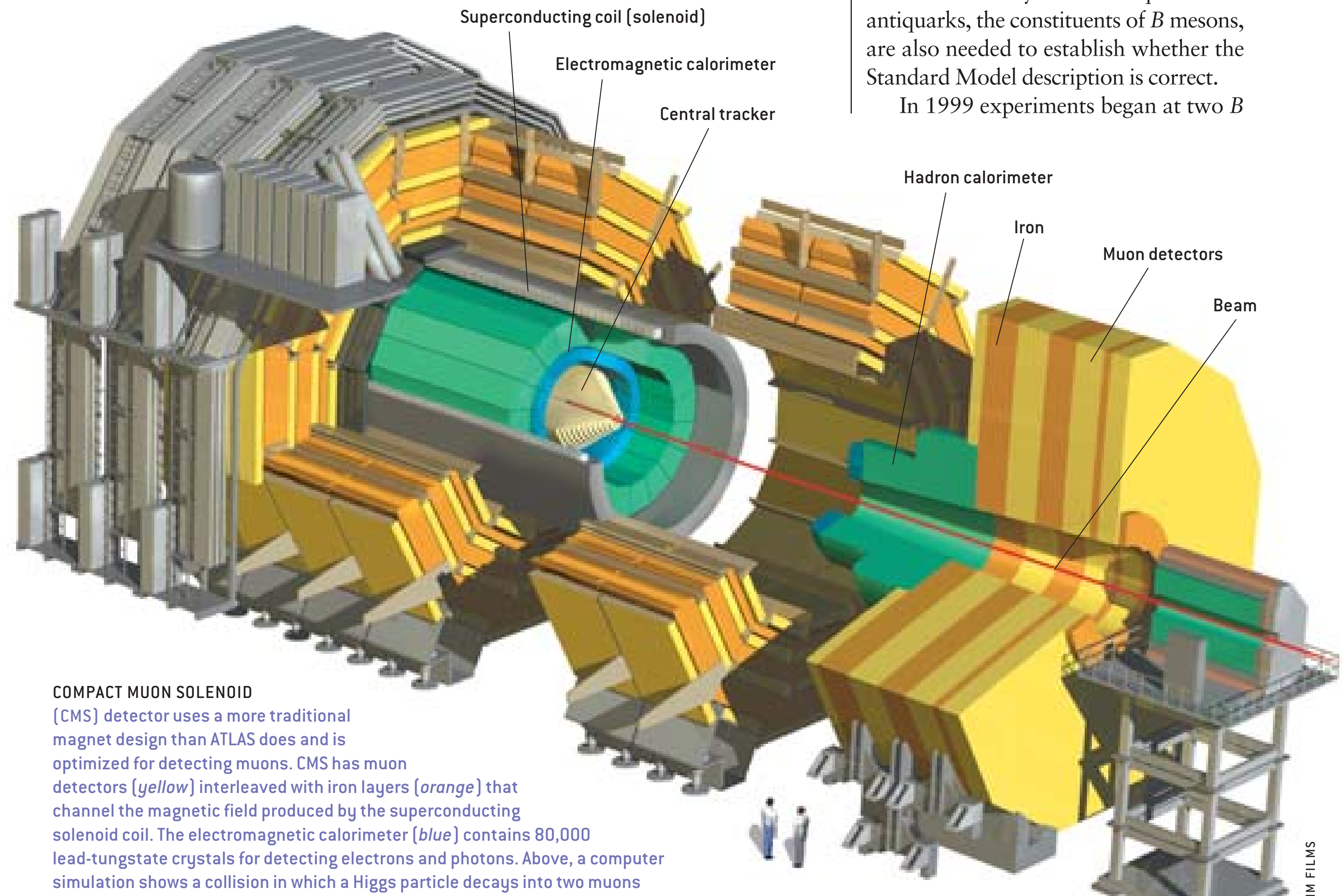
There is good evidence that experiments at CERN have already created a quark-gluon plasma. Over the coming years, Brookhaven National Laboratory's Relativistic Heavy Ion Collider (RHIC) has a good chance of studying QGP in detail by packing 10 times more energy per nucleon into its collisions than CERN does. The LHC will extend this by a further factor of 30. The higher energy at the LHC will complement the more varied range of experiments at RHIC, guaranteeing a thorough study of an important phase in the universe's early evolution.

B mesons, the subject of LHCb's investigations, could help tell us why the universe is made of matter instead of equal amounts of matter and antimatter. Such an imbalance can arise only if heavy quarks and antiquarks decay into their lighter cousins at different rates. The Standard Model can accommodate this phenomenon, called CP violation, but probably not enough of it to account completely for the dominance of matter in the universe. Physicists observed CP violation in the decay of strange quarks in the 1960s, but data on heavy "bottom" quarks and antiquarks, the constituents of *B* mesons, are also needed to establish whether the Standard Model description is correct.

In 1999 experiments began at two *B*



CERN



SLIM FILMS

COMPACT MUON SOLENOID (CMS) detector uses a more traditional magnet design than ATLAS does and is optimized for detecting muons. CMS has muon detectors (*yellow*) interleaved with iron layers (*orange*) that channel the magnetic field produced by the superconducting solenoid coil. The electromagnetic calorimeter (*blue*) contains 80,000 lead-tungstate crystals for detecting electrons and photons. Above, a computer simulation shows a collision in which a Higgs particle decays into two muons (the tracks at about "4 o'clock") and two jets of hadrons (at about "11 o'clock").

HOW THE HIGGS FIELD GENERATES MASS

"Empty" space, which is filled with the Higgs field, is like a roomful of people chatting quietly.

A particle crossing that region of space is like a celebrity arriving . . .

. . . and attracting a cluster of admirers who impede his progress—he acquires "mass."

HOW HIGGS PARTICLES ARE CREATED

Energy from a particle collision can be like a rumor crossing the room . . .

. . . creating a similar cluster that is self-sustaining, analogous to a Higgs particle itself.

factories in California and Japan that can produce tens of millions of *B* mesons a year. These experiments have observed the CP violation predicted by the Standard Model in one *B* meson decay mode. The high luminosity of the LHC beams can churn out a *trillion B* mesons a year for LHCb. This will allow much higher precision studies in a wider variety of circumstances and perhaps uncover crucial exotic decay modes too rare for the other factories to see clearly.

A Laboratory for the World

SCIENTIFIC EXPERIMENTS as ambitious as the LHC project are too expensive to be palatable for any one country. Of course, international collaboration has always played a role in particle physics, scientists being attracted to the facilities best suited to their research interests, wherever situated. As detectors have become larger and costlier, the size and geographic spread of the collaborations that built them have grown correspondingly. (It was the need to facilitate communication between the LEP collaborations that stimulated the invention of the World Wide Web by Tim Berners-Lee at CERN.)

The LHC accelerator originally had funding only from CERN's (then) 19 European member states, with construction to occur in two phases on a painfully slow timetable—a poor plan scientifically and more expensive in toto than a faster, single-phase development. Fortunately, additional funds from other countries (which will provide some 40 percent of the LHC's users) will speed up the project. Contributions of money or labor have been agreed to by Canada, India, Israel, Japan, Russia and the U.S. For example, Japan's KEK laboratory will supply 16 special focusing magnets. The U.S., with more than 550 scientists already involved, will furnish the largest national group; accelerator components will be designed and fabricated by Brookhaven, Fermilab and Lawrence Berkeley National Laboratory.

Furthermore, 5,000 scientists and engineers in more than 300 universities and research institutes in 50 countries on six continents are building the ATLAS and CMS detectors. When possible, components will be built in the participating institutions, close to students (who get great training by working on such projects) and in collaboration with local industries. The data analysis will also be dispersed. It will be a formidable challenge to manage these projects, with their stringent technical requirements and tight schedules, while maintaining the democracy and freedom for scientific initiatives that are essential for research to flourish.

Until now, CERN has been primarily a European laboratory. With the LHC, it is set to become a laboratory for the world. Already its 7,000 scientific users amount to more than half the world's experimental particle physicists. In 1994 John Peoples, Jr., then director of Fermilab, summed it up nicely: "For 40 years, CERN has given the world a living demonstration of the power of international cooperation for the advancement of human knowledge. May CERN's next 40 years bring not only new understanding of our Universe, but new levels of understanding among nations." SA

MORE TO EXPLORE

The Particle Century. Edited by Gordon Fraser. Institute of Physics, 1998.

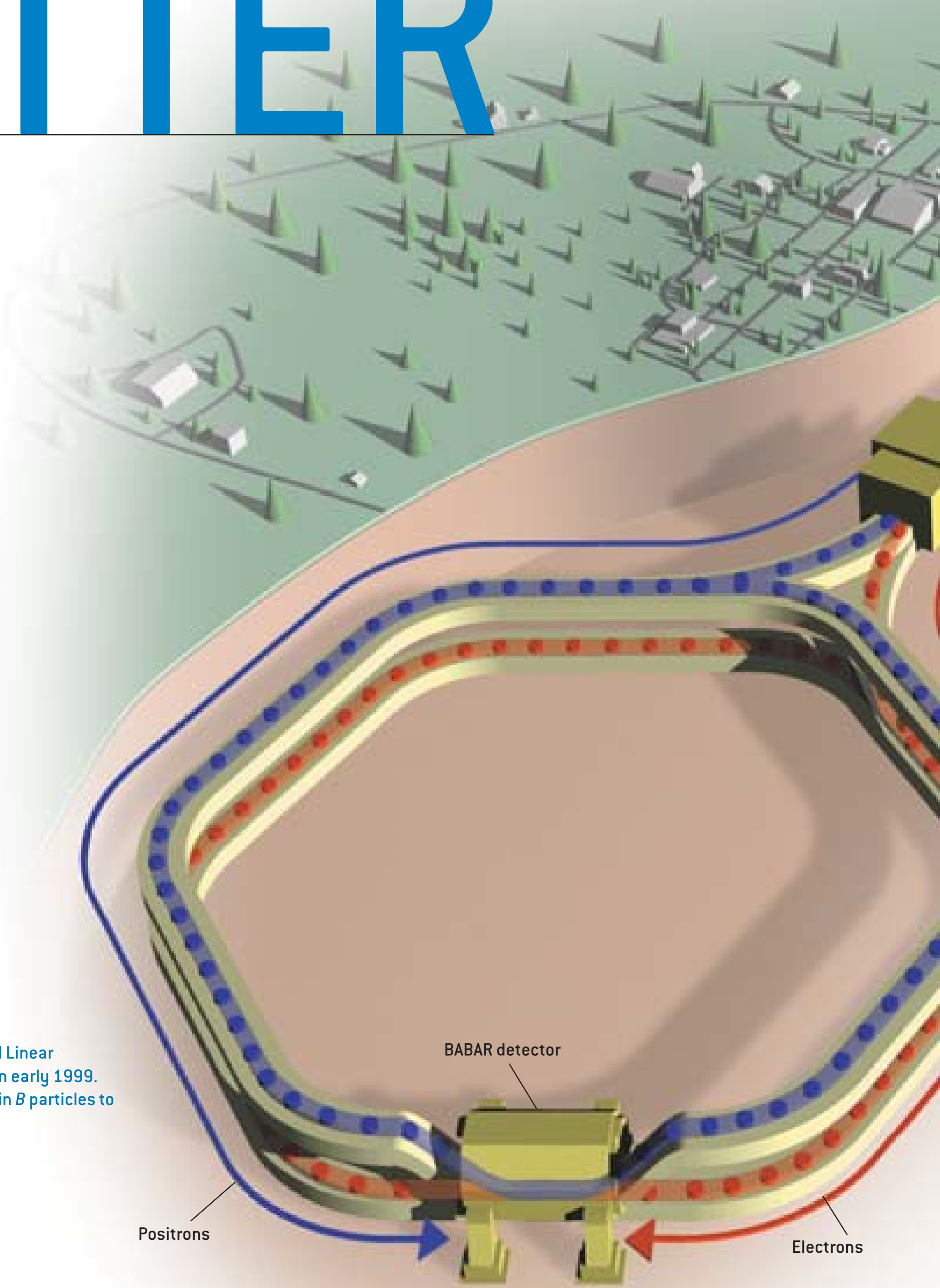
Supersymmetry. Gordon Kane. Perseus Publishing, 2000.

Links to home pages for all four LHC experiments are on the CERN Web site at **www.cern.ch/CERN/Experiments.html**

Two other excellent sites are **http://pdg.lbl.gov/atlas/atlas.html** and **www.particleadventure.org**

GEORGES BOIXADER *CERN*

the asymmetry between MATTER



B FACTORY, constructed at the Stanford Linear Accelerator Center, began taking data in early 1999. It examines violations of charge parity in *B* particles to set the stage for 21st-century physics.

and ANTI-MATTER

New accelerators are searching for violations in a **FUNDAMENTAL SYMMETRY OF NATURE,** throwing open a window to physics beyond the known

By Helen R. Quinn and Michael S. Witherell

As far as humans can see into the universe, an essential imbalance strikes the eye. Stars, planets, rocks, asteroids—everything is made of matter. Essentially no antimatter is evident.

Is this imbalance the result of an accident, a chance occurrence during the birth of the universe? Or is it an inevitable outcome of some asymmetry in the laws of nature? Theorists believe that the excess of matter comes from fundamental disparities in how matter and antimatter behave. These differences amount to violations of a symmetry called charge-parity reversal, or CP.

After years of effort, experimental and theoretical physicists have found a natural way for CP symmetry to be broken within the prevailing theory of particle physics, called the Standard Model. Curiously, the amount of CP violation the model predicts is too small to explain the matter excess in the universe.

This finding is a vital clue that not all is well with the Standard Model: unknown factors are very likely at play. Two new accelerators in California and in Japan have begun to probe violations of CP, with the aim of understanding whether the Standard Model needs to be revamped or replaced. These accelerators, which produce enormous numbers of particles called *B* mesons, are known as asymmetric *B* factories. They are the latest tool in the search for physics beyond the Standard Model.

Everything known about the elementary properties of matter is encapsulated in the Standard Model. It describes all the hundreds of observed particles and their interactions in terms of a few types of fundamental constituents: six quarks and six leptons. (The leptons are light particles, such as the electron, the neutrino and their relatives.) In addition, each quark and each lepton comes with an antiparticle, which has the same mass but the opposite sign for some quantum numbers, such as electric charge. These ingredients are arranged in three generations of increasing mass [*see box on opposite page*], the first of which provides the primary constituents of matter.

The Standard Model describes three kinds of interactions among particles: the familiar electromagnetic force as well as the strong and the weak forces. (For objects of such low mass, gravity is too weak to be of interest.) Strong interactions confine quarks, which are never seen alone, within composite objects such as protons. Weak interactions cause instability—in particular, the slow decays of the more massive quarks and leptons into objects of lower mass. All these forces are transmitted by specialized particles that also appear in the Standard Model: the photon, the gluon, and the *W* and *Z* bosons. Last, the theory requires an as yet unobserved Higgs particle, whose interactions are held responsible for the masses of the quarks and leptons as well as for much of their behavior.

Essential to the story of CP violation is a family of composite objects called mesons. A meson contains one quark and one antiquark, in an equal mixture of matter and antimatter. Of great significance is the set of mesons called kaons, or *K* mesons, which contain a strange quark or antiquark along with up or down quarks and antiquarks. Similar in many respects are the *B* mesons, which contain a bottom quark or antiquark paired with up or down partners.

Beyond the Standard

DESPITE ITS MANIFOLD successes in describing the behavior of matter, deep questions remain about the Standard Model. Physicists do not understand the mechanisms that determine the model's 18 parameters. For the theory to describe the world as we know it, some of those parameters must have very finely tuned values, and no one knows why those values would apply. More fundamentally, we do not understand why the model describes nature at all—why, for instance, should there be exactly three generations of leptons and quarks, no more or less? Finally, aspects of the theory that involve the Higgs particle are all untested. The Large Hadron Collider at CERN, the European laboratory for particle physics near Geneva, will allow the Higgs to be observed if its properties are as predicted by the Standard Model. The Higgs is believed to lie behind most of the mysteries of the Standard Model, including the violation of CP symmetry.

A theory of physics is said to have a symmetry if its laws apply equally well even after some operation, such as reflection, transforms parts of the physical system. An important example is the operation called parity reversal, denoted by P. This operation turns an object into its mirror reflection and rotates it 180 degrees about the axis perpendicular to the mirror [*see box on page 64*]. In mathematical terms, parity reverses the vectors associated with the object.

A theory has P symmetry if the laws of physics in the parity-reversed world are the same as they are in the real world. Par-

SLIM FILMS (*preceding pages and opposite page*)

Updated from the October 1998 issue

Deep questions remain about the Standard Model of everything known about matter.

ticles such as leptons and quarks can be classified as right- or left-handed depending on the sense of their internal rotation, or spin, around their direction of motion. If P symmetry holds, right-handed particles behave exactly the same as left-handed ones.

The laws of electrodynamics and the strong interactions are the same in a parity-reflected universe. But in a famous 1957 experiment, Chien-Shiung Wu of Columbia University and her collaborators found that the weak interactions are very different for particles of different handedness. Peculiarly, only left-handed particles can decay by means of the weak interaction; right-handed ones cannot. Moreover, it had long been held that there are no right-handed neutrinos, only left-handed ones. Because neutrinos have only weak interactions with the rest of the universe, this asymmetry is attributed to the weak force. So the weak force violates P.

Another basic symmetry of nature is charge conjugation, or C. This operation changes the quantum numbers of every particle into those of its antiparticle. Charge symmetry is also violated in weak interactions: antineutrinos were long considered only right-handed, not left-handed.

Theorists combine C and P to get the operation CP, which turns all particles into their antiparticles and also reverses the direction of all vectors. When subjected to CP, the left-handed neutrino becomes a right-handed antineutrino. Not only does the right-handed antineutrino exist, but its interactions with other particles are the same as they are for left-handed neutrinos. So although charge and parity symmetry are individually broken by neutrinos, in combination their dictates would seem to be obeyed.

Much to the surprise of physicists, the story of CP turned out to be far from simple. A mathematical theorem proved in 1917 by German mathematician Emmy Noether states that every symmetry implies the existence of a related quantity that is conserved, or immutable. For instance, the fact that spacetime is the same in all directions—that is, has rotational symmetry—leads to the conservation of angular momentum. Noether's theorem implies that if charge parity were an exact symmetry of nature, then a quantity called CP number would be conserved.

CP Violated

A PARTICLE and its antiparticle moving in opposite directions with equal energies form a pair with charge-parity symmetry.

PARTICLES OF THE STANDARD MODEL

THE PRIMARY CONSTITUENTS of matter, quarks and leptons, are divided into generations. The first generation contains up and down quarks and antiquarks, as well as the electron, a neutrino and their antiparticles. Ordinary matter is made almost exclusively of first-generation particles: an atom's nucleus contains protons and neutrons, themselves made of up and down quarks. The other generations occurred in the early universe, may still exist in hot environments such as neutron stars and are routinely observed in accelerators.

In addition, the Standard Model contains several particles that transmit force as well as a mysterious and unobserved particle called the Higgs. In the Standard Model the Higgs is responsible for the masses of all particles and for violations in charge-parity symmetry.

—H.R.Q. and M.S.W.

TRANSMITTERS OF FORCE

WEAK BOSONS	PHOTON	GLUON	HIGGS
w	γ	g	H

CONSTITUENTS OF MATTER

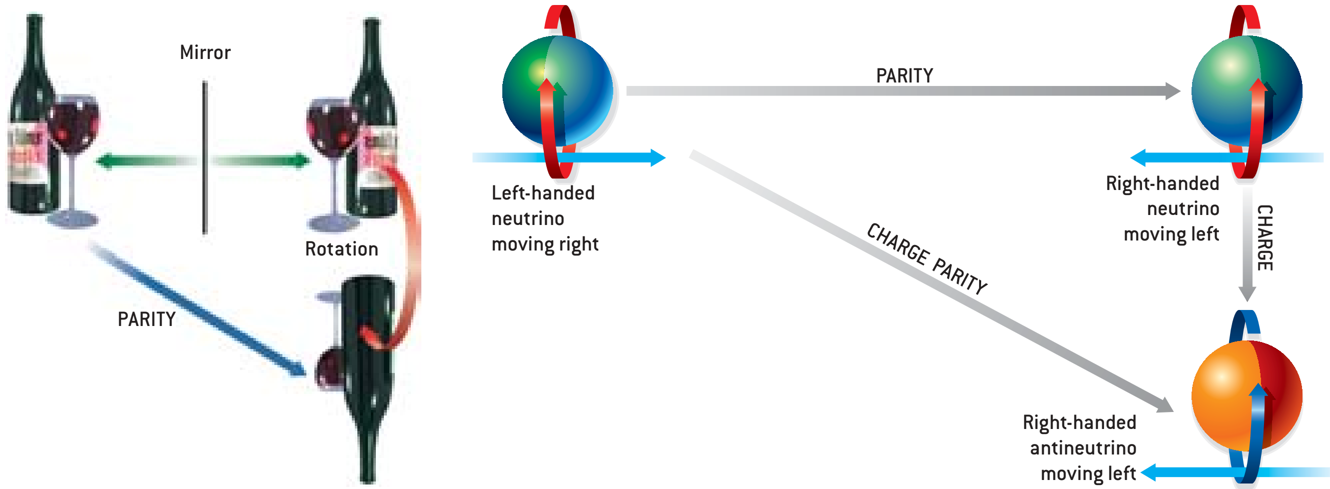
	PARTICLE	SYMBOL	CHARGE	MASS (GeV/c^2)
FIRST GENERATION				
Quarks	Up	u	+2/3	0.03
	Down	d	–1/3	0.06
Leptons	Electron	e^-	–1	0.0005
	Electron-neutrino	ν_e	0	?
SECOND GENERATION				
Quarks	Charm	c	+2/3	1.3
	Strange	s	–1/3	0.14
Leptons	Muon	μ^-	–1	0.106
	Muon-neutrino	ν_μ	0	?
THIRD GENERATION				
Quarks	Top	t	+2/3	174
	Bottom	b	–1/3	4.3
Leptons	Tau	τ	–1	1.7
	Tau-neutrino	ν_τ	0	?

