Cold dark matter, illustrated here in blue and black, is shown forming halos around galaxies and connecting the giant collection of galaxies known as the Great Wall. The picture is an artist's conception, as cold dark matter has never been observed. Its existence is called for by one theory of how gravity caused the growth of large-scale structures such as those pictured here. The latest simulation "experiments" on massively parallel computers have sufficient speed and accuracy to test the cold-dark-matter scenario.

EXPERIMENTAL COSMOLOGY

and the puzzle of large-scale structures

Wojciech H. Zurek and Michael S. Warren

A critical limitation in astrophysics is the impossibility of testing theories with controllable and repeatable experiments. Cosmologists face the worst version of this problem. Not only are the data restricted to observations of uncontrolled events, but also the experiment was performed only once—there is only one universe! Moreover, only a fraction of the universe that is in principle observable (given the finite speed of light and the finite age of the universe) is accessible in practice to observation. Cosmology has suffered from having too little hard data and too much freedom to build theoretical models, so that even the most basic questions—about the size of the universe, the age of the universe, the kinds of matter in the universe—are still only partially answered.

This situation is changing through advances in observation and computation. Instruments such as the Hubble Space Telescope and the Cosmic Background Explorer satellite are providing new data, some relating to the very early history of the universe, that help constrain cosmological theories. Furthermore, computer simulations carried out on the most advanced massively parallel machines now contain enough physics and are sufficiently accurate to predict detailed consequences of the many proposed models. Here we present high-resolution computer simulations that address one of the outstanding puzzles in modern cosmology: How did the observed distribution of galaxies—the so-called large-scale structure of the universe—arise?

Cosmologists once thought that the matter distribution on large scales is fairly uniform, that galaxies like ours are sprinkled evenly throughout the cosmos. Over the last few decades, however, observations out to distances of billions of light-years* have revealed a "froth-like" structure to the universe. Large numbers of galaxies are grouped into clusters, and those clusters appear to be interconnected by thin sheets and filaments of galaxies surrounding large low-density cells, or "voids," approximately a hundredth of the radius of the universe in size. Figure 1 illustrates some of these large-scale structures, from individual galaxies with bright components on the order of 10 kiloparsecs in radius (a parsec is about 3.3 light-years), to clusters of galaxies on scales of 1 to 10 megaparsecs, and finally to sheets or filaments of galaxy clusters that surround large voids on scales of 30 to 50 megaparsecs. By comparison, the current size of the observable universe is over a hundred times larger, roughly 10,000 megaparsecs across.

In confronting the data on large-scale structure, the task of a cosmologist is somewhat similar to the task of a prosecutor trying to prove that the accused (a cosmological model) is indeed guilty of the crime (the creation of the large-scale structure), while his observational colleagues are trying either to support the case by supplying the observational evidence (which is nearly always circumstantial) or to side with the defense and provide the alibi. The crime has, of course, happened only once, and over the years many suspects have been rounded up. It is conceivable, though, that the guilty party has not yet come under suspicion. However, many possible models have already been exonerated.

Consider the idea that nearly all the matter in the universe is contained in visible, luminous stars and that gravity is the force primarily responsible for the clumping of matter into structures of various sizes. After all, that idea does indeed describe our solar system. But on galactic and larger scales, that natural model has been "found innocent." As we shall see below, it cannot account for the internal dynamics of spiral galaxies. Instead those galaxies must contain about ten times more mass than has so far been observed to explain the rapid motion of the stars inside them.

More important for our discussion, it is difficult to explain how gravitational forces alone produced large-scale structure unless we assume the presence of still greater amounts of invisible mass—what is usually referred to as dark matter. It seems that there is neither enough mass in the luminous matter nor enough time since the beginning of the universe for primordial fluctuations in the density of that matter to have grown under the influence of gravity into the structures now observed. The size of the primordial fluctuations, as inferred from inhomogeneities in the cosmic background radiation, is just not large enough. Thus, observations on galactic and larger scales present clear evidence that the universe contains much more mass than "meets the eye."

What is this dark matter that dominates the mass content of the universe? Astrophysicists have responded to this puzzle with a variety of ideas. Some postulate that the dark matter is more or less ordinary (that is, made of the same stuff as our sun and the planets) but, for some reason, simply invisible. Calculations of the primordial synthesis of the light elements, however, place an upper limit on the amount of ordinary matter that might be present, a limit that is probably too low for ordinary matter alone to explain large-scale structure. Other astrophysicists suggest that neutrinos with a small, but non-negligible, mass are the main ingredient of dark matter. The latter postulate leads to the so-called hot-dark-matter (HDM) models: Massive but not-too-heavy neutrinos remain relativistic or "hot" (they move with velocities comparable to the velocity of light), until fairly late in the history of the universe. Their movements wipe out density perturbations on galactic scales and thereby determine a certain predictable course for the development of structure, which turns out to be too slow to match observations.

Perhaps the best-defined model, and the one we have tested through computer simulation, is the so-called cold-dark-matter (CDM) model. The main ingredient of dark matter in the CDM model is not specified beyond the requirements that it be very weakly interacting with ordinary matter and suffi-

*These distances and all the other absolute distances in this article are uncertain by approximately a factor of 2 for various reasons, primarily the uncertainty in the Hubble parameter. Relative distances can however be expressed in a manner independent of those uncertainties by using the redshift, which will be introduced in the following section.

