The Mystery of the Cosmological Constant

According to theory, the constant, which measures the energy of the vacuum, should be much greater than it is. An understanding of the disagreement could revolutionize fundamental physics

by Larry Abbott

What determines the structure of space and time in the universe? According to Einstein's general theory of relativity, the geometric properties of space are related to the density of energy (and momentum) in the universe. To understand the structure of spacetime, therefore, we must identify potentially relevant sources of energy and evaluate their contributions to the total energy (and momentum) density. The most obvious energy sources that come to mind are ordinary matter and radiation. A much less obvious source of energy that can have an enormous impact on the structure of the universe is empty space itself: the vacuum.

The notion that the vacuum can be a source of energy may seem counterintuitive. But present theories of elementary particles and forces not only allow for a nonzero vacuum energy density but also strongly suggest that it should have a large value. Is the vacuum energy density really as large as these theories appear to suggest it is?

The answer is most emphatically no. The geometric structure of the universe is extremely sensitive to the value of the vacuum energy density. So important is this value that a constant proportional to the vacuum energy density has been defined. It is called the cosmological constant. If the vacuum energy density, or equivalently the cosmological constant, were as large as theories of elementary particles suggest, the universe in which we live would be dramatically different, with properties we would find both bizarre and unsettling. What has gone wrong with our theories? We do not know the an-

swer to this question at present. Indeed, a comparison of our theoretical and experimental understanding of the cosmological constant leads to one of the most intriguing and frustrating mysteries in particle physics and relativity today.

Most people are unaccustomed to the idea that the vacuum might have a nonzero energy density: How can a unit volume of empty space contain energy? The answer in part lies in the fact that, according to quantum mechanics, physical quantities tend to fluctuate unavoidably. Even in the apparent quiet of the vacuum state pairs of particles are constantly appearing and disappearing. Such fluctuations contribute energy to the vacuum.

The notion of a vacuum energy is also unfamiliar because that energy cannot be detected by normal techniques. Energies are usually determined by measuring the change in the energy of a system when it is modified in some way, or by measuring a difference in energy between two systems. For example, we might measure the energy released when two chemicals react. Because of this, energy as we normally define it is a relative quantity. The energy of any state of a system only has meaning in relation to some other state.

By convention, energies are often measured in relation to the vacuum. When it is defined in this way, the vacuum automatically has zero energy in relation to itself. The traditional approach will not work if we want to discuss the energy of the vacuum in an absolute and significant way. We must use a different technique to measure its value.

The only way to establish an absolute measure of energy is by using gravity. In general relativity, energy is the source of gravitational fields in the same way that electric charge is the source of electric fields in the Maxwell theory of electromagnetism. An energy density of any kind, including that produced by fluctuations in the vacuum, generates a gravitational field that reveals itself as a change in the geometry of spacetime. The gravitational field of the earth, for instance, is produced by its rest energy, which equals the mass of the earth multiplied by the square of the speed of light (as given by the famous formula $E = mc^2$). The gravitational field produces a small distortion in the spacetime geometry near the earth, resulting in the attractive force that pulls us all toward the ground. In general relativity the energy density of the vacuum has an absolute meaning, and it can be determined by measuring the gravita-

