Dark Matter in the Universe

(from: "Particle Physics in the Cosmos," a collection of readings from Scientific American Magazine.)

1

More matter exists than is seen. The motions of stars and galaxies indicate where some of it is; theory suggests there is far more. What and where is it? Particle physics and astrophysics are yielding clues.

• • •

Lawrence M. Krauss

What is the universe made of? What kind of matter is commonest, how much is there and how is it distributed? These questions, always a focus of cosmology, have become even more intriguing over the past few years as evidence has piled up to support the proposition that most of the mass in the universe is dark—invisible to any existing telescope or other observational device—and new developments in both high-energy physics and astrophysics have made possible new predictions of the makeup and distribution of this possibly exotic form of matter.

There is already overwhelming evidence that the visible matter within galaxies may account for less than 10 percent of the galaxies' actual mass: the rest, not yet directly detectable by observers on the earth, is probably distributed within and around each galaxy. Theoretical considerations now suggest this may be only the tip of the cosmic "iceberg" of dark matter: much greater amounts of dark matter may be distributed throughout the universe, perhaps in configurations entirely independent of the distribution of galaxies. It may be that this mass can be accounted for only by the existence of new kinds of matter.

The question of dark matter—how much of it there is, how it is distributed and what it is made of—is intimately linked to questions about the overall structure and evolution of the universe: because dark matter is probably the dominant form of mass in the universe, it must have affected the evolution of the features observable today. Questions of structure in turn depend for their answers on a deep bond that has formed between macrophysics and microphysics, the bodies of knowledge that respectively describe interactions on the largest scale (that of the universe as a whole) and the smallest scale (that of the fundamental particles that make up all matter).

This bond is provided by the observation that the universe is expanding. If we are bold enough to extrapolate the expansion backward by between 10 and 20 billion years, the cosmological and microscopic scales begin to merge, because at the earliest times those structures now observed on the largest scales occupied regions having characteristic distances and energies on scales that are typically associated with the processes governing the interactions of fundamental particles. Since the structure remaining on the largest scales observable today reflects the imprint of those processes, it is natural to expect the resolution of the dark-matter question to come in part from advances in the understanding of the physics of high-energy particles.

At present a number of testable predictions for the nature of both the dark matter and the primordial structures in the early universe have been proposed. Future developments, both theoretical and observational, will help to decide issues ranging from how and when galaxies and stars first formed to what kinds of symmetries underlie the interactions of particles at very high energies. Ultimately the debate about dark matter may help to answer a question as old as human inquiry: What will be the fate of the universe?

Ever since the early 1930's, when Edwin P. Hubble confirmed that the universe is expanding, it has been natural to ask whether the expansion will eventually halt. The answer depends on two factors: how fast the universe is currently expanding and how strongly the force of gravity, determined by the average density of mass within the universe, holds that mass together. A high mass density would cause a strong gravitational attraction.

According to the general theory of relativity, there is a relation between the magnitudes of these two factors and the mean curvature of the universe. If the average mass density is so small compared with the expansion rate that the universe will continue to expand at a finite rate forever, the universe is said to be open. If the density is high enough to halt the expansion and cause the universe to contract again, the universe is said to be closed. If the gravitational attraction is precisely strong enough to continue to slow the expansion but not strong enough to close the universe, the universe is said to be flat. The shape of space will affect the shape of geometric objects and the reference length, which is the difference between any two regions of the expanding universe (see Figure 1). Over small distances, such as on the earth, these measures would not be noticeable. Strong theoretical arguments support the proposition that the universe is actually flat, even though in order to be flat it would have to contain much more mass than has yet been observed, either directly or indirectly.

Because the observable universe is highly uniform in all directions, its rate of expansion can be described in terms of a single parameter, which is known as the Hubble constant even though it is actually a slowly varying function of time. The Hubble constant is the average speed with which any two regions of the universe are moving apart from each other divided by the distance between them (see Chapter 2, Figure 7).

For any given measurement of the Hubble constant, it is easy to determine the mass density that would correspond to a flat universe. Measurements of the Hubble constant, however, depend on a variety of uncertain measurements. The Hubble constant is generally determined by measuring the velocity at which various objects are receding from the earth and gauging their distance by such techniques as estimating their intrinsic brightness and comparing that with their brightness as seen from the earth.

Because those measurements are highly uncertain, there is a spread of about a factor of two in current determinations of the universe's rate of expansion. As an upper limit, objects one megaparsec (about 3.26 million light-years) apart are on the average receding from one another at a speed somewhat less than about 100 kilometers per second. At that rate the average mass density that would result in a flat universe is about 2×10^{-29} gram per cubic centimeter, which is roughly equivalent to the mass of 10 hydrogen atoms per cubic meter of space.

How is it possible to determine how much mass actually exists? One method for finding at least a lower limit is simply to add up the total amount of visible matter. Since what can be measured directly is not mass but luminosity, some amount of interpretation is necessary in translating observations into putative mass densities. When the observed distribution and luminosity of stellar objects and diffuse gas are taken in combination with theoretical estimates of their masses, it seems that the mass-to-luminosity ratio of the luminous matter associated with galaxies is a few times the mass-to-luminosity ratio of the sun. Given this estimate and estimated lower limits on the Hubble constant, the average density of luminous matter in the universe is less than about 2 percent of the density needed to halt the universe's expansion.

It has been known since as early as 1933, however, that clusters of galaxies may contain a significant proportion of nonluminous mass. In that year Fritz Zwicky of the California Institute of Technology was analyzing the individual velocities of galaxies within the Coma cluster. He found many galaxies were moving so quickly that the cluster as a whole should tend to fly apart unless there was more mass to hold it together than the luminous mass alone. Other evidence indicated the cluster was stable, and so Zwicky concluded that the cluster must contain nonluminous matter.