REVERSAL OF CHARGE AND PARITY

SYMMETRIES ARE VITAL to the study of physics, and few symmetries are more intriguing than the combination of charge and parity. Charge reversal gives the opposite sign to quantum numbers such as electric charge, changing a particle to its antiparticle. Parity reversal reflects an object and also rotates it by 180 degrees (equivalent to changing the arrow on all vectors associated with the object).

The laws of classical mechanics and electromagnetism are invariant under either of these operations, as are the strong interactions of the Standard Model. The weak interactions, however, are changed by the reversal of either charge or parity.

For many years, it appeared that parity and charge flipped in succession ("charge parity") were invariant even for weak interactions. Experiments in 1964 shattered this illusion, posing the puzzle of why nature looks different when reflected in the charge-parity mirror. —*H.R.Q. and M.S.W.*



The CP operation does not change the system (taken as a whole), except that its mathematical representation acquires an overall factor: the CP number.

Either C or P, if acting twice on a system, returns it to the original state. This property is expressed as $C^2 = P^2 = 1$ (where 1, the identity operation, imparts no change at all). As a result, the CP number can be only +1 or –1. If nature has perfect charge-parity symmetry, then according to Noether's theorem no physical state with CP number –1 can transform into a state with CP number +1.

Consider the electrically neutral kaons. The K^0 consists of a down quark and an antistrange quark, whereas the anti-K^0 consists of an antidown quark and a strange quark. Because CP transposes quarks and antiquarks, it would turn each kaon into the other instead of leaving it unchanged. Therefore, neither of these kaons has a definite CP number. Theorists can, however, construct a pair of kaons with definite CP numbers by superposing the wave functions for K^0 and anti-K^0. According to the rules of quantum mechanics, these mixtures correspond to real particles and have definite mass and lifetime.

The conservation of CP number would explain an odd detail: the two "combination" kaons, though apparently similar, differ in their life spans by a factor of about 500 [*see illustration on page 67*]. The kaon with CP number +1 can change to two pions, a state that has the same CP number. This decay proceeds rapidly because the kaon is massive enough to yield two pions readily. But the kaon with CP number –1 can decay only to another state with CP number –1: three pions. This latter breakdown takes time, because the kaon has barely enough mass to generate three pions. So when physicists found a long-lived kaon in addition to a short-lived one, they acquired strong evidence that the combination kaons obeyed CP symmetry.

This tidy picture was shattered in 1964, when in a groundbreaking experiment at Brookhaven National Laboratory on Long Island, James Christenson, James Cronin, Val Fitch and René Turlay observed that about one out of every 500 of the long-lived kaons (those with CP number –1) decays into two pions. If CP were an exact symmetry of nature, it would forbid such a decay. Few experiments in particle physics have produced a result as surprising as this one. Theorists found it hard to see why CP symmetry should be broken at all and even harder

THE AUTHORS

HELEN R. QUINN and *MICHAEL S. WITHERELL* bring complementary skills to the study of charge-parity violation. Quinn is a theorist whose contributions include a demonstration of how the strong, weak and electromagnetic forces can become unified at high energies, as well as an explanation of why strong interactions conserve CP symmetry. She is currently involved in designing tests of the Standard Model being carried out at the *B* factories and has been a staff scientist at the Stanford Linear Accelerator Center since 1976. Witherell is an experimenter who led a Fermilab effort to study the decays of charm mesons, for which he obtained an award from the American Physical Society in 1990. He was instrumental in helping to build the silicon vertex detector for the BABAR experiment. He has been director of Fermilab since 1999 and was elected to the National Academy of Sciences in 1998.

to understand why any imperfection should be so small.

In 1972 Makoto Kobayashi and Toshihide Maskawa of Nagoya University in Japan showed that charge parity could be violated within the Standard Model if three or more generations of quarks exist. As it happened, only two generations of quarks—the first, containing the up and down, and the second, with the strange and charm—were known at the time. So this explanation began to gain currency only when Martin L. Perl and others at the Stanford Linear Accelerator Center (SLAC) spied τ (tau) leptons, the first particles of the third generation, in 1975. Two years later experimenters at Fermi National Accelerator Laboratory in Batavia, Ill., found the bottom quark. But only in 1995, when the top quark was nailed down, also at Fermilab, was the third generation completed.

Skewing the Universe

IT IS IMAGINABLE that the universe was born skewed—that is, having unequal numbers of particles and antiparticles. Such an initial imbalance, however, would be quickly eliminated if the early universe contained any processes that could change baryon number—the number of matter particles minus the number of antimatter particles. (In extensions of the Standard Model called Grand Unified Theories, such processes would have been very common soon after the big bang.) Theorists prefer the alternative scenario, in which particles and antiparticles were equally numerous in the early universe, but the former came to dominate as the universe cooled.

Soviet physicist (and dissident) Andrei Sakharov pointed out three conditions necessary for this asymmetry to develop. First, processes that do not conserve baryon number must exist. Second, during the expansion there must be some stage when the universe is not in thermal equilibrium. (When in thermal equilibrium, all states of equal energy contain equal populations of particles, and because particles and antiparticles have equal mass or energy, they would be generated at the same rate.) Third, CP symmetry—essentially, the symmetry between matter and antimatter—must be violated. Otherwise any process that changes the amount of matter would be balanced by a similar effect for antimatter.

The prevailing theory holds that when the universe was born, the quantum field associated with the Higgs particle was everywhere zero. Then, somewhere in the universe, a bubble developed, inside which the Higgs field assumed its current nonzero value. Outside the bubble, particles and antiparticles had no mass; once inside, however, they interacted with the Higgs field to acquire mass. But as the bubble grew, particles and antiparticles were swept through its surface at unequal rates because of CP violation. Any imbalances between matter and antimatter thus created outside the bubble were quickly corrected by processes that changed baryon number.

Such processes were extremely rare inside the bubble, however, so the imbalance was frozen in. By the time the bubble had expanded to occupy the entire universe, it contained more particles than antiparticles. Eventually the universe cooled to a point at which particles and antiparticles could no longer be generated in collisions but would annihilate when they found one another.

Unfortunately, when theorists calcu-

BABAR DETECTOR at Stanford University, built by a collaboration from nine nations, is designed to capture the decays of *B* mesons in charge-parity studies.

STANFORD LINEAR ACCELERATOR CENTER

Experimenters will identify almost every particle emerging from the decays of *B* mesons.

late how much of an imbalance between matter and antimatter this mechanism can create, it comes out too small—by orders of magnitude. This failure suggests that there must be other ways in which CP symmetry breaks down and hence that the Standard Model may be incomplete.

A fruitful place to search for more violations is most likely among the *B* mesons. The Standard Model predicts the various decays of the B^0 and the anti-B^0 to be highly asymmetric. A B^0 contains a down quark bound to an antibottom quark, whereas the anti-B^0 consists of an antidown quark and a bottom quark. The *B* mesons behave much like the kaons discussed earlier: the observed *B* mesons consist of mixtures of the B^0 and anti-B^0.

Consider the evolution of a B^0 meson produced at a certain instant. Some time later an observer has a certain probability of finding the same particle and also of finding its antiparticle. This peculiar meson state, oscillating between a given quark-antiquark combination and its antiparticle, is a remarkable illustration of quantum mechanics at work.

The Bottom Line

TO STUDY CP violation, experimenters need to study decays of B^0 into those final states that have a definite CP number. Such decays should proceed at a different rate for a particle that is initially B^0 compared with one that is initially anti-B^0. This difference will indicate the extent of CP violation in the system. But rather than resulting in the one-in-1,000 effect seen in K^0 decays, the predicted asymmetry for B^0 decays grows so large that one decay rate can become several times larger than the other.

Models other than the Standard often have additional sources of CP violation—sometimes involving extra Higgs particles—in general offering any value for imbalances in B^0 decays. Thus, measuring the pattern of asymmetries will provide a clear test of the predictions.

When the bottom quark was discovered, its mass was measured at around five giga-electron-volts (GeV), or about five times the mass of a proton. Theorists calculated that it would take a little more than 10 GeV of energy to produce two *B* mesons (because the added down or antidown quarks are very light). In the early 1980s at Cornell University, operators of an electron-positron collider—which accelerates electrons and positrons into head-on collisions—tuned it so that an electron-positron pair would release an energy of 10.58 GeV on annihilating. As had been predicted, this burst of energy preferentially converts to *B* mesons, providing a rich source of the particles. About one in four annihilations results in a *B* meson and its antiparticle, leaving behind no other particles in the aftermath.

At SLAC in 1983 experimenters found an unexpectedly long lifetime, about 1.5 picoseconds, for the *B* meson. The extended life improved the chances that a B^0 would turn into an anti-B^0 before decaying, making CP-violating asymmetries easier to observe. Furthermore, in 1987 experimenters at the Electron Synchrotron Laboratory (DESY) in Hamburg, Germany, measured this "mixing" probability at 16 percent, making it likely that the asymmetries would be far larger than those for the K^0. Still, these large asymmetries occur in relatively rare decays of the *B* mesons. For a true study of CP violation, a great number of *B* mesons would be needed.

In 1988 at a workshop in Snowmass, Colo., the major topic of interest was the Higgs particle. A group of participants also discussed CP violation, especially in *B* mesons, and determined that a favorable way to study the *B* mesons would be with an electron-positron collider tuned to 10.58 GeV, in which the electron and positron beams had different energies. This rather unusual feature would facilitate the measurement of a *B* meson's life span. Experimenters identify the point of birth and the point of death (that is, decay) of a *B* meson from traces of particles in the detector. Dividing the distance between these two points by the calculated velocity of the meson yields its life span. But an ordinary electron-positron collider at 10.58 GeV produces two *B* mesons that are almost at rest; the small distances they move are hard to measure.

Pier Oddone of Lawrence Berkeley National Laboratory had pointed out that if the electrons and positrons have different energies, the B^0 mesons that are produced move faster. For instance, if the electron beam has an energy of 9.0 GeV and the positron beam an energy of 3.1 GeV, the B^0 mesons move at half the speed of light, traveling about 250 microns (about one hundredth of an inch)

PARTICLE			ANTIPARTICLE		
NAME	SYMBOL	CONTENT	NAME	SYMBOL	CONTENT
Proton	p	u u d	Antiproton	$\bar{p}$	$\bar{u}$ $\bar{u}$ $\bar{d}$
Neutron	n	u d d	Antineutron	$\bar{n}$	$\bar{u}$ $\bar{d}$ $\bar{d}$
Pi-plus	π^+	u $\bar{d}$	Pi-minus	π^-	$\bar{u}$ d
Pi-zero	π^0	u $\bar{u}$ + d $\bar{d}$	Anti-pi-zero	$\bar{\pi}^0$	u $\bar{u}$ + d $\bar{d}$
K-zero	K^0	d $\bar{s}$	Anti-K-zero	$\bar{K}^0$	$\bar{d}$ s
B-zero	B^0	d $\bar{b}$	Anti-B-zero	$\bar{B}^0$	$\bar{d}$ b

COMPOSITE PARTICLES are either baryons (such as the proton and the neutron), made up of three quarks, or mesons, made up of one quark and one antiquark. The most common meson is a pion, containing up and down quarks and antiquarks. *K* mesons and *B* mesons, important to the study of charge-parity violation, contain strange and bottom quarks (or antiquarks), respectively.

SLIM FILMS

before they decay. Such a distance can yield a reasonably accurate measure of the lifetime.

An accelerator with two separate rings delivering different energies to the electrons and positrons would fit the task. Each ring would have to deliver intense beams of particles, obtaining a high rate of collisions. Such a machine came to be called an asymmetric *B* factory: asymmetric because of the different beam energies, and *B* factory because of the large numbers of *B* mesons it would produce.

Teams at several laboratories developed designs that could generate about 30 million pairs of *B* mesons a year. In 1993 the U.S. Department of Energy and the Japanese agency Monbusho approved two proposals for construction: one at SLAC in California and the other at KEK, the High Energy Accelerator Research Organization in Tsukuba, Japan. The SLAC project uses the existing linear tunnel to accelerate the positrons and electrons. These are then circulated in separate rings constructed in a 25-year-old tunnel and set to collide at a point of crossing. The accelerator construction cost $177 million. The Japanese project also employs extant tunnels—those that previously housed the Tristan collider.

Physicists and engineers are operating a large experiment that can identify the rare decays of a *B* meson and measure their positions to within the requisite 80 microns. This accuracy is obtained by using the silicon microstrip technology that helped to unearth the top quark [see "The Discovery of the Top Quark," by Tony M. Liss and Paul L. Tipton; Scientific American, September 1997]. Experimenters aim to identify almost every particle that emerges from the decays of the *B* mesons in order to isolate the rare events that shed light on charge-parity questions.

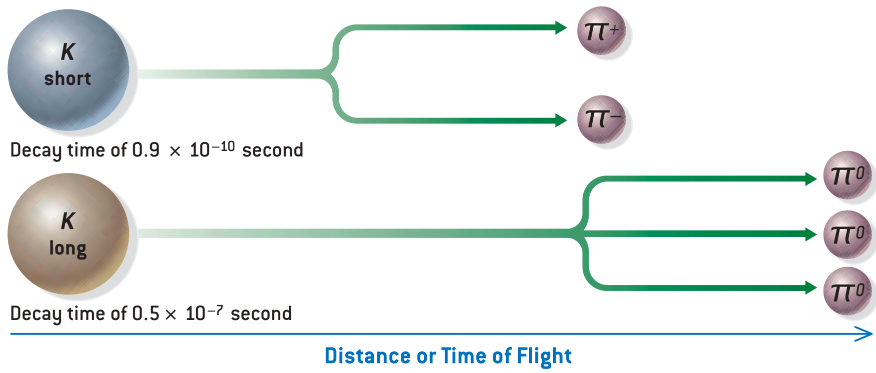
In the BABAR detector at SLAC, the silicon microstrip is the innermost layer, forming a cylinder roughly 30 centimeters in diameter and 60 centimeters long. Outer layers measure energy, velocity and penetration power for each particle created, allowing physicists to reconstruct the original events. More than 500 participants—including both of us—from 72 institutions in nine nations built the detector and shared its cost of $110 million. (It was, in fact, to facilitate international collaborations of this kind that the World Wide Web was invented at CERN.) The BELLE collaboration established to build the Japanese facility is also international in scope, with members from 10 countries. Both *B* factories were completed in 1999.

Other kinds of violations of charge parity, less predictable than the quantum-mechanical mixing, should also occur in *B* decays. The Cornell collider and detector have been upgraded to search for such effects. A number of experiments on *B* physics are either in planning stages or already under way at proton accelerators around the world. Both types of colliders will provide crucial, and complementary, evidence on CP violation.

The *B* factories could definitively tell researchers that the Standard Model concept works and then help to determine its remaining parameters. Alternatively, they could show that the model's predictions cannot fit the data no matter what the choice of parameters. Indeed, the results could rule out entire classes of models beyond the Standard Model, thus helping theorists to zero in on a successor. And if all goes well, we may even come to understand why our world is made exclusively of matter.

Authors' note: Since this article was written, the two *B* factories have begun achieving better than expected data collection rates, and their researchers have reported many results. So far the Standard Model reigns supreme. One significant CP-violating asymmetry between *B* decays and anti-*B* decays has been observed; it is consistent with the predictions of that theory. Many other measurements are under way, but most results are not yet at the level of precision needed to probe the theory further. Like any good factory, these facilities must work hard for years to achieve their total expected output.

Exciting recent results in neutrino physics have invalidated one statement we made; these data indicate that neutrinos have mass, which means that right-handed neutrinos must exist, along with the left-handed ones. Likewise, there must be left-handed antineutrinos. The masses are tiny and do not affect any other statements made in this article. They do, however, open up another possible place where CP violation can occur, in the neutrino and the quark sectors. Physicists are just beginning to explore these possibilities. SA



NEUTRAL KAONS, or *K* mesons, are observed to have two very different life spans. One type of kaon decays quickly into two pions, whereas the other decays slowly into three pions. The different behavior comes from the two kaons' having opposite charge-parity symmetry. On rare occasions, however, the second type of kaon also decays to two pions, proving that charge parity can be violated.

SLIM FILMS

MORE TO EXPLORE

The Character of Physical Law. Richard Feynman. MIT Press, 1965.

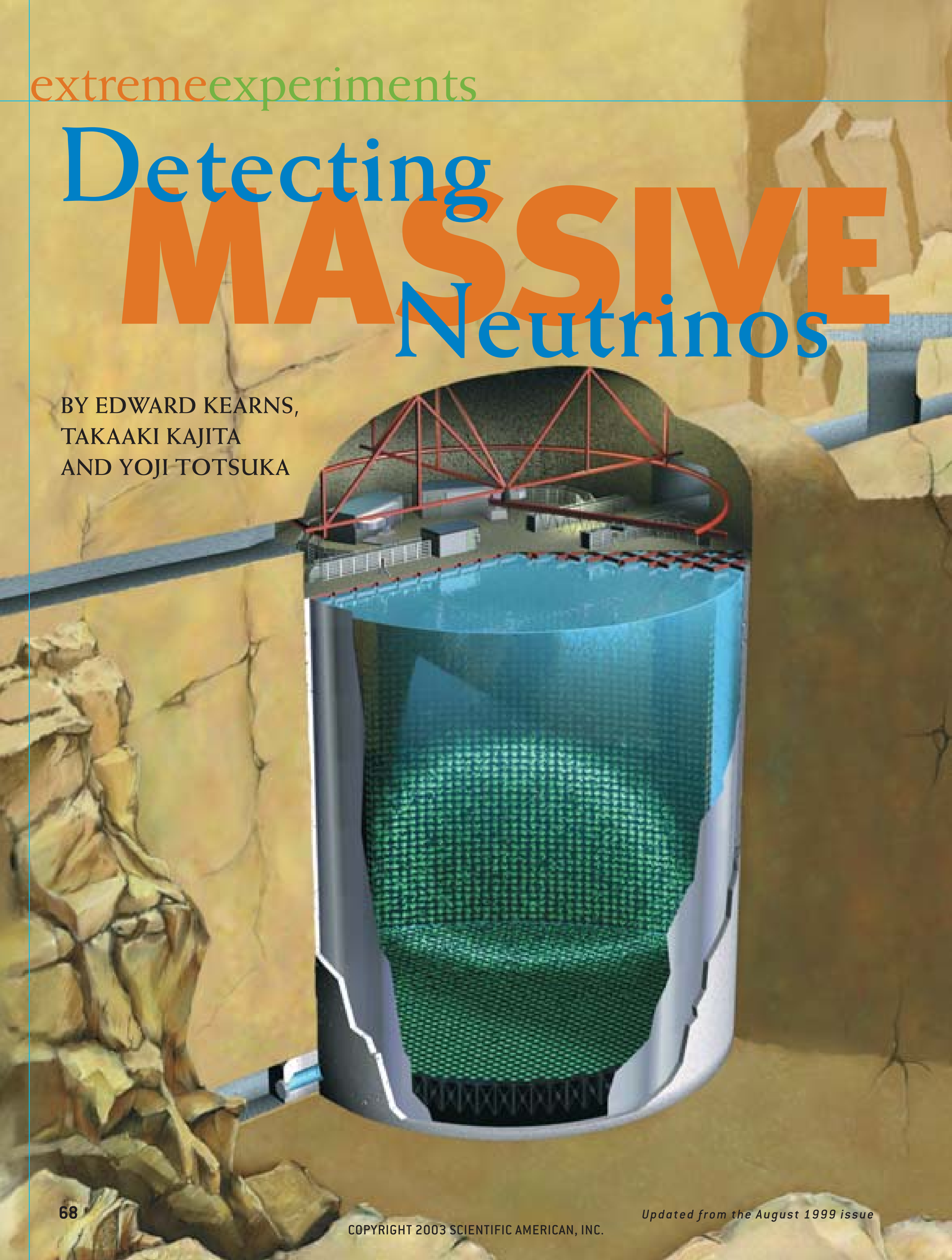
Fearful Symmetry: The Search for Beauty in Modern Physics. A. Zee. Macmillan Publishing, 1986.

The Physics of Time Reversal. Robert G. Sachs. University of Chicago Press, 1987.

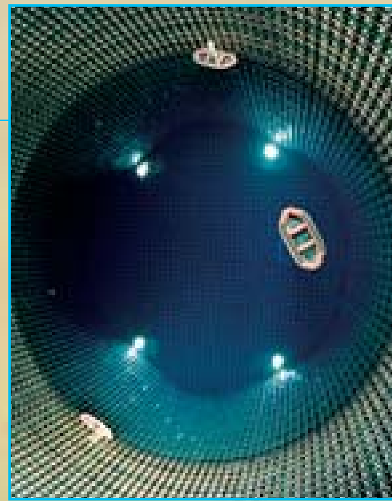
The New Ambidextrous Universe: Symmetry and Asymmetry from Mirror Reflections to Superstrings. Martin Gardner. W. H. Freeman and Company, 1990.