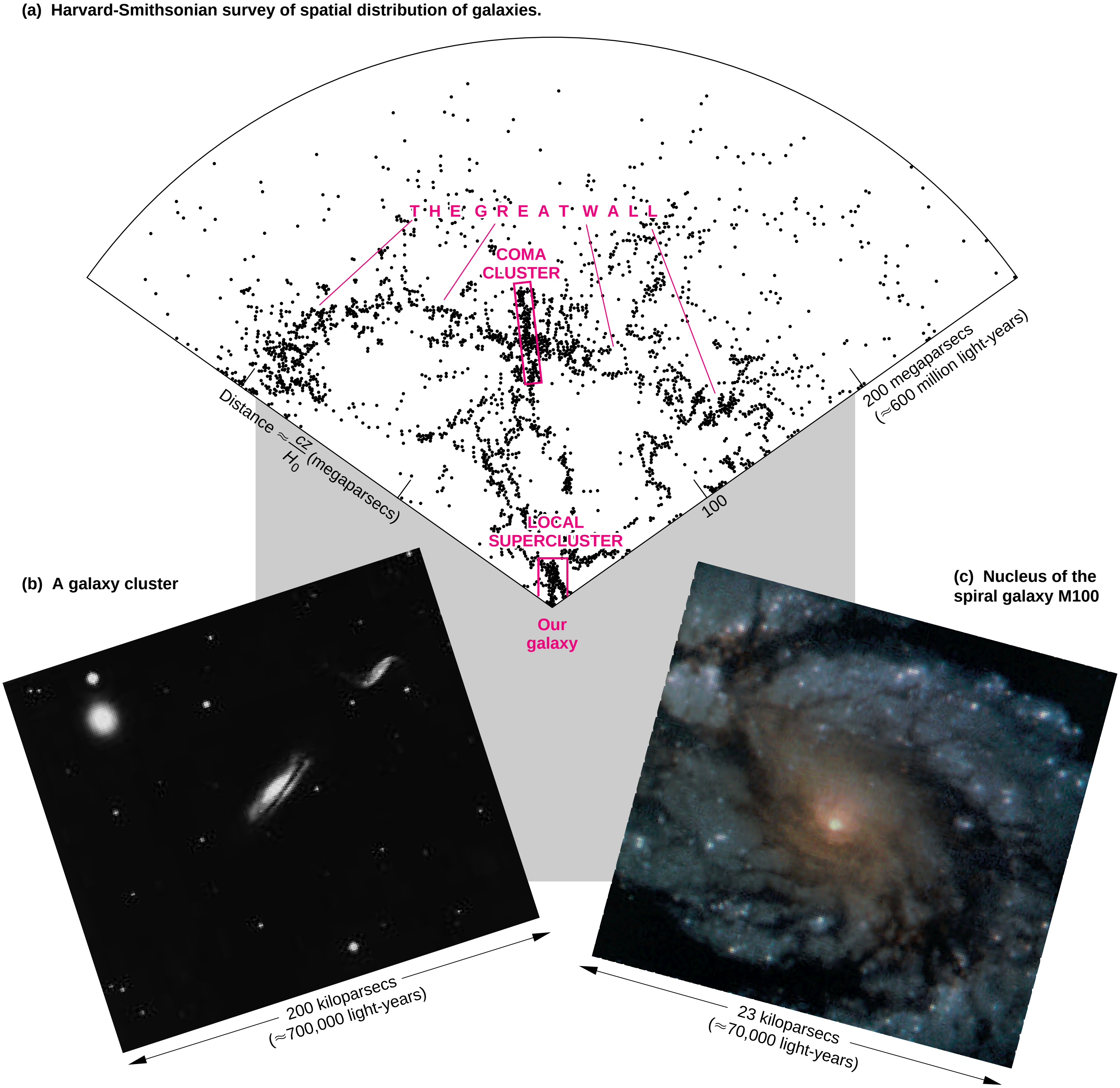


Figure 1. Inhomogeneity of the Universe

The distribution of matter in the universe is marked by clumps and voids on all scales. (a) The spatial distribution of galaxies according to a survey by the Harvard-Smithsonian Center for Astrophysics. Each point represents a galaxy. The distance from the vertex (our galaxy) to the outer edge of the region surveyed is roughly 300 megaparsecs. The distances plotted are inferred from an observed measure of distance called redshift, z, through the use of Hubble's Law—distance = v/H_0, where $v \approx cz$ for objects with small redshifts—as will be discussed later in the main text. Here c is the velocity of light and H_0 is Hubble's constant. The elongated clump of galaxies in the center of the figure includes the rich cluster in the constellation Coma Berenices. It is part of the "Great Wall," the large sheet of galaxies running across the image at intermediate distances, which includes more than half the galaxies in the figure. The Great Wall and the voids beside it are among the largest known structures in the universe. (Figure adapted by permission of Margaret J. Geller, Harvard-Smithsonian Center for Astrophysics.) (b) A galaxy cluster—a structure of intermediate size—in the constellation Leo. (Photograph courtesy of Palomar/California Institute of Technology.) (c) The nucleus of a single spiral galaxy, catalogued as M100, with a radius on the order of 10 kiloparsecs. Galaxies are the smallest objects we consider in our study of large-scale structure. (Photograph courtesy of NASA/Space Telescope Science Institute.)

ciently massive to become "cold," or nonrelativistic (moving at velocities small compared with the velocity of light), very early in the history of the universe, before the universe was about a thousand years old. The model assumes, as an initial condition, a simple, scale-independent spectrum of primordial density fluctuations in the cold dark matter that are postulated to grow under the influence of gravity to form the large-scale structures that we now see.

The only important free parameter in the CDM model is the size of the assumed primordial density fluctuations. Information about those fluctuations is imprinted in the cosmic background radiation, the oldest detectable remnant of the early universe. That sea of microwave photons, which fills all space, is thought to have remained essentially unscattered since the time when radiation decoupled from matter and the universe became transparent to radiation. It would therefore have retained inhomogeneities present at the time of decoupling, when, according to standard Big Bang Cosmology, the universe was about 10,000 years old.

The size of those inhomogeneities was recently determined by data from the Cosmic Background Explorer satellite. Known as COBE (pronounced to rhyme with Toby) the satellite is dedicated to measuring various properties of the cosmic background radiation. The 1992 COBE data show variations in the temperature of the cosmic background radiation of about one part in 10^5 on distance scales of around 1 billion light years (hundreds of megaparsecs), depending on the spatial direction of the measurements. The small temperature fluctuations are a direct measure of the amplitudes of the fluctuations in the matter density present when the cosmic background originated. Thus the size of the temperature differences measured by COBE fixes the size of the primordial density fluctuations in the CDM model.

Our contribution has been to incorporate the new COBE data into the initial conditions of the CDM model and then determine, through high-resolution, state-of-the-art simulations, the CDM predictions for the growth of structure through time. Our simulations keep track of the long-range gravitational forces among 17 million point masses with a precision sufficient to resolve density contrasts of six orders of magnitude on scales that differ by as much as four orders of magnitude. Our program has undergone continual development since it received the 1992 Gordon Bell Prize for practical parallel-processing research. It is described in "A Fast Tree Code for Many-Body Problems," immediately following this article.

The high resolution that we have achieved has allowed much more detailed comparison with observations than was previously possible. Our results discussed on page 78 suggest that the CDM model, recently abandoned by many, should still be taken seriously.

Before presenting those comparisons, we will explain the initial conditions of the simulations and the basic assumptions of the CDM model through a review of standard Big Bang cosmology. The CDM model and all other recent models of structure formation are defined within that basic framework, a framework that has been strengthened by all recent observations.

The Expansion of the Universe

Big Bang cosmology grew out of the most far-reaching and well-accepted cosmological observation: The universe is expanding. In other words, to any observer in the universe, distant objects such as galaxies appear to be receding from the observer at velocities v proportional to l, the distance from the object to the observer. This relation, called Hubble's law, was discovered by Edwin Hubble in 1929. It is written

$$v = H_0 l.$$

The Hubble "constant," H_0, is shorthand for $H(t_0)$, which means the value of the Hubble parameter $H(t)$ measured now, at $t = t_0$. Although Hubble's law was deduced from observations carried out from our vantage point, the planet Earth, it is generalized to the cosmos by assuming that at any given time the universe looks qualitatively the same to all observers regardless of their locations and the directions in which they look. This assumption, called the Cosmological Principle, is fundamental to the study of cosmology.

Figure 2 shows a traditional way of visualizing the expansion of the universe. The three-dimensional universe is replaced by an analogous two-dimensional surface, the surface of a sphere. As the sphere expands, any observer confined to the surface sees other points receding with speeds proportional to their distances from the observer, as observed by Hubble.

Hubble discovered this proportionality, and thus the expansion of the universe, by measuring both the apparent velocities v of nearby galaxies and their distances l. Measuring the distances to galaxies is difficult, involving a ladder of distance measurements from nearby stars (tens of parsecs away) to more distant stars in our own galaxy (up to 10 kiloparsecs away) to stars in nearby galaxies (1 to 10 megaparsecs away) and on to more distant galaxies. The process remains problematic, so that even today uncertainties in the distance measurements yield estimates of H_0 ranging from 50 to 100 kilometers per second per megaparsec. A convenient