Zwicky set an important precedent by showing

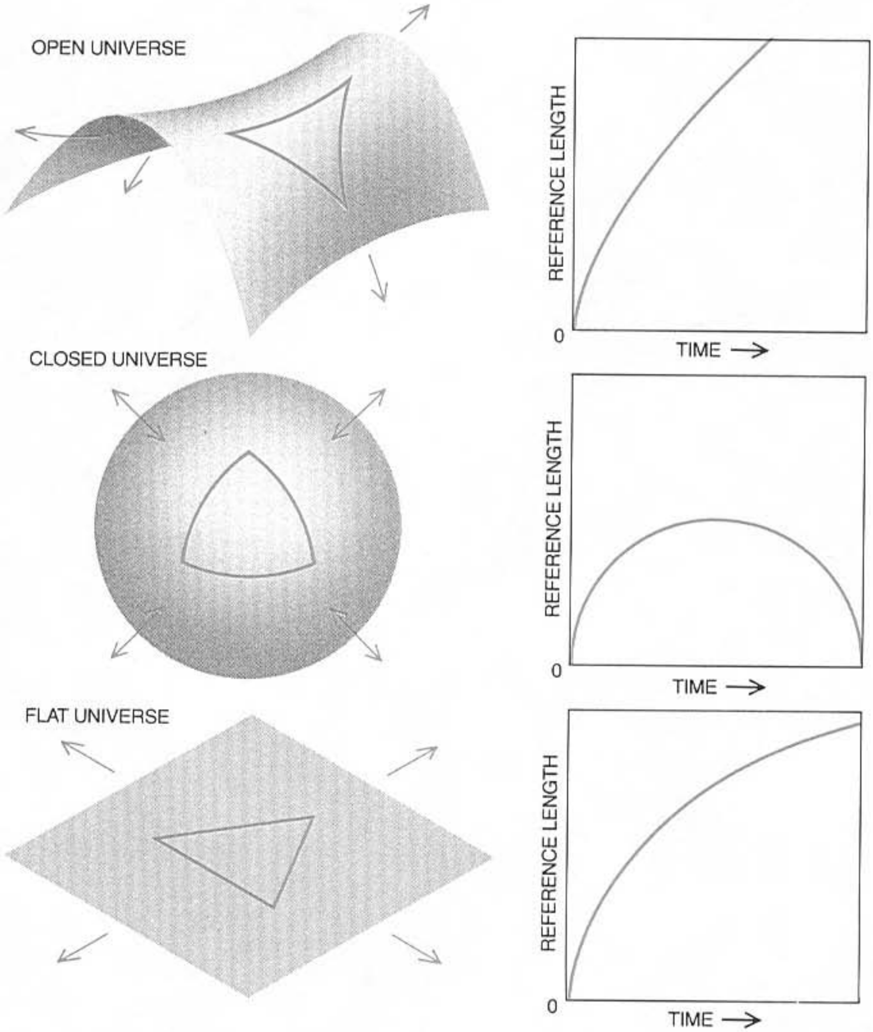

Figure 1 CURVATURE OF THE UNIVERSE. In an open universe geometry is analogous to the surface of a saddle, a triangle has an angle-sum of less than 180 degrees and a reference length between regions continues to increase. A closed universe closes in much the same way as the surface of a sphere, the angles of a triangle add up to more than 180 degrees and a reference length eventually begins to decrease. In a flat universe geometry is analogous to a plane, the sum of the angles in a triangle is exactly 180 degrees and expansion continues but slows asymptotically.

that dark matter can in principle be detected indirectly by its gravitational effects. In recent years investigators have shown convincingly that similar techniques can detect the presence of dark matter in structures on scales ranging from the immediate solar neighborhood through galaxies and clusters of galaxies to superclusters made up of thousands of galaxies.

The best-documented evidence for the presence of dark matter is based on the velocities of rotation of spiral galaxies [see "Dark Matter in Spiral Galaxies," by Vera C. Rubin; SCIENTIFIC AMERICAN, June, 1983]. The Doppler frequency shift makes it possible to determine how quickly a light-emitting object is moving toward or away from an observer and how fast the arms of a spiral galaxy are rotating. A stellar object emits light at characteristic frequencies determined by its composition. If the object is mov-

Figure 2 SPIRAL GALAXY M31 (ANDROMEDA) reveals the presence of dark matter by the motion of its outer arms, which rotate about the galactic center faster than they would be expected to if the galaxy's visible, luminous matter represented most of its mass. (Palomar Observatory Photograph.)

ing away from the observer, the wavelength of the observed light appears to be lengthened. This is called a red shift, because longer-wavelength light is redder. Nearly all galaxies are moving away from the earth because the universe is expanding. To an observer on earth the wavelength of the light from the spiral galaxy therefore appears to be lengthened, or red-shifted (see Figure 3). By comparing the red shifts of the galactic center and the arms, the rate of rotation of any part of either arm can be determined. It is then possible to infer the distribution of mass in the galaxy.

The velocity of rotation of an object in a stable, gravitationally bound system, such as a spiral galaxy, depends in part on its distance from the center of rotation. According to Newton's laws, the orbital velocity of objects far from a central concentration of mass should drop off in proportion to the reciprocal of the square root of their distance from the center of rotation. In extensive surveys of stars and hot gas in the outer regions of spiral galaxies, several groups have shown that the rotational velocities of these objects remain constant, rather than dropping off, out to distances greater than 30 kiloparsecs from the galactic core. It had already been suggested by Jeremiah P. Ostriker and P. James E. Peebles of Princeton University that there must be some unseen mass in spiral galaxies, because otherwise gravitational instabilities would cause the galaxies to collapse into barshaped formations. The stability of spiral galaxies, as well as the rates of rotation of their outer arms, could be explained if the galaxies were each embedded in a large, roughly spherical distribution of dark matter.

There is other dynamical evidence for dark matter, on scales both larger and smaller than the scale of individual galaxies. The evidence is obtained not from measurements of rotational velocities but from measurements of the random individual velocities of objects within gravitationally bound systems. A well-known theorem of classical mechanics called the virial theorem establishes a

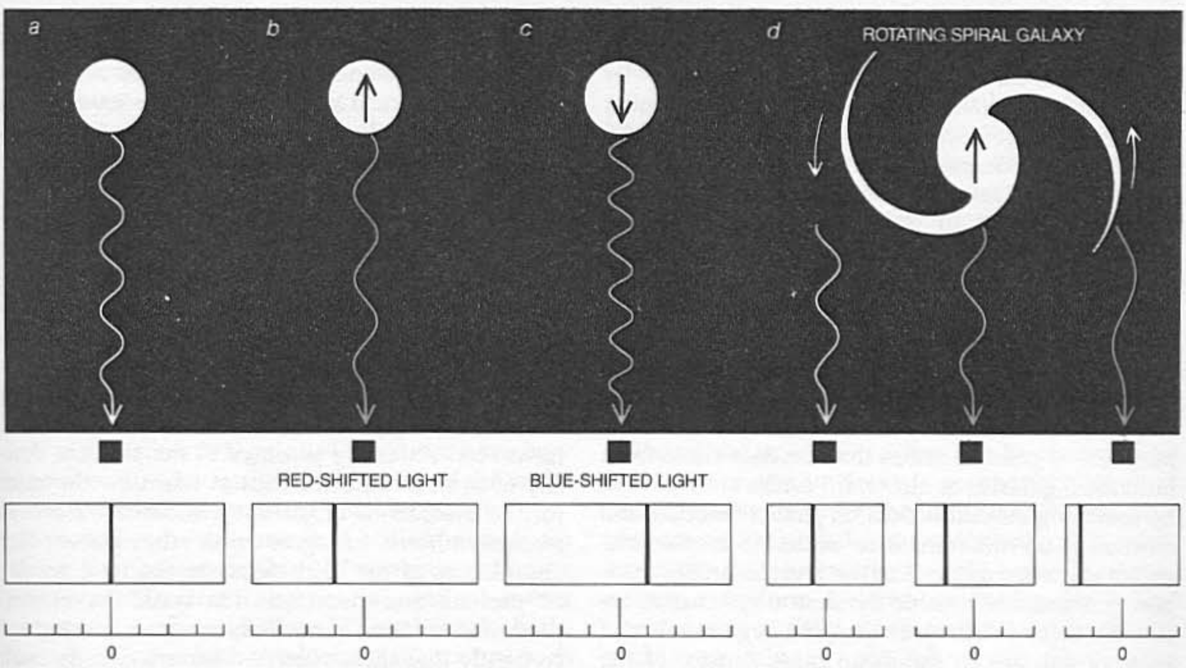

Figure 3 DOPPLER FREQUENCY SHIFT. A stellar object emits light at frequencies determined by its composition (*a*). If the object is moving away from the observer (*b*), the wavelength of the observed light is lengthened, or red-shifted. If the object is moving toward the observer (*c*), the wavelength of the light is shortened, or blue-shifted. The light from the center of a galaxy that is moving away from the earth is red-shifted (*d, center*). One arm of the spinning galaxy (*d, left*) will not be moving away from the earth as quickly as the galactic center, so its light will be less red-shifted. The other arm will be moving away more quickly (*d, right*), so its light will be even more red-shifted.