Detecting MASSIVE Neutrinos

BY EDWARD KEARNS, TAKAAKI KAJITA AND YOJI TOTSUKA



Updated from the August 1999 issue



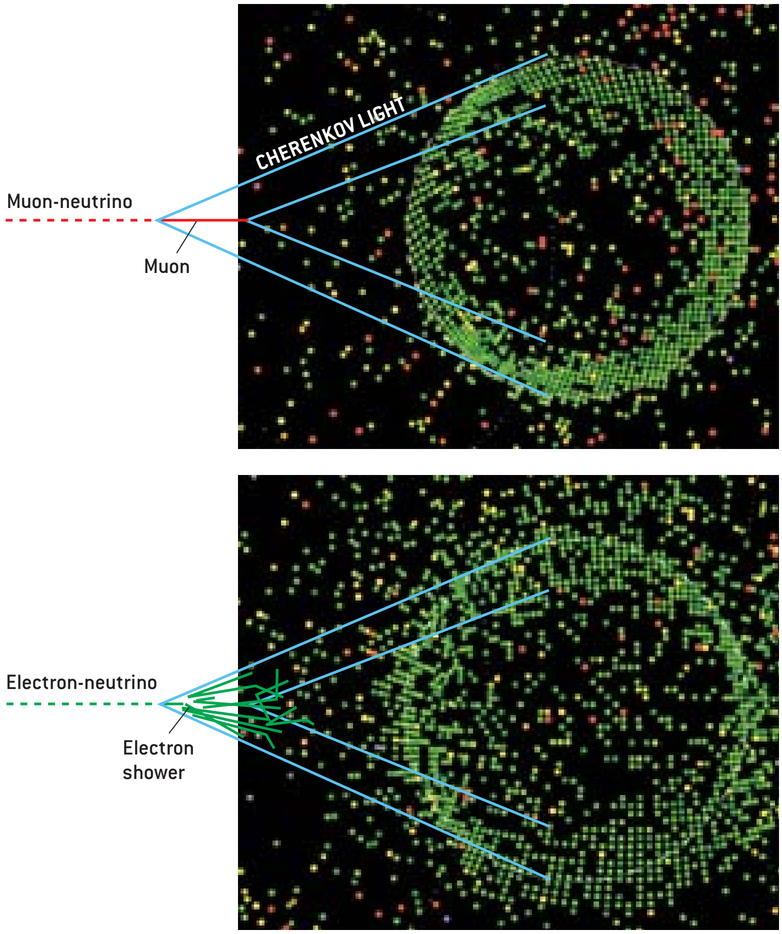
INSTITUTE FOR COSMIC RAY RESEARCH (ICRR), UNIVERSITY OF TOKYO

SUPER-KAMIOKANDE DETECTOR resides in an active zinc mine inside Mount Ikenoyama. Its stainless-steel tank contains 50,000 tons of ultrapure water so transparent that light can pass through 70 meters of it before losing half its intensity (for a swimming pool that figure is a few meters). The water is monitored by 11,000 photomultiplier tubes that cover the walls, floor and ceiling. Each tube is a handblown, evacuated glass bulb half a meter in diameter. The tubes register conical flashes of Cherenkov light, each of which signals a rare collision of a high-energy neutrino and an atomic nucleus in the water. Technicians in inflatable rafts clean the bulbs while the tank is filled (*right*).

A giant detector in the heart of Mount Ikenoyama in Japan has demonstrated that neutrinos metamorphose in flight, strongly suggesting that these ghostly particles have mass

One man's trash is another man's treasure. For a physicist, the trash is "background"—some unwanted reaction, probably from a mundane and well-understood process. The treasure is "signal"—a reaction that we hope will reveal new knowledge about the way the universe works. Case in point: over the past two decades, several groups have been hunting for the radioactive decay of the proton, an exceedingly rare signal (if it occurs at all) buried in a background of reactions caused by elusive particles called neutrinos. The proton, one of the main constituents of the atom, seems to be immortal. Its decay would be a strong indication of processes described by the Grand Unified Theories that many believe lie beyond the extremely successful Standard Model of particle physics. Huge proton-decay detectors were placed deep underground, in mines or tunnels around the world, to escape the constant rain of particles called cosmic rays. But no matter how deep they went, these devices were still exposed to penetrating neutrinos produced by the cosmic rays.

The first generation of proton-decay detectors, operating from 1980 to 1995, saw no signal, no signs of proton decay—but along the way the researchers found that the supposedly mundane



CONES OF CHERENKOV LIGHT are emitted when high-energy neutrinos hit a nucleus and produce a charged particle. A muon-neutrino (*top*) creates a muon, which travels perhaps one meter and projects a sharp ring of light onto the detectors. An electron, produced by an electron-neutrino (*bottom*), generates a small shower of electrons and positrons, each with its own Cherenkov cone, resulting in a fuzzy ring of light. Green dots indicate light detected in the same narrow time interval.

neutrino background was not so easy to understand.

One such experiment, Kamiokande, was located in Kamioka, Japan, a mining town about 250 kilometers (155 miles) from Tokyo (as the neutrino flies). Scientists there and at the IMB experiment, located in a salt mine near Cleveland, Ohio, used sensitive detectors to peer into ultrapure water, waiting for the telltale flash of a proton decaying.

Such an event would have been hidden, like a needle in a small haystack, among about 1,000 similar flashes caused by neutrinos interacting with the water's atomic nuclei. Although no proton decay was seen, the analysis of those 1,000 reactions uncovered a real treasure—tantalizing evidence that the neutrinos were unexpectedly fickle, changing from one species to another in midflight. If true, that phenomenon was just as exciting and theory-bending as proton decay.

Neutrinos are amazing, ghostly particles. Every second, 60 billion of them, mostly from the sun, pass through each square centimeter of your body (and of everything else). But because they seldom interact with other particles, generally all 60 billion go through you without so much as nudging a single atom. In fact, you could send a beam of such neutrinos through a light-year-thick block of lead, and most of them would emerge unscathed at the end. A detector as large as Kamiokande catches only a tiny fraction of the neutrinos that pass through it every year.

Neutrinos come in three flavors, corresponding to their three charged partners in the Standard Model: the electron and its heavier relatives, the muon and the tau particle. An electron-neutrino interacting with an atomic nucleus can produce an electron; a muon-neutrino makes a muon; a tau-neutrino, a tau. For most of the seven decades since neutrinos were first posited, physicists have assumed that they are massless. But if they can change from one flavor to another, quantum theory indicates that they most likely have mass. And in that case, these ethereal particles could collectively outweigh all the stars in the universe.

A Bigger Neutrino Trap

AS IS SO OFTEN the case in particle physics, the way to make progress is to build a bigger machine. Super-Kamiokande, or Super-K for short, took the basic design of Kamiokande and scaled it up by about a factor of 10 [*see illustration on preceding pages*]. An array of light-sensitive detectors looks in toward the center of 50,000 tons of water whose protons may decay or get struck by a neutrino. In either case, the reaction creates particles that are spotted by means of a flash of blue light known as Cherenkov light, discovered by Pavel A. Cherenkov in 1934. Much as an aircraft flying faster than the speed of sound produces a shock wave of sound, an electrically charged particle (such as an electron or a muon) emits Cherenkov light when it exceeds the speed of light in the medium in which it is moving. This motion does not violate Einstein's theory of relativity, for which the crucial velocity is *c*, the speed of light in a vacuum. In water, light propagates 25 percent slower than *c*, but other highly energetic particles can still travel almost as fast as *c* itself. Cherenkov light is emitted in a cone along the flight path of such particles.

In Super-K, the charged particle generally travels just a few meters and the Cherenkov cone projects a ring of light onto the wall of photon detectors [*see il-*

THE AUTHORS

EDWARD KEARNS, TAKAAKI KAJITA and *YOJI TOTSUKA* are members of the Super-Kamiokande Collaboration. Kearns, professor of physics at Boston University, and Kajita, professor of physics at the University of Tokyo, lead the analysis team that studies proton decay and atmospheric neutrinos in the Super-Kamiokande data. Totsuka recently became the director of KEK, Japan's national particle physics laboratory, after serving as spokesperson for Super-K since its inception.

DAVID FIERSTEIN (*preceding pages*); ICRR, UNIVERSITY OF TOKYO (*this page*)

lustration on opposite page*]. The size, shape and intensity of this ring reveal the properties of the charged particle, which in turn tell us about the neutrino that produced it. We can distinguish the Cherenkov patterns of electrons from those of muons: the electrons generate a shower of particles, leading to a fuzzy ring quite unlike the crisper circle from a muon. From the Cherenkov light we also measure the energy and direction of the electron or muon, which are decent approximations of the energy and direction of the neutrino.

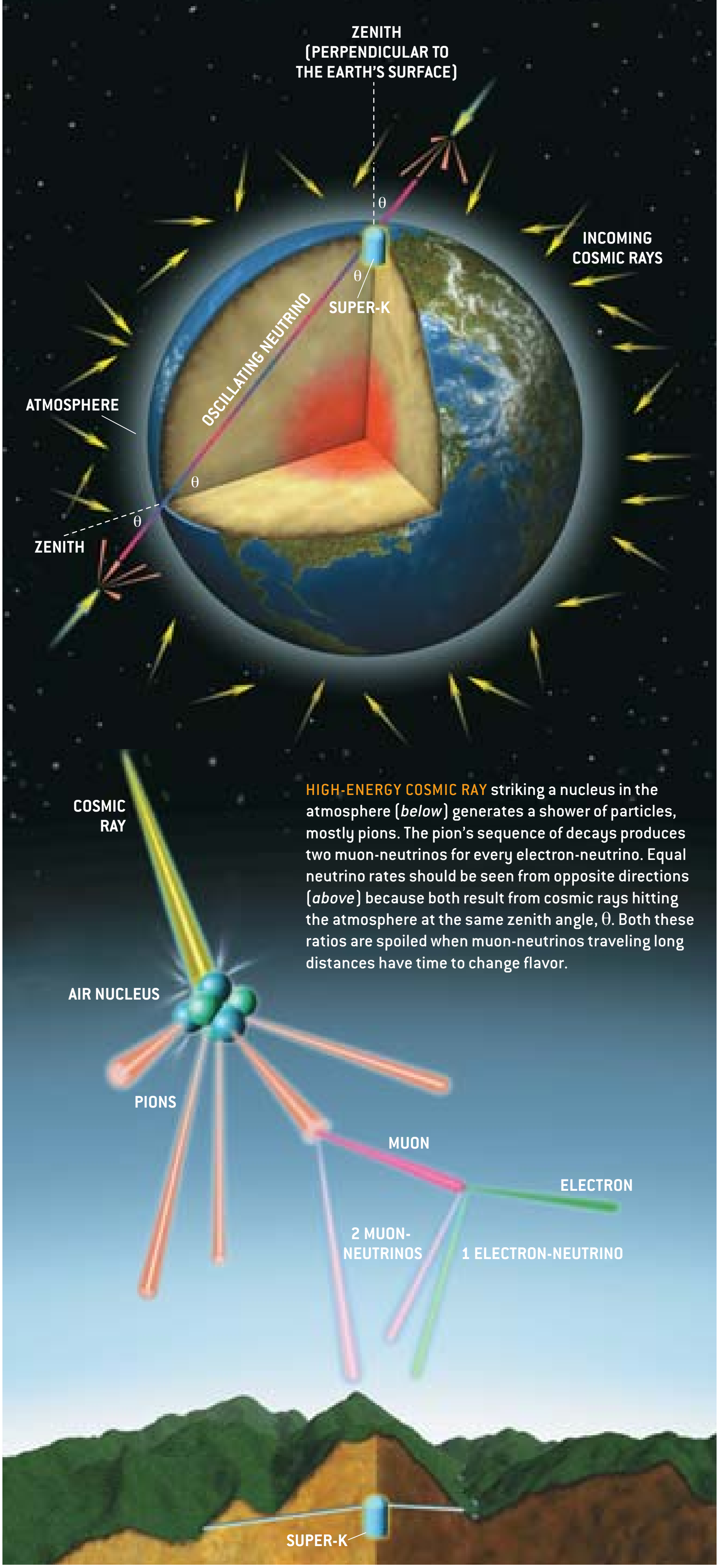
Super-K cannot easily identify the third type of neutrino, the tau-neutrino. Such a neutrino can interact with a nucleus and make a tau particle only if it has enough energy. A muon is about 200 times as heavy as an electron; the tau, about 3,500 times. The muon mass is well within the range of atmospheric neutrinos, but only a tiny fraction are at tau energies, so most tau-neutrinos in the mix will pass through Super-K undetected.

One of the most basic questions experimenters ask is "How many?" We have built a beautiful detector to study neutrinos, and the first task is simply to count how many we see. The related question is "How many did we expect?" To answer that, we must analyze how the neutrinos are produced.

Super-K monitors atmospheric neutrinos, which are born in the spray of particles when a cosmic ray strikes the top of our atmosphere. The incoming projectiles (called primary cosmic rays) are mostly protons, with a sprinkling of heavier nuclei such as helium or iron. Each collision generates a shower of secondary particles, mostly pions and muons, which decay during their short flight through the air, creating neutrinos [*see illustration at right*]. We know roughly how many cosmic rays hit the atmosphere each second and roughly how many pions and muons are made in each collision, so we can predict how many neutrinos to expect.

Tricks with Ratios

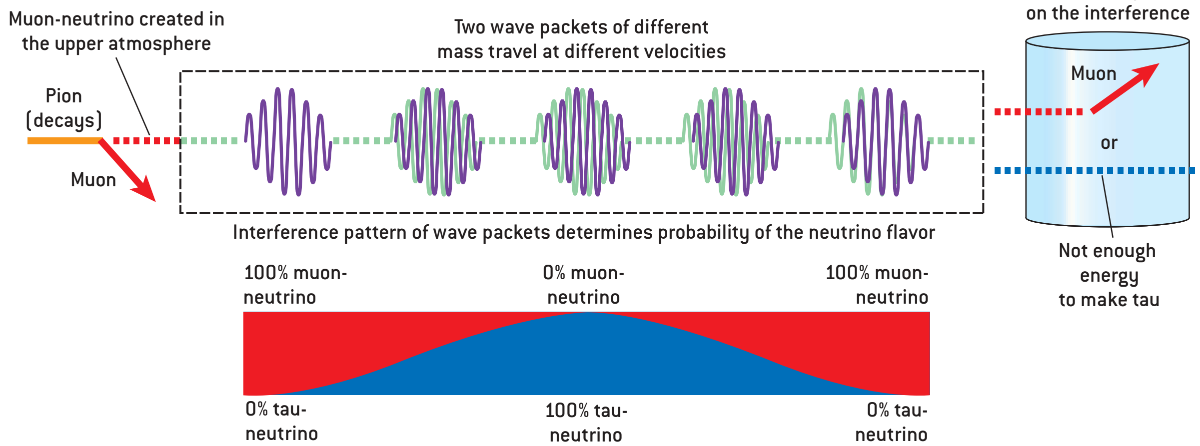
UNFORTUNATELY, this estimate is only accurate to 25 percent, so we take advantage of a common trick: often the ratio of two quantities can be better determined than either quantity alone. For



HIGH-ENERGY COSMIC RAY striking a nucleus in the atmosphere (*below*) generates a shower of particles, mostly pions. The pion's sequence of decays produces two muon-neutrinos for every electron-neutrino. Equal neutrino rates should be seen from opposite directions (*above*) because both result from cosmic rays hitting the atmosphere at the same zenith angle, θ. Both these ratios are spoiled when muon-neutrinos traveling long distances have time to change flavor.

DAVID FIERSTEIN

HOW QUANTUM WAVES MAKE A NEUTRINO OSCILLATE



WHEN A PION DECAYS (*top left*), it produces a neutrino. Described quantum-mechanically, the neutrino is apparently a superposition of two wave packets of different mass (*purple* and *green*). The wave packets propagate at different speeds, with the lighter wave packet getting ahead of the heavier one. As this proceeds, the waves interfere, and the interference pattern controls what flavor neutrino—muon (*red*) or tau (*blue*)—is most likely to be detected at any point along the flight path (*bottom*). Like all quantum effects, this is a game of chance, with the chances heavily favoring a muon-neutrino close to where it was produced. But the probabilities oscillate back and forth, favoring the tau-neutrino at just the right distance and returning to favor the muon-neutrino farther on. When the neutrino finally interacts in the detector (*top right*), the quantum dice are rolled. If the outcome is muon-neutrino, a muon is produced. If chance favors the tau-neutrino, and the neutrino does not have enough energy to create a tau particle, Super-K detects nothing. —*E.K., T.K. and Y.T.*

LAURIE GRACE

Super-K, the key is the sequential decay of a pion to a muon and a muon-neutrino, followed by the muon's decay to an electron, an electron-neutrino and another muon-neutrino. No matter how many cosmic rays are falling on the earth's atmosphere, or how many pions they produce, there should be about two muon-neutrinos for every electron-neutrino. The calculation is more complicated than that, but the final predicted ratio is accurate to 5 percent, providing a much better benchmark.

After counting neutrinos for almost two years, the Super-K team has found that the ratio of muon-neutrinos to electron-neutrinos is about 1.3 to 1 instead of the expected 2 to 1. Even if we stretch our assumptions about the flux of neutrinos, how they interact with the nuclei and how our detector responds, we cannot explain such a low ratio—unless neutrinos are changing from one type into another.

We can play the ratio trick again to test this surprising conclusion. The clue to our second ratio is to ask how many neutrinos should arrive from each possible direction. Primary cosmic rays fall on the earth's atmosphere almost equally from all directions, with only two effects spoiling the uniformity. First, the earth's magnetic field deflects some cosmic rays, especially the low-energy ones, skewing the pattern of arrival directions. Second, cosmic rays that skim the earth at a tangent make neutrino showers that do not descend deep into the atmosphere, and these can develop differently from those that plunge straight in from above.

But geometry saves us: if we "look" up into the sky at some angle from the vertical and then down into the ground at the same angle, we should "see" the same number of neutrinos coming from each direction. Both sets of neutrinos are produced by cosmic rays hitting the atmosphere at the same angle; it is just that in one case the collisions happen overhead and in the other they are partway around the world. To use this fact, we select neutrino events of sufficiently high energy (so their parent cosmic ray was not deflected by the earth's magnetic field) and then divide the number of neu-

1930

Wolfgang Pauli rescues conservation of energy by hypothesizing an unseen particle that takes away energy missing from some radioactive decays.

1933

Enrico Fermi formulates the theory of beta-decay incorporating Pauli's particle, now called the neutrino ("little neutral one" in Italian).

1956

Frederick Reines (*center*) and Clyde Cowen first detect the neutrino using the Savannah River nuclear reactor.

1962

At Brookhaven, the first accelerator beam of neutrinos proves the distinction between electron-neutrinos and muon-neutrinos.

1969

Raymond Davis, Jr., first measures neutrinos from the sun, using 600 tons of cleaning fluid in a mine in Homestake, S.D.

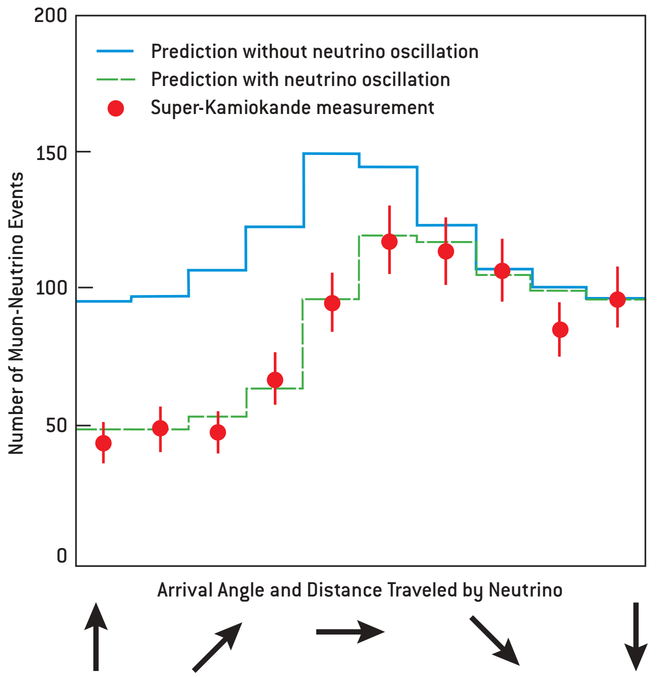
trinos going up by the number going down. This ratio should be exactly 1 if none of the neutrinos are changing flavor.

We saw equal numbers of high-energy electron-neutrinos going up and going down, as expected, but only half as many upward muon-neutrinos as downward ones. This finding is the second indication that neutrinos are changing identity. Moreover, it provides a clue to the metamorphosis. The upward muon-neutrinos cannot be turning into electron-neutrinos, because there is no excess of upward electron-neutrinos. That leaves the tau-neutrino. The muon-neutrinos that become tau-neutrinos pass through Super-K without interaction, without detection.

Fickle Flavor

THE ABOVE TWO RATIOS are good evidence that muon-neutrinos are transforming into tau-neutrinos, but why should neutrinos switch flavor at all? Quantum physics describes a particle moving through space by a wave: in addition to properties such as mass and charge, the particle has a wavelength, can diffract, and so on. Furthermore, a particle can be the superposition of two waves. Now suppose that the two waves correspond to slightly different masses. Then, as the waves travel along, the lighter wave gets ahead of the heavier one, and the waves interfere in a way that fluctuates along the particle's trajectory [*see box on opposite page*]. This interference has a musical analogue: the beats that occur when two notes are almost but not exactly the same.

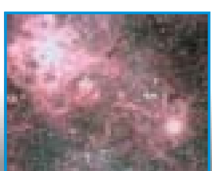
In music this effect makes the volume oscillate; in quantum physics what oscillates is the probability of detecting one type of neutrino or another. At the outset the neutrino appears as a muon-neutrino with a probability of 100 percent. After traveling a certain distance, it looks like a tau-neutrino with 100 percent probability. At other positions, it could be either a muon-neutrino or a tau-neutrino.



NUMBER OF HIGH-ENERGY MUON-NEUTRINOS seen arriving on different trajectories at Super-K clearly matches a prediction incorporating neutrino oscillations (*green*) and does not match the no-oscillation prediction (*blue*). Upward-going neutrinos (*plotted toward left of graph*) have traveled far enough for half of them to change flavor and escape detection.

This oscillation sounds like bizarre behavior for a particle, but another familiar particle performs similar contortions: the photon, the particle of light. Light can occur in a variety of polarizations, including vertical, horizontal, left circular and right circular. These do not have different masses (all photons are massless), but in certain optically active materials, light with left circular polarization moves faster than right circular light. A photon with vertical polarization is actually a superposition of these two alternatives, and when it is traversing an optically active material its polarization will rotate (that is, oscillate) from vertical to horizontal and so on, as its two circular components go in and out of sync.