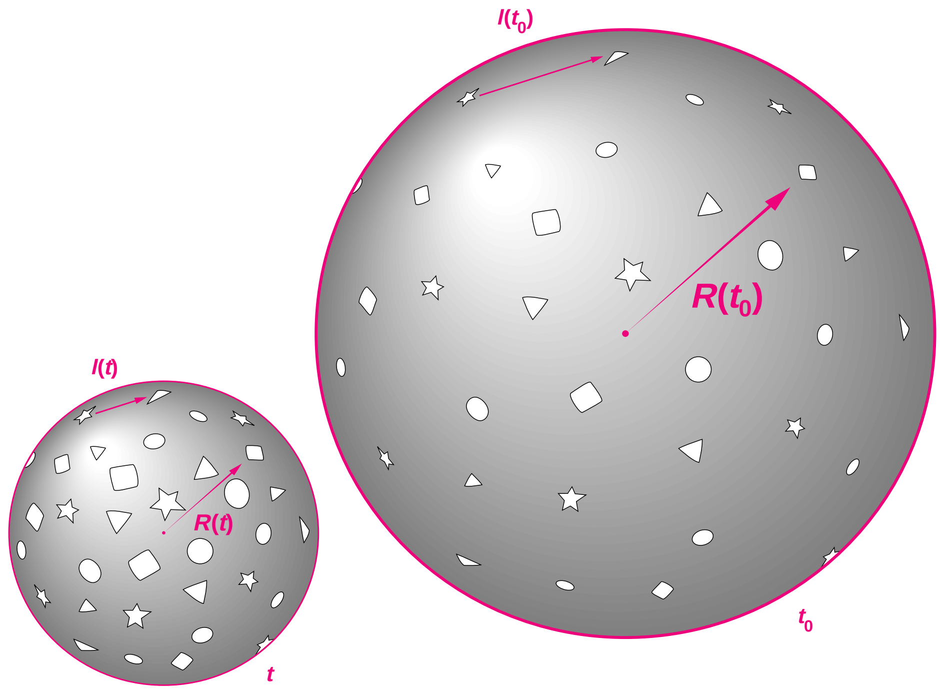


Figure 2. The Expansion of the Universe

The Hubble flow and the expansion of the universe are depicted here in a spherical fragment of the real universe. The size of the sphere, $R(t)$, changes with time, as do the distances $l(t)$ between the objects (such as galaxies) inside. If all of the objects started expanding from the same point at different velocities v, then their present locations will depend on how fast they are moving. This dependence yields a proportionality relation, $l = v/H$, which is the content of Hubble's law. The velocity of a galaxy can usually be inferred from the redshift, z, of the light emitted by it, and when $z \ll 1$ the velocity is given by $v = zc$. Because the light travels at a finite velocity, the redshift is also a measure of the "look-back time," which in turn is related to the relative size of the universe at the time when the photons were emitted: Expansion "stretches" the wavelengths of the photons and the size of the universe at the same rate, so that the photons emitted at the instant t and detected now, at t_0, will be redshifted—the photon wavelengths will be longer by the redshift factor $1 + z = R(t_0)/R(t)$. It is convenient to define a universal scale factor $a(t)$—the factor by which all of the distances in the real universe need to be rescaled to account for the effect of the universal expansion. Thus $R(t_0)/R(t) = l(t_0)/l(t) = a(t_0)/a(t)$. By definition $a(t_0) = 1$. Thus $a(t) = 1/[1 + z(t)]$.

way to represent this uncertainty is to define a parameter h such that $H_0 = h\,100$ (km/s)/Mpc, where $1/2 \leq h \leq 1$ (Mpc is the abbreviation for a megaparsec). To ensure that our calculations do not assume a universe too young to be consistent with other predictions of astrophysics, we have taken the value of h to be 1/2, or the value of H_0 to be 50 (km/s)/Mpc, throughout this article and in our simulations.*

In contrast to the distance, the apparent velocity of a receding galaxy is reliably and easily determined by measuring the Doppler shift—in this case the redshift—of the light that it emits. More specifically, one identifies the absorption and emission lines in the spectrum of the galaxy and measures the shift of those features toward longer wavelengths, in the direction from blue toward red. (Here again is an application of the Cosmological Principle: One is assuming that the matter (atoms) and the wavelengths of the light it emits are the same throughout the universe.) The size of the redshift z is related to the apparent recession velocity v through the Doppler-shift formula:

$$1+z \equiv \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \left(\frac{1+v/c}{1-v/c}\right)^{1/2},$$

in which λ_{emitted} is the wavelength emitted by the galaxy, $\lambda_{\text{observed}}$ is the wavelength of the observed signal, and c is the speed of light. For galaxies with small redshifts, or with small recession velocities $v \ll c$, the redshift $z \approx v/c$, or the recession velocity is given by

$$v \approx zc.$$

Thus the velocities of nearby galaxies are directly proportional to their observed redshifts. The formula applies even to the most distant galaxies shown in Figure 1. They have redshifts of 0.05 and by Hubble's law are at distances of $l \approx zc/H_0$, or 300 megaparsecs for $h = 1/2$.

In contrast, the most distant object observed to date is a quasar with a redshift of nearly 5. The full Doppler-shift formula implies that that object is rushing away from us at the enormous velocity of 0.95c or about 285,000 kilometers per second, and, by Hubble's law, is at a distance of approximately 5700 megaparsecs.

In general, the higher the redshift of an object, the greater is its distance from us, the faster is its recession velocity, and the longer its signal took to reach us. The observed redshift of an object thus indicates the "look-back time," a fact that is used again and again to interpret cosmological observations and to piece together the history of the universe.

Cosmologists typically relate the redshift not to the look-back time but rather to the relative size of the universe, because the latter relationship is simpler. The redshift of a light signal (more exactly, $1 + z$) is inversely pro-

*The new corrective optics recently installed in the Hubble Space Telescope should help put the controversy about the value of H_0 to rest by skipping several of the rungs in the ladder of distance measurements.

portional to the relative size, or scale, of the universe at the time that signal originated. Here the size of the universe is measured by the distances between nearby objects, $l(t)$, or by the radius, $R(t)$, of a spherical fragment of the universe. As depicted in Figure 2, both quantities increase with the universal expansion. The relative size—the universal scale factor—is then defined as the ratio

$$a(t) \equiv l(t)/l(t_0) \equiv R(t)/R(t_0).$$

In other words, the universal scale factor is defined to be one at present, $a(t_0) = 1$, and its value decreases to zero as we go back in time.

The inverse proportionality between the redshift and the universal scale factor can be derived by reinterpreting the origin of the redshift as due to the overall expansion of the universe. As the universe expands, the distances between objects co-moving with the universal expansion increase proportionally to $R(t)$. In addition, the wavelengths of photons traversing the universe at velocity c are stretched by the expansion in proportion to $R(t)$. The ratio of wavelengths in the redshift formula can therefore be written as

$$\frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \frac{R(t_0)}{R(t)},$$

where t_0 is now, the time at which the signal is observed, and t is the time at which the signal originated. Rewriting this equation in terms of the universal scale factor shows that the redshift increases as the scale factor decreases:

$$1 + z(t) = \frac{R(t_0)}{R(t)} = \frac{1}{a(t)}.$$

This equation means, for example, that when we observe a quasar with a redshift of 5 (or $1 + z = 6$), we are looking back in time to when distances between astronomical objects were approximately $1/(1 + z) = 1/6$ of their present values. According to standard Big Bang cosmology, the cosmic background photons, which contain the oldest imprint of large-scale structure, have been redshifted by a factor of about 1000 ($z \approx 1000$), which means that those photons decoupled from matter at the time when the scale factor was only one-thousandth of its present size. Our simulations of the clumping of matter into large-scale structures begin later, at $z \approx 100$, or when the scale factor was approximately one-hundredth of its present size.

The Big Bang and the Definition of Ω

Extrapolating the expansion of the universe all the way back in time suggests that initially any point was arbitrarily close to any other point, or $R(t{=}0) \approx 0$ and $a(t{=}0) \approx 0$. Such a picture is consistent with solutions to the equations that arise when general relativity is applied to the entire cosmos; it is also consistent with observations. Therefore cosmologists now generally believe that the universe expanded and cooled from an initial state of extremely high temperature and density. Cosmologists can trace the history back no farther than the time when the density of the universe was the so-called Planck density (roughly 5×10^{33} grams/centimeter^3). At that density quantum-gravity effects begin to dominate. The initial explosive expansion is called the Big Bang.

How much time has elapsed since the Big Bang? We need to estimate the age of the universe to determine the time available for the development of large-scale structure. The simplest estimate is made by assuming that the Hubble parameter has remained constant at its present value of H_0 and thus the speed of expansion has also remained constant since the Big Bang. Then t_0, the time since the Big Bang, is approximated by

$$t_0 \approx \frac{\text{distance expanded}}{\text{speed of expansion}}$$

and, by Hubble's law, that time is

$$t_0 \approx \frac{1}{H_0} \approx 20 \text{ billion years}$$

where we have taken the value of the Hubble constant to be 50 (km/s)/Mpc.

The true age of the universe must be somewhat less than 20 billion years because the initial expansion at $t = 0$ must have slowed down over time due to the mutual gravitational attraction of the total energy content, or equivalent mass content, in the universe. The theory of relativity relates all forms of energy E, even pure radiation, to an equivalent mass m through the formula $E = mc^2$. Thus both matter and radiation are sources of gravitation and contributed to the slowing down of the initial expansion. As the universe has expanded, the Hubble parameter $H(t)$ has been decreasing continuously to its present value H_0.*

The time history of the expansion is typically described in terms of the Hubble parameter where

*For H_0 = 50 (km/s)/Mpc and $\Omega = 1$ (see below), the age of the universe is 14 billion years, approximately the age of the oldest globular clusters inferred from stellar evolution and primordial abundances of the light elements. A larger value of H_0, or equivalently h, would imply a younger universe and could thus lead to a contradiction, or at the very least, indicate a low value for Ω.