UNIVERSE with a large cosmological constant would be vastly different from the existing one. Here an artist has painted a scene as it might appear if the constant were as large as theoretical estimates suggest it could be. The illustration is based on a positive value for the constant on the order of $1/(1 \text{ kilometer})^2$. With such a value the structure of space would be so distorted that the radiation from distant objects would be redshifted, or shifted toward longer wavelengths. The farther an object is from an observer, the greater the red shift would be. A spectral blue object about a kilometer away would look red; objects more than a kilometer or so away would have such large red shifts that they would be invisible. Distant objects would appear spatially distorted.
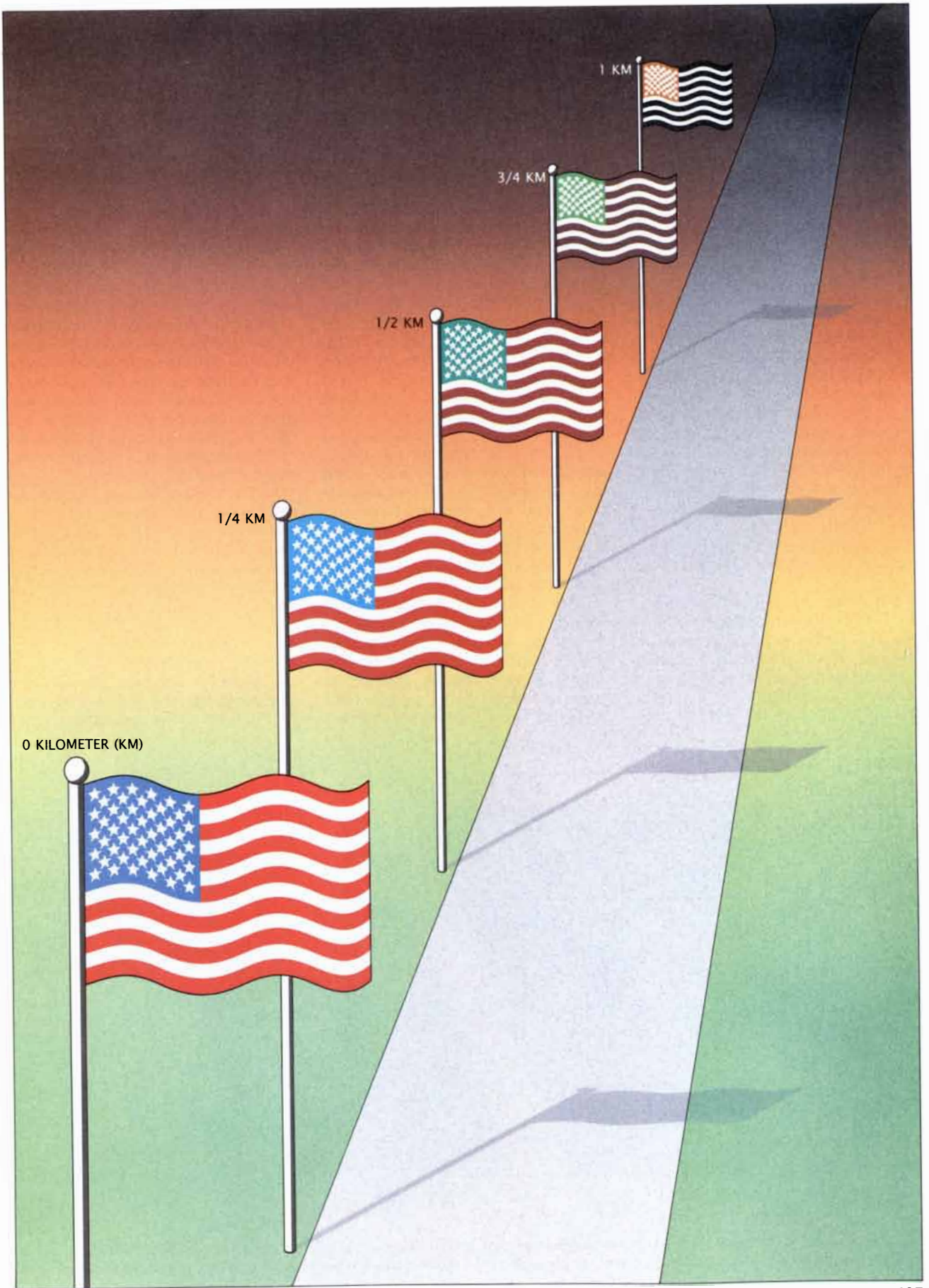
tional field produced not by matter but by the vacuum itself.

Of course, determining the energy density of the vacuum is tantamount to determining the cosmological constant, since one is proportional to the other. It turns out that the cosmological constant can be assigned units of 1 over distance squared. In other words, the square root of the reciprocal of the cosmological constant is a distance. This distance has a direct physical meaning. It is the length scale over which the gravitational effects of a nonzero vacuum energy density would have an obvious and highly visible effect on the geometry of space and time. By studying the geometric properties of the universe over length scales on the order of that distance, the value of the cosmological constant can be measured.

Physicists have been struggling with the issue of the cosmological constant for more than 70 years. The constant was first introduced by Einstein in 1917 in an attempt to eliminate two "problems" in his original formulation of the general theo-

ry of relativity. First, he thought that without a cosmological constant the general theory could not account for a homogeneous and isotropic universe: one that looks much the same everywhere. (It is remarkable that Einstein even cared about such matters in 1917, since at the time there was no evidence that the universe was homogeneous and isotropic, which indeed it is.) Unfortunately Einstein's reasoning was incorrect. In 1922 Alexander A. Friedmann showed that the general theory does allow for a homogeneous and isotropic universe, although not a static one: the universe must be expanding (or contracting). Subsequent astronomical observations have convincingly demonstrated that models based on Friedmann's work accurately describe the large-scale structure of the universe.

Einstein was also dissatisfied with his original formulation because the theory did not provide an explanation of inertia. He believed that by adding a cosmological constant he might produce a theory capable of relating the inertial properties of mat-

ter directly to the distribution of energy and momentum in the universe, in a manner first suggested by the Austrian physicist and philosopher Ernst Mach. The hope was dashed soon after Einstein's paper appeared by an argument advanced by the Dutch physicist Willem de Sitter, who discovered the spacetime we shall discuss.

After such an ignominious start it is not surprising that in 1923 Einstein wrote, perhaps somewhat bitterly, "away with the cosmological term." As we shall see, it has not been so easy to eliminate the cosmological constant—it has survived to frustrate many theoretical physicists since Einstein. George Gamow has written that Einstein felt "the introduction of the cosmological term was the biggest blunder he ever made in his life," but once introduced by Einstein "the cosmological constant...rears its ugly head again and again."