For neutrino oscillations of the type we see at Super-K, no "optically active" material is needed; a sufficient mass difference between the two neutrino components will cause flavor oscillations whether the neutrino is passing through air, solid rock or pure vacuum. When a neutrino arrives at Super-K, the amount it has oscillated depends on its energy and the distance it has traveled. For downward muon-neutrinos, which have traveled at most a few dozen kilometers, only a small fraction of an oscillation cycle has



1987

Neutrino astronomy: the IMB and Kamiokande proton-decay experiments detect 19 neutrinos from Supernova 1987A in the Large Magellanic Cloud.

1989

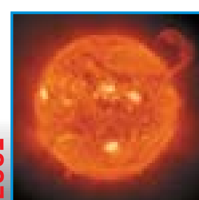
The Z^0 decay rate is precisely measured at SLAC and CERN, showing there are only three active neutrino generations.

1998

Super-K assembles evidence of neutrino oscillation using atmospheric neutrinos.

2000

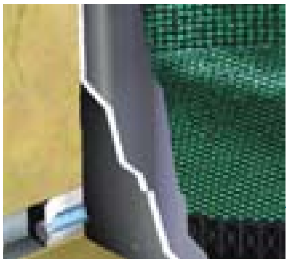
Tau-neutrino events are detected at Fermilab, completing the distinction of neutrinos by flavor begun in 1962.



2001

The solar neutrino puzzle is solved, with key evidence from the Sudbury Neutrino Observatory experiment in Ontario.

LAURIE GRACE (*graph*); TIMELINE BY HEIDI NOLAND; AIP EMILIO SEGRÈ VISUAL ARCHIVES, SEGRÈ COLLECTION (*Pauli*); AIP EMILIO SEGRÈ VISUAL ARCHIVES, PHYSICS TODAY COLLECTION (*Reines*); COURTESY OF THE INSTITUTE FOR ADVANCED STUDY, PRINCETON, N.J. (*Davis*); DAVID MALIN *Anglo-Australian Observatory* (*SN1987A*); ICRR (*Super-Kamiokande*); COURTESY OF SOHO/EXTREME ULTRAVIOLET IMAGING TELESCOPE CONSORTIUM (*sun*)



Neutrinos could account for a mass nearly equal to that of all the stars combined.

taken place, so the neutrinos' flavor is slightly shifted, and we are nearly certain to detect their original muon-neutrino flavor. The upward muon-neutrinos, produced thousands of kilometers away, have gone through so many oscillations that on average only half of them can be detected as muon-neutrinos. The other half pass through Super-K as undetectable tau-neutrinos.

This description is just a rough picture, but the arguments based on the ratio of flavors and the up/down event rate are so compelling that neutrino oscillation is now widely accepted as the most likely explanation for our data. We have also done more detailed studies of how the number of muon-neutrinos varies according to the neutrino energy and the arrival angle. We compare the measured number against what is expected for a wide array of possible oscillation scenarios (including no oscillations). The data look quite unlike the no-oscillation expectation but match well with neutrino oscillation for certain values of the mass difference and other physical parameters [*see illustration on preceding page*].

With about 5,000 events from our first two years of experimentation, we were able to eliminate any speculation that the anomalous numbers of atmospheric neutrinos could be just a statistical fluke. But it is still important to confirm the effect by looking for the same muon-neutrino oscillation with other experiments or techniques. Detectors in Minnesota and Italy have provided some verification, but because they have measured fewer events they do not offer the same certainty.

Corroborating Evidence

FURTHER SUPPORT comes from studies of a different atmospheric neutrino interaction: collisions with nuclei in the rock around our detector. Electron-neutrinos again produce electrons and subsequent showers of particles, but these are absorbed in the rock and never reach Super-K's cavern. High-energy muon-neutrinos make high-energy muons, which can travel through many meters of rock and enter our detector. We count such muons from upward-traveling neutrinos—downward muons are masked by the background of cosmic-ray muons that penetrate Mount Ikenoyama.

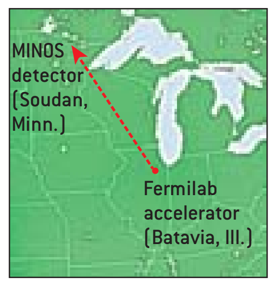
We can count upward-traveling muons arriving on trajectories that range from directly up to nearly horizontal. These paths correspond to neutrino travel distances (from production in the atmosphere to the creation of a muon near Super-K) as short as 500 kilometers (the distance to the edge of the atmosphere when looking horizontally) and as long as 13,000 kilometers (the diameter of the earth, looking straight down). We find that the numbers of muon-neutrinos of lower energy that travel a long distance are more depleted than higher-energy muon-neutrinos that travel a short distance. This behavior is exactly what we expect from oscillations, and careful analysis produces neutrino parameters similar to those from our first study.

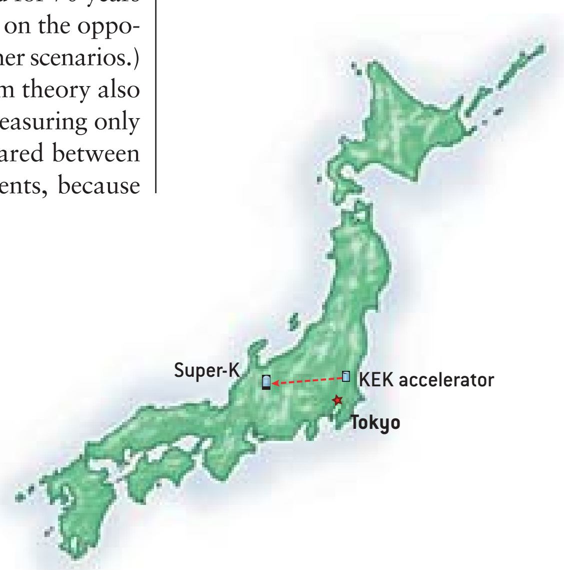
If we consider just the three known neutrinos, our data tell us that muon-neutrinos are changing into tau-neutrinos. Quantum theory says that the underlying cause of the oscillation is almost certainly that these neutrinos have mass—although it has been assumed for 70 years that they do not. (The box on the opposite page mentions some other scenarios.)

Unfortunately, quantum theory also limits our experiment to measuring only the difference in mass-squared between the two neutrino components, because that is what determines the oscillation wavelength. It is not sensitive to the mass of either one alone. Super-K's data give a mass-squared difference somewhere between 0.001 and 0.01 electron volt (eV) squared. Given the pattern of masses of other known particles, it is likely that one neutrino is much lighter than the other, which would mean that the mass of the heavier neutrino is in the range of 0.03 to 0.1 eV. What are the implications?

First, giving neutrinos a mass does not wreck the Standard Model. The mismatch between the mass states that make up each neutrino requires the introduction of a set of so-called mixing parameters. A small amount of such mixing has long been observed among quarks, but our data imply that neutrinos need a much greater degree of mixing—an important piece of information that any successful new theory must accommodate.

Second, 0.05 eV is still very close to





LONG-BASELINE neutrino oscillation experiments are planned in Japan and the U.S. Beams of neutrinos from accelerators will be detected hundreds of kilometers away. The experiments should confirm the oscillation phenomenon and precisely measure the constants of nature that control it.

LAURIE GRACE

zero, compared with other particles. (The lightest of those is the electron, with a mass of 511,000 eV.) So the long-held belief that neutrinos have zero mass is understandable. But theoreticians who wish to build a Grand Unified Theory, which would elegantly combine all the forces except gravity at enormously high energies, also take note of this relative lightness of neutrinos. They often employ a mathematical device called the seesaw mechanism, which actually predicts that such a small but nonzero neutrino mass is natural. Here the mass of some extremely heavy particle, perhaps at the Grand Unified mass scale, provides the leverage to separate the very light neutrinos from the quarks and leptons that are a billion to a trillion times heavier.

Another implication is that the neutrino should be considered in the bookkeeping of the mass of the universe. For some time, astronomers have been trying to tabulate how much mass is found in luminous matter, such as stars, and in ordinary matter that is difficult to see, such as brown dwarfs or diffuse gas. The mass can also be measured indirectly from the orbital motion of galaxies and the rate of expansion of the universe. The direct accounting falls short of these indirect measures by a factor of 20. The neutrino mass suggested by our result is too small to resolve this mystery by itself. Nevertheless, neutrinos created during the big bang permeate space and could account for a mass nearly equal to the combined mass of all the stars. They could have affected the formation of large astronomical structures, such as galaxy clusters.

Finally, our data have an immediate implication for two new experiments. Based on the earlier hints from smaller detectors, many physicists have decided to stop relying on the free but uncontrollable neutrinos from cosmic rays and instead are creating them with high-energy accelerators. Even so, the neutrinos must travel a long distance for the oscillation effect to be observed. So the neutrino beams are aimed at a detector hundreds of kilometers away. One detector, MINOS, is being built in a mine in Soudan, Minn., to study neutrinos sent from the Fermilab accelerator near Batavia, Ill., 730 kilometers away on the outskirts of Chicago.

OTHER PUZZLES, OTHER POSSIBILITIES

PARTICLE PHYSICISTS have been busy sorting out other indications of neutrino mass. For more than 30 years, scientists have been capturing electron-neutrinos generated by nuclear fusion processes in the sun. These experiments have always counted fewer neutrinos than the best models predict.

Super-K has also counted solar neutrinos and finds only about 50 percent of the number expected. If solar neutrinos are changing flavor, this deficit is understandable, because at solar energies Super-K responds to the electron flavor and mostly ignores those transformed into the muon or tau flavors. The Sudbury Neutrino Observatory (SNO) in Ontario, however, which uses 1,000 tons of heavy water, has recently achieved a breakthrough in proving this change. The heavy water allows SNO to measure the total number of neutrinos (electron, muon and tau) as well as the number of electron-neutrinos alone, and it shows that the total is much greater. The accounting seems to balance.

It appears that the mass splitting associated with solar neutrinos is much smaller than that for atmospheric neutrinos. This fits a picture in which the three flavors of neutrinos are spread over three distinct neutrino masses. But the picture doesn't allow for the hint of neutrino oscillation, suggesting much larger masses, detected at Los Alamos National Laboratory. Some exotic explanations are waiting in the wings while Fermilab checks this signature.

Physicists are also checking the theory that transforms solar neutrinos. In a cavern in the same zinc mine as Super-K, a detector has been built that uses 1,000 tons of mineral oil doped with a chemical that emits light in response to the neutrino reaction. This detector counts electron-neutrinos from more than two dozen Japanese nuclear power reactors, from 80 to 400 kilometers away. The results are being compared with a precise model of how many neutrinos are expected from each reactor. This experiment should pin down the detailed particle physics revealed by solar neutrinos.

Overall, our picture of neutrinos is just coming into focus. More clarity will rely on more ambitious projects. Later in this decade Super-K will be exposed to a beam of neutrinos from a much more intense accelerator being built near Japan's Pacific coast. The goal is to verify that muon-neutrinos change flavor to tau- and electron-neutrinos in a proportion that fits our newfound expectations. Follow-up measurements may reveal the role of neutrinos in the matter-antimatter imbalance of the universe. Or we may be presented with new puzzles to solve. *—E.K., T.K. and Y.T.*

Of course, a good atmospheric neutrino detector is also a good accelerator neutrino detector, so in Japan we are using Super-K to monitor a beam of neutrinos created at the KEK accelerator laboratory 250 kilometers away. Unlike atmospheric neutrinos, the beam can be turned on and off and has a well-defined energy and direction. Most important, we have placed a detector similar to Super-K near the origin of the beam to characterize the muon-neutrinos before they oscillate. We are essentially using the ratio (again) of the counts near the source to those far away to cancel uncertainty and verify the effect. Since 1999, neutrinos from the first long-distance artificial neutrino beam have passed under the mountains of Japan, with 50,000 tons of Super-K capturing a small handful. Exactly how many are being captured will be the next chapter in this story. SA

MORE TO EXPLORE

The Search for Proton Decay. J. M. LoSecco, Frederick Reines and Daniel Sinclair in *Scientific American,* Vol. 252, No. 6, pages 54–62; June 1985.

The Elusive Neutrino: A Subatomic Detective Story. Nickolas Solomey. Scientific American Library, W. H. Freeman and Company, 1997.

Official Super-Kamiokande Web site: **www-sk.icrr.u-tokyo.ac.jp/doc/sk/**

K2K Long Baseline Neutrino Oscillation Experiment Web site: **neutrino.kek.jp/**

Super-Kamiokande at Boston University Web site: **hep.bu.edu/~superk/**

TABLETOP LASER fires terawatt pulses 10 times a second, striking a thin cloth in the foreground. The photograph is a triple exposure to accommodate the range of intensities.

Focusing light
with the power of
1,000 HOOVER DAMS
onto a point
the size of
A CELL NUCLEUS
accelerates
electrons
to the speed
of light in a
femtosecond

EXTREME LIGHT

By Gérard A. Mourou and Donald Umstadter

The dream of intensifying light is as old as civilization.

Legend has it that Archimedes focused the sun's rays with a giant mirror to set the Roman fleet afire at Syracuse in 212 B.C. Although that story is a myth, it is true that around 200 B.C. another Greek, Diocles, invented the first ideal focusing optic, a parabolic mirror. Two millennia later, mirrors and quantum mechanics were put together to make the most versatile of high-intensity light sources: the laser.

The epitome of high-power lasers is Nova, which operated at Lawrence Livermore National Laboratory from 1985 to 1999. Named for the brilliance of an exploding star, Nova was one of the largest lasers ever built. Ten parallel chains of laser amplifiers occupied a 90-meter enclosure; mirrors made from 180-kilogram blocks of glass directed the beams to targets for nuclear fusion and other experiments. Nova was fired no more than a few times each day to avoid overheating. Clearly, it marshaled a lot of energy to achieve its ultrahigh power.

Yet power is the *rate* at which energy is delivered, so another approach to ultrahigh power is to release a modest amount of energy in an extremely short time. Nova's usual pulses were relatively long by the standards of today's ultrafast lasers—three nanoseconds—and each one required kilojoules of energy. By using pulses of one ten-thousandth their duration, a new type of laser that fits on a tabletop can deliver power similar to Nova's [see "Ultrashort-Pulse Lasers: Big Payoffs in a Flash," by John-Mark Hopkins and Wilson Sibbett; SCIENTIFIC AMERICAN, September 2000]. For example, an ultrahigh-power laser that delivers a mere joule in a pulse lasting 100 femtoseconds (10^{-13} second) achieves 10 trillion watts (10^{13} W, or 10 terawatts), more than the output of all the world's power plants combined.

These compact lasers can fire a hundred million shots a day and can concentrate their power onto a spot the size of a micron, producing the highest light intensities on earth. Associated with these gargantuan power densities are the largest electric fields ever produced, in the range of a trillion volts per centimeter. Such intense laser light interacting with matter re-creates the extreme physical conditions that can be found only in the cores of stars or in the vicinity of a black hole: the highest temperatures, 10^{10} kelvins; the largest magnetic fields, 10^9 gauss; and the largest acceleration of particles, 10^{25} times the earth's gravity.

Costing $1 million instead of several hundred million dollars, these lasers are helping to bring "big science" back to standard university laboratories and to countries with limited research budgets. Dozens of such systems have been built throughout the world in the past few years, for use in research in several subfields of physics, including nuclear physics, astrophysics, high-energy particle physics and general relativity. This new breed of laser has already spawned applications, such as x-ray lasers, ultracompact particle accelerators and precision medical radiography. It also shows great promise for radiation therapy and improvements in nuclear fusion power generation.

Overview/*Extreme Light*

- A method of laser amplification invented in the mid-1980s has enabled a new generation of tabletop lasers that produce very brief pulses of extremely intense light.
- Light of such high intensity interacts with matter in new ways, directly propelling electrons to nearly the speed of light in femtoseconds. The lasers can accelerate particles at 10,000 times the rate of standard accelerators.
- Potential applications include high-resolution medical imaging, inexpensive precision radiation therapy, nuclear fusion, and research in numerous subfields of physics.

The Trick

IN THE FIVE YEARS after the invention of the laser in 1960, tabletop lasers advanced in a series of technological leaps to reach a power of one gigawatt (10^9 W). For the next 20 years, progress was stymied and the maximum power of tabletop laser systems did not grow. The sole way to increase power was

WILLIAM PELLETIER *Photo Services, Inc.*, COURTESY OF THE UNIVERSITY OF MICHIGAN (*page 76*)

Tabletop ultrahigh-intensity lasers are bringing "big science" back to standard university laboratories.

to build ever larger lasers. Trying to operate beyond the limiting intensity would create unwanted nonlinear effects in components of the laser, impairing the beam quality and even damaging the components. Only in 1985 was this optical damage problem circumvented, with the introduction of chirped pulse amplification (CPA), a technique developed by the research group led by one of us (Mourou). Tabletop laser powers then leaped ahead by factors of 10^3 to 10^5.

"Chirping" a signal or a wave means stretching it in time. In chirped pulse amplification, the first step is to produce a short pulse with an oscillator and stretch it, usually 10^3 to 10^5 times as long [*see illustration below*]. This operation decreases the intensity of the pulse by the same amount. Standard laser amplification techniques can now be applied to this pulse. Finally, a sturdy device, such as a pair of diffraction gratings in a vacuum, recompresses the pulse to its original duration—increasing its power 10^3 to 10^5 times beyond the amplifier's limit. A typical example would begin with a seed pulse lasting 100 femtoseconds and having 0.2 nanojoule of energy. We stretch it by a factor of 10^4 to a nanosecond (reducing its power from about two kilowatts to 0.2 watt) and amplify it by 10 orders of magnitude to two joules and two gigawatts. Recompressing the pulse to 100 femtoseconds increases the power to 20 terawatts. Without this method, sending the original two-kilowatt pulse through a tabletop amplifier would have destroyed the amplifier—unless we increased the amplifier's cross-sectional area 10^4 times and dispersed the beam across it. The CPA technique makes it possible to use conventional laser amplifiers and to stay below the onset of nonlinear effects.

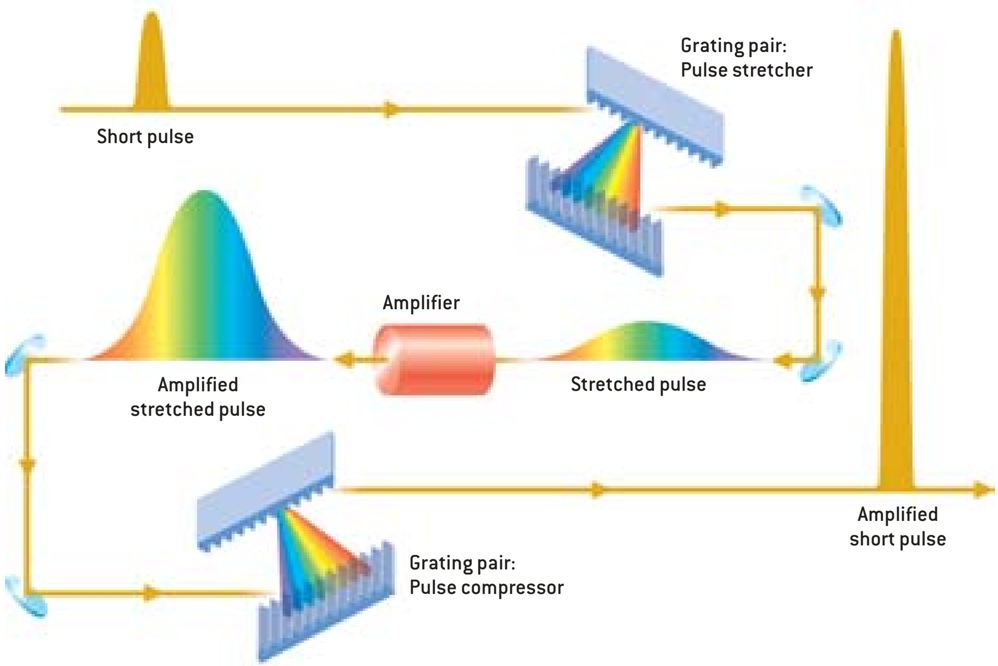
Perfecting CPA was not as straightforward as it sounds. Typical devices used to stretch or compress pulses generally do not do so in an exactly linear fashion, and the result will be spoiled if the characteristics of the chirper and the compressor are not closely matched.

A further increase in light intensities has occurred in the past few years with the development of corrective optics that allow laser beams to be focused onto much smaller spots. That advance and further improvements in pulse compression have resulted in pulses that have the maximum possible intensity for a given energy of light.

These increases in power and intensity in the 1990s opened up a new regime of interactions between light and matter,

CHIRPED PULSE AMPLIFICATION

THE KEY to tabletop ultrahigh-intensity lasers is a technique called chirped pulse amplification. An initial short laser pulse is stretched out ("chirped") by a factor of about 10^4, for instance, by a pair of diffraction gratings. The stretched pulse's intensity is low, allowing it to be amplified by a small laser amplifier. A second pair of gratings recompresses the pulse, boosting it to 10^4 times the peak intensity that the amplifier could have withstood.



JUAN VELASCO

known as relativistic optics, in which the light accelerates electrons close to the speed of light. Prior to CPA, this regime could be reached only by very large and expensive laser systems.

Relativistic Optics

OPTICS IS THE STUDY of how electrons respond to light. That definition may not sound like what many people think of as optics—light reflecting off mirrors or being refracted by the water of a swimming pool. Yet all the optical properties of a material are a consequence of how light interacts with electrons in the material.