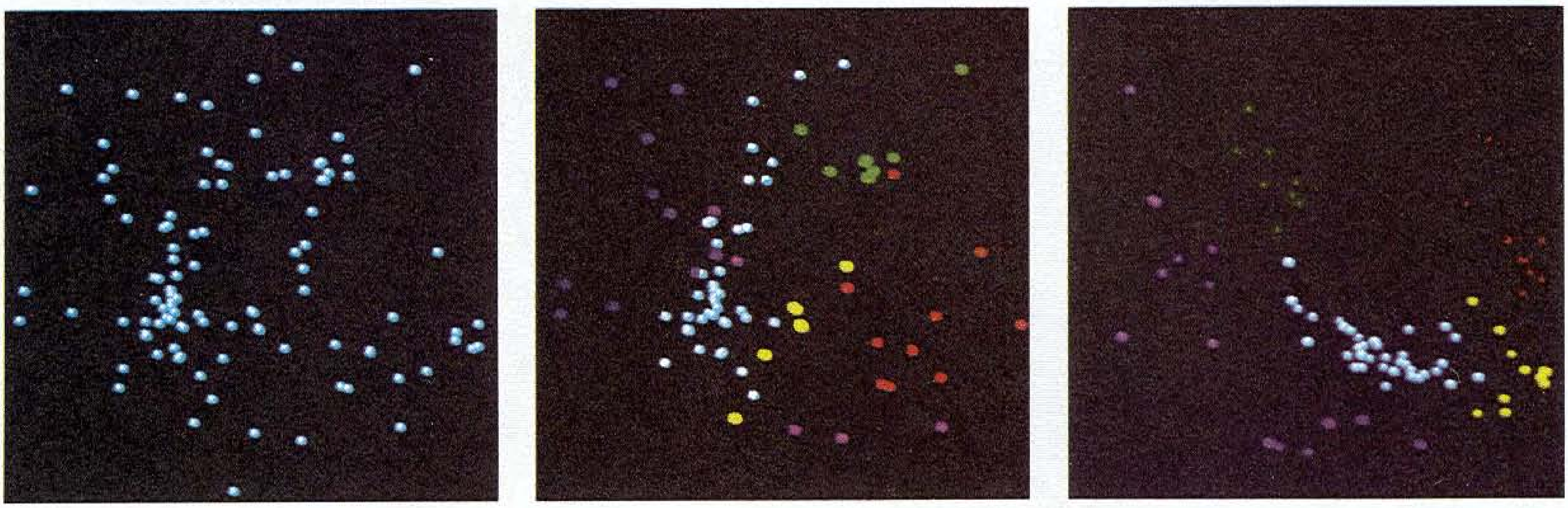

Figure 4 CLUSTER OF GALAXIES IN CANCER, as these computer-generated views by Michael J. Kurtz show, is not a single dynamical system. From the earth (*left*) the cluster seems to be a roughly spherical system in apparent equilibrium. Analysis showed that the cluster is made up of several groups of galaxies separated in space (*center*, *colors*). A rotated view (*right*) shows the separation of the various groups more clearly. Within each group the relative velocities of galaxies are much lower than the relative velocities of the group, indicating there is less mass in the system as a whole than previously estimated.

relation between the average kinetic and gravitational potential energies of objects in stable, gravitationally bound systems that have reached dynamical equilibrium. It should therefore be possible to estimate the total mass of such a system (which is related to its total gravitational potential energy) by measuring the relative velocities of a large number of pairs of objects within the system. This method has yielded evidence of dark matter in a wide variety of systems, ranging from dwarf spheroidal galaxies as small as 10^7 solar masses to clusters of galaxies as large as 10^{15} solar masses. On the largest scales probed by this kind of analysis (regions within roughly a megaparsec of galaxies) the average mass densities are no larger than about 20 percent of the density needed to close the universe.

Another method, pioneered by Peebles and his co-workers, relies on statistical analysis of large numbers of galaxies rather than on data taken from individual galaxies or clusters. Peebles showed that by amassing statistical data on galactic motion and clustering on different size scales it is possible, under the assumption that the regions probed contain gravitationally stable dynamical systems, to relate the mean relative velocity of a large number of pairs of galaxies to the mean mass density of the universe.

It is striking that all the available methods, including those I have discussed and several I have not mentioned, yield essentially the same result: if the distribution of galaxies traces the distribution of mass in the universe, then the universe contains less than about 20 to 30 percent of the mean mass density that would be necessary for closure.

Even if galaxies are not good tracers of mass, or if somehow all the analyses have involved systematic errors, there is still good reason to believe that at any rate the total amount of ordinary mass (mass consisting mainly of protons and neutrons) in the universe accounts for no more than about 20 percent of the amount that would be required for closure. The evidence comes for the most part from the theoretical framework that explains the process of nucleosynthesis, in which various cosmically abundant light elements and isotopes were first formed.

Nucleosynthesis of light elements occurred primarily in the first few minutes of the universe's existence. The process of nucleosynthesis would have been extremely sensitive to the absolute density of protons and neutrons at that time. In order for the predictions of current theoretical models of nucleosynthesis to agree with the present-day abundances of the light elements, the total density of protons and neutrons that could have been present at the time of nucleosynthesis is constrained so tightly that these particles' current density must be less than about 20 percent of the density required for closure. Thus it seems that if the universe is closed, at least 80 percent of the total mass in it is made up of some other kind of matter.

Since such fundamental theoretical arguments limit the amount of normal mass in the universe to 20 percent of the critical density, and since observational evidence suggests that the mass density associated with galaxies and clusters of galaxies is about that amount, why should cosmologists not assume the universe is in fact open? It is by no means impossible to imagine a form in which enough normal matter to explain the dynamics of galaxies and clusters could remain unseen. Why, then, is there a need to postulate any other form of mass? Why is there a larger dark-matter problem?

Two theoretical barriers stand in the way of the simple assumption that most or all of the mass in the universe is composed of normal matter and that the mean density is only 20 percent of the critical amount. The first barrier is set by a combination of the theory of galaxy formation and observations of the background of microwave radiation that pervades the cosmos.

It is generally assumed that galaxies eventually formed when regions of the early universe that were denser than the average condensed under the force of gravity until they separated from the background expansion to form isolated bound systems. For roughly 100,000 years after the big bang, ordinary matter could not condense in this way. Ordinary matter was still too hot for its constituent particles to have combined into electrically neutral atoms, and so it consisted of independent charged particles. Because ordinary matter was ionized in this way, its microscopic motion was strongly influenced by background fields of electromagnetic radiation: matter and radiation were coupled. Regions of ordinary matter that were denser than surrounding regions and smaller than the horizon size (the distance a light ray could have traveled since the big bang, and therefore the maximum distance over which physical systems could be in causal contact; see Chapter 3, Figure 18) could not have condensed further, because the "pressure" of the radiation combated the attracting force of gravity.