At the present time we would appear to be in an excellent position to address the issue of the cosmological constant, because we possess one of the most successful physical theories ever developed, namely the standard model. The standard model is the rather unimaginative name given to a collection of theories that successfully describes all the known elementary particles and their interactions. The remarkable ability of the standard model to interpret and predict
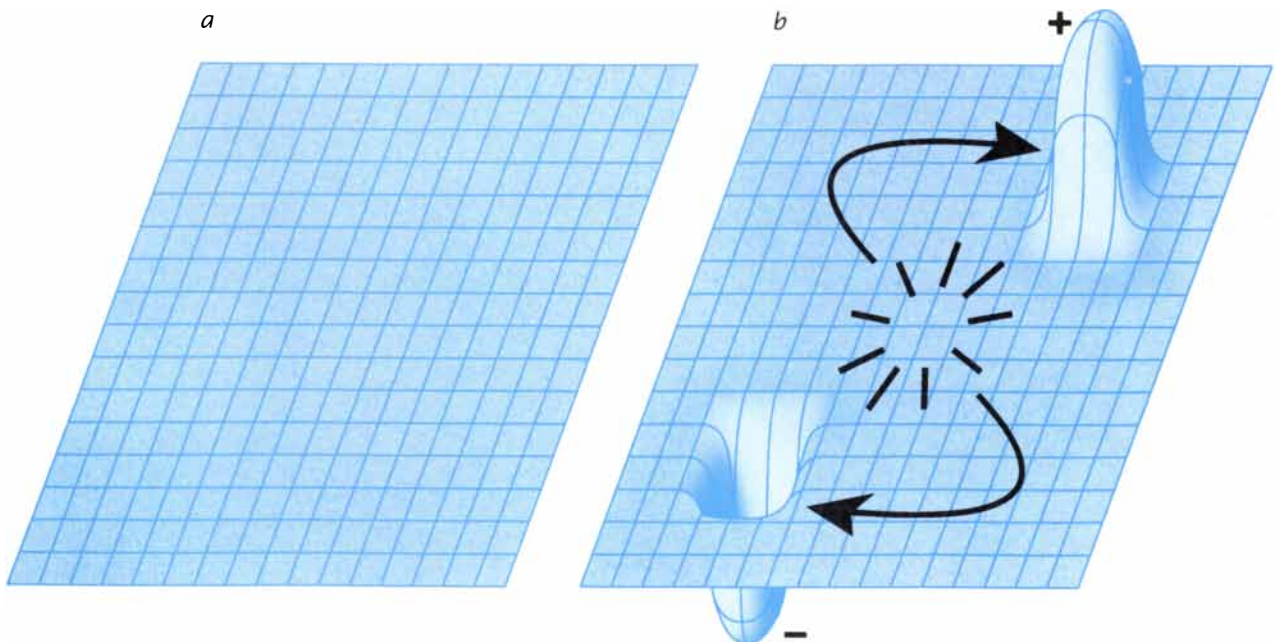
QUANTUM FLUCTUATIONS are among the phenomena that contribute to the energy density of the vacuum (*a*). According to quantum mechanics, the values of physical quantities tend to fluctuate unavoidably. As a consequence pairs of so-called virtu-

the results of an enormous range of particle-physics experiments leaves it unchallenged as a model for particle physics (at least up to the highest energies accessible to current particle accelerators).

The standard model is a quantum field theory. This means that for every distinct type of fundamental particle in nature there exists a corresponding field in the model used to describe the properties and interactions of that particle. Thus in the standard model there is an electron field, a field for the photon (the electromagnetic field) and a field for each of the known particles.

The standard model depends on a fairly large number of free parameters: numbers that must be determined by experiment and fed into the theory before definite predictions can be made. Examples of free parameters include the values of the masses of the particles and numbers characterizing the strengths of their interactions. Once the numbers have been determined the model can be used to predict the results of further experiments, and it can be tested on the basis of its predictions. In the past such tests have been spectacularly successful.

The free parameters of the standard model will play a central role in our discussion. Although the standard model is highly successful, the fact that it depends on such a large number of free parameters seriously limits its predictive power. The model, for example, predicts that an additional particle called the top quark remains to be discovered, but is unable to provide a value for its mass, because this is another free parameter of the theory. A key challenge in particle physics today is to develop a more powerful theory based on a smaller number of free parameters that nonetheless incorporates all the successes of the standard model. Such a theory would be able to determine the values of some of the parameters that cannot be predicted by the standard model. In their search for such a theory, physicists are constantly looking for relations among the parameters of the standard model that might reveal a deeper structure. As we shall see, the cosmological constant will provide us with such a relation, but in this case we shall get more than we bargained for.

In the standard model, as in any quantum field theory, the vacuum is defined as the state of lowest energy, or more properly as the state of least energy density. This does not imply that the energy density of the vacuum is zero, however. The energy density can in fact be positive, negative or zero depending on the values of various parameters in the theory. Regardless of its value, there are many complex processes that contribute to the total vacuum energy density.

In essence the total energy density of the vacuum is the sum of three types of terms. First there is the bare cosmological constant: the value the cosmological constant would have if none of the known particles existed and if the only force in the universe were gravity. The bare cosmological constant is a free parameter that can be determined only by experimentally measuring the true value of the cosmological constant.