Figure 3. Evolution of the Universe

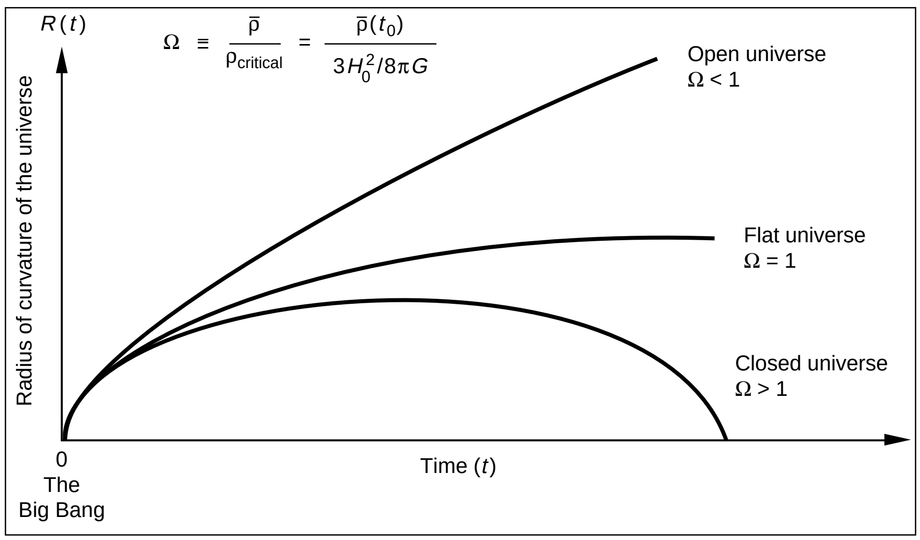
Each curve shows a possible history of the size of the universe as a function of time, depending on whether the average equivalent mass density $\bar{\rho}$ of the universe is greater than, equal to, or less than the critical value, ρ_{critical}, or equivalently on whether Ω is greater than, equal to, or less than 1. The curves can be obtained by using the Newtonian approximation to predict the motion of a spherical fragment of the universe of radius $R(t)$ and containing mass M equal to $4\pi R^3\bar{\rho}/3$. The Newtonian approximation can be derived from Einstein's equations when $GM/Rc^2 \ll 1$, that is, when $R(t)$ is large compared to the Schwarzschild radius of the mass M. Then the radius of the sphere satisfies the familiar equation

$$\frac{d^2R}{dt^2} = -\frac{GM}{R^2}, \qquad (1)$$

which can be used in turn to derive the equally familiar statement of conservation of energy in a gravitating system:

$$\left(\frac{dR}{dt}\right)^2 = \frac{2GM}{R} + C. \qquad (2)$$

The constant C on the right-hand side appears in the course of the integration involved in the transition from the first equation to the second. Its sign determines the balance between the kinetic energy of expansion (proportional to the left-hand side of this equation) and the potential energy (on the right-hand side). By dividing both sides of equation 2 by R^2 (to express the right-hand side in terms of the density $\bar{\rho}(t)$ rather than mass M) and re-expressing $R(t)$ as $R_0 a(t)$, one eventually arrives at the equation



$$H(t)^2 \equiv \frac{1}{a^2}\left(\frac{da}{dt}\right)^2 = \frac{8\pi G\bar{\rho}(t)}{3} - \frac{kc^2}{R^2}. \qquad (3)$$

Here k is the constant C from equation 1, rescaled so that it can assume only the values ± 1 and 0. It is known as the curvature constant. It should be emphasized that the above derivation is perfectly correct, and not merely a Newtonian analogue: The condition $R \gg GM/c^2$ can always be satisfied by adopting a sufficiently small radius. Each curve is a solution to the equation obtained by using one of the different values of k (and by assuming that matter, not radiation, makes the dominant contribution to the energy density). Regardless of the value of k and the composition of the universe, solutions to the equation have $R = 0$ at $t = 0$; that is, they entail a Big Bang at the beginning of the universe. Also, the equation implies that $k = +1$ if $\Omega \equiv \bar{\rho}(t)/\rho_{\mathrm{critical}}(t) > 1$, where $\rho_{\mathrm{critical}}(t) \equiv 3(H(t))^2/8\pi G$. Likewise $k = 0$ if $\Omega = 1$, and $k = -1$ if $\Omega < 1$. The value of k, or equivalently the value of Ω, determines whether the universe is open, flat, or closed, as explained in the main text.

$$H(t) \equiv \frac{1}{a(t)}\frac{da}{dt} = \frac{1}{R(t)}\frac{dR}{dt},$$

and $a(t)$ is the universal scale factor introduced above. The relevant differential equation from general relativity is presented in the caption for Figure 3. This equation yields three different types of universe, depending on the average density of the universe, or more exactly, on the value of Ω, the dimensionless density parameter. This parameter is defined as the ratio of the average equivalent mass density in the universe, $\bar{\rho}(t)$, to a critical density:

$$\Omega \equiv \frac{\bar{\rho}(t)}{\rho_{\text{critical}}(t)}$$

where

$$\rho_{\text{critical}}(t) \equiv \frac{3(H(t))^2}{8\pi G},$$

and G is Newton's gravitational constant. Figure 3 shows the three possible histories of the expansion of the universe depending on whether the value of Ω is greater than, equal to, or less than one.

In Newtonian cosmology, which applies when most energy is in the form of matter, Ω has a simple interpretation. It is just the ratio of the magnitude of the gravitational potential (or binding) energy of matter to the kinetic energy of the universal expansion:

$$\Omega = \frac{|E_{\text{potential}}|}{E_{\text{kinetic}}}.$$

Thus, if the potential energy is less than the kinetic energy ($\Omega < 1$), galaxies have enough energy to escape the pull of gravity and will travel to infinity. Alternately, if the potential energy is greater than the kinetic energy ($\Omega >$ 1), galaxies are gravitationally bound, the Hubble expansion will eventually stop, and the universe will contract back on itself.

The same criterion defines the dividing line between two entirely different types of universes allowed by general relativity. As shown in Figure 3, if $\Omega < 1$, the universe is said to be open; it is spatially infinite and will continue expanding forever. If $\Omega > 1$, the universe is said to be closed; it is spatially finite (being curved like the sphere in Figure 2) and will eventually stop expanding and begin to contract [$H(t)$ will eventually become negative]. A third solution also exists. Namely, if Ω is exactly equal to 1, the universe is said to be flat; it is infinite in space and time, but differs from an open universe in that the recession speeds of galaxies eventually approach zero rather than a positive constant.*

Observational evidence suggests that Ω should lie within the generous bounds $0.2 < \Omega < 2.0$. This range is obtained from observations that usually allow one to estimate Ω more directly than by comparing the present average density of the universe, $\bar{\rho}(t_0)$, with the critical density:

$$\rho_{\text{critical}}(t_0) \equiv \frac{3H_0^2}{8\pi G} \sim 10^{-29}\, h^2 \text{ g/cm}^3.$$

This number, uncertain by a factor of 4 because of the uncertainty in the value of the Hubble constant, corresponds to the mass density of a few hydrogen atoms per cubic meter.

Determinations of Ω tend to yield larger values when they employ observations on larger scales. Thus Ω inferred from galactic scales tends to be on the order of 0.1 or less, while clusters of galaxies (scales of a few megaparsecs) indicate an Ω of 0.2 or more. On still larger scales of 30 to 50 megaparsecs, Ω as large as 1 is necessary to explain coherent flows of galaxies such as the "Great Attractor."

Cosmologists must assume a value for Ω to investigate the development of structure in the expanding universe. Although the present density of luminous matter, $\bar{\rho}_{\text{luminous}}(t_0)$, appears to be less than one percent of the present critical density, that is $\Omega_{\text{luminous}} <$ 0.01, most theorists—for both esthetic and theoretical reasons—take the case of the flat universe ($\Omega = 1$) very seriously. In that case, most of the mass in the universe must be in the form of dark matter.

*Note that if Ω is equal to 1 at any time, it remains constant at that value. Otherwise its value changes with time, but because the total energy of the universe is conserved, Ω cannot change from being less than 1 to being greater than 1 or vice-versa.