Light is a wave composed of coupled electric and magnetic fields oscillating in synchrony at very high frequencies. The electric and magnetic fields oscillate perpendicular to each other and perpendicular to the direction the light is traveling [*see illustration below*]. When an electron encounters a light wave of ordinary power, the electric field of the wave exerts a force on the electron and makes it oscillate. The electron oscillates parallel to the electric field and at the same frequency, but it does not necessarily oscillate in phase with the light wave. Depending on how the electron is bound to the atoms of the material, its oscillations may lag behind or lead those of the light wave. The amplitude and phase of these electron oscillations in turn determine how the light wave propagates through the material and thereby confer on the material its optical properties.

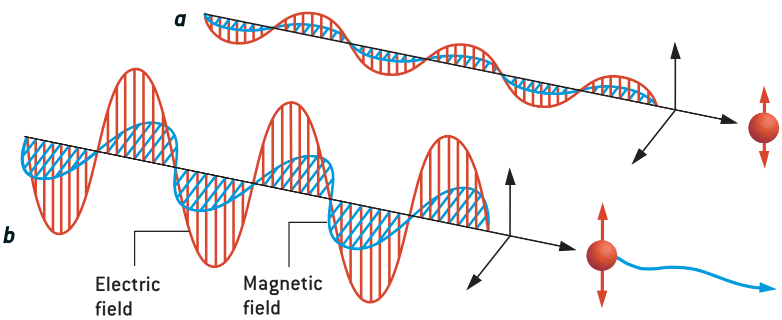
In classical optics the amplitudes are small enough that the electrons' oscillation velocities are always very small compared with the speed of light. With the advent of laser intensities above 10^{18} watts per square centimeter, however, the electrons' oscillation velocities approach the speed of light, and relativistic effects fundamentally change the electrons' response to the light.

First, a high velocity increases the mass of an electron, which affects the amplitude and phase of its oscillations. More important, the magnetic field of the light wave starts to play a role. A

LIGHT INTERACTING WITH MATTER

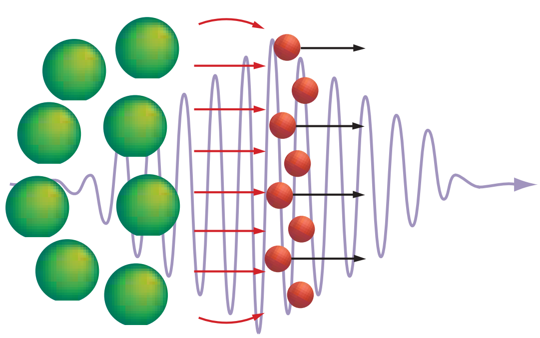
RELATIVISTIC OPTICS

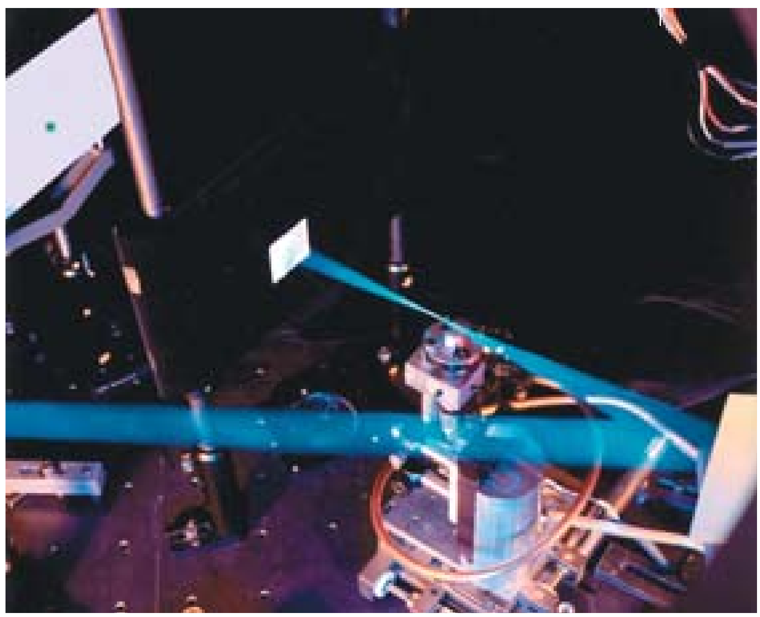
FOR LIGHT of ordinary intensity (*a*), the light's electric field (*red waves*) makes electrons oscillate at relatively low speeds. At extremely high intensities (*b*), the electrons oscillate at nearly the speed of light, and the light's magnetic field (*blue waves*) makes them fly forward with very high momentum.



WAKE-FIELD ACCELERATION

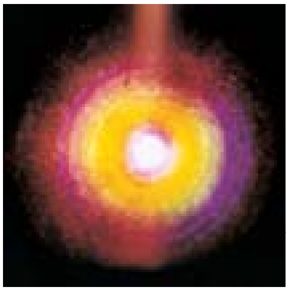
HIGH-INTENSITY LIGHT striking a plasma (*below*) pushes the electrons to very high speeds, leaving the heavier positive ions (*green*) behind and producing a powerful electric field (*red lines*) between these separated charges. This separation of charges and the associated electric field trails along in the wake of the light and can accelerate other charged particles to very high energies.





ULTRAHIGH-INTENSITY LASER PULSE (*added in blue*) focused on a helium gas jet by a parabolic mirror accelerates electrons from the gas to 60 million electron volts in one millimeter. A fluorescent screen (*upper left*) detects the high-energy electron beam.

NINA FINKEL (*illustrations*); WILLIAM PELLETIER *Photo Services, Inc.*, AND DONALD UMSTADTER, COURTESY OF THE UNIVERSITY OF MICHIGAN (*photograph*)



Small ultrahigh-power lasers might work like spark plugs, igniting thermonuclear fusion at power plants.

magnetic field exerts a force on an electric charge only when the charge is moving. In the regime of classical optics the magnetic force is negligible. But for electron oscillation velocities near the speed of light, it curls the paths of the electrons and gives them tremendous momentum in the direction of the light beam. This effect plays a central role in relativistic optics.

The interaction of light with atomic nuclei can usually be ignored because protons are almost 2,000 times as massive as electrons and therefore oscillate much less. But at high enough intensities, the light starts moving protons around at relativistic velocities as well. That regime may be called nuclear optics because of the great variety of nuclear processes, such as fusion, that can occur.

"0 to 60" (MeV) in a Millimeter

THE MOST OBVIOUS APPLICATION of the relativistic force of an ultraintense laser beam is to accelerate particles. Charged-particle accelerators have numerous uses, ranging from television tubes to cancer therapy to the study of the fundamental forces of the universe. What they all have in common is that the particles, such as electrons or protons, are accelerated by electric or magnetic fields. Although light waves in the regime of classical optics can have electric fields as strong as those near bolts of lightning, these fields are not effective for accelerating particles on their own, because they oscillate transversely. In contrast, when an ultraintense pulse of light strikes a plasma (a gas of electrons and positive ions), it propels the electrons forward at close to the speed of light, as we described above.

That is not the end of the story. The plasma's positive ions, being thousands of times heavier than the electrons, are left behind. This separation of positive and negative charges produces a large electric field, which can be used to accelerate other particles. The region of high electric field travels through the plasma as a wave, trailing in the wake of the light pulse. Charged particles are accelerated to high energy in laser wake fields just as dolphins gain energy by swimming in phase with the water wave in the wake of a ship. Such a laser wake-field accelerator was first proposed in 1979 by Toshiki Tajima and John M. Dawson, both then at the University of California at Los Angeles.

The process of converting the oscillating electric field of the light pulse into a wake field that points always in one direction is called rectification, by analogy with rectifiers in electronics that convert alternating current (AC) to direct current (DC). Conventional accelerators, such as the three-kilometer-long one at the Stanford Linear Accelerator Center (SLAC), use metal cavities to rectify radio-frequency waves to repeatedly "kick" charged particles along the beam line. (Radio waves are electromagnetic waves just like light but having much lower frequencies and longer wavelengths.) The Stanford accelerator has to be three kilometers long to achieve its target particle energies because the accelerating field of each cavity is limited. The field could be increased by using radio waves of shorter wavelength and greater intensity, but both of these properties are limited by the cavity: the cavity size limits the wavelength, and high intensities cause electronic breakdown (sparking) of the metal cavity walls. Laser wake-field accelerators avoid these limits by eliminating the cavity. With the highest-intensity pulses, particles might be accelerated directly, the same way that relativistic electrons are generated by the beam, allowing the plasma to be dispensed with.

In the past few years, laser-driven electron and proton accelerators have produced beams with energies greater than 50 million electron volts (MeV), comparable to a single stage (a few meters long) of a conventional accelerator. The laser system achieves the same energy in a millimeter.

Prompt acceleration with high gradients has advantages. For example, one of us (Umstadter) has demonstrated electron beams of a few million electron volts whose "brightness" (in essence, the concentration of particles in the beam) exceeds that of beams made by conventional accelerators, mainly because the charges bunched in one pulse of the beam have less time to blow it apart by its own electrostatic forces. In addition, researchers have shown that low-cost laser accelerators are suitable for many of the same applications as conventional accelerators, such as producing short-lived radioisotopes used in medical diagnostics and generating neutron and positron beams for studies of materials.

The laser systems create beams that have a relatively broad spread of particle energies, however, which is undesirable for some applications. Also, conventional systems routinely chain together numerous accelerator stages, as in SLAC's three-kilometer collider and the seven-kilometer-circumference main ring of the Tevatron at Fermilab. Current research on laser accelera-

THE AUTHORS

GÉRARD A. MOUROU and *DONALD UMSTADTER* were among the founders of the National Science Foundation–sponsored Center for Ultrafast Optical Science at the University of Michigan at Ann Arbor. Mourou is director of the center and professor of electrical engineering; Umstadter is associate professor of both nuclear and electrical engineering. When they are not accelerating particles with intense lasers, they can be found accelerating down ski slopes to "ultrahigh" speeds.

RADIOGRAPH OF A RAT shows the very high resolution that can be achieved by using x-rays generated from a tiny spot of plasma at the focus of a tabletop ultrahigh-intensity laser.

tor systems is concentrated on reducing the beam's energy spread and achieving multistaging to increase the beam's energy. Researchers are also exploring the use of innovative waveguides to increase the distance over which the wake field keeps accelerating particles.

We don't expect laser accelerators to replace conventional accelerators at high-energy particle physics facilities such as the Tevatron. Rather they complement and augment present-day systems and have characteristics that make them useful for specific applications and new types of experiments. One such niche could be the acceleration of unstable particles.

The Tevatron represents the high-energy frontier today: colliding protons with energies at the TeV level. Its successor, CERN's Large Hadron Collider, will also use protons. Such collisions are very complicated and messy because protons are agglomerations of strongly interacting particles called quarks and gluons. Electrons and positrons have a more elementary structure than protons and consequently produce much "cleaner" collisions, which allow more detailed, higher-precision studies. But accelerating them runs into a problem: the lightweight electrons and positrons lose too much of their energy to so-called synchrotron radiation as they travel around the curves of a circular accelerator.

One solution will be to accelerate muons, which are 200 times as heavy as electrons and thereby suffer synchrotron losses a billionth the amount. Unfortunately, muons are unstable and decay in just over two microseconds on average. High-intensity lasers could be used to accelerate muons very close to the speed of light in a fraction of that fleeting lifetime. At that point, relativistic time dilation helps out, extending the muons' lifetime in proportion to the energy achieved and providing more time for a conventional accelerator to take over. The benefit of prompt laser acceleration would be even greater for particles such as pions, which decay in a mere 26 nanoseconds on average.

Another new type of particle physics experiment enabled by ultrahigh-power lasers is the gamma-gamma collider. Gamma rays are extremely high energy photons or, equivalently, extremely high frequency light—beyond x-rays on the spectrum. A high-power laser beam colliding with a high-energy electron beam produces a narrow beam of gamma rays. In essence, the laser's photons rebound off the electrons in a process called Compton scattering. The energy of the gamma rays depends mostly on the energy of the electron beam: a 250-giga-electron-volt (GeV) electron beam knocks the photons from around 1 eV (visible light) to about 200 GeV.

When two such gamma-ray beams collide, the interactions are even cleaner than electron-positron or muon-antimuon collisions. The process is the reverse of matter-antimatter annihilation, in which particles merge and become a flash of radiation: instead pairs of particles and antiparticles burst into life out of a clash of photons. Only with ultrahigh-intensity lasers, however, are there enough photons in each pulse to produce a significant number of gamma-gamma collisions. In 1997 researchers from the University of Rochester, Princeton University, the University of Tennessee and SLAC demonstrated a variant of this system and produced electron-positron pairs by colliding gamma rays and laser photons. Today every linear particle collider has plans to conduct gamma-gamma experiments, which complement the research possible with the usual electron-positron collisions.

Finding and Curing Cancer

BY GENERATING HIGHLY PENETRATING radiation such as x-rays or particle beams, laser-driven charged-particle accelerators may also be used for cancer diagnosis and therapy. X-rays, of course, have been a diagnostic workhorse for a century. Conventional x-ray tubes accelerate electrons in an electric field that is set up between a cathode and an anode. When they strike the anode, the electrons are violently decelerated, which produces copious x-ray emissions. The resolution is limited by the size of the x-ray source, in this case the anode, which is generally about 100 microns across. The smallest tumor detectable by such a system is about a millimeter in diameter.

An ultrahigh-intensity laser, however, can produce x-rays simply by being focused onto an appropriate metal target. The beam accelerates electrons near the surface of the metal to high energies. These electrons are decelerated by their passage through the volume of the metal, once again emitting copious x-rays. Focusing the laser to a spot a few microns across makes an extremely small x-ray source, allowing detection of very small clumps of cancerous cells so that treatment can begin at

COURTESY OF LUND LASER CENTER, SWEDEN

The lasers can produce x-rays that could detect very small clumps of cancerous cells, so treatment could begin earlier.

a much earlier stage in a tumor's development. In principle, resolution of a micron—a little larger than the wavelength of the driving laser—is possible. Research groups at Stanford University, Lund University in Sweden and the National Institute of Scientific Research in Quebec have already demonstrated these x-ray systems.

Precision delivery of energy is also of great importance for radiation therapy. The goal is to maximize the dose delivered to the tumor while minimizing harm to surrounding healthy tissues. When treating tumors in such sensitive areas as the brain or the spinal cord, the ability to deposit controlled amounts of energy in small, distinct areas is critical. Particles such as protons and carbon ions are particularly well suited to this task. Unlike electrons and photons, these heavier particles suffer only minimal lateral scattering, so a beam remains narrow. The particles lose energy at a steady, very low rate along their track and then dump most of their energy at the end of it. For a specific initial energy, this dissipation occurs at a well-defined range through the tissue. Consequently, such heavier ions have much better accuracy than electrons and photons for delivering a dose to deep-seated tumors.

Clinical trials of particle-based therapy using proton and carbon beams are under way in several countries. One of the chief obstacles to wide-scale use of the technology, however, is the high cost of conventional particle accelerators. For example, the Heavy Ion Medical Accelerator in Chiba, Japan, cost almost $300 million to build. It can treat only about 200 patients a year, a small fraction of the cases that could benefit from this form of cancer therapy. At the present time, laser-driven accelerators are able to achieve ion energies that are about a factor of five too low and have too great a spread of energies. But if those two problems can be overcome, ion radiotherapy will be possible at much lower cost and thus available to many more cancer patients.

Power for Fusion?

A PULSE FROM AN ULTRAHIGH-INTENSITY laser delivers as much power as all the world's power generators. In the future, that equation may be turned around, with such lasers becoming an essential component of nuclear fusion power plants *supplying* some of the world's power needs. Controlled nuclear fusion for power generation has been pursued for decades and has remained frustratingly out of reach. A method that has gained favor in recent years is inertial-confinement fusion, in which capsules of fuel—such as mixtures of deuterium and tritium (heavy isotopes of hydrogen)—are hit from all sides simultaneously by dozens or hundreds of intense laser pulses. The lasers compress and heat the capsules to the extreme densities and temperatures at which the deuterium and tritium nuclei fuse together to form helium and release large amounts of energy. The huge Nova laser at Livermore was one of the leading experimental devices used in research toward that goal.

Tabletop ultrahigh-intensity lasers cannot supply enough total energy to drive thermonuclear fusion, but in conjunction with their Nova-size cousins, they may bring the process much closer to economic and technical feasibility. Achieving the conditions needed to ignite fusion by compressing the capsules requires an extraordinarily symmetrical implosion process. The tiniest imperfections lead to worthless fizzles. In the new technique, proposed by researchers at Livermore, the large lasers will still do the hard work of compressing the fuel to high density but do not have to achieve the full ignition temperature as well. Instead, near the point of maximum density, an ultrashort pulse of ions accelerated by a compact, ultrahigh-power CPA laser strikes the imploding capsule, playing a role like a spark plug in an automobile engine: the pulse creates an intense hot spot, igniting a wave of fusion that burns across the rest of the pellet. This method should reduce the immensely difficult technical requirements of igniting fusion by implosion alone, and it should significantly increase the ratio of energy produced to energy used.

Some of the fundamentals of the fast-ignition technique were recently demonstrated by researchers from Rutherford Appleton Laboratory in Oxfordshire, England, and Osaka University in Japan. But as is always the case in fusion research, much more must be accomplished to prove the method's practicality for economical power generation. Whether or not that particular application becomes the stuff of legend, ultrahigh-intensity light has a future that is spectacular and diverse beyond the wildest dreams of Archimedes and Diocles. SA

MORE TO EXPLORE

Terawatt Lasers Produce Faster Electron Acceleration. D. Umstadter in *Laser Focus World,* pages 101–104; February 1996.

Ultrahigh-Intensity Lasers: Physics of the Extreme on a Tabletop. G. A. Mourou, C.P.J. Barty and M. D. Perry in *Physics Today,* Vol. 51, No. 1, pages 22–28; January 1998.

Review of Physics and Applications of Relativistic Plasmas Driven by Ultra-Intense Lasers. D. Umstadter in *Physics of Plasmas,* Vol. 8, No. 5, pages 1774–1785; May 2001.

High Field Science Research at the University of Michigan Center for Ultrafast Optical Science: **www.eecs.umich.edu/USL-HFS/**

A WORMHOLE WOULD APPEAR as a spherical opening to a distant part of the cosmos. In this doctored photograph of Times Square, the wormhole allows New Yorkers to walk to the Sahara with a single step. Although such a wormhole does not break any known laws of physics, it would require unrealistic amounts of negative energy.

NEGATIVE WORMHOLES

The construction of wormholes and warp drive But the same laws of physics that allow this

Updated from the January 2000 issue

ENERGY, *and* WARP DRIVE

By Lawrence H. Ford and Thomas A. Roman

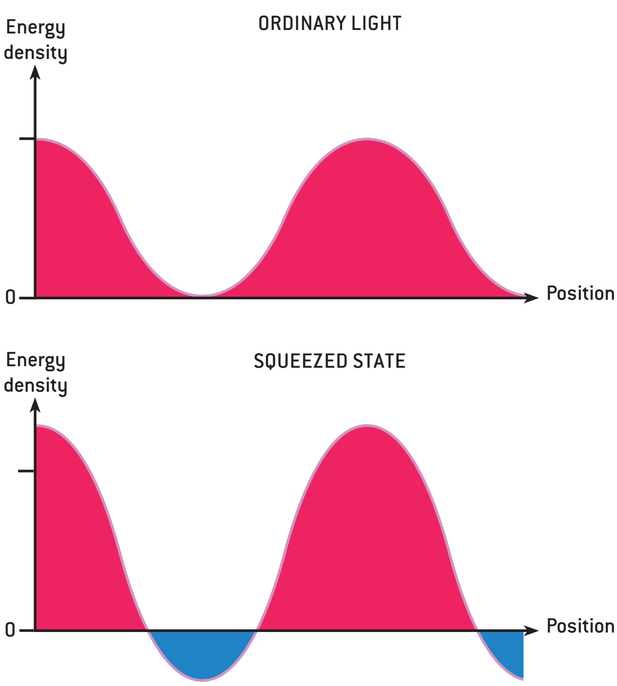
would require a very unusual form of energy.
"negative energy" also appear to limit its behavior

Can a region of space contain less than nothing?

Common sense would say no; the most one could do is remove all matter and radiation and be left with a vacuum. But quantum physics has a proven ability to confound intuition, and this case is no exception. A region of space, it turns out, can contain less than nothing. Its energy per unit volume—the energy density—can be less than zero.

Needless to say, the implications are bizarre. According to general relativity, Einstein's theory of gravity, the presence of matter and energy warps the geometric fabric of space and time. What we perceive as gravity is the spacetime distortion produced by normal, positive energy or mass. But when negative energy or mass—so-called exotic matter—bends spacetime, all sorts of amazing phenomena might become possible: traversable wormholes, which could act as tunnels to otherwise distant parts of the universe; warp drive, which would allow for faster-than-light travel; and time machines, which might permit journeys into the past. Negative energy could even be used to make perpetual-motion machines or to destroy black holes.

For physicists, these ramifications set off alarm bells. The potential paradoxes of backward time travel—such as killing your grandfather before your father is conceived—have long been explored in science fiction, and the other consequences of exotic matter are also problematic. They raise a question of fundamental importance: Do the laws of physics that permit negative energy place any limits on its behavior? We and others have discovered that nature imposes stringent constraints on the magnitude and duration of negative energy, which (unfortunately, some would say) appear to render the construction of wormholes and warp drives very unlikely.



WAVES OF LIGHT ordinarily have a positive or zero energy density at different points in space (*top*). But in a so-called squeezed state, the energy density at a particular instant in time can become negative at some locations (*bottom*). To compensate, the peak positive density must increase.

Double Negative

BEFORE PROCEEDING, we should draw attention to what negative energy is not. It should not be confused with antimatter, which has positive energy. When an electron and its antiparticle, a positron, collide, they annihilate. The end products are gamma rays, which carry positive energy. If antiparticles were composed of negative energy, such an interaction would result in a final energy of zero. One should also not confuse negative energy with the energy associated with the cosmological constant, postulated in inflationary models of the universe. In the latter case, there is negative pressure but positive energy. (Some authors call this exotic matter; we reserve that term for negative energy densities.)