The second type of contribution to the total energy density of the vacuum arises in part from quantum fluctuations. The fields in the standard model, such as the electron field, experience fluctuations even in the vacuum. Such fluctuations manifest themselves as pairs of so-called virtual particles, which appear spontaneously, briefly interact and then disappear. (Each pair of virtual particles consists of a particle and its corresponding antiparticle, such as the electron and the positron, which have identical masses but opposite electric charges.) Although virtual particles cannot be detected by a casual glance at empty space, they have measurable impacts on physics, and in particular they contribute to the vacuum energy density. The contribution made by vacuum fluctuations in the standard model depends in a complicated way on the
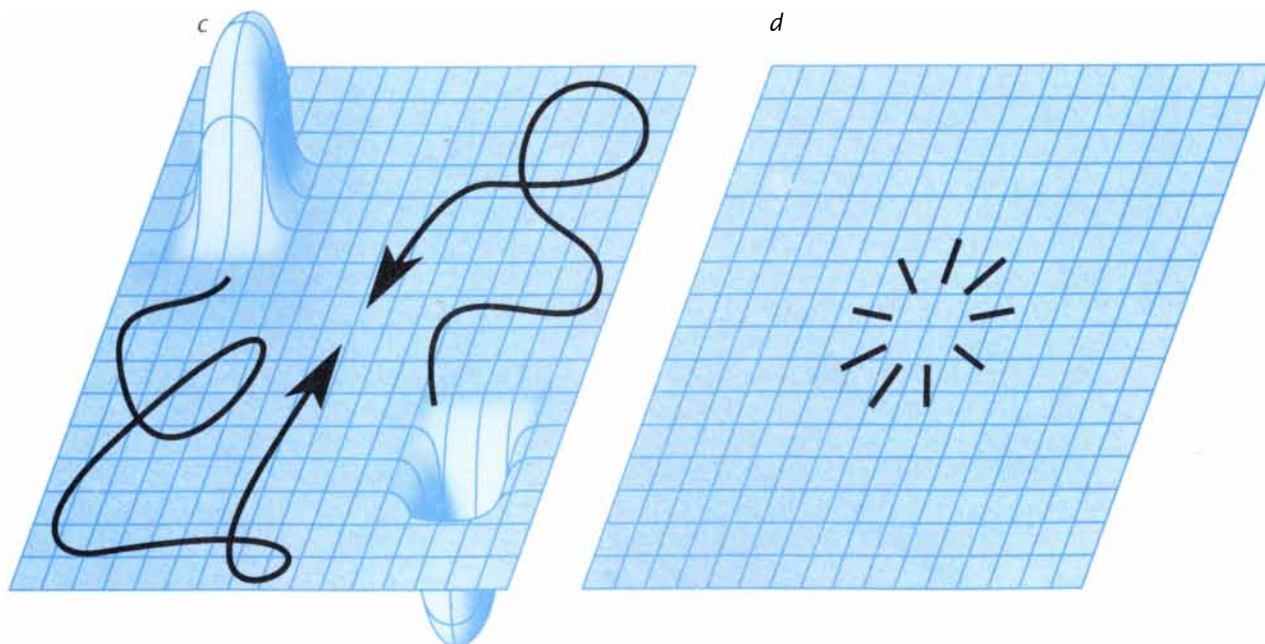
al particles can appear spontaneously in the vacuum (*b*), interact briefly (*c*) and then disappear (*d*). Here fluctuations are depicted in an abstract and highly symbolic manner. Each pair of virtual particles consists of a particle and corresponding antiparticle.

109

masses and interaction strengths of all the known particles.

The second type of term also depends on at least one additional field known as the Higgs field, which represents a massive particle, the Higgs boson, that has not yet been detected. The Higgs field should have a particularly dramatic effect on the energy density of the vacuum state [see "The Higgs Boson," by Martinus J. G. Veltman; SCIENTIFIC AMERICAN, November, 1986].

The last type of term that must be included is essentially a fudge factor representing the contributions to the vacuum energy density from additional particles and interactions that may exist but we do not yet know about. The value of this term is of course unknown.

The cosmological constant is determined by adding together the three terms we have discussed. Our ability to predict its value using the standard model is frustrated by the existence of the bare cosmological constant—a free parameter that can be determined only by carrying out the very measurement we are attempting to predict—and by the sensitivity of the vacuum energy to unknown physics. All is not lost, however, at least not yet. Although all the terms that go into making up the cosmological constant depend in a complicated way on all the parameters of the standard model, the values of many of the terms can be fairly accurately estimated. The constituents of protons and neutrons, the "up" and "down" quarks, contribute an amount of about $1/(1 \text{ kilometer})^2$ to the cosmological constant, for instance, and the Higgs field contributes an even larger amount, roughly $1/(10 \text{ centimeters})^2$.

Each of the terms that contributes to the cosmological constant depends on the parameters of the standard model in a distinct and independent way. If we assume that the parameters of the standard model are really free and independent (an assumption we are continually checking in our search for deeper structure), it seems unlikely that these apparently unrelated terms would cancel one another. As a consequence it seems reasonable to assume that the total cosmological constant will be at least as large as or larger than the individual terms we can compute. Such an argument is too crude to predict whether the cosmological constant should be positive or negative, but we would conservatively estimate that its magnitude should be at least $1/(1 \text{ kilometer})^2$, that it could well be something on the order of $1/(10 \text{ centimeters})^2$ and perhaps that it is even larger. In other words, we expect the gravitational effects of a nonzero vacuum energy density to appear as distortions in spacetime geometry over distances of one kilometer or less.