Eventually the universe had cooled enough for oppositely charged particles to combine, rendering normal matter electrically neutral, and so matter decoupled from radiation. The thermal background-radiation bath to which the matter had been coupled was then free to cool as the universe expanded, and it now constitutes the well-known cosmic microwave background radiation, which fills the universe. Observations have shown that this background radiation is isotropic—the same in all directions—to within a very high degree of accuracy.

Since gravity is a universally attractive force, any initial fluctuations, or small variations, in the density of ordinary matter in the early universe would have tended to grow after the force of radiation pressure no longer acted against the force of gravity. Thus it is presumed that the universe became (and is becoming) clumpier with time and that galaxies, whose cores now have densities more than one million times the average background density, began in fluctuations whose densities were much closer to the background value.

How large were the initial fluctuations? Because of the limited data currently available on large-scale structures, and because of the mathematical difficulties inherent in describing analytically the evolution of systems as dense as galaxies, it is extremely difficult to work backward from the current state of the universe to determine the precise nature of the initial fluctuations. An easier approach is to assume some initial pattern of fluctuations, simulate the growth and evolution of that pattern and compare the result with present-day observations. In this approach the cosmologist is guided by both lower and upper limits on the size and nature of the initial fluctuations. First, they must have been extreme enough (that is, the ratio between the local overdensity in the region of the fluctuation and the average density in space must have been large enough) for fluctuations on the scale corresponding to galactic sizes to have condensed to form galaxies by today. Second, the fluctuations must have been of small enough amplitude for them not to have left an anisotropy in the background radiation larger than the measured upper limit.

These two conditions appear to be mutually inconsistent if the universe is composed mainly of normal matter. Between the time when normal matter became decoupled from radiation and the time when the fluctuations that would become galaxies collapsed to form isolated, gravitationally bound systems, the initially small fluctuations in density could grow only at a well-defined rate. Fluctuations large enough to have had sufficient time to form self-bound systems would have led to an anisotropy on the background radiation more than an order of magnitude greater than the observational upper bounds. In other words, there has not been enough time, since decoupling, for galaxies to form gravitationally from variations in density small enough not to have left observable traces in the background radiation.

This conclusion depends on two widely held assumptions, namely that the microwave background has not been significantly disturbed since the time of decoupling and that gravity alone led to the formation of galaxies. Unless either of these standard assumptions is false (as various investigators have suggested), it appears that some new form of matter is necessary, one that could have begun to condense gravitationally earlier than normal matter could have.

There is a second and more fundamental reason to suppose the universe is not dominated by normal matter having a density of only about 20 percent of the critical density. This reason, now called the flatness problem, was first pointed out by R. H. Dicke of Princeton and Peebles. The essential point is that any deviation from an exactly flat universe should tend to increase linearly with time. If the universe had had even a small nonzero curvature at the time of nucleosynthesis, the deviation from flatness would by today have increased by a factor of about 10^{12}. Since the mass density in the present-day universe is within a factor of 10 of the mass density of a closed universe (in other words, since the universe is relatively close to being flat), at nucleosynthesis the universe must have been either exactly flat or curved to an extremely small degree: it must have been flat to an accuracy of within one part in a million million.

If the universe is measurably curved today, cosmologists must accept the miraculous fact that this is so for the first time in the 10^{10}-year history of the universe; if it had been measurably nonflat at much earlier times, it would be much more obviously curved today than it is. This line of reasoning suggests that the observable universe is essentially exactly flat: that it contains precisely the critical density of mass. Since normal matter probably accounts for only 20 to 30 percent of the critical density, some form of more exotic matter is probably present.

The next logical question is: Why is the universe exactly flat? In 1980 Alan H. Guth, now at the Massachusetts Institute of Technology, proposed an answer. It took the form of a model of the evolution of the early universe based on ideas in particle physics that had only recently been proposed.

Guth drew on the work of Howard Georgi and Sheldon Lee Glashow of Harvard University. In 1974 the two investigators proposed that three of the fundamental forces of nature—the so-called strong, weak and electromagnetic forces—are different aspects of a single, "unified" force. At sufficiently high energies the three forces should be exactly symmetrical: they should behave identically. At energies comparable to those observed now on the earth, on the other hand, the three forces can behave quite differently (see Chapter 4, "A Unified Theory of Elementary particles and Forces," by Howard Georgi). The temperature of the early universe, soon after the big bang, was initially high enough for the symmetry of the three forces to be manifest. As the universe cooled below the critical energy at which the symmetries relating the forces can be maintained, the preferred configuration of the universe became one in which the symmetry was "broken." The effect of this symmetry breaking was that the forces appeared distinct from one another.

(A simple example of this type of behavior is found in ferromagnets. At sufficiently high temperatures a piece of iron is not magnetized: the spins of all the electrons, each of which causes a small magnetic field, point in random, different directions. Below a certain critical temperature, however, it may be energetically more favorable for all the spins to point in one direction, aligning their magnetic fields and creating a permanent magnet. The direction of the magnetic field in the magnet represents a unique direction, and so the symmetry of the former configuration, in which no direction was special, is broken.)

According to Guth's idea, which was later extended by Andrei D. Linde of the P. N. Lebedev Physical Institute in Moscow and by Paul J. Steinhardt and Andreas Albrecht of the University of Pennsylvania, the abrupt breaking of symmetry could have caused the universe to "inflate" rapidly: the universe could have expanded exponentially, growing by more than 28 orders of magnitude in less than 10^{-30} second. After the period of rapid inflation the universe could have reverted to its normal, nonexponential expansion, which is observed today (see Chapter 11, "The Inflationary Universe," by Alan H. Guth and Paul J. Steinhardt).

It is the rapid inflation of the universe, according to this model, that caused the observable regions of space to become flat, in much the same way as inflating a balloon makes its surface appear flatter; after inflation the part of the universe observed today would necessarily appear flat.

In addition to its resolution of the flatness problem, the inflationary-universe scenario is remarkably successful in other ways. In particular, it is the only model consistently tying the initial conditions that caused the universe's expansion to the laws of microphysics. The inflationary model also makes it possible to calculate, from first principles, quantities whose values had previously been assumed or inferred. For example, the model remarkably predicts the shape of the spectrum of primordial density fluctuations (the functional relation between the amplitude of fluctuations and their scale size) to be precisely the shape that had been suggested earlier on phenomenological grounds. The wide acceptance by many cosmologists of the predictions of the inflationary-universe model indicates the deep impact particle theory is having on modern cosmology.

In solving the flatness problem, the inflationary model makes the dark-matter problem more urgent. If the universe is flat, then most of the mass in the universe is probably not normal matter, and most of it has not yet been detected in any way, even indirectly.