Formation of Structure in a Cold-Dark-Matter Universe

A major event in the history since the Big Bang was the shift from the radiation-dominated era, when the universe was so hot that most of the energy was in the form of radiation rather than matter, to the matter-dominated era, when the universe had cooled down enough that most of the energy was in the form of matter. The main constituents of matter then were the two most stable baryons, namely protons and neutrons, enough electrons to balance the charge on the protons, and perhaps, as proposed by the CDM model, an exotic brand of noninteracting, cold, dark matter. [Note that during the first few minutes after the Big Bang, nuclear synthesis reactions had caused most neutrons to combine with protons (hydrogen nuclei) to form helium nuclei.] The transition to a matter-dominated universe (described in "Big Bang Cosmology and the Microwave

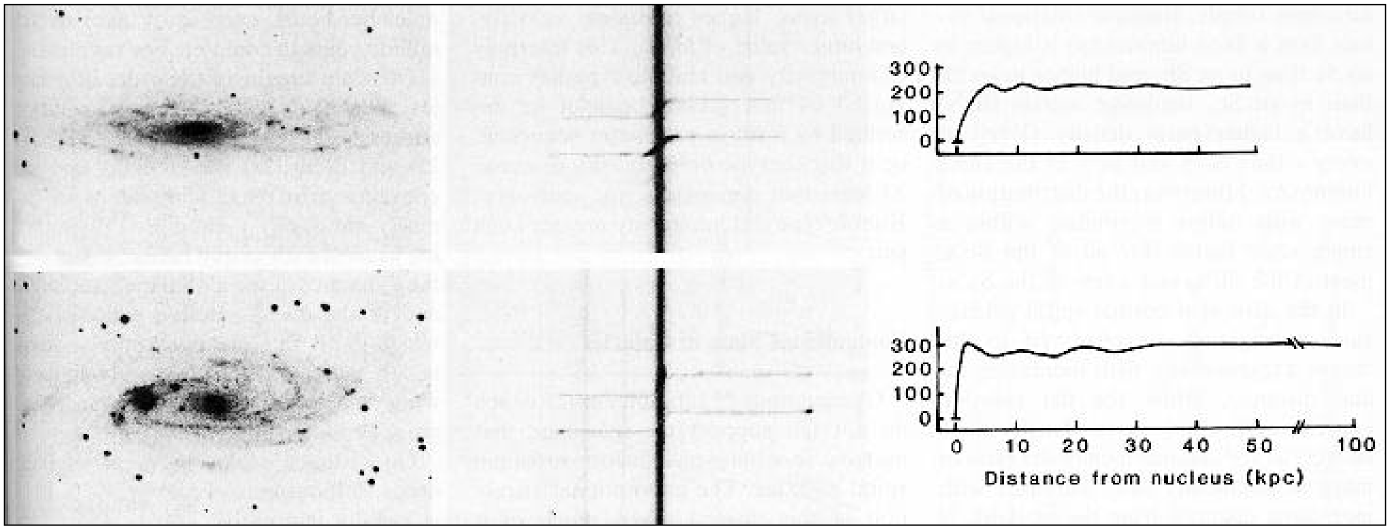


Figure 4. Evidence for the Existence of Dark Matter

The figure shows observations of two galaxies from which their rotations can be inferred; the inferred rotations provide evidence for the existence of dark matter. On the left are photographs of the galaxies NGC 801 and UGC 2885. The images in the center are photographs of their spectra made by allowing light from the galaxies to diffract through a horizontal slit. Therefore each photograph is the spectrum of a horizontal strip through the center of the galaxy; vertical position indicates wavelength. The heavy vertical lines in the center are the spectra of the galactic nuclei, which are highly luminous at all wavelengths. The step-shaped horizontal lines show the wavelengths of spectral lines of hydrogen and nitrogen as a function of distance from the center. Their shapes result from the Doppler blueshifts and redshifts due to the rotations of the galaxies; the wavelengths of these lines are shorter on the side of the galaxy rotating toward us and longer on the side rotating away from us. Thus the average orbital speeds of stars as a function of radius can be deduced from the displacements of those lines as a function of distance from the center of the galaxies. The resulting "rotation curves," combining data from both sides of the galaxies, are on the right. (The vertical axis shows the rotational velocity in kilometers second^{-1}) Note that beyond a few kiloparsecs the velocity is nearly constant. This velocity distribution is typical of spiral galaxies. An entirely different velocity distribution, decreasing to zero as $r^{-1/2}$, would be expected if the masses of the galaxies were concentrated at their centers, as are their distributions of luminous matter. The only convincing explanation of the observed velocity distribution is that the galaxies are embedded in massive halos of unobservable "dark matter." (From an article by Vera Rubin, The rotation of spiral galaxies, *Science* 220: 1339–1344, 24 June 1983. Copyright AAAS. The photographs of the galaxies were made by B. Carney using the 4-meter telescope at Kitt Peak National Observatory.)

Background"), about ten thousand years after the Big Bang, marks the time when, at least in principle, the force of mutual gravitation attraction could begin causing matter to collapse into the large-scale structures we see.

A later major event was recombination, when nuclei of hydrogen and helium combined with electrons to form stable neutral atoms. At that time the background radiation, consisting of photons in thermal equilibrium with the matter, decoupled from the neutral atoms and expanded freely to become the cosmic background radiation now observed. As explained in the sidebar "Big Bang Cosmology and the Microwave Background," any density perturbation on a distance scale λ present at the time of recombination, t_{recomb}, should be imprinted on the cosmic background as a temperature variation on a distance scale $\lambda/a(t_{\text{recomb}})$. Thus the cosmic background provides a direct measure of the matter distribution at the time of recombination and an observational constraint on models of large-scale structure formation.

So far the resolution of the instruments on the COBE satellite limits measurements of inhomogeneity to very large distance scales, too large to differentiate the various models of large-scale structure formation. Nevertheless, density inhomogeneity was seen and its amplitude was large enough to be consistent with the idea that gravity was decisively responsible for thee formation of structure.

Initial conditions of the CDM model. The standard CDM model postulates very specific initial conditions for the development of large-scale structure. First it assumes that $\Omega = 1$, as suggested by inflationary models of the very early universe. Second, in line with the upper limit on the density of ordinary, or baryonic, matter allowed by primordial nucleosynthesis calculations (see "Big Bang Cosmology and the Microwave Background"), the standard CDM model assumes that 95 percent of the matter is nonbaryonic dark matter that is essentially noninteracting with ordinary matter. It also assumes that the dark matter was cold, or moving at nonrelativistic speeds, by the end of the radiation-dominated era. Finally, it postulates a scale-free spectrum of primordial density fluctuations—similar to the spectrum predicted by inflationary models. However, nonbaryonic cold dark matter has never been detected. So why should anyone take these assumptions about dark matter seriously?

Evidence for dark matter. The existence of dark matter (not necessarily cold) was first proposed by Fritz Zwicky in 1933 to explain how high-velocity galaxies observed in the very dense cluster known as the Coma cluster could remain gravitationally bound. More mass must exist in that cluster than was visible as luminous matter. In the 1970s a similar type of observation was made on the scale of single galaxies. Figure 4 shows the circular components of the orbital velocities of stars and gas clouds in two spiral galaxies deduced from very careful redshift measurements; the data are plotted as a function of radius r from the centers of the galaxies. In each galaxy the circular velocities, $v_{circ}(r)$, outside a radius of a few kiloparsecs are all approximately equal to a constant, v_{const}. Similar results are found in all spiral galaxies. On the other hand, quite a different prediction follows from the concentration of luminous matter toward the centers of those galaxies. The centrifugal force on a star orbiting the galactic center at a radius r must equal the gravitational force, or

$$\frac{mv_{circ}^2}{r} = \frac{GmM(<r)}{r^2},$$

where m is the mass of the star and $M(<r)$ is the mass of the galaxy inside the radius r. If the mass of the galaxy is distributed similarly to the luminous matter, one would conclude that the velocities of stars in roughly circular orbits far from the center, $v_{circ}(r)$, are proportional to $r^{-1/2}$, as is the case for planets in our solar system. But in spiral galaxies (including our own—see Figure 4) this is never the case!

To resolve the contradiction, astrophysicists postulated the existence of invisible (dark) matter distributed such that $M(<r)$ increases approximately with the radius:

$$M(<r) \approx \frac{v_{const}^2 r}{G} \propto r.$$

This distribution is less concentrated toward the center of the galaxy than the distribution of luminous matter and provides the additional gravitational force needed to keep the stars in orbit at the observed velocities.

The dark matter in a typical bright galaxy would form an invisible halo, as shown in the opening spread of this article, and would have to be at least ten times more massive than the luminous matter to explain the observed motions of stars in the galaxy. The estimate of Ω would therefore increase from about 0.01 (deduced from luminous matter) to about 0.1. All that matter might be made of baryons and still not violate the constraints on baryon density provided by primordial nucleosynthesis. Indeed, recent observations of a phenomenon called gravitational microlensing suggest that dark matter in the form of old stars or dense planet-like objects, presumably made of baryonic matter, is present in galactic halos.