It does not require any sophisticated experimentation to show that the theoretical estimate we have just
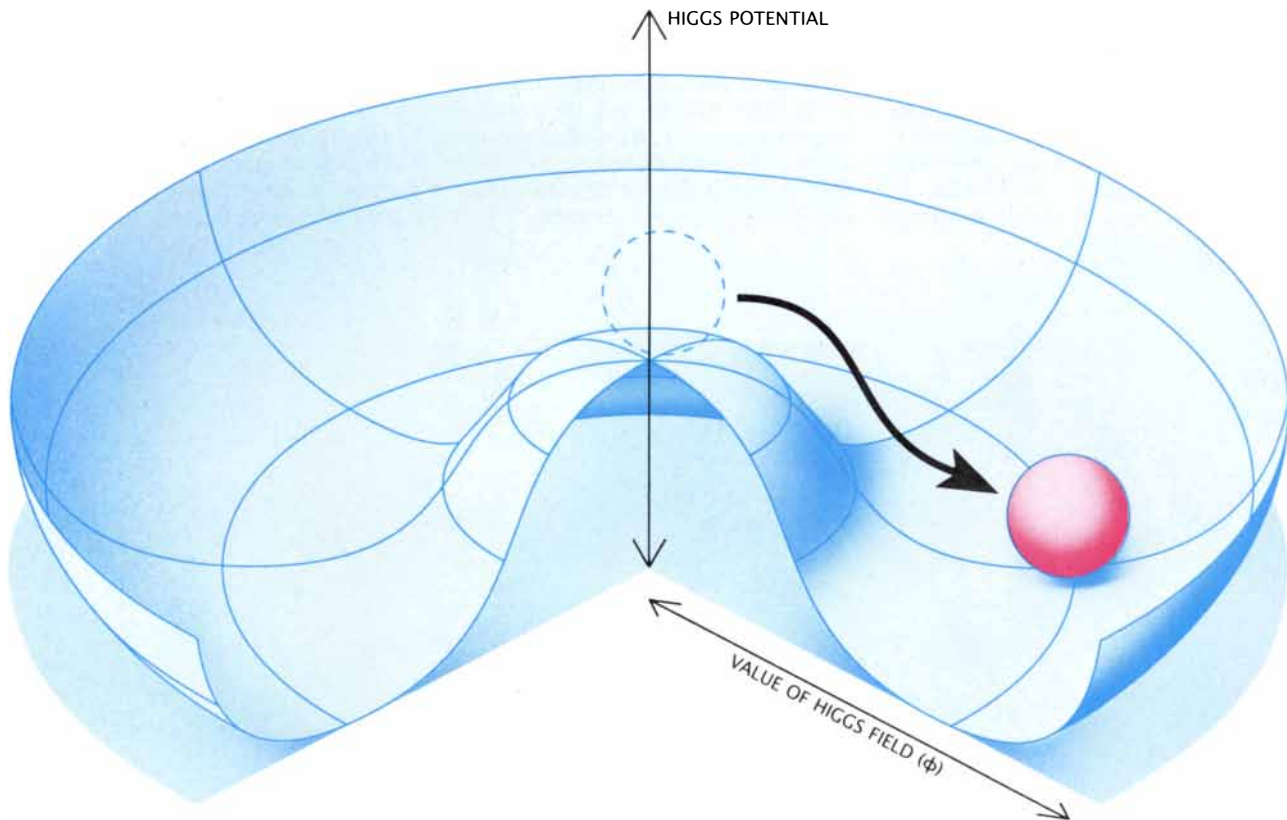
HIGGS FIELD, if it exists, would make a particularly large contribution to the energy density of the vacuum. The Higgs field is the conjectured field corresponding to the particle called the Higgs boson, which is thought to give rise to particle masses. Here the Higgs potential—the part of the vacuum energy density that depends on the value of the Higgs field—is plotted against the value of the field, ϕ. Although the Higgs potential is completely symmetric about the vertical axis, the vacuum must break the symmetry by choosing a certain position in the trough (*ball*). Such a selection is known as spontaneous symmetry breaking, and it plays a key role in the standard model: the theory that describes elementary particles and their interactions.

given is wildly wrong. We all know that ordinary Euclidean geometry provides a perfectly adequate description of space over distances much greater than one kilometer. While walking around the block none of us has ever noticed large distortions in the spacetime structure of our neighborhood. If the magnitude of the cosmological constant were as large as our standard model estimate, ordinary Euclidean geometry would not be valid over distance scales of one kilometer or even less. If the cosmological constant were negative with a magnitude of $1/(1 \text{ kilometer})^2$, then the sum of the angles of a triangle with sides on the order of one kilometer would be significantly less than 180 degrees, and the volume of a sphere of radius one kilometer would be significantly greater than $4\pi/3$ cubic kilometers.

A positive cosmological constant of order $1/(1 \text{ kilometer})^2$ would have even more bizarre consequences. If the cosmological constant were that large, we would not be able to see objects more than a few kilometers away from us owing to the tremendous distortions in spacetime structure. In addition, if we walked farther than a few kilometers away from home to see what the rest of the world looked like, the gravitational distortion of spacetime would be so great that we could never return home no matter how hard we tried.

What if the cosmological constant is nonzero but quite small? In this case we would have to look over large distances to see its effects on spacetime structure. Of course, we cannot draw triangles the size of the universe and measure their angles, but we can observe the positions and motions of distant galaxies. By carefully charting the distribution and velocities of distant galaxies, astronomers can deduce the geometric structure of the spacetime in which they exist and move.