The concept of negative energy is not pure fantasy; some of its effects have even been produced in the laboratory. They arise from Heisenberg's uncertainty principle, which requires that the energy density of any electric, magnetic or other field fluctuate randomly. Even when the energy density is zero on average, as in a vacuum, it fluctuates. Thus, the quantum vacuum can never remain empty in the classical sense of the term; it is a roiling sea of "virtual" particles spontaneously popping in and out of existence. In quantum theory, the usual notion of zero energy corresponds to the vacuum with all these fluctua-

THE AUTHORS

LAWRENCE H. FORD and *THOMAS A. ROMAN* have collaborated on negative energy issues for more than a decade. Ford received his Ph.D. from Princeton University in 1974 under John Wheeler, one of the founders of black hole physics. He is now professor of physics at Tufts University and works on problems in both general relativity and quantum theory, with a special interest in quantum fluctuations. Roman received his Ph.D. in 1981 from Syracuse University under Peter Bergmann, who collaborated with Albert Einstein on unified field theory. He is currently professor of physics at Central Connecticut State University. His interests include the implications of negative energy for a quantum theory of gravity.

tions. So if one can somehow contrive to dampen the undulations, the vacuum will have less energy than it normally does—that is, less than zero energy.

As an example, researchers in quantum optics have created special states of fields in which destructive quantum interference suppresses the vacuum fluctuations. These so-called squeezed vacuum states involve negative energy. More precisely, they are associated with regions of alternating positive and negative energy. The total energy averaged over all space remains positive; squeezing the vacuum creates negative energy in one place at the price of extra positive energy elsewhere. A typical experiment involves laser beams passing through nonlinear optical materials. The intense laser light induces the material to create pairs of light quanta, photons. These photons alternately enhance and suppress the vacuum fluctuations, leading to regions of positive and negative energy, respectively.

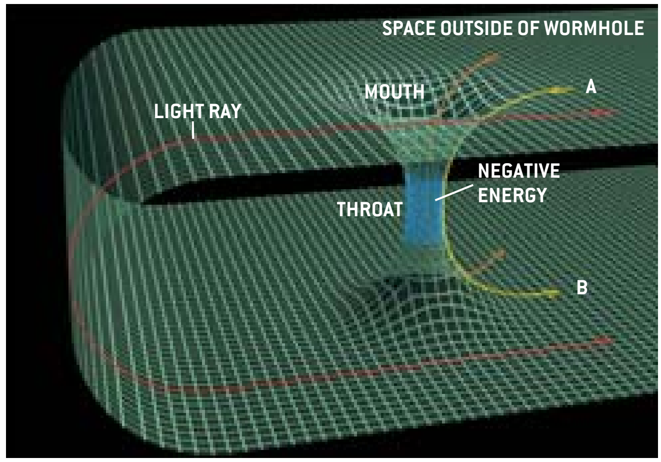
Another method for producing negative energy introduces geometric boundaries into a space. In 1948 Dutch physicist Hendrik B. G. Casimir showed that two uncharged parallel metal plates alter the vacuum fluctuations in such a way as to attract each other. The energy density between the plates was later calculated to be negative. In effect, the plates reduce the fluctuations in the gap between them; this creates negative energy and pressure, which pulls the plates together. The narrower the gap, the more negative the energy and pressure, and the stronger the attractive force. The Casimir effect has been measured by Steve K. Lamoreaux of Los Alamos National Laboratory and by Umar Mohideen of the University of California at Riverside and his colleague Anushree Roy. Other groups have recently confirmed these experiments and have even begun to explore the role of the effect in nanotechnology. Similarly, in the 1970s Paul C. W. Davies and Stephen A. Fulling, then at King's College at the University of London, predicted that a moving boundary, such as a moving mirror, could produce a flux of negative energy.

For both the Casimir effect and squeezed states, researchers have measured only the indirect effects of negative energy. Direct detection is more difficult but might be possible using atomic spins, as Peter G. Grove, then at the British Home Office, Adrian C. Ottewill, then at the University of Oxford, and one of us (Ford) suggested in 1992.

Gravity and Levity

THE CONCEPT OF NEGATIVE ENERGY arises in several areas of modern physics. It has an intimate link with black holes. In 1974 Stephen W. Hawking of the University of Cambridge made his famous prediction that black holes evaporate by emitting radiation [see "The Quantum Mechanics of Black Holes," by Stephen W. Hawking; Scientific American, January 1977]. A black hole radiates energy at a rate inversely proportional to the square of its mass. Although the evaporation rate is large only for subatomic-size black holes, it provides a crucial link between the laws of black holes and the laws of thermodynamics. The Hawking radiation allows black holes to come into thermal equilibrium with their environment.

At first glance, evaporation leads to a contradiction. The



WORMHOLE ACTS AS A TUNNEL between two different locations in space. Light rays traveling from A to B can enter one mouth of the wormhole, pass through the throat and exit the other mouth—a journey that would take much longer if they had to go the long way around. At the throat must be negative energy (*blue*), the gravitational field of which allows converging light rays to begin diverging. (This diagram is a two-dimensional representation of three-dimensional space. The mouths and throat of the wormhole are actually spheres.) A wormhole could also connect two different points in time (*not shown here*).

MICHAEL GOODMAN

black hole's horizon is a one-way street; energy can only flow inward. So how can a black hole radiate energy outward? Because energy must be conserved, the production of positive energy—which distant observers see as the Hawking radiation—is accompanied by a flow of negative energy into the hole. Here the negative energy is produced by the extreme spacetime curvature near the hole, which disturbs the vacuum fluctuations. In this way, negative energy is required for the consistency of the unification of black hole physics with thermodynamics.

The black hole is not the sole curved region of spacetime where negative energy seems to play a role. Another is the wormhole—a hypothesized type of tunnel that connects one region of space and time to another. Physicists used to think that wormholes exist only on the very finest scales, bubbling in and out of existence like virtual particles.

But in the late 1980s various researchers—notably Michael S. Morris and Kip S. Thorne, both of the California Institute of Technology, and Matt Visser of Washington University—found that certain wormholes could in fact be made large enough for a person or spaceship. Someone might enter the mouth of a wormhole stationed on Earth, walk a short distance inside the wormhole and exit the other mouth in, say, the Andromeda galaxy. The catch is that traversable wormholes require negative energy. Because negative energy is gravitationally repulsive, it would prevent the wormhole from collapsing.

For a wormhole to be traversable, it ought to (at bare minimum) allow signals, in the form of light rays, to pass through it. Light rays entering one mouth of a wormhole are converging, but to emerge from the other mouth, they must defocus—in other words, they must go from converging to diverging somewhere in between [*see illustration above*]. This defocusing requires negative energy. Whereas the curvature of space produced by the attractive gravitational field of ordinary matter acts like a converging lens, negative energy acts like a diverging lens.

No Dilithium Needed

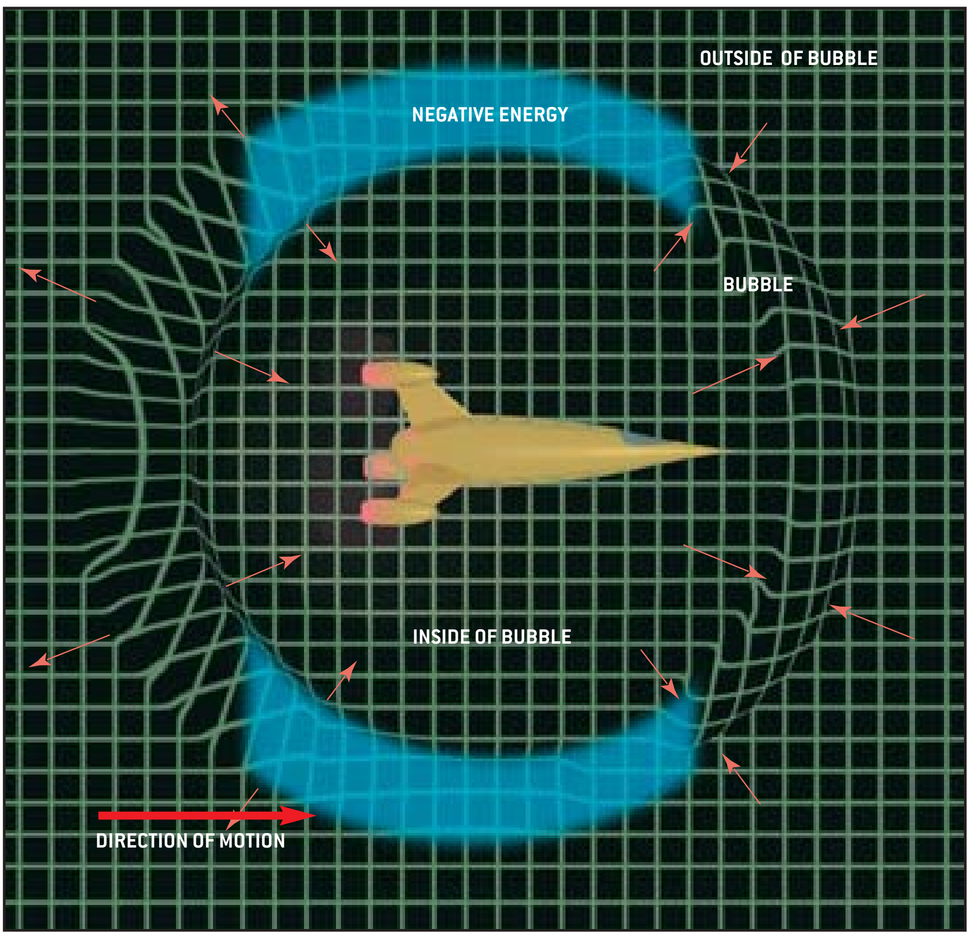
SUCH SPACETIME CONTORTIONS would enable another staple of science fiction as well: faster-than-light travel. In 1994 Miguel Alcubierre Moya, then at the University of Wales in Cardiff, discovered a solution to Einstein's equations that has many of the desired features of warp drive. It describes a spacetime bubble that transports a starship at arbitrarily high speeds relative to observers outside the bubble. Calculations show that negative energy is required.

Warp drive might appear to violate Einstein's special theory of relativity. But special relativity says that you cannot outrun a light signal in a fair race in which you and the signal follow the same route. When spacetime is warped, it might be possible to beat a light signal by taking a different route, a shortcut. The contraction of spacetime in front of the bubble and the expansion behind it create such a shortcut [*see illustration below*].

One problem, pointed out by Sergei V. Krasnikov of the Central Astronomical Observatory in Pulkovo, Russia, is that the interior of the warp bubble is causally disconnected from its forward edge. A starship captain on the inside cannot steer the bubble or turn it on or off; some external agency must set it up ahead of time. To get around this problem, Krasnikov proposed a "superluminal subway," a tube of modified spacetime (not the same as a wormhole) connecting Earth and a star. Within it, superluminal travel in one direction is possible. During the outbound journey at sublight speed, a spaceship crew would create such a tube. On the return, members could travel through it at warp speed. Like warp bubbles, the subway involves negative energy. It has since been shown by Ken D. Olum of Tufts University, by Visser, together with Bruce Bassett of Oxford and Stefano Liberati of the International School for Advanced Studies in Trieste, Italy, and by Sijie Gao and Robert M. Wald of the University of Chicago that any faster-than-light travel requires negative energy.

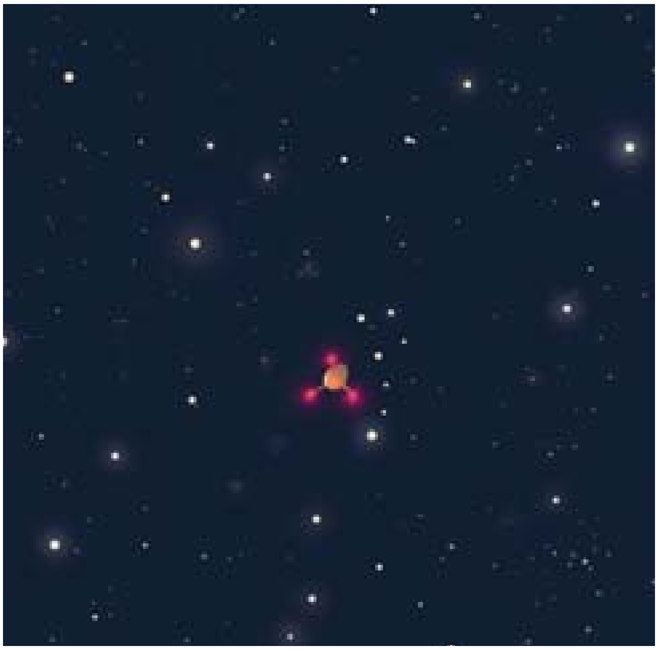
If one can construct wormholes or warp drives, time travel might become possible. The passage of time is relative; it depends on the observer's velocity. A person who leaves Earth in a spaceship, travels at near light speed and returns will have aged less than someone who remains on Earth. If the traveler manages to outrun a light ray, perhaps by taking a shortcut through a wormhole or a warp bubble, he may return before he left. Morris, Thorne and Ulvi Yurtsever, then at Caltech, proposed a wormhole time machine in 1988, and their paper has stimulated much research on time travel since. In 1992 Hawking proved that any construction of a time machine in a finite region of spacetime inherently requires negative energy.

Negative energy is so strange that one might think it must violate some law of physics. Before and after the creation of equal



SPACETIME BUBBLE is the closest that modern physics comes to the "warp drive" of science fiction. It can convey a starship at arbitrarily high speeds. Spacetime contracts at the front of the bubble, reducing the distance to the destination, and expands at its rear, increasing the distance from the origin (*arrows*). The ship itself stands still relative to the space immediately around it; crew members do not experience any acceleration. Negative energy (*blue*) is required on the sides of the bubble.

MICHAEL GOODMAN



VIEW FROM THE BRIDGE of a faster-than-light starship as it heads in the direction of the Little Dipper (*above*) looks nothing like the star streaks typically depicted in science fiction. As the velocity increases (*right*), stars ahead of the ship (*left column*) appear ever closer to the direction of motion and turn bluer in color. Behind the ship (*right column*), stars shift closer to a position directly astern, redden and eventually disappear from view. The light from stars directly overhead or underneath remains unaffected. (Illustration based on calculations by Chad Clark, William A. Hiscock and Shane L. Larson of Montana State University.)

MICHAEL GOODMAN

amounts of negative and positive energy in previously empty space, the total energy is zero, so the law of conservation of energy is obeyed. But there are many phenomena that conserve energy yet never occur in the real world. A broken glass does not reassemble itself, and heat does not spontaneously flow from a colder to a hotter body. Such effects are forbidden by the second law of thermodynamics. This general principle states that the degree of disorder of a system—its entropy—cannot decrease on its own without an input of energy. Thus, a refrigerator, which pumps heat from its cold interior to the warmer outside room, requires an external power source. Similarly, the second law also forbids the complete conversion of heat into work.

Negative energy potentially conflicts with the second law. Imagine an exotic laser, which creates a steady outgoing beam of negative energy. Conservation of energy requires that a byproduct be a steady stream of positive energy. One could direct the negative energy beam off to some distant corner of the universe while employing the positive energy to perform useful work. This seemingly inexhaustible energy supply could be used to make a perpetual-motion machine, thereby violating the second law. If the beam were directed at a glass of water, it could cool the water while using the extracted positive energy to power a small motor—providing refrigeration with no need for external power. These problems arise from the unrestricted separation of negative and positive energy.

Unfettered negative energy would also have profound consequences for black holes. When a black hole forms, general relativity predicts the formation of a singularity, a region where the gravitational field becomes infinitely strong. At this point, all known laws of physics are unable to say what happens next. This inability is a profound failure of the current mathematical description of nature. So long as the singularity is hidden within an event horizon, however, the damage is limited. The description of nature everywhere outside of the horizon is unaffected. For this reason, Roger Penrose of Oxford proposed the cosmic censorship hypothesis: there can be no naked singularities, unshielded by event horizons.

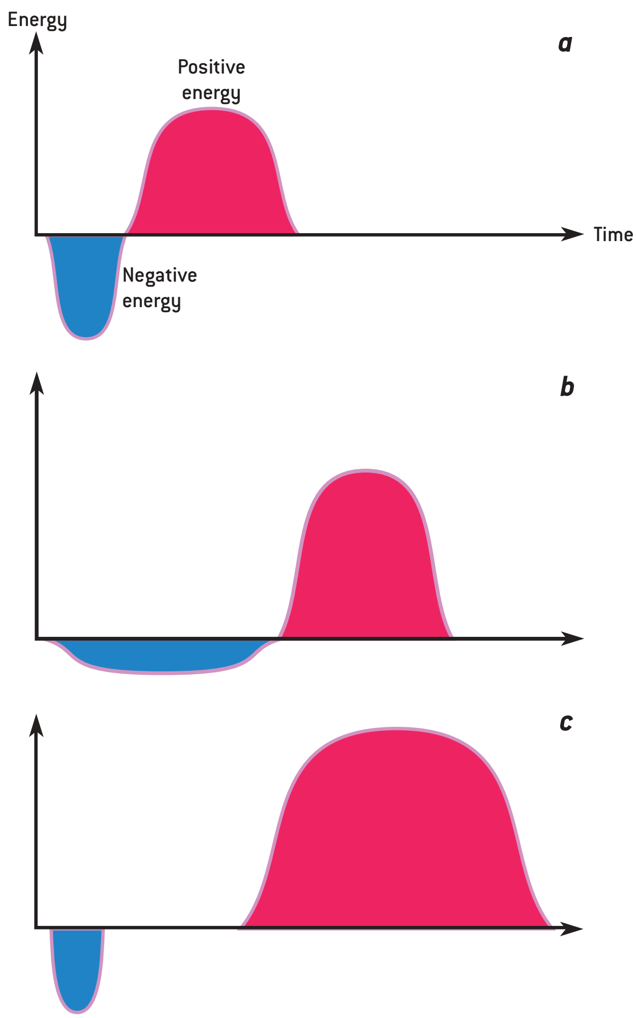
For special types of charged or rotating black holes—known as extreme black holes—even a small increase in charge or spin or a decrease in mass could theoretically destroy the horizon and convert the hole into a naked singularity. Attempts to charge up or spin up these black holes using ordinary matter seem to fail. One might instead envision producing a decrease in mass by shining a beam of negative energy down the hole, without altering its charge or spin, subverting cosmic censorship. One might create such a beam, for example, using a moving mirror.

In principle, it would require only a tiny amount of negative energy to produce a dramatic change in the state of an extreme black hole. Therefore, this might be the scenario in which negative energy is the most likely to produce macroscopic effects.

Not Separate and Not Equal

FORTUNATELY (or not, depending on your point of view), although quantum theory allows the existence of negative energy, it also appears to place strong restrictions—known as quantum inequalities—on its magnitude and duration. These inequalities were first suggested by Ford in 1978. Over the past decade they have been proved and refined by us and others, including Éanna E. Flanagan of Cornell University, Michael J. Pfenning, then at Tufts, Christopher J. Fewster and Simon P. Eveson of the University of York in England, and Edward Teo of the National University of Singapore.

The inequalities bear some resemblance to the uncertainty principle. They say that a beam of negative energy cannot be arbitrarily intense for an arbitrarily long time. The permissible magnitude of the negative energy is inversely related to its temporal or spatial extent. An intense pulse of negative energy can last for a short time; a weak pulse can last longer. Furthermore, an initial negative energy pulse must be followed by a larger pulse of positive energy [*see illustration below*]. The larger the magnitude of the negative energy, the nearer its positive energy counterpart must be. These restrictions are independent of the details of how the negative energy is produced. One can think of negative energy as an energy loan. Just as a debt is negative money that has to be repaid, negative energy is an energy deficit.



PULSES OF NEGATIVE ENERGY are permitted by quantum theory but only under three conditions. First, the longer the pulse lasts, the weaker it must be (*a*, *b*). Second, a pulse of positive energy must follow. The magnitude of the positive pulse must exceed that of the initial negative one. Third, the longer the time interval between the two pulses, the larger the positive one must be—an effect known as quantum interest (*c*).

In the Casimir effect, the negative energy density between the plates can persist indefinitely, but large negative energy densities require a very small plate separation. The magnitude of the negative energy density is inversely proportional to the fourth power of the plate separation. Just as a pulse with a very negative energy density is limited in time, very negative Casimir energy density must be confined between closely spaced plates. According to the quantum inequalities, the energy density in the gap can be made more negative than the Casimir value, but only temporarily. In effect, the more one tries to depress the energy density below the Casimir value, the shorter the time over which this situation can be maintained.

When applied to wormholes and warp drives, the quantum inequalities imply that such structures must be limited to submicroscopic sizes or, if they are macroscopic, the negative energy must be confined to incredibly thin bands. In 1996 we showed that a submicroscopic wormhole would have a throat radius of no more than 10^{-32} meter. This is just slightly larger than the Planck length, 10^{-35} meter, the smallest distance that has meaning. We found that it is possible to model wormholes of macroscopic size but only at the price of confining the negative energy to an extremely thin band around the throat. In one model, a throat radius of one meter requires the negative energy to be a band no thicker than 10^{-21} meter, a millionth the size of a proton. Visser has estimated that the negative energy required for this wormhole has a magnitude equivalent to the total energy generated by 10 billion stars in one year. The situation does not improve for larger wormholes. For the same model, the maximum thickness of the negative energy band is proportional to the cube root of the throat radius. Even if the throat radius is increased to one light-year, the negative energy must still be confined to a region smaller than a proton radius, and the total required increases linearly with the throat size.

It seems that wormhole engineers face daunting problems. They must find a mechanism for confining large amounts of negative energy to extremely thin volumes. So-called cosmic strings, hypothesized in some cosmological theories, involve very large energy densities in long, narrow lines. But all known physically reasonable cosmic-string models have positive energy densities.