What might this exotic, undetected matter be made of? One of the earliest proposals was that the dark matter is composed of neutrinos. First postulated in order to solve problems involving the conservation of energy and momentum in nuclear decay, neutrinos interact very weakly with normal matter and are thus extremely difficult to detect. Nevertheless, three kinds of neutrino, called the electron neutrino, the muon neutrino and the tau neutrino, have now been found experimentally. It was originally proposed that neutrinos were massless, but there is no theoretical reason for supposing they might not have some mass. Stringent experimental limits have nonetheless been set on the maximum possible neutrino mass, and it is very small indeed. The strongest constraint is on the electron neutrino, which must have a mass less than about 10,000 times smaller than the mass of the electron.

As dark-matter candidates, neutrinos have two strong advantages over other contenders. First of all, they are known to exist. Second, the calculations that have been so successful in describing primordial nucleosynthesis also suggest that light neutrinos must be abundant in the universe today. When big-bang nucleosynthesis started, at temperatures greater than 10^{10} degrees Kelvin (degrees Celsius above absolute zero), light neutrinos were kept in thermal equilibrium with matter by the weak interaction and were therefore as abundant as photons. Thus, as R. Cowsik of the Tata Institute of Fundamental Research in India and J. McLelland of the University of Melbourne first estimated, if neutrinos have approximately the same present-day density as the photons that make up the background radiation, and if they have a mass in the range of one ten-thousandth to one hundred-thousandth the mass of the electron, they could account for enough mass to close the universe. (The estimate was later confirmed by more detailed calculations.)

This point became particularly relevant in 1980 when V. A. Lubimov and his collaborators at the Institute of Theoretical and Experimental Physics in Moscow announced they had found evidence that the electron neutrino has a mass within that range. On the basis of this result it seemed neutrinos were ideal candidates to be the dominant mass in the universe. Since then, however, the likelihood that light neutrinos are the dark matter has become much smaller. In the first place, there are many outstanding experimental questions about the Soviet result; as a matter of fact, a recent finding by a group at the Swiss Institute for Nuclear Research appears to contradict it. In addition a great deal of work by astrophysicists has shown that theoretical pictures of a universe dominated by light neutrinos are not as compatible with observation as it once seemed.

The first such theoretical evidence came in 1979 from investigations by Scott D. Tremaine and James E. Gunn, then both at Caltech. They noted that, for reasons based partly on the Pauli exclusion principle, neutrinos in the relevant mass range could not condense sufficiently to be dark matter on scales much smaller than galaxies. The existence of dark matter on such scales has since been demonstrated convincingly by observations of dwarf spheroidal galaxies.

This work does not preclude the possibility that neutrinos are the dark matter on larger scales. Nev-

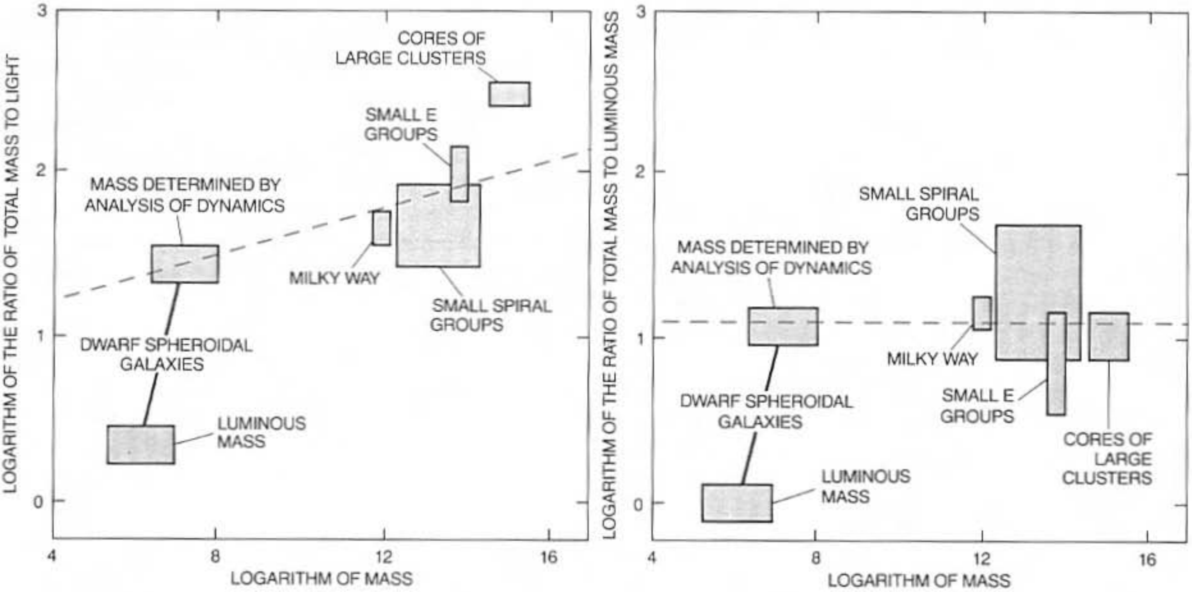

Figure 5 STRUCTURES ON DIFFERENT SIZE SCALES have different mass-to-light ratios (*left graph*), but about the same ratio of total mass to luminous mass (*right graph*), which seems to be constant, indicating that larger structures do not have proportionally more mass than smaller ones.

ertheless, such a proposal seems incompatible with substantial recent theoretical work describing the evolution of the early universe, which has demonstrated that the large-scale gravitational clustering (the clustering of galaxies and of clusters of galaxies) likely to occur in a neutrino-dominated universe does not seem to resemble the clustering actually observed.

In a neutrino-dominated universe the first structures to form would not be on the size scale of galaxies but rather on the scale of clusters of galaxies or even superclusters (clusters of clusters of galaxies). Unlike normal matter, neutrinos in the early universe were not coupled to electromagnetic radiation. Even so, for some time they were not able to clump together appreciably because, being extremely light they moved at relativistic speeds, and relativistic objects are not bound gravitationally except by very highly condensed objects such as black holes.

As the universe expanded, neutrinos cooled until they slowed down and became nonrelativistic. At the same time, the radiation background continued to cool to mean energies below those of the nonrelativistic neutrinos. Shortly before the time at which normal matter decoupled from electromagnetic radiation, neutrinos having masses in the appropriate range to close the universe would have become nonrelativistic and would have begun to make up the primary component of the energy density of the universe. Analytic calculations show that only after this time could they have clumped together gravitationally. At any earlier times, fluctuations on scales smaller than the horizon would have been broken up because the neutrinos, being relativistic, would not have been bound to dense regions.

Thus the first scale on which fluctuations could have grown in a neutrino-dominated universe is the scale of the horizon distance at the time when neutrinos could begin clumping gravitationally. This distance scale corresponds to the size of superclusters, not that of galaxies. Soon after it had decoupled, normal matter would have been drawn into the gravitational potential wells caused by clumps of neutrinos. These supercluster-size formations might then have fragmented into galaxies.

That scenario of a neutrino-dominated universe is attractive in many ways. It would have led to a system of filament-shaped superclusters and large "voids" (regions empty of matter) that resemble features identified in current surveys of large-scale clustering [see "Very Large Structures in the Universe," by Jack O. Burns: SCIENTIFIC AMERICAN, July, 1986]. In addition, the fact that gravitationally bound formations of neutrinos could begin to grow earlier than systems composed of normal matter indicates that the initial density fluctuations in the universe could have been small enough to be at least marginally consistent with measurements of the background radiation's isotropy.