It has long been recognized that the dominant source of gravitational distortion in the spacetime geometry of the universe at large scales appears to be the energy density of matter and not that of the vacuum. Although the energy density of matter and that of the vacuum both affect the geometric structure of the universe, they do so in different and distinguishable ways. Numerous observations have shown that the galaxies in the universe are moving away from one another, a fact that is one of the cornerstones of the expanding
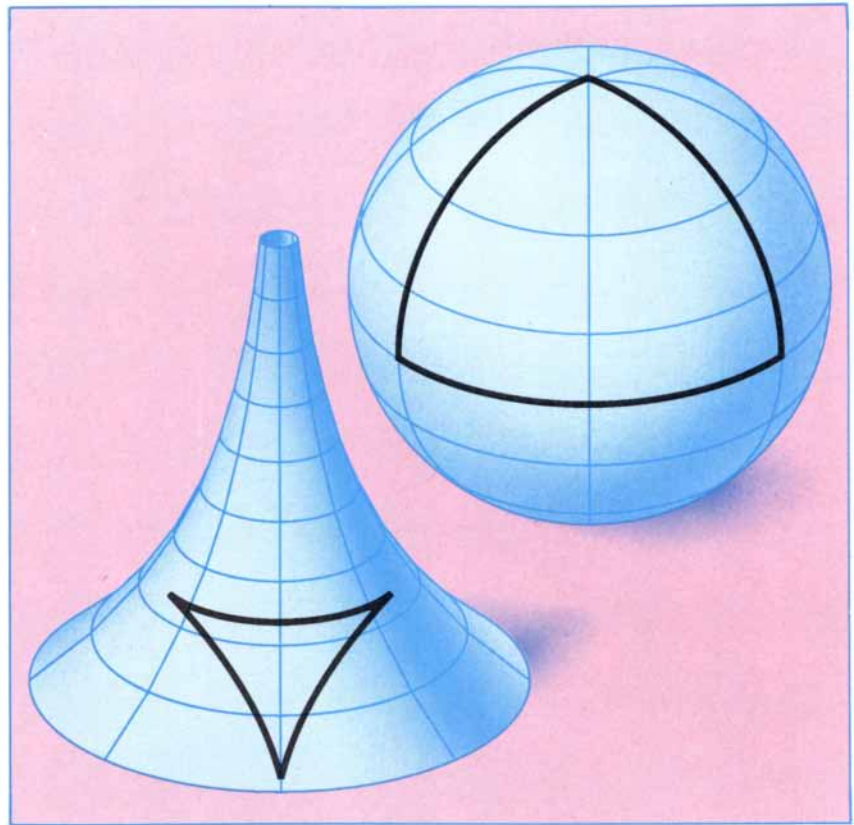
GEOMETRIC DISTORTIONS produced by a nonzero cosmological constant can affect both space and time. Here the effect on spatial geometry is shown, assuming that the distortions are independent of time. A negative cosmological constant would produce a space with negative constant curvature (*left*); a positive constant would produce positive constant curvature (*right*). (The positive case corresponds to the illustration on page 107.) In a space with negative curvature the sum of the angles of a triangle would be less than 180 degrees; with positive curvature the sum would be greater than 180.

universe in the "big bang" cosmology currently accepted. The ordinary gravitational attraction among galaxies tends to slow this expansion. As the galaxies get farther away from one another their gravitational attraction weakens, and so the rate at which the expansion slows decreases with time. Thus the effect of ordinary matter on the expansion of the universe is to decelerate the expansion at an ever decreasing rate.

What effects would a nonzero cosmological constant have on the expansion rate of the universe? A negative cosmological constant would tend to slow the expansion of the galaxies, but at a rate that is constant, not decreasing with time. A positive cosmological constant, on the other hand, would tend to make the galaxies accelerate away from one another and increase the expansion rate of the universe. Comprehensive studies of the expansion rates of distant galaxies show no evi-

dence for either a positive or a negative cosmological constant.

A good example of how astronomers can measure the geometry of the universe and look for a nonzero cosmological constant is provided by the recently published work of Edwin D. Loh and Earl J. Spillar of Princeton University. Their survey counts the numbers of galaxies in regions of a specific size at various locations in space. If we assume that on the average the number of galaxies per unit volume is the same everywhere, then by counting galaxies in a region we are estimating the volume of that region. By measuring volumes of regions far from us we are determining the relation between distance and volume over very large scales and at earlier times, since the light from distant galaxies takes a long time to reach us—billions of years in the case of this survey.

Although such surveys contain many subtle sources of potential error, the results differ so startlingly