On scales of megaparsecs, the translational motions of galaxies, especially those in clusters of galaxies, seem to imply the existence of still more dark matter, enough to bring the value of Ω to 0.2 ± 0.1. On the largest scales observed (30 to 100 megaparsecs), the densities deduced from comparison of the translational motions with the distribution of matter suggest that Ω must be at least as large as 0.3, and probably close to 1, which is the value assumed in the CDM model. Perhaps the most compelling argument in support of a large amount of cold dark matter is that its existence may well explain the observed structure of the universe with the fewest assumptions and the most natural physics.

The composition of cold dark matter. What might be the composition of cold dark matter? Particle theorists have helped attack this interesting problem by offering up a whole list of possible candidates. Those candidates, such as axions and photinos, have been predicted in theoretical models that extend the standard model of particle physics. They are called weakly interacting massive particles (WIMPS) and have both properties necessary to constituents of cold dark matter: First, they would interact very, very weakly with the rest of matter and radiation; in fact they would be practically undetectable. Second, they would have sufficiently large masses (typically, much larger than the mass of the proton) that their thermal velocities near the end of the radiation-dominated era would have been far slower than the speed of

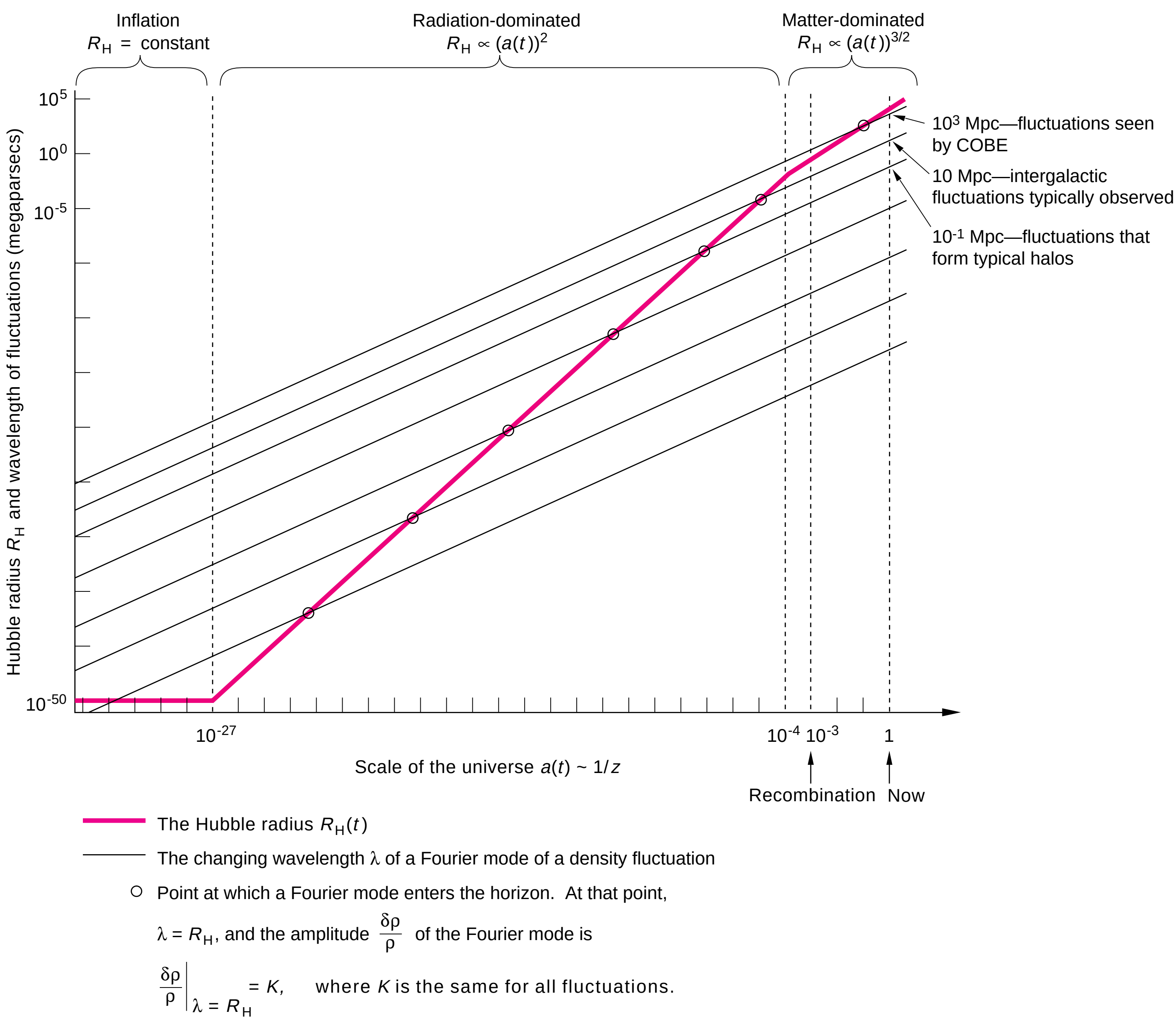


Figure 5. Horizon Crossing and the "Scale-Free" Spectrum

The log-log plot shows the wavelengths of the Fourier modes of the primordial density fluctuations, λ, as they stretch out because of the expansion of the universe. Here the expansion is measured by the scale factor, $a(t)$. Also shown is the Hubble radius, $R_H(t)$ (red), as a function of $a(t)$. As noted in the sidebar "Big Bang Cosmology and the Microwave Background," $a(t)$ increases exponentially with time (as $\exp(H_{inflation}t)$ during inflation, as $t^{1/2}$ during the radiation-dominated era, and as $t^{2/3}$ during the matter-dominated era. Since the Hubble radius is defined as $R_H(t) \equiv c/H(t) \equiv ca(t)/(da/dt)$, it is equal to the constant $c/H_{inflation}$ during inflation; it increases as $(a(t))^2$ during the radiation-dominated era, and it increases as $(a(t))^{3/2}$ during the matter-dominated era. Therefore after inflation the wavelength of each mode increases more slowly than $R_H(t)$. The wavelength of each mode crosses the Hubble radius at a particular time (circled on the graph), depending on the initial wavelength of that mode. At that time the fluctuation is said to "enter the horizon." The Harrison-Zel'dovich spectrum is defined to be scale-free in the sense that each mode has the same amplitude at the time it enters the horizon, or $\delta\rho/\rho\,|_{\lambda = R_H(t)} = K$, where K is the same for all modes.

light.* Density fluctuations in cold dark matter on galactic and smaller scales would therefore not be wiped out by the free streaming of the cold-matter particles from more dense to less dense regions. Consequently, those fluctuations could begin to grow under the influence of gravity at the start of the matter-dominated era, as postulated by the CDM model.

Primordial fluctuations in cold dark matter. The power spectrum is a convenient and succinct way of characterizing density perturbations in many (but not all) cosmological models. Density fluctuations as a function of position **r** can always be expressed in terms of a sum of elementary sinusoidal ripples, that is, through the Fourier expansion of $\delta\rho(\mathbf{r}) \equiv \rho(\mathbf{r}) - \bar{\rho}$, the deviation of the density from its average:

$$\delta\rho(\mathbf{r}) \propto \sum_{\mathbf{k}} a_{\mathbf{k}} \cos(\mathbf{k} \cdot \mathbf{r} + \phi_{\mathbf{k}}).$$

Above, **k** are the wavevectors of the ripples, $a_{\mathbf{k}}$ are the amplitudes of different modes corresponding to different wavelengths $\lambda = 2\pi/|\mathbf{k}|$, and $\phi_{\mathbf{k}}$ are the phases. In a large class of cosmological models (including CDM), it is assumed that phases do not matter, that they are random and uncorrelated between different modes. When this is true, one can focus solely on the amplitudes $a_{\mathbf{k}}$ and characterize the fluctuations by their average values, that is, by the power spectrum:

$$P(k) = \langle a_{\mathbf{k}}^2 \rangle,$$

where $k \equiv |\mathbf{k}|$. In addition, the deviations of density from the average, $\delta\rho(\mathbf{r})$ at various points **r**, turn out to have a Gaussian distribution.