These attractive features led Carlos S. Frenk of the University of Cambridge, Simon D. M. White of the University of Arizona and Marc Davis of the University of California at Berkeley and, independently, Joan Centrella of Drexel University and Adrian L. Melott of the University of Chicago to develop numerical models investigating the details of gravitational clumping in a neutrino-dominated universe. The investigators encountered serious difficulties when they tried to re-create the clustering that has actually been observed. Essentially they found that in a neutrino-dominated universe the fragmenting of clusters into galaxies and the formation of galaxies would have to have occurred relatively recently (when the universe was at least half its present age) in order to match the currently observed level of clustering. This conclusion is hard to reconcile with the existence of such structures as quasars, which formed in much earlier eras.

In general, the major problem with neutrino-dominated cosmology is that in order for galaxies to have condensed by the present time, structures on much larger scales would have to be much less diffuse than the observed large-scale structures actually are, because structure on the scales of galaxies and superclusters would have formed contemporaneously. Well-defined large-scale clustering would also cause difficulties in matching the predicted random velocities of galaxies in clusters to the observed velocities. For these and other reasons a neutrino-dominated universe now seems implausible.

A way out of the problems with neutrino models seems clear: find models in which galaxies can form significantly earlier than larger structures do. This suggests the need for what has become known as cold dark matter: dark matter that was so cold (that is, moving so slowly) that it was nonrelativistic significantly earlier than neutrinos were and could therefore cluster gravitationally much earlier.

The time at which a class of particles becomes nonrelativistic is a key factor in determining the size of structures that can be formed by that class of particles. At times before the particles become nonrelativistic, structures on scales smaller than the horizon would break up. Hence, in order for galaxies to form before larger structures, cold dark matter would have to have been nonrelativistic by the time the horizon reached the scale size of galaxies.

Ever since the problems with neutrino-dominated theories became clear, a great deal of effort has gone into the analysis of cosmology dominated by cold dark matter, and almost all the results have been positive. Because density fluctuations can grow earlier, the initial fluctuations need not be as large and so any conflict with the observed isotropy of the background radiation is eliminated. Moreover, because cold dark matter could have clumped on smaller scales than neutrinos could have, it might account for the excess mass in such small structures as dwarf galaxies.

Detailed analytical and numerical investigations are most encouraging. For example, it has been shown that the presence of cold dark matter in the early universe could account in detail for the shape and structure of many types of galaxies. More generally, Frenk and George Efstathiou of Cambridge, along with Davis and White, have shown numerically that clustering on large scales in a universe dominated by cold dark matter can match well with most of the observed features of the actual clustering.

There is still at least one obstacle that apparently prevents complete agreement between theory and observation if the universe is exactly closed and dominated by cold dark matter: Where is the matter? Apparently it can cluster readily on galactic scales, but, as I have described, there is no evidence for a critical density on such scales. One solution to the problem is to assume that galaxies themselves are not good indicators of where most of the high concentrations of mass are: that much of the cold dark matter lies in regions uncorrelated with the locations of these luminous systems. It could well be that galaxies represent statistically rare events, and that most of the mass in the universe has not ever condensed to form galaxies. Examining the clustering of galaxies would then give a biased value for the actual mass density of the universe. The implications of this proposal have been studied in detail, and it appears to lead to scenarios that agree well with most aspects of the observed clustering (with some notable exceptions). Moreover, current work by Frenk and his collaborators suggests that scenarios in which galaxies are statistically rare might arise more naturally from gravitational clustering than had previously been supposed.

The cold-dark-matter hypothesis has forged a strong link between particle physics and cosmology. At a time when cosmologists were deciding some form of cold dark matter was necessary, high-energy physicists were independently proposing the possible existence of new, exotic particles within the framework of various unified theories. As it happens, several of the particles proposed to fill theoretical gaps in high-energy physics could also serve quite naturally as the cosmologists' cold dark matter. These particles have the disadvantage that they have not been observed; unlike neutrinos, they are at this point merely theoretical constructs. Nevertheless, they have the virtue that their existence was proposed independently from cosmology: they were suggested as solutions to quite different problems in particle theory, and yet each of them, for entirely different reasons, could act as cold dark matter.

Among the most attractive candidates on the market today are particles called axions. The existence of axions follows naturally from a theoretical approach developed to explain a special relation that links, in the theory of strong interactions between quarks, the two kinds of symmetry known as charge conjugation and parity.

An interaction is said to be symmetrical under charge conjugation if the interaction would "look" the same were every particle to be replaced by its antiparticle (which has the opposite charge). An interaction is symmetrical under parity if it would look the same when mirror-reflected. The interactions governed by the strong nuclear force (the force that bind quarks together to form protons and neutrons) appear to be symmetrical to a very high degree under a special combination of charge conjugation and parity: the interactions look much the same if all the particles are replaced with their antiparticles and the entire interaction is mirror-reflected. Theoretically this special combination of symmetries need not hold true. The equations governing the strong interactions include several terms that could in principle grossly violate the combination of symmetries.

In 1977 Roberto D. Peccei and Helen R. Quinn, then both at Stanford University, suggested a way to explain why the combination of symmetries is obeyed so well. Their solution was to introduce a new kind of symmetry—a relation between the forms of different fundamental forces that is manifest at sufficiently high energies but is broken at low energies. It was later pointed out by Frank Wilczek of the University of California at Santa Barbara and Steven Weinberg of the University of Texas at Austin that the fact that the Peccei-Quinn symmetry breaks indicates the existence of a new, very light particle. The new particle is the axion. Much recent theoretical work has refined the original model and increased the temperature at which the Peccei-Quinn symmetry is expected to be broken. One of the big surprises to result is that, because the existence of axions depends on symmetry breaking, an axion "background field" might form in the universe, much as a background electric field would exist if the universe were not charge-symmetric (that is, if it did not contain equal numbers of positive and negative charges). Although axions are themselves very light, calculations show that the background field as a whole could clump in much the same way as heavier, nonrelativistic particles would, making the background field an ideal candidate for dark matter.

Another candidate for cold dark matter comes from the theoretical framework known as supersymmetry. In the theory of supersymmetry, for every particle now known there exists a "supersymmetric partner": a particle identical in most respects except spin. Such particles have not yet been observed in the laboratory, and so they must have large masses. Simple models suggest that supersymmetric partners could behave, in their interactions with normal matter, much like very heavy neutrinos. The most promising dark-matter candidate of the supersymmetric partners is the supersymmetric partner of the photon, which is called the photino. Calculations done by me and by others have shown that photinos in the mass range of from one to 50 times the mass of the proton could naturally have sufficient cosmic abundance to close the universe today. Although this proposal has generated a great deal of excitement recently, I should note that the models predicting the existence of photinos lead to other cosmological predictions that are hard to reconcile with observations.

A final candidate, related to the hit parade of cold-dark-matter candidates, is not a particle at all. It is a structure called a cosmic string. Cosmic strings are extended topological defects that might have arisen from symmetry breaking in the early universe. They would take the form of long, thin tubes of constant and very great energy density winding through the universe. Much work has gone into showing that cosmic strings could have evolved in such a way that their total energy density would be less than that required to close the universe. Nevertheless, in a universe dominated by cold dark matter and containing strings, the mechanism of galaxy formation, although it is quite different from mechanisms in standard cold-dark-matter models, might still lead to clustering that matched observations.