111
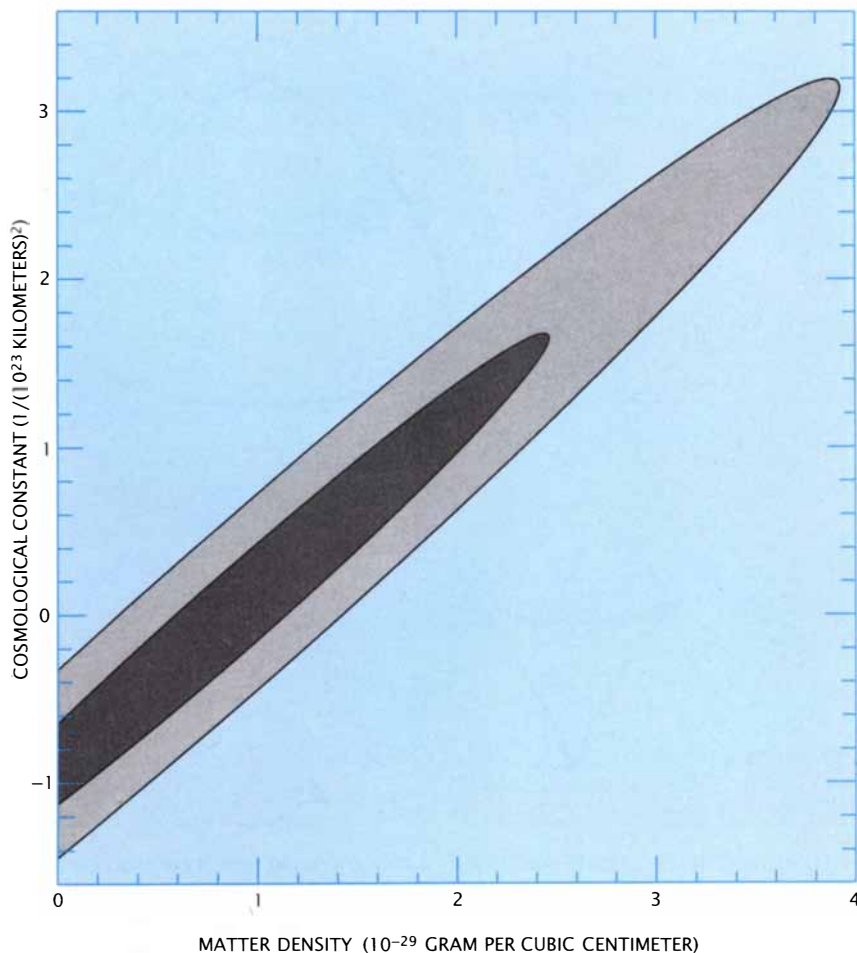
COSMOLOGICAL CONSTANT has been probed by counting the number of galaxies in regions of the universe and thereby determining the geometry of those regions. The graph plots allowed values of the cosmological constant versus the matter density of the universe. (The black area corresponds to values that are allowed with a confidence of 67 percent; the gray area is a region of 95 percent confidence.) The units are approximate, but the graph shows that the magnitude of the cosmological constant must be less than about $1/(10^{23}$ kilometers$)^2$, some 46 orders of magnitude smaller than the value predicted on the basis of the standard model. The graph is from an analysis by Edwin D. Loh of Princeton University, based on work with Earl J. Spillar, also of Princeton.

stant add up to a quantity more than 46 orders of magnitude smaller than the individual terms in the sum. In other words, the small value of the cosmological constant is telling us that a remarkably precise and totally unexpected relationship exists among all the parameters of the standard model, the bare cosmological constant and unknown physics.

A relationship among the free parameters of the standard model is just what we seek in our quest to discover deeper and more predictive theories. How could such a complex relationship among what we thought were free and unconstrained parameters arise, and what does it mean?

In answering this question it is well to keep in mind two examples from an earlier period in the history of physics. In the mid-19th century the speed of light had been measured and theories existed describing electric and magnetic phenomena, but it had not yet been shown that light propagation is an electromagnetic effect. Several physicists noticed, however, a curious relation between the speed of light and two parameters that enter into the equations for electric and magnetic phenomena. In modern notation what they noticed was that the electromagnetic permittivity constant ϵ_0 and the magnetic permeability constant μ_0 could be combined in the form $\sqrt{(1/\epsilon_0\mu_0)}$, yielding a quantity that is numerically equal to the measured velocity of light (at least within the rather large experimental errors of that time).

The workers appreciated the fact that this was either a miraculous numerical coincidence or evidence of a fundamental and as yet undiscovered relation between electromagnetic phenomena and light. James Clerk Maxwell was also aware of this numerical curiosity, and it served as an important inspiration for him in showing, through the set of equations now bearing his name, that the propagation of light is indeed profoundly related to electric and magnetic phenomena.

Does the remarkable relation among the parameters of the standard model implied by the small value of the cosmological constant suggest that a wonderful unifying theory awaits our discovery? Before jumping to such a conclusion, I should like to relate another example from the history of electromagnetic theory.

After Maxwell had incorporated light propagation into electromag-

from our theoretical estimate that errors of a factor of two or even 10 are fairly insignificant. All galactic surveys agree that there is no evidence for any spacetime distortions due to a nonvanishing cosmological constant out to the farthest distances accessible to astronomers, about 10 billion light-years, or 10^{23} kilometers. This implies that the magnitude of the cosmological constant must be smaller than $1/(10^{23}$ kilometers$)^2$. Our theoretical estimate suggesting a magnitude larger than $1/(1$ kilometer$)^2$ is incorrect by, at the least, an astonishing factor of 10^{46}. Few theoretical estimates in the history of physics made on the basis of what seemed to be reasonable assumptions have ever been so inaccurate.