Warp drives are even more tightly constrained. In Alcubierre's model, a warp bubble traveling at 10 times the speed of light must have a wall thickness of no more than 10^{-32} meter. A bubble large enough to enclose a starship 200 meters

LISA BURNETT

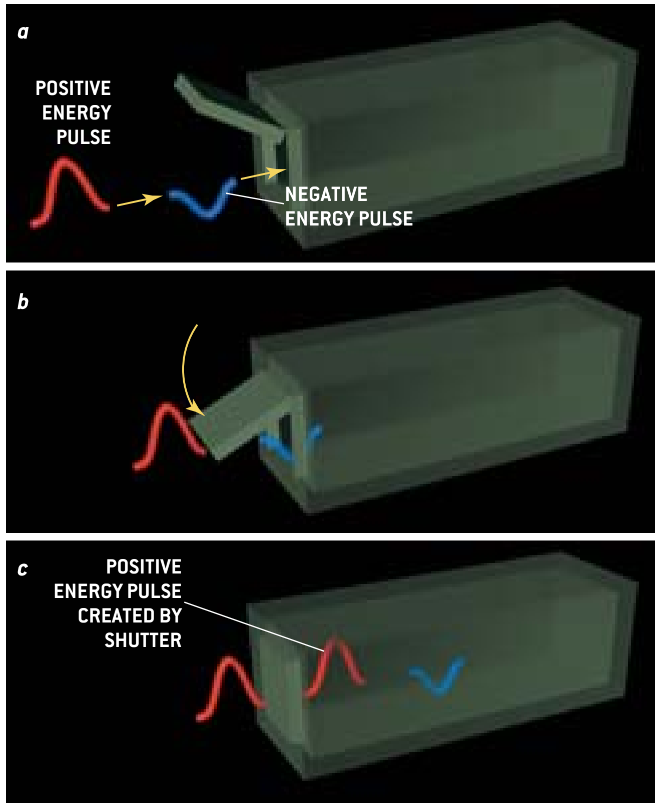
across would require a total amount of negative energy equal to 10 billion times the mass of the observable universe. Similar constraints apply to Krasnikov's superluminal subway. A modification of Alcubierre's model was constructed in 1999 by Chris Van Den Broeck of the Catholic University of Louvain in Belgium. It requires much less negative energy but places the starship in a curved spacetime bottle whose neck is about 10^{-32} meter across, a difficult feat. These results would seem to make it rather unlikely that one could construct wormholes and warp drives using negative energy generated by quantum effects.

Cosmic Flashing and Quantum Interest

THE QUANTUM INEQUALITIES prevent violations of the second law. If one tries to use a pulse of negative energy to cool a hot object, it will be quickly followed by a larger pulse of positive energy, which reheats the object. A weak pulse of negative energy could remain separated from its positive counterpart for a longer time, but its effects would be indistinguishable from normal thermal fluctuations. Attempts to capture or split off negative energy from positive energy also appear to fail. One might intercept an energy beam by, say, using a box with a shutter. By closing the shutter, one might hope to trap a pulse of negative energy before the offsetting positive energy arrives. But the very act of closing the shutter creates an energy flux that cancels out the negative energy it was designed to trap [*see illustration at right*].

We have shown that there are similar restrictions on violations of cosmic censorship. A pulse of negative energy injected into a charged black hole might momentarily destroy the horizon, exposing the singularity within. But the pulse must be followed by a pulse of positive energy, which would convert the naked singularity back into a black hole, a scenario we have dubbed cosmic flashing. The best chance to observe cosmic flashing would be to maximize the time separation between the negative and positive energy, allowing the naked singularity to last as long as possible. But then the magnitude of the negative energy pulse would have to be very small, according to the quantum inequalities. The change in the mass of the black hole caused by the negative energy pulse would get washed out by the normal quantum fluctuations in the hole's mass, which are a natural consequence of the uncertainty principle. The view of the naked singularity would thus be blurred, so a distant observer could not unambiguously verify that cosmic censorship had been violated.

Recently we, Frans Pretorius (then at the University of Victoria in British Columbia), and Fewster and Teo have all shown that the quantum inequalities lead to even stronger bounds on negative energy. The positive pulse that follows an initial negative pulse must do more than compensate for the negative pulse; it must overcompensate. The overcompensation increases with the time interval between the pulses. Therefore, the negative and positive pulses can never be made to cancel exactly. The positive energy must always dominate—an effect known as quantum interest. If negative energy is thought of as an energy loan, the loan must be repaid with interest. The longer the loan period or the larger the loan amount, the greater the interest. Furthermore, the larger the loan, the smaller the maximum allowed loan period. Nature is a shrewd banker and always calls in its debts.



ATTEMPT TO CIRCUMVENT the quantum laws that govern negative energy inevitably ends in disappointment. The experimenter intends to detach a negative energy pulse from its compensating positive energy pulse. As the pulses approach a box (*a*), the experimenter tries to isolate the negative one by closing the lid after it has entered (*b*). Yet the very act of closing the lid creates a second positive energy pulse inside the box (*c*).

MICHAEL GOODMAN

The concept of negative energy touches on gravitation, quantum theory and thermodynamics. The interweaving of all these parts of physics illustrates the tight logical structure of the laws of nature. Negative energy seems to be required to reconcile black holes with thermodynamics. On the other hand, quantum physics prevents unrestricted production of negative energy, a phenomenon that would violate the second law of thermodynamics. Whether these restrictions are also features of some deeper underlying theory, such as quantum gravity, remains to be seen. Nature no doubt has more surprises in store. SA

MORE TO EXPLORE

Black Holes and Time Warps: Einstein's Outrageous Legacy. Kip S. Thorne. W. W. Norton, 1994.

Quantum Field Theory Constrains Traversable Wormhole Geometries. L. H. Ford and T. A. Roman in *Physical Review D,* Vol. 53, No. 10, pages 5496–5507; May 15, 1996.

The Unphysical Nature of Warp Drive. M. J. Pfenning and L. H. Ford in *Classical and Quantum Gravity,* Vol. 14, No. 7, pages 1743–1751; July 1997. Available at **xxx.lanl.gov/abs/gr-qc/9702026**

Time Machines: Time Travel in Physics, Metaphysics, and Science Fiction. Second edition. Paul J. Nahin. AIP Press, Springer-Verlag, 1999.

The Quantum Interest Conjecture. L. H. Ford and T. A. Roman in *Physical Review D,* Vol. 60, No. 10, Article No. 104018; November 15, 1999.

Plenty of Room, Indeed

There is plenty of room for practical innovation at the nanoscale. But first, scientists have to understand the unique physics that governs matter there

By Michael Roukes

Back in December 1959, future Nobel laureate Richard Feynman gave a visionary and now oft-quoted talk entitled "There's Plenty of Room at the Bottom." The occasion was an American Physical Society meeting at the California Institute of Technology, Feynman's intellectual home then and mine today. Although he didn't intend it, Feynman's 7,000 words were a defining moment in nanotechnology, long before anything "nano" appeared on the horizon.

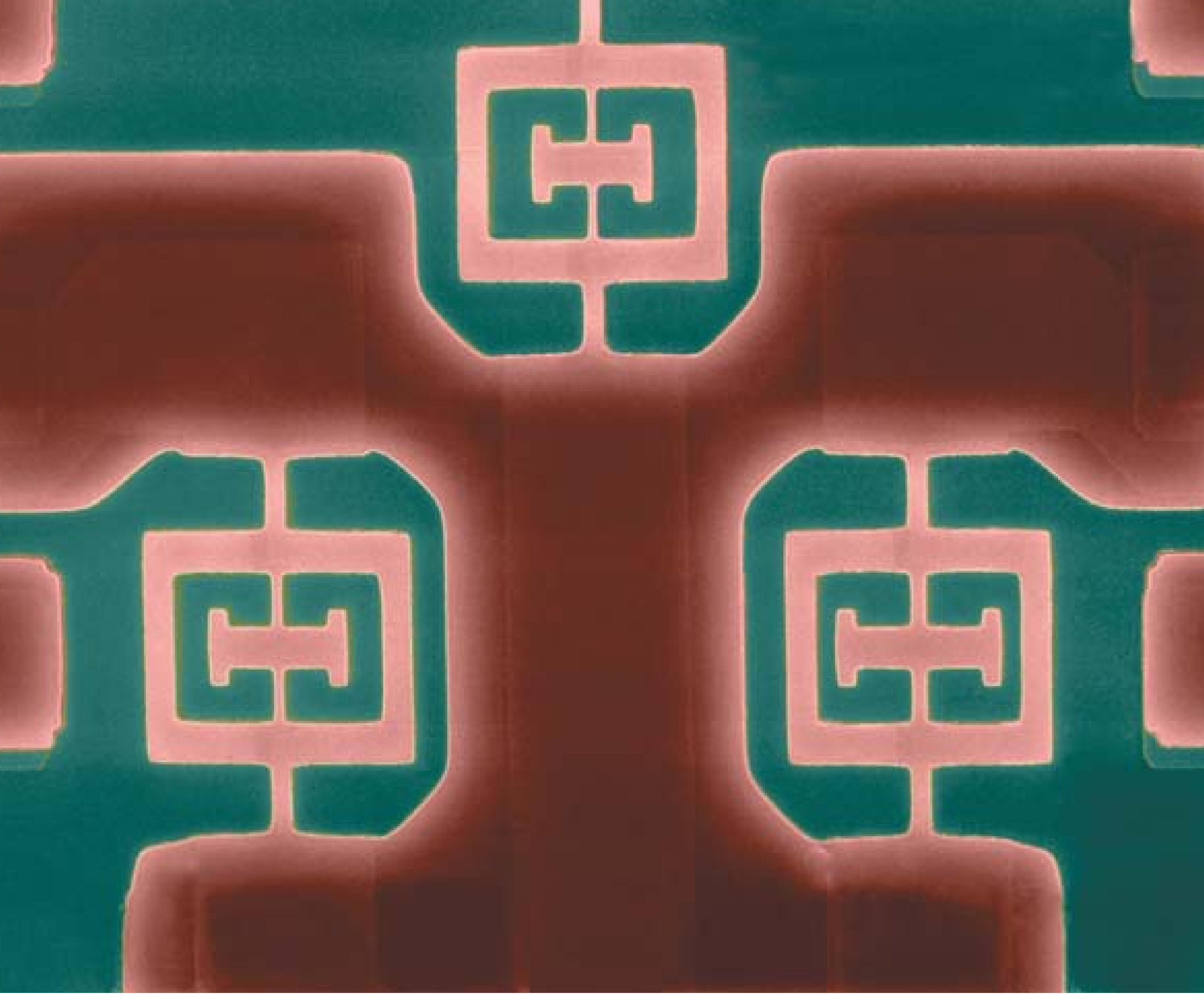
"*What I want to talk about,*" he said, "*is the problem of manipulating and controlling things on a small scale.... What I have demonstrated is that there is room—that you can decrease the size of things in a practical way. I now want to show that there is plenty of room. I will not now discuss how we are going to do it, but only what is possible in principle.... We are not doing it simply because we haven't yet gotten around to it.*"

The breadth of Feynman's vision is staggering. In that lecture 44 years ago he anticipated a spectrum of scientific and technical fields that are now well established, among them electron-beam and ion-beam fabrication, molecular-beam epitaxy, nanoimprint lithography, projection electron microscopy, atom-by-atom manipulation, quantum-effect electronics, spin electronics (also called spintronics) and microelectromechanical systems (MEMS). The lecture also projected what has been called the "magic" Feynman brought to everything he turned his singular intellect toward. Indeed, it has profoundly inspired my two decades of research on physics at the nanoscale.

Today there is a nanotechnology gold rush. Nearly every major funding agency for science and engineering has announced its own thrust into the field. Scores of researchers and institutions are scrambling for a piece of the action. But in all honesty, I think we have to admit that much of what invokes the hallowed prefix "nano" falls a bit short of Feynman's mark.

We've only just begun to take the first steps toward his grand vision of assembling complex machines and circuits atom by atom. What can be done now is extremely rudimentary. We're certainly nowhere near being able to commercial-

Updated from the September 2001 issue



M. J. MURPHY, D. A. HARRINGTON AND M. L. ROUKES *California Institute of Technology;* COLORIZATION BY FELICE FRANKEL

NOVEL NANOTECH DEVICES, such as these nanoelectromechanical resonators, are enabling scientists to discover the laws of physics that regulate the unique properties of matter at the mesoscale.

ly mass-produce nanosystems—integrated multicomponent nanodevices that have the complexity and range of functions readily provided by modern microchips. But there is a fundamental science issue here as well. It is becoming increasingly clear that we are only *beginning* to acquire the detailed knowledge that will be at the heart of future nanotechnology. This new science concerns the properties and behavior of aggregates of atoms and molecules, at a scale not yet large enough to be considered macroscopic but far beyond what can be called microscopic. It is the science of the *mesoscale,* and until we understand it, practical devices will be difficult to realize.

Scientists and engineers readily fashion nanostructures on a scale of one to a few hundred nanometers—small indeed, but much bigger than simple molecules. Matter at this mesoscale is often awkward to explore. It contains too many atoms to be easily understood by the straightforward application of quantum mechanics (although the fundamental laws still apply). Yet these systems are not so large as to be completely free of quantum effects; thus, they do not simply obey the classical physics governing the macroworld. It is precisely in this intermediate domain, the mesoworld, that unforeseen properties of collective systems emerge.

Researchers are approaching this transitional frontier using complementary top-down and bottom-up fabrication methods. Advances in top-down nanofabrication techniques, such as electron-beam lithography (used extensively by my own research group), yield almost atomic-scale precision, but achieving success, not to mention reproducibility, as we scale down to the single-digit-nanometer regime becomes problematic. Alternatively, scientists are using bottom-up techniques for *self-assembly* of atoms. But the advent of preprogrammed self-assembly of arbitrarily large systems—with complexity comparable to that built every day in microelectronics, in

It is becoming increasingly clear that we are only beginning to acquire the detailed knowledge that will be at the heart of future nanotechnology.

MEMS and (of course) by Mother Nature—is nowhere on the horizon. It appears that the top-down approach will most likely remain the method of choice for building really complex devices for a good while.

Our difficulty in approaching the mesoscale from above or below reflects a basic challenge of physics. Lately, the essence of Feynman's "Plenty of Room" talk seems to be taken as a license for laissez-faire in nanotechnology. Yet Feynman never asserted that "anything goes" at the nanoscale. He warned, for instance, that the very act of trying to *"arrange the atoms one by one the way we want them"* is subject to fundamental principles: *"You can't put them so that they are chemically unstable, for example."*

Accordingly, today's scanning probe microscopes can move atoms from place to place on a prepared surface, but this ability does not immediately confer the power to build complex molecular assemblies at will. What has been accomplished so far, though impressive, is still quite limited. We will ultimately develop operational procedures to help us coax the formation of individual atomic bonds under more general conditions. But as we try to assemble complex networks of these bonds, they certainly will affect one another in ways we do not yet understand and, hence, cannot yet control.

Feynman's original vision was clearly intended to be inspirational. Were he observing now, he would surely be alarmed when people take his projections as some sort of gospel. He delivered his musings with characteristic playfulness as well as deep insight. Sadly for us, the field that would be called nanotechnology was just one of many that intrigued him. He never really continued with it, returning to give but one redux of his original lecture, at the Jet Propulsion Laboratory in 1983.

New Laws Prevail

IN 1959, AND EVEN in 1983, the complete physical picture of the nanoscale was far from clear. The good news for researchers is that, by and large, it still is! Much exotic territory awaits exploration. As we delve into it, we will uncover a panoply of phenomena that we must understand before practical nanotechnology will become possible. The past two decades have seen the elucidation of entirely new, fundamental physical principles that govern behavior at the mesoscale. Let's consider three important examples.

In the fall of 1987 graduate student Bart J. van Wees of the Delft University of Technology and Henk van Houten of the Philips Research Laboratories (both in the Netherlands) and their collaborators were studying the flow of electric current through what are now called quantum-point contacts. These are narrow conducting paths within a semiconductor, along which electrons are forced to flow [*see illustration on page 96*]. Late one evening van Wees's undergraduate assistant, Leo Kouwenhoven, was measuring the conductance through the constriction as he varied its width systematically. The research team was expecting to see only subtle conductance effects against an otherwise smooth and unremarkable background response. Instead there appeared a very pronounced, and now characteristic, staircase pattern. Further analysis that night revealed that plateaus were occurring at regular, precise intervals.

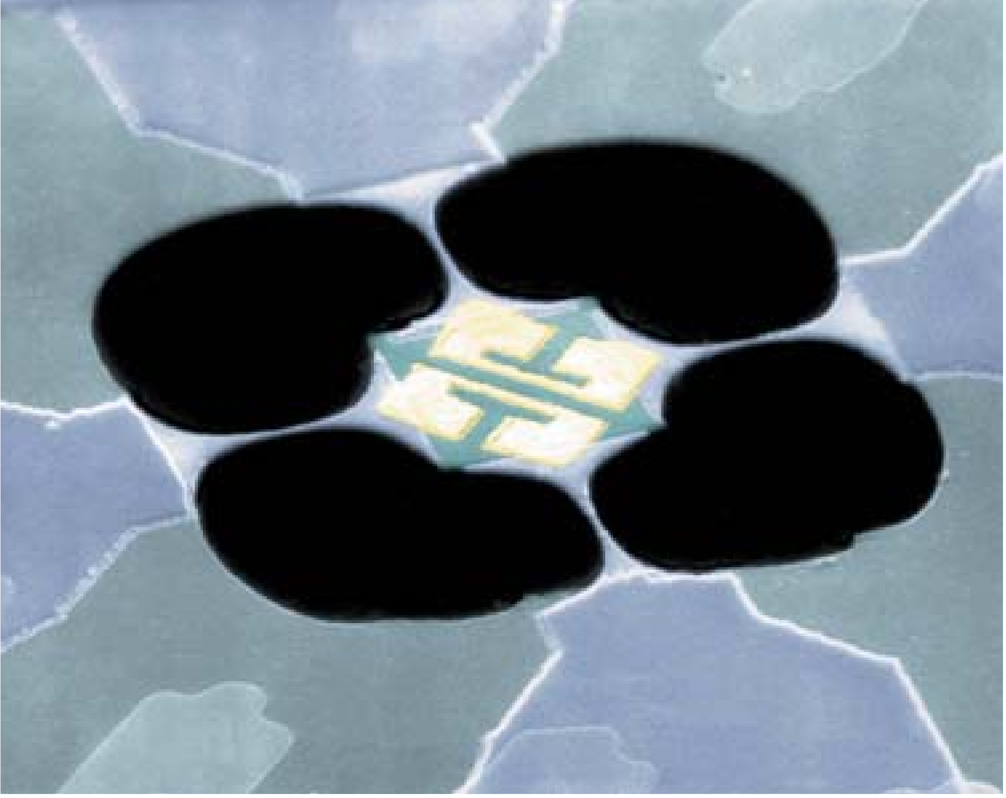
David Wharam and Michael Pepper of the University of Cambridge observed similar results. The two discoveries represented the first robust demonstrations of the *quantization of electrical conductance*. This is a basic property of small conductors that occurs when the wavelike properties of electrons are coherently maintained from the "source" to the "drain"—the input to the output—of a nanoelectronic device.

Feynman anticipated, in part, such odd behavior: *"I have thought about some of the problems of building electric circuits on a small scale, and the problem of resistance is serious...."* But the experimental discoveries pointed out something truly new and fundamental: quantum mechanics can completely govern the behavior of small electrical devices.

Direct manifestations of quantum mechanics in such devices were envisioned back in 1957 by Rolf Landauer, a theoretician at IBM who pioneered ideas in nanoscale electronics and in the physics of computation. But only in the

Overview/*Nanophysics*

- Smaller than macroscopic objects but larger than molecules, nanotechnological devices exist in a unique realm—the mesoscale—where the properties of matter are governed by a complex and rich combination of classical physics and quantum mechanics.
- Engineers will not be able to make reliable or optimal nanodevices until they comprehend the physical principles that prevail at the mesoscale.
- Scientists are discovering mesoscale laws by fashioning unusual, complex systems of atoms and measuring their intriguing behavior.
- Once we understand the science underlying nanotechnology, we can fully realize the prescient vision of Richard Feynman: that nature has left plenty of room in the nanoworld to create practical devices that can help humankind.



NANOBRIDGE DEVICE allowed Caltech physicists to first observe the quantization of thermal conductance—a fundamental limit to heat flow in minute objects. Four holes (*black*) etched into a silicon nitride membrane defined an isolated thermal reservoir (*central green square*) suspended by four narrow bridges. One gold transducer (*yellow*) electrically heated this reservoir; the second measured its temperature. Thin superconducting films (*blue*) on top of the bridges electrically connected the transducers to off-chip instrumentation but carried no heat. The reservoir therefore cooled only through the silicon nitride bridges, which were so narrow that they passed only the lowest-energy heat waves.

K. C. SCHWAB, E. A. HENRIKSEN AND M. L. ROUKES *Caltech;* COLORIZATION BY M. L. ROUKES

mid-1980s did control over materials and nanofabrication begin to provide access to this regime in the laboratory. The 1987 discoveries heralded the heyday of "mesoscopia."

A second significant example of newly uncovered mesoscale laws that have led to nascent nanotechnology was first postulated in 1985 by Konstantin Likharev, a young physics professor at Moscow State University working with postdoctoral student Alexander Zorin and undergraduate Dmitri Averin. They anticipated that scientists would be able to control the movement of single electrons on and off a "coulomb island," a conductor weakly coupled to the rest of a nanocircuit. This could form the basis for an entirely new type of device, called a single-electron transistor. The physical effects that arise when putting a single electron on a coulomb island become more robust as the island is scaled downward. In very small devices, these single-electron charging effects can completely dominate the current flow.