What makes all these dark-matter candidates so intriguing at present is the prospect that each of them may well be detected, directly or indirectly, in the near future. Experiments are possible that would rule out or, what is more significant, confirm various ones of the hypotheses. A positive result in any of these experiments would yield invaluable information about the evolution of large-scale structure in the universe and about the fundamental structure of matter, and it might provide a unique mechanism for probing the sequence of events that occurred during the first few seconds of the big-bang explosion itself.

Pierre Sikivie of the University of Florida was the first to point out that cosmic axions, although they interact with other matter extremely weakly, might be detected in microwave cavities (cavities in which electromagnetic radiation in microwave frequencies resonates). A background field of axions oscillating together might produce electromagnetic radiation that could in principle be detected in a microwave device. Wilczek, John Moody of the University of California at Santa Barbara, Donald E. Morris of the Lawrence Berkeley Laboratory and I have investigated this detection scheme in detail and have proposed refinements and alternative schemes. The sensitivity necessary to detect cosmic axions appears to be near the limit of modern technology, although the technology itself is improving rapidly.

Heavy dark-matter candidates, such as photinos, might be detected in several ways. Recently I suggested, as several other workers did independently, that heavy dark-matter candidates in the galactic halo could be captured in the cores of the sun and the earth, where they would accumulate. There, as later calculations have shown, they could collide with their antiparticles (which could also be captured) in annihilation reactions that could produce light neutrinos. The light neutrinos might then escape from the sun's or the earth's core and be measured in large underground detectors. The degree to which such a flux of light neutrinos has not yet been observed puts limits on the masses and densities of heavy dark-matter candidates.

Recently it has been pointed out that heavy dark-matter particles might also be detected directly by devices that are sensitive to very small deposits of energy in very large volumes of material. A variety of new detectors of this type have recently been proposed. One device, put forward by Blas Cabrera of Stanford, Wilczek and me, is designed to measure a small increase in the temperature of a large sample of ultracold silicon or of another pure crystalline material. The increase in temperature would occur when sound waves, or phonons, produced by impinging dark-matter particles, scattered and randomized. Work by Cabrera, Barbara Neuhauser and Jeffrey C. Martoff at Stanford suggests that the sound waves themselves could perhaps be detected directly.

In one class of possible detectors (see Figure 6), when an impinging dark-matter particle scatters off the nucleus of a silicon atom (6.1), it causes a set of phonons to spread throughout the material (6.2). Phonons arriving at the silicon's surface will have a distinctive pattern (6.3), which will depend on the location and intensity of the original collision. One detector configuration might detect individual phonons in the pattern as they impinge on the surface of the crystal. To do so the silicon could be overlaid with strips of two layers of superconducting aluminum sandwiching a layer of aluminum oxide (6.3a). In superconductors electrons are bound together in pairs called Cooper pairs. An incoming phonon might break apart a Cooper pair, and if the layers are kept at different voltages, the freed electrons might "tunnel" from one layer to the other, forming an electrical current (6.3b). Or, investigators could measure the rise in the silicon's temperature after the initial energetic phonons had dissipated into a uniform background of random thermal vibrations (6.4). Then the detector could consist of a thin film of a material whose electrical resistance increases sharply with temperature (6.4a). A change in temperature of the sample as a whole (6.4b) could be determined from the change in resistance.

Even cosmic strings may soon be detectable, either by their direct gravitational effects on the light from distant quasars and the microwave back-

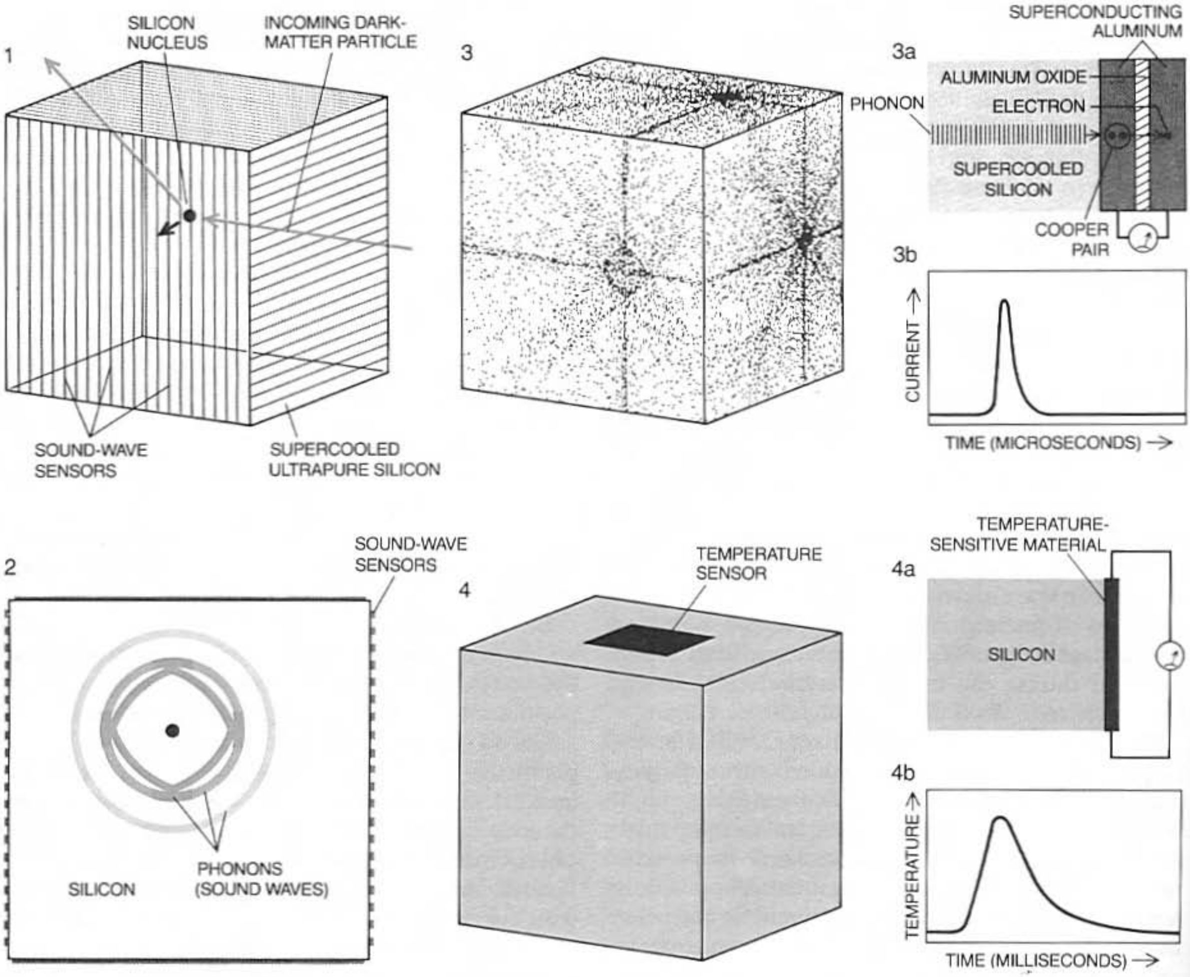

Figure 6 DETECTION DEVICES for dark-matter particles might be made of very pure silicon crystals cooled to within one degree of absolute zero. Such crystals could react measurably to extremely small deposits of energy.

ground (concentrations of energy as dense as cosmic strings should create gravitational fields that would bend light appreciably) or indirectly by measurement of the gravity waves or other radiation they should emit as they evolve.