The stupendous failure we have experienced in trying to predict the value of the cosmological constant is far more than a mere embarrassment. Recall that the basic assumption we used to obtain our estimate was that there are no unexpected cancellations among the various terms in the sum determining the total energy density of the vacuum. This expectation was based on the assumed independence of the free parameters of the standard model. Clearly this assumption is spectacularly wrong. There must in fact be a miraculous conspiracy occurring among both the known and the unknown parameters governing particle physics, so that the many terms making up the cosmological con-

112

netic theory it was generally assumed that light waves traveled through a medium known as the ether. Using an interferometer, Albert A. Michelson and Edward W. Morley attempted to measure the velocity of the earth as it traveled through the ether. They found that the relative velocity was zero: the velocity of the earth and the velocity of the ether were identical. This is another relationship involving what was then thought to be a fundamental parameter of nature, namely the velocity of the ether. Did the discovery point the way to a unified theory relating a fundamental property of electromagnetism to the motion of the earth?

Although the idea that the ether drifted with the earth was suggested, the zero result of the Michelson-Morley experiment is actually explained by Einstein's special theory of relativity, which showed that the conception of the ether being used in that era was inconsistent with the symmetries of space and time. No theory providing a fundamental relation between the velocity of the ether and something as idiosyncratic as the velocity of the earth has survived. That is hardly surprising. The velocity of the earth is affected by many things—the shape and size of its orbit around the sun, the mass of the sun and the motion of the sun in the galaxy, for instance—that seem completely unrelated to issues in the theory of electromagnetism. There is no fundamental relation between the velocity of the ether and the velocity of the earth because the ether itself as the 19th-century theorists imagined it does not even exist.

In both examples a surprising relation between parameters of nature foreshadowed dramatic and revolutionary new discoveries. We have every reason to believe the mysterious relation implied by the vanishingly small value of the cosmological constant indicates that discoveries as important as these remain to be made. The two examples we have considered are quite different. The first relation, which involves two parameters of electromagnetism and one from light propagation, is what physicists today would call a "natural" relation: one that involves a small number of well-known parameters. The existence of a natural relation may indicate that a unifying theory exists, and, more important, it suggests that such a theory can be discovered.

The second example, in which the velocity of the ether was related to the velocity of the earth, is what today would be called an "unnatural" relation: one that involves many parameters, some of which are unknown or even unknowable. It seems unlikely, for instance, that we will ever know and understand all the many factors that determine what the velocity of the earth is in relation to the distant galaxies. Any unified theory developed to account for an unnatural relation would have to explain the values of many known and unknown parameters all at once. It seems quite unlikely that such a theory could be discovered even if it did exist.

Our example indicates that an unnatural relation suggests a deep misunderstanding about the essence of what is being measured and related, rather than the existence of an underlying unified theory. As a consequence an unnatural relation may point to an even more dramatic revolution in our thinking than a natural one would.

If we discount the possibility that the vanishingly small value of the cosmological constant is accidental, we must accept that it has profound implications for physics. Before we launch into constructing new unified models, however, we must face the dilemma that the relation implied by the vanishing of the cosmological constant is unnatural. The miraculous cancellations required to produce an acceptably small cosmological constant depend on all the parameters relevant to particle physics, known and unknown. To predict a zero (or small) value for the cosmological constant, a unified theory would face the imposing task of accounting for every parameter affecting particle physics. Even worse, achieving a sufficiently small cosmological constant requires that extremely precise (one part in 10^{46} or more) cancellations take place; the parameters would have to be predicted by the theory with extraordinary accuracy before any improvement in the situation regarding the cosmological constant would even be noticeable. Constructing such a theory, even if it does exist, seems to be an awesome if not impossible task.

Although certain theories of the "ether drift" variety have been proposed, most efforts concerning the cosmological constant now focus on finding the underlying misunderstanding, the missing piece of the standard model or the misconception about the vacuum, which once understood will either eliminate the problem or at least turn it into a natural one. As long as the problem of the cosmological constant remains unnatural, the only hope we have for finding a solution is to stumble on an all-encompassing theory capable of accounting for all particle-physics parameters with nearly perfect accuracy. If we can change the relation required to produce an acceptably small vacuum energy density into a natural one, then, even though we have not yet accounted for its value, we at least reduce the issue of the cosmological constant to a more manageable problem involving a reasonable number of known parameters that only have to be predicted with a moderate degree of accuracy. There is little to report to date about this effort. In spite of a lot of hard work and creative ideas we still do not know why the cosmological constant is so small.

Even though nature does not, in the words of Aristotle, "abhor a vacuum," perhaps it does abhor a vacuum that is not empty. By introducing the ether in the early days of electromagnetic theory, Maxwell and others cluttered the vacuum with a hypothetical fluid that had complex properties. Michelson and Morley showed that this view of the vacuum was inconsistent with experimental reality, and Einstein showed that it was inconsistent with the symmetries of the universe.

Quantum field theories also fill the emptiness of the vacuum, this time with quantum fluctuations and fields rather than ether. These modern forms of clutter are consistent with the special theory of relativity, but they seem to cause problems when they are viewed in the framework of the general theory. With the mystery of the cosmological constant, perhaps we are again paying the price for dumping too much into the vacuum. The standard model, which has a large number of fluctuating quantum fields including a Higgs field, is a particularly egregious polluter of the vacuum. There is no doubt that the resulting theory is a beautiful and highly successful structure, but it may be based on a conception of the vacuum or of spacetime that is flawed. It is our challenge to repair that faulty foundation without destroying the towering edifice we have built on it.

113