Cold-dark-matter cosmology starts with the assumption that primordial density perturbations are Gaussian and their power spectrum has the form

$$\mathrm{P}(k) \propto k,$$

which is known as the Harrison-Zel'dovich spectrum. This spectrum follows approximately from the scale invariance of the process of inflation. (Inflation is a postulated period of exponential expansion at very early times, which solves certain cosmological problems related to causality.) This scale-free spectrum was actually proposed before inflationary models on the basis of an elegant feature: In a universe with $\mathrm{P}(k) \propto k$ and $\Omega = 1$, each density fluctuation has the same amplitude at the time it enters the horizon of the observable universe, that is, at the time its wavelength λ is equal to c times the age of the universe, or approximately equal to the Hubble radius, $c/H(t)$.

Figure 5 shows the increase in wavelength of each mode with time due to the expansion of the universe; note that the horizon increases faster than the wavelengths of the modes, so as time goes on, larger and larger modes fulfill the criterion that their wavelengths are about equal to $c/H(t)$, at which times they enter the observable universe. Such a universe has the appealing feature of being cosmologically scale-invariant: At any time the only important scale is defined by the size of the horizon at that time.

Linear growth of density fluctuations. The initial density fluctuations postulated by the CDM model would grow under the influence of gravity because, in a distribution of matter that is nearly uniform in density except for small "ripples," the gravity of denser regions tends to attract more mass from nearby regions, so the ripples become larger. As they become larger, they become even more effective at attracting matter, and so they continue to grow, possibly becoming progenitors of galaxies or galaxy clusters. In the CDM model and most models other than HDM, gravitational collapse occurred on small scales first and then on larger scales. Thus globular clusters formed before galaxies, and galaxies before clusters of galaxies.

Until the time of recombination and for a long time afterward, the growth of the density fluctuations can be modeled analytically because their average amplitude, or $\delta\rho/\bar{\rho}$, was small enough for the growth to be linear. Linear growth is calculated simply by computing the independent growth of each mode—the evolution of the power spectrum with time. In this linear regime the fate of each mode depends on whether it enters the horizon in the radiation-dominated or in the matter-dominated era.

Modes that enter the horizon in the radiation-dominated era are, in effect, ripples in the density of a plasma that is dominated by photons. The pressure of the plasma prevents gravitational development of these ripples and causes them to oscillate as sound waves. This state of affairs persists until t_{eq}, the time at which the energy density of matter equals the energy density of radiation. Only after t_{eq}, when the radiative contribution to the energy density becomes negligible, can density perturbations begin to grow. Thus, the growth of the modes with wavelengths smaller than the horizon at t_{eq} is delayed, and therefore stunted. As shown in Figure 5, these modes include the ones that develop into galaxies.

By contrast, the modes that enter the horizon in the matter-dominated era, well after t_{eq}, begin to grow immediately as a result of gravitational attraction. Thus by the time of recombination, when the fluctuations made a lasting imprint in the cosmic background, the the shape of the power spectrum $P(k)$

*Axions are an exception: They are less massive, but move slowly for other reasons.

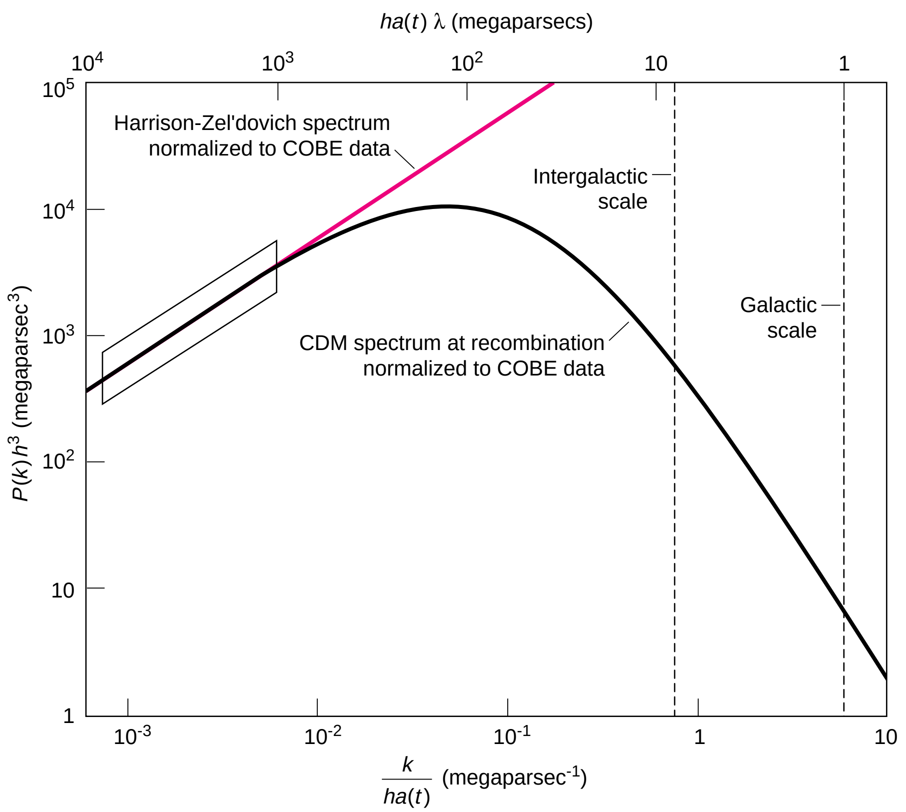


Figure 6. The Density-fluctuation Spectrum of Cold Dark Matter

Plotted is the density-fluctuation power spectrum, $P(k)$, of cold dark matter at the time of recombination. (For comparison with present structures, the horizontal scales give the wavenumbers k and wavelengths λ of the modes now, when the universe is $a(t_0)/a(t_{recomb}) \approx 1000$ times bigger.) The spectrum is normalized to agree with the COBE observations; the parallelogram at left shows the range allowed by the COBE data. Also plotted is the Harrison-Zel'dovich spectrum, $P(k) \propto k$ (gray), with the same normalization. As discussed in the text, the CDM spectrum has less power at small scales (large k) than the Harrison-Zel'dovich spectrum, because fluctuations on those scales enter the horizon during the radiation-dominated era, when they cannot grow. The relatively small scale labeled "intergalactic" is roughly the scale used for observations of the galaxy distribution (also shown in Figure 5). The comparison of the galaxy distribution to the CDM spectrum will be discussed in the text.

had changed from the primordial Harrison-Zel'dovich spectrum: Modes with wavelengths less than $c/H(t_{eq})$ had less power relative to modes with wavelengths greater than $c/H(t_{eq})$ than was the case in the primordial spectrum (see Figure 6).

Details of the CDM power spectrum depend somewhat on the assumed fraction of the matter that consists of baryons because the baryonic density distribution is affected by interactions with radiation during the radiation-dominated era, but for reasonable values of Ω_{baryon} those effects are minor. The crucial assumption determining the CDM power spectrum is that the dark matter is composed of cold particles at the start of the matter-dominated era—when the gravity of dark matter begins to matter!

Early computer simulations of nonlinear growth. Given the power spectrum of density fluctuations, $P(k)$, at the time of recombination, one can, in principle, compute the subsequent evolution of the matter distribution in the CDM universe for comparison with observations. The computation requires very large-scale computer simulations for two reasons. First, on scales of tens of megaparsecs and smaller, the observed amplitudes of density fluctuations $\delta\rho/\bar{\rho}$ are on the order of 1 and larger, and so their growth must have been nonlinear. Second, in order to resolve inhomogeneities on the relatively small scales of galaxies, 10 kiloparsecs, and at the same time show the distribution and motion of matter on the largest observed scales, many tens of megaparsecs and greater, the computation must involve a very large number of massive particles.