The solution of the dark-matter question could have broad effects on many areas of physics and astronomy. At stake are fundamental notions about both cosmology and particle physics, and it is fitting that each field—often by provoking active debate in the other—has played an important role in the symbiotic evolution of this area of research.

It is important to recognize, however, that cosmology is in many ways in its infancy. There are comparatively few experimental and observational data available for theorists to work with, and so dramatic changes in the field are possible and much of the standard wisdom may be in error. The point is well illustrated by several new results that arose as this article was being written, any of which may have a profound effect on the field.

One new observational result is found in the preliminary analysis of a deep-sky survey being made by Margaret J. Geller, John P. Huchra and their collaborators at the Harvard-Smithsonian Astrophysical Observatory. It seems that nearby galaxies are clustered in filmlike surfaces that surround nearly spherical voids—a structure resembling that of soapsuds or foam bubbles. This remarkable observation, which could completely revise cosmologists' picture of large-scale structure, suggests that forces other than those of gravity are perhaps at work in determining the present-day large-scale structure.

In another new development, work done independently by Tremaine (now at the Canadian Institute of Theoretical Astrophysics) and J. Anthony Tyson at AT&T Bell Laboratories suggests that galactic rotation curves may not be flat indefinitely but rather may drop off at radiuses beyond about 30 kiloparsecs. The work implies that whatever makes up the dark matter may interact more strongly with normal matter than the cold dark matter would be expected to.

Finally, recent data on the motions with respect to the microwave background of very large-scale regions of matter have provided evidence that these regions are moving, together, with an extremely large drift velocity. No current theory of large-scale structure can explain this apparent phenomenon. New measurements such as these, as well as the possibility of detecting the dark matter itself, may soon revolutionize the accepted picture of the universe.

POSTSCRIPT

The dark-matter problem provides an explicit and timely demonstration of the "cosmic connection" between particle physics and cosmology that has blossomed in the past decade and that is the subject of this reprint volume. Many of the developments described in the subsequent articles here have impacted, in one way or another, our present perspective of the dark-matter issue.

Two factors form the basis of this intellectual connection. First, there is a distinct possibility that most or all of the dark matter in the universe is some exotic form of matter. This matter could be composed either of new types of elementary particles or made from coherent configurations of the fields associated with new forces between elementary particles. Beyond this, the distribution and abundance of dark matter today can reflect the nature of elementary-particle interactions at high energy. One of the most pressing problems in cosmology is to explain the origin of the observed macrostructure of the universe. This explanation is inevitably linked to our understanding of how the initial conditions of the big-bang expansion developed—a question that depends ultimately on our understanding of microphysics.

Several developments, which have taken place in the short time since my article appeared, attest to both the speed with which the interface between cosmology and particle physics is developing and the potential for radical changes.

Eugene Loh and Earl Spiller at Princeton apparently provided the first empirical support for the flatness of the universe. They claim to have measured the distances of about a thousand galaxies within a radius of about six billion light years—a fair fraction of the observed universe. Counting the number of galaxies at a given distance is one way to probe the curvature of space, because volume changes with distance in a way that depends on curvature. Their data appear to be consistent with a flat universe—allowing the possibility of three to five times greater mean density of matter than implied by all previous work. Since its appearance, the particular method of analysis they used has been seriously questioned by other workers on both experimental and theoretical grounds. While it seemed to signal the beginnings of agreement between theory and observation, it now appears as if this method cannot give definitive results.

Next, Davis and his collaborators are continuing their numerical simulations of clustering in a cold-dark-matter-dominated universe. The most recent results are very encouraging. Structures resembling, on a fine scale, the observed galactic mass distributions can form in a statistical ensemble that, on large scales, could reconcile a flat universe with observation. Moreover, a structure recently observed by Geller and collaborators bears a resemblance to certain (but not all) features of the fascinating bubble-like structure on large scales. At the same time, Alan Dressler and his colleagues have continued to argue that a large region of galaxies is falling coherently in the direction of the Hydra-Centaurus cluster. The large mass agglomeration required to produce this is not easily explained by any cold-dark-matter model.

The search also continues for direct and indirect signatures of cold-dark-matter candidates. New calculations of stellar production of light axions in the sun, red giants, white dwarfs, neutron stars and the recent supernova in the Large Magellanic Cloud have tightened the bounds on these candidates. Also, more refined calculations of the signal in the atmospheric neutrino background for the annihilation products of dark-matter candidates captured in the sun, earth and galactic halo continue to be produced. Proton-decay detectors, discussed in Chapter 7, "The Search for Proton Decay," by J. M. LoSecco, Fredrick Reines and Daniel Sinclair, have not yet seen such a signal, thereby strongly constraining a variety of heavy weakly interacting particles (WIMPs) as cold-dark-matter candidates. These constraints also seem to be supported by recent analyses from ultra-low background experiments on double beta decay of nuclei, which are sensitive directly to the scattering of very heavy dark-matter candidates. They have as yet yielded no such signal. As a result stable neutrinos with masses in excess of 15 to 20 GeV may be ruled out as candidates for galactic halo dark matter. Finally, at present at least 10 different groups around the world are building ultracold detectors of various types to directly probe for WIMPs, which should be operational soon.

The suggestion, made by myself and independently by several other people, that WIMPs captured in the sun could give rise to the observed apparent shortfall of solar neutrinos detected on earth has led to a brief flurry of activity. Recent work by John Faulkner and collaborators has suggested that this could result in a number of other observed solar features. Unfortunately, as I and my collaborators demonstrated, when analyzed in detail the capture of WIMPs from the galactic halo yields abundances that are generically too small for this mechanism to involve standard dark-matter candidates. Thus, even more exotic objects would be required.

There has also been a great deal of activity recently on the formation and evolution of large-scale structure in a cosmic-string-inhabited universe. This has focused on the ability of large cosmic strings to evolve via the formation of loops, which can subsequently decay quickly enough to yield acceptable mass densities today. Inside cosmic strings fundamental forces can appear in the same symmetric configuration they could have had at the high temperatures in the early universe. The different possibilities result in different possible observable consequences today. These scenarios could in principle explain galaxy formation and also yield general features of observed galaxy clustering.

Finally, several groups recently followed up on a possible loophole in the constraints from nucleosynthesis on the abundance of baryonic matter today. They suggest that under certain model dependent conditions, a flat universe dominated by normal matter would be consistent with the observed abundances of light elements resulting from primordial nucleosynthesis. If this were true all the dark matter could in principle be baryonic. Alternatively, another group has suggested that the decay of some heavy particle well after the standard nucleosynthesis epoch might alter primordial abundances in a way that would also allow the dark matter to be baryonic. Both these scenarios appear to suffer problems predicting the correct primordial abundance of the isotopes of lithium. However, the matter is still unresolved.