Such considerations are becoming increasingly important technologically. Projections from the International Technology Roadmap for Semiconductors, prepared by long-range thinkers in the industry, indicate that by 2014 the minimum feature size for transistors in computer chips will decrease to 20 nanometers. At this dimension, each switching event will involve the equivalent of only about eight electrons. Designs that properly account for single-electron charging will become crucial.

By 1987 advances in nanofabrication allowed Theodore A. Fulton and Gerald J. Dolan of Bell Laboratories to construct the first single-electron transistor [*see illustration on page 98*]. The single-electron charging they observed, now called the coulomb blockade, has since been seen in a wide array of structures. As experimental devices get smaller, the coulomb blockade phenomenon is becoming the rule, rather than the exception, in weakly coupled nanoscale devices. This is especially true in experiments in which electric currents are passed through individual molecules. These molecules can act like coulomb islands by virtue of their weak coupling to electrodes leading back to the macroworld. Using this effect to advantage and obtaining robust, reproducible coupling to small molecules (in ways that can actually be engineered) are among the important challenges in the new field of molecular electronics.

In 1990, against this backdrop, I was at Bell Communications Research studying electron transport in mesoscopic semiconductors. In a side project, my colleagues Larry M. Schiavone and Axel Scherer and I began developing techniques that we hoped would elucidate the quantum nature of *heat* flow. The work required much more sophisticated nanostructures than the planar devices used to investigate mesoscopic electronics. We needed freely suspended devices, structures possessing full three-dimensional relief. Ignorance was bliss; I had no idea the experiments would be so involved that they would take almost a decade to realize.

The first big strides were made after I moved to Caltech in 1992, in a collaboration with John M. Worlock of the University of Utah and two successive postdocs in my group. Thomas S. Tighe developed the methods and devices that generated the first direct measurements of heat flow in nanostructures. Subsequently, Keith C. Schwab revised the design of the suspended nanostructures and put in place ultrasensitive superconducting instrumentation to interrogate them at ultralow temperatures, at which the effects could be seen most clearly.

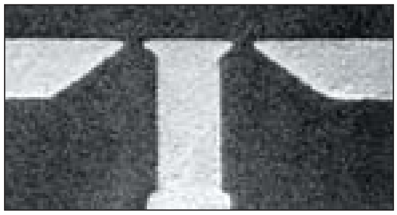
In the late summer of 1999 Schwab

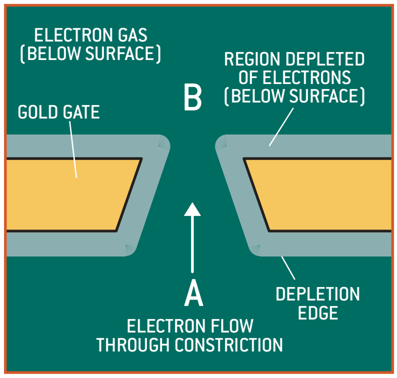
THE AUTHOR

MICHAEL ROUKES, professor of physics at the California Institute of Technology, heads a group studying nanoscale systems. Among the holy grails his team is chasing are a billionfold improvement in present-day calorimetry, which would allow observation of the individual heat quanta being exchanged as nanodevices cool, and a quadrillionfold increase in the sensitivity of magnetic resonance imaging, which would enable complex biomolecules to be visualized with three-dimensional atomic resolution.

ONE STEP AT A TIME

QUANTIZATION OF ELECTRICAL CONDUCTANCE





In 1987 Bart J. van Wees and his collaborators at the Delft University of Technology and Philips Research Laboratories (both in the Netherlands) built a novel structure (*micrograph*) that revealed a basic law governing nanotech circuits. Gold gate electrodes (*bright areas*) were placed atop a semiconductor substrate (*dark background*). Within the substrate, a planar sheet of charge carriers, called a two-dimensional electron gas, was created about 100 nanometers below the surface. The gates and the gas acted like the plates of a capacitor.

When a negative voltage bias was applied to the gates, electrons within the gas underneath the gates, and slightly beyond the gates' periphery, were pushed away. (The diagram shows this state.) When increasing negative voltage was applied, this "depletion edge" became more pronounced. At a certain threshold, carriers on either side of the constriction (*between points A and B*) became separated, and the conductance through the device was zero. From this threshold level, conductance did not resume smoothly. Instead it increased in stepwise fashion, where the steps occurred at values determined by twice the charge of the electron squared, divided by Planck's constant. This ratio is now called the electrical conductance quantum, and it indicates that electric current flows in nanocircuits at rates that are quantized.

finally began observing heat flow through silicon nitride nanobridges [*see illustration on preceding page*]. Even in these first data the fundamental limit to heat flow in mesoscopic structures emerged. The manifestation of this limit is now called the thermal conductance quantum. It determines the maximum rate at which heat can be carried by an individual wavelike mechanical vibration, spanning from the input to the output of a nanodevice. It is analogous to the electrical conductance quantum but governs the transport of heat.

This quantum is a significant parameter for nanoelectronics; it represents the ultimate limit for the power-dissipation problem. In brief, all "active" devices require a little energy to operate, and for them to operate stably without overheating, we must design a way to extract the heat they dissipate. As engineers try continually to increase the density of transistors and the clock rates (frequencies) of microprocessors, the problem of keeping microchips cool to avoid complete system failure is becoming monumental. This will only become further exacerbated in nanotechnology.

Considering even this complexity, Feynman said, "*Let the bearings run dry; they won't run hot because the heat escapes away from such a small device very, very rapidly.*" But our experiments indicate that nature is a little more restrictive. The thermal conductance quantum can place limits on how effectively a very small device can dissipate heat. What Feynman envisioned can be correct only if the nanoengineer designs a structure so as to take these limits into account.

From the three examples above, we can arrive at just one conclusion: we are only starting to unveil the complex and wonderfully different ways that nanoscale systems behave. The discovery of the electrical and thermal conductance quanta and the observation of the coulomb blockade are true discontinuities—abrupt changes in our understanding. Today we are not accustomed to calling our discoveries "laws." Yet I have no doubt that electrical and thermal conductance quantization and single-electron-charging phenomena are indeed among the universal rules of nanodesign. They are new laws of the nanoworld. They do not contravene but augment and clarify some of Feynman's original vision. Indeed, he seemed to have anticipated their emergence: "*At the atomic level, we have new kinds of forces and new kinds of possibilities, new kinds of effects. The problems of manufacture and reproduction of materials will be quite different.*"

We will encounter many more such discontinuities on the path to true nanotechnology. These welcome windfalls will occur in direct synchrony with advances in our ability to observe, probe and control nanoscale structures. It would seem wise, therefore, to be rather modest and circumspect about forecasting nanotechnology.

The Boon and Bane of Nano

THE NANOWORLD is often portrayed by novelists, futurists and the popular press as a place of infinite possibilities. But as you've been reading, this domain is not some ultraminiature version of the Wild West. *Not* everything goes down there; there are *laws*. Two concrete illustrations come from the field of nanoelectromechanical systems (NEMS), in which I am active.

Part of my research is directed toward harnessing small mechanical devices for sensing applications. Nanoscale structures appear to offer revolutionary potential; the smaller a device, the more susceptible its physical properties to alteration. One example is resonant detectors, which are frequently used for sensing mass. The vibrations of a tiny mechanical element, such as a small cantilever, are intimately linked to the element's mass, so the addition of a minute

B. J. VAN WEES *Delft University of Technology (micrograph)*; NINA FINKEL *(diagram)*

The difficulties in communication between the nanoworld and the macroworld represent a central issue in the development of nanotechnology.

amount of foreign material (the "sample" being weighed) will shift the resonant frequency. Work in my lab by then postdoc Kamil Ekinci shows that nanoscale devices can be made so sensitive that "weighing" individual atoms and molecules becomes feasible.

But there is a dark side. Gaseous atoms and molecules constantly adsorb and desorb from a device's surfaces. If the device is macroscopic, the resulting fractional change in its mass is negligible. But the change can be significant for nanoscale structures. Gases impinging on a resonant detector can change the resonant frequency randomly. Apparently, the smaller the device, the less stable it will be. This instability may pose a real disadvantage for various types of futuristic electromechanical signal-processing applications. Scientists might be able to work around the problem by, for example, using arrays of nanomechanical devices to average out fluctuations. But for individual elements, the problem seems inescapable.

A second example of how "not everything goes" in the nanoworld relates more to economics. It arises from the intrinsically ultralow power levels at which nanomechanical devices operate. Physics sets a fundamental threshold for the minimum operating power: the ubiquitous, random thermal vibrations of a mechanical device impose a "noise floor" below which real signals become increasingly hard to discern. In practical use, nanomechanical devices are optimally excited by signal levels 1,000-fold or a million-fold greater than this threshold. But such levels are still a millionth to a billionth the amount of power used for conventional transistors.

The advantage, in some future nanomechanical signal-processing system or computer, is that even a million nanomechanical elements would dissipate only a millionth of a watt, on average. Such ultralow power systems could lead to wide proliferation and distribution of cheap, ultraminiature "smart" sensors that could continuously monitor *all* of the important functions in hospitals, in manufacturing plants, on aircraft, and so on. The idea of ultraminiature devices that drain their batteries extremely slowly, especially ones with sufficient computational power to function autonomously, has great appeal.

But here, too, there is a dark side. The regime of ultralow power is quite foreign to present-day electronics. Nanoscale devices will require entirely new system architectures that are compatible with amazingly low power thresholds. This prospect is not likely to be received happily by the computer industry, with its overwhelming investment in current devices and methodology. A new semiconductor processing plant today costs more than $1 billion, and it would probably have to be retooled to be useful. But I am certain that the revolutionary prospects of nanoscale devices will eventually compel such changes.

RICHARD FEYNMAN predicted the rise of nanotechnology in a landmark 1959 talk at Caltech. "The principles of physics," he said, "do not speak against the possibility of maneuvering things atom by atom." But he also anticipated that unique laws would prevail; they are finally being discovered today.

JOE MUNROE *Ohio Historical Society*, COURTESY OF THE ARCHIVES, CALIFORNIA INSTITUTE OF TECHNOLOGY

Monumental Challenges

CERTAINLY A HOST of looming issues will have to be addressed before we can realize the potential of nanoscale devices. Although each research area has its own concerns, some general themes emerge. Two challenges fundamental to my current work on nanomechanical systems, for instance, are relevant to nanotechnology in general.

Challenge I: Communication between the macroworld and the nanoworld. NEMS are incredibly small, yet their motion can be far smaller. For example, a nanoscale beam clamped on both ends vibrates with minimal harmonic distortion when its vibration amplitude is kept below a small fraction of its thickness. For a 10-nanometer-thick beam, this amplitude is only a few nanometers. Building the requisite, highly efficient transducers to transfer information from such a device to the macroworld involves reading out information with even greater precision.

Compounding this problem, the natural frequency of the vibration increases as the size of the beam is decreased. So to track the device's vibrations usefully, the ideal NEMS transducer must be capable of resolving extremely small displacements, in the picometer-to-femtometer (trillionth to quadrillionth of a meter) range, across very large bandwidths, extending into the microwave range. These twin requirements pose a truly monumental challenge, one much more significant than those faced so far in MEMS work. A further complication is that most of the methodologies from MEMS are inapplicable; they simply don't scale down well to nanometer dimensions.

In each new regime, some wonderful scientific phenomenon emerges. But then a thorny host of underlying, equally unanticipated problems appear.

These difficulties in communication between the nanoworld and the macroworld represent a generic issue in the development of nanotechnology. Ultimately, the technology will depend on robust, well-engineered information transfer pathways from what are, in essence, individual macromolecules. Although the grand vision of futurists may involve self-programmed nanobots that need direction from the macroworld only when they are first wound up and set in motion, it seems more likely that most nanotechnological applications realizable in our lifetimes will entail some form of reporting up to the macroworld and feedback and control back down. The communication problem will remain central.

Orchestrating such communication immediately invokes the very real possibility of collateral damage. Quantum theory tells us that the process of measuring a quantum system nearly always perturbs it. This can hold true even when we scale up from atoms and molecules to nanosystems comprising millions or billions of atoms. Coupling a nanosystem to probes that report back to the macroworld always changes the nanosystem's properties to some degree, rendering it less than ideal. Introducing the transducers required for communication will do more than just increase the nanosystem's size and complexity. They will also necessarily extract some energy to perform their measurements and can degrade the nanosystem's performance. Measurement always has its price.

Challenge II: Surfaces. As we shrink MEMS to NEMS, device physics becomes increasingly dominated by the surfaces. Much of the foundation of solid-state physics rests on the premise that the surface-to-volume ratio of objects is infinitesimal, meaning that physical properties are always dominated by the physics of the bulk. Nanoscale systems are so small that this assumption breaks down completely.

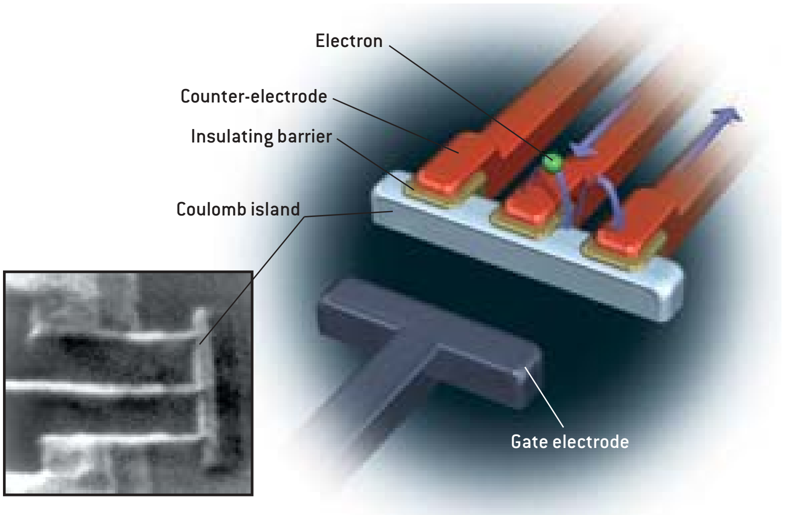
For example, mechanical devices patterned from single-crystal, ultrapure materials can contain very few (even zero) crystallographic defects and impurities. My initial hope was that, as a result, there would be only very weak damping of mechanical vibrations in monocrystalline NEMS. But as we shrink mechanical devices, we repeatedly find that acoustic energy loss seems to increase in proportion to the increasing surface-to-volume ratio. This result clearly impli-

TAKING CHARGE

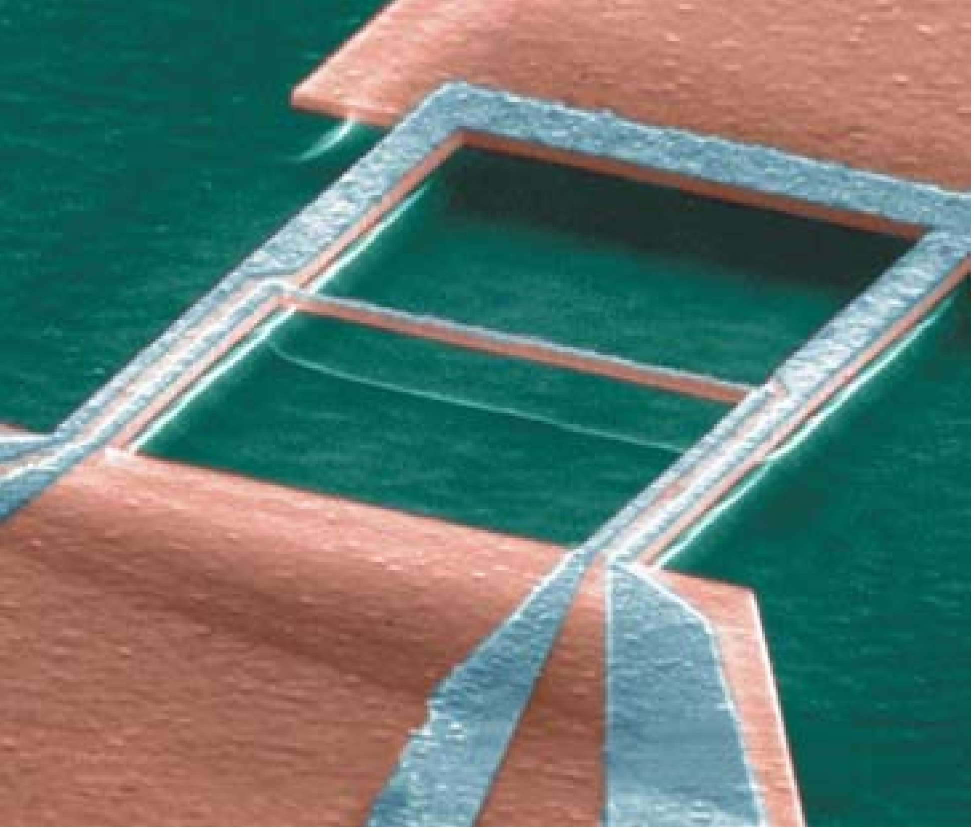
SINGLE ELECTRONICS

Advances in nanofabrication allowed Theodore A. Fulton and Gerald J. Dolan to build a single-electron transistor at Bell Laboratories in 1987 (*micrograph*). In this structure, the controlled movement of individual electrons through a nanodevice was first achieved. At its heart was a coulomb island, a metallic electrode isolated from its counter-electrodes by thin insulating oxide barriers (*diagram*). The counter-electrodes led up to the macroscale laboratory instrumentation used to carry out the experiments. An additional gate electrode (*visible in the diagram but not the micrograph*) was offset from the coulomb island by a small gap; it allowed direct control of the charge introduced to the island. Electric current flowed through the device from one counter-electrode to another, as in a conventional circuit, but here it was limited by the stepwise hopping of electrons onto and off the coulomb island.

Fulton and Dolan's experiments demonstrate both the fundamental physics of single-electron charging and the potential of these devices as ultrasensitive electrometers: instruments that can easily detect individual electron charges. Circuits that switch one electron at a time could someday form the basis for an entirely new class of nanoelectronics. The advent of such single electronics, however, also presages problems that will have to be faced as conventional electronic circuits are shrunk to the nanoscale.



T. A. FULTON AND G. J. DOLAN *Bell Laboratories* (*micrograph*); BRYAN CHRISTIE (*diagram*)



NANOMECHANICAL AMPLIFIER overcomes the vexing problem of communication with the macroworld by providing up to 1,000-fold amplification of weak forces. Two suspended bridges of monocrystalline silicon carbide (*left* and *right*) support the central crossbridge, to which the signal force is applied. Thin-film electrodes (*silver*) atop these structures provide very sensitive readouts of nanoscale motion.

cates surfaces in the devices' vibrational energy-loss processes. In a state-of-the-art silicon beam measuring 10 nanometers wide and 100 nanometers long, more than 10 percent of the atoms are at or next to the surface. It is evident that these atoms will play a central role, but understanding precisely how will require a major, sustained effort.

In this context, nanotube structures, which have been heralded lately, look ideal. A nanotube is a crystalline, rodlike material perfect for building the miniature vibrating structures of interest to us. And because it has no chemical groups projecting outward along its length, one might expect that interaction with "foreign" materials at its surfaces would be minimal. Apparently not. Although nanotubes exhibit ideal characteristics when shrouded within pristine, ultrahigh vacuum environments, samples in more ordinary conditions, where they are exposed to air or water vapor, evince electronic properties that are markedly different. Mechanical properties are likely to show similar sensitivity. So surfaces definitely *do* matter. It would seem there is no panacea.

Payoff in the Glitches

FUTURISTIC THINKING is crucial to making the big leaps. It gives us some wild and crazy goals—a holy grail to chase. And the hope of glory propels us onward. Yet the 19th-century chemist Friedrich August Kekulé once said, "Let us learn to dream, gentlemen, then perhaps we shall find the truth.... But let us beware of publishing our dreams before they have been put to the proof by the waking understanding."

This certainly holds for nanoscience. While we keep our futuristic dreams alive, we also need to keep our expectations realistic. It seems that every time we gain access to a regime that is a factor of 10 different—and presumably "better"—two things happen. First, some wonderful, unanticipated scientific phenomenon emerges. But then a thorny host of underlying, equally unanticipated new problems appear. This pattern has held true as we have pushed to decreased size, enhanced sensitivity, greater spatial resolution, higher magnetic and electric fields, lower pressure and temperature, and so on. It is at the heart of why projecting forward too many orders of magnitude is usually perilous. And it is what should imbue us with a sense of humility and proportion at this, the beginning of our journey. Nature has already set the rules for us. We are out to understand and employ her secrets.

Once we head out on the quest, nature will frequently hand us what initially seems to be nonsensical, disappointing, random gibberish. But the science in the glitches often turns out to be even more significant than the grail motivating the quest. And being proved the fool in this way can truly be the joy of doing science. If we had the power to extrapolate everything correctly from the outset, the pursuit of science would be utterly dry and mechanistic. The delightful truth is that, for complex systems, we do not, and ultimately probably cannot, know everything that is important.

Complex systems are often exquisitely sensitive to a myriad of parameters beyond our ability to sense and record—much less control—with sufficient regularity and precision. Scientists have studied, and in large part already understand, matter down to the fundamental particles that make up the neutrons, protons and electrons that are of crucial importance to chemists, physicists and engineers. But we still cannot deterministically predict how arbitrarily complex assemblages of these three elemental components will finally behave en masse. For this reason, I firmly believe that it is on the foundation of the experimental science under way, in intimate collaboration with theory, that we will build the road to true nanotechnology. Let's keep our eyes open for surprises along the way! SA

MORE TO EXPLORE

Nanoelectromechanical Systems Face the Future. Michael Roukes in *Physics World,* Vol. 14, No. 2; February 2001. Available at **physicsweb.org/article/world/14/2/8**

The author's group: **www.its.caltech.edu/~nano**

Richard Feynman's original lecture "There's Plenty of Room at the Bottom" can be found at **www.its.caltech.edu/~feynman**