The early computer studies, initiated about ten years ago right after the CDM model was proposed, were inconclusive because they could not include enough particles. Typically, about ten thousand very massive particles represented all the matter in a region 10 to 100 megaparsecs across, and a single particle with a mass on the order of 10^{12} solar masses represented all the stars and dark matter in a galactic halo. Thus there was no convincing criterion for deciding which of those particles were prospective galactic halos that would become luminous due to star formation and which of those particles represented matter that would remain dark. Nevertheless those early "one-particle-per-galaxy" simulations gave hope that the CDM cosmology was right: The value of the normalization of the power spectrum could be chosen so that the simulations reproduced the observed relative motions of galaxies on small intergalactic scales (on the order of 1 megaparsec). They also reproduced the observed spatial distribution of galaxies provided one introduced a bias, an ad hoc assumption that the nonuniformity of galaxies exaggerates the nonuniformity of dark matter by some factor b greater than 1. More formally, if we define the number of galaxies in a given volume as N and the total mass in that volume as M, then on average

$$\delta N/\overline{N} = b\ \delta M/\overline{M}.$$

Those early studies required a bias b of 2.5, or equivalently only particles in very dense regions (on scales of 1 megaparsec) were identified as prospective galaxies. Later, smaller values of the bias and correspondingly larger values for the normalization of the power spectrum had to be adopted to account for observations on scales around 50 megaparsecs such as the famous "Great Attractor"—an observed flow of galaxies, including ours—and the distribution of rich clusters of galaxies called Abell clusters.

The introduction of a bias to match simulations with the observed spatial distribution of galaxies is not unreasonable, because the spatial distribution of galaxies is determined not just by the distribution of total matter in the universe, which is modeled by simulations, but also by the highly uncertain process of star formation. The data that more directly reflect the underlying matter distribution, and that therefore must be matched by simulations, are the observed peculiar velocities of galaxies, motions that are due not to the expansion of the universe, but to gravitational forces on galaxies from inhomogeneities in the local mass distribution, including dark matter. Therefore peculiar velocities have long been considered a better measure of mass inhomogeneity than correlations of galaxy positions.

Since galaxies are too far away for the detection of their motion transverse to our line of sight, only the components along our line of sight of their velocities—the components that affect redshift—are observed. Peculiar velocities show up as deviations from Hubble's law—provided there is some way to determine the distance to galaxies independent of velocity, or redshift, measurements. The naive method is to assume that galaxies near each other in the sky are members of the same cluster; then differences in their redshifts would arise from differences in their peculiar velocities rather than from differences in distance.* The actual procedure for determining peculiar velocities is a sophisticated statistical application of the same principle. The analysis yields the component along the line between two galaxies of the inferred peculiar-velocity difference between them. That component is called the pairwise radial velocity. The distribution of peculiar velocities is characterized by the standard deviation of pairwise radial velocities, which is written as σ_v. Measurements seemed to indicate that for pairs of galaxies on the order of 1 megaparsec apart, σ_v is 300 to 400 kilometers per second.

Before the 1992 COBE results, scientists would vary the normalization of the power spectrum to achieve agreement between the peculiar velocities predicted by the simulation and those observed and then vary the bias to match the observed spatial distribution of luminous matter. Since none of the simulations were able to achieve high resolution on both large and small scales simultaneously, the results were ambiguous.

*These differences in radial velocities are responsible for the distorted appearance of clusters in galaxy maps based on redshift, such as Figure 1a. The Coma Berenices cluster in that figure probably has an approximately spherical shape, but the peculiar velocities of its galaxies give them widely different redshifts, elongating the plotted spatial distribution of the cluster into a "finger of God" pointing directly at Earth.

Cold Dark Matter, Large Scales, and COBE

The freedom to vary the normalization of the power spectrum and "fine-tune" the bias disappeared in early 1992 after the announcement that the COBE satellite had detected microwave-background fluctuations of a few parts in 10^6 on scales of 1000 megaparsecs and higher. Since the shape of the CDM fluctuation spectrum is determined, the COBE measurements fix the normalization constant and thus the entire spectrum.*

The COBE results also determine the bias b between the distribution of matter and the distribution of luminous matter. One extrapolates the CDM spectrum to the present and compares the resulting spectrum to the present amplitude of density fluctuations. The result for scales around $10/h$ megaparsecs (the intergalactic scales shown on Figures 5 and 6) is $\delta\rho/\rho \approx 1$ with a standard deviation of about 20 percent. One can compare that predicted amplitude with the quantity conventionally used to report the distribution of luminous matter, namely, the amplitude $\delta\rho/\rho$ on the scale of $8/h$ megaparsecs, known as σ_8. That is,

$$\sigma_8 \equiv \frac{\delta\rho}{\rho}\Big|_{\lambda\,=\,8/h\ \mathrm{Mpc}}.$$

*There is still some freedom in the baryonic content of the universe and, perhaps more important, in the exponent of the primordial power spectrum of density fluctuations, which would be exactly scale-invariant only if it were generated by endless inflation.

The observed value of σ_8 for luminous matter is also about 1; thus the COBE data exclude the possibility of a significant bias.

High-resolution state-of-the-art simulations. Spurred by the COBE observations, we recently carried out large computer simulations of structure formation in a CDM universe. The simulations were large enough to resolve the formation of prospective galaxies on kiloparsec scales and to model the structure and motions of sheets and filaments of matter on scales four orders of magnitude larger.

The simulations involve either 9 or 17 million point particles moving under the influence of mutual gravitational attraction in an expanding spherical fragment of the universe. (Since the particles have no internal degrees of freedom, and the only interaction in the simulation is gravity, no energy is dissipated.) The fragment expands during the simulation to a final diameter of 100 or 250 megaparsecs at a redshift of $z = 0$, which corresponds to the present time, $t = t_0$. For numerical reasons the simulation starts at a redshift z between 50 and 100, that is, at a time well after the time of recombination ($z = 1000$) but before the onset of nonlinear growth. To determine the density-perturbation spectrum at the start of the simulation, we use linear theory to calculate the spectrum as a function of z from the CDM spectrum at recombination under the assumption that the fluctuations grow linearly with the scale factor. The universe is assumed to be flat, the Hubble parameter is assumed to be 50 (km/s)/Mpc, and the normalization of the spectrum in each of the runs is selected to bracket the value suggested by the COBE measurements. We cannot do this directly, but rather fine-tune the normalization to yield the desired value of σ_8 at $z = 0$.

The spectrum of fluctuations at the starting value of z is built into the simulation by arranging the particles in a regular array and assigning them different masses selected at random to match the predetermined spectrum. To include the effects of the universal expansion, the particles are given initial velocities in accordance with Hubble's law. The simulation then follows the motion of this collection of particles. At each timestep, the changes in the position and velocity of each particle are found by first calculating the gravitational force on each particle produced by the other particles and then integrating Newton's second law of motion over the duration of each timestep. The forces are calculated according to the familiar formula for Newtonian gravity.

The use of Newtonian gravity together with initial conditions in accordance with Hubble's law may appear to be a poor man's version of the general model for a relativistic universe, but it can be shown (see the caption of Figure 3) that, to a good approximation, a region of the universe really does behave according to that description provided that the region is matter-dominated and is significantly smaller than the Hubble radius c/H_0. Nevertheless, only after general relativity was understood did scientists become bold enough to apply Newton's laws to the cosmos.

In a standard simulation of N particles, the time needed to calculate the gravitational force between each pair of particles and to sum up the total force on each particle in each timestep is proportional to N^2. When N is in the millions, as in our simulations, that time is prohibitively long even on the most powerful computers. We have been able to reduce the time significantly by replacing pairwise interactions between distant particles with the well-known multipole approximation for the gravitational force exerted by a group of particles on a distant particle. The approximation is implemented by using a hierarchical algorithm called a treecode in which the problem domain is subdivided into cubes of decreasing size, the first terms of the multipole expansion are computed for the particles in each cube, and the multipole approximation of the gravitational force is applied whenever the accompanying errors are negligible. Treecodes calculate forces involving millions of particles thousands of times faster than conventional algorithms. Their execution time depends on the number of particles as $N \log N$ rather than N^2, a very significant reduction when N is large.

Our treecode and its potential applications in other fields are described in "A Fast Tree Code for Many-Body Problems." The simulations presented here were run on the Intel Touchstone Delta, a parallel supercomputer owned partly by the Laboratory and installed at Caltech.

Figure 7 illustrates the development of structure in the simulation. A typical final distribution of particles appears in Figure 8. Many of the clumps shown in those figures are made up of hundreds of particles and have masses in the range characteristic of a galactic halo. We interpret those clumps as prospective galactic halos; the final distribution of halos is shown in Figure 9, and a detailed distribution of both particles and halos in a small area is shown in Figure 10.

Once the halos are identified, we analyze local halo dynamics including internal rotation, collapse, and merger. At the same time, since each simulation generates approximately 10,000 such halos, we can work with statistically significant numbers in investigating the relative motions and overall spatial distribution of halos on scales of tens of megaparsecs. The size of the halos and their spatial distribution are not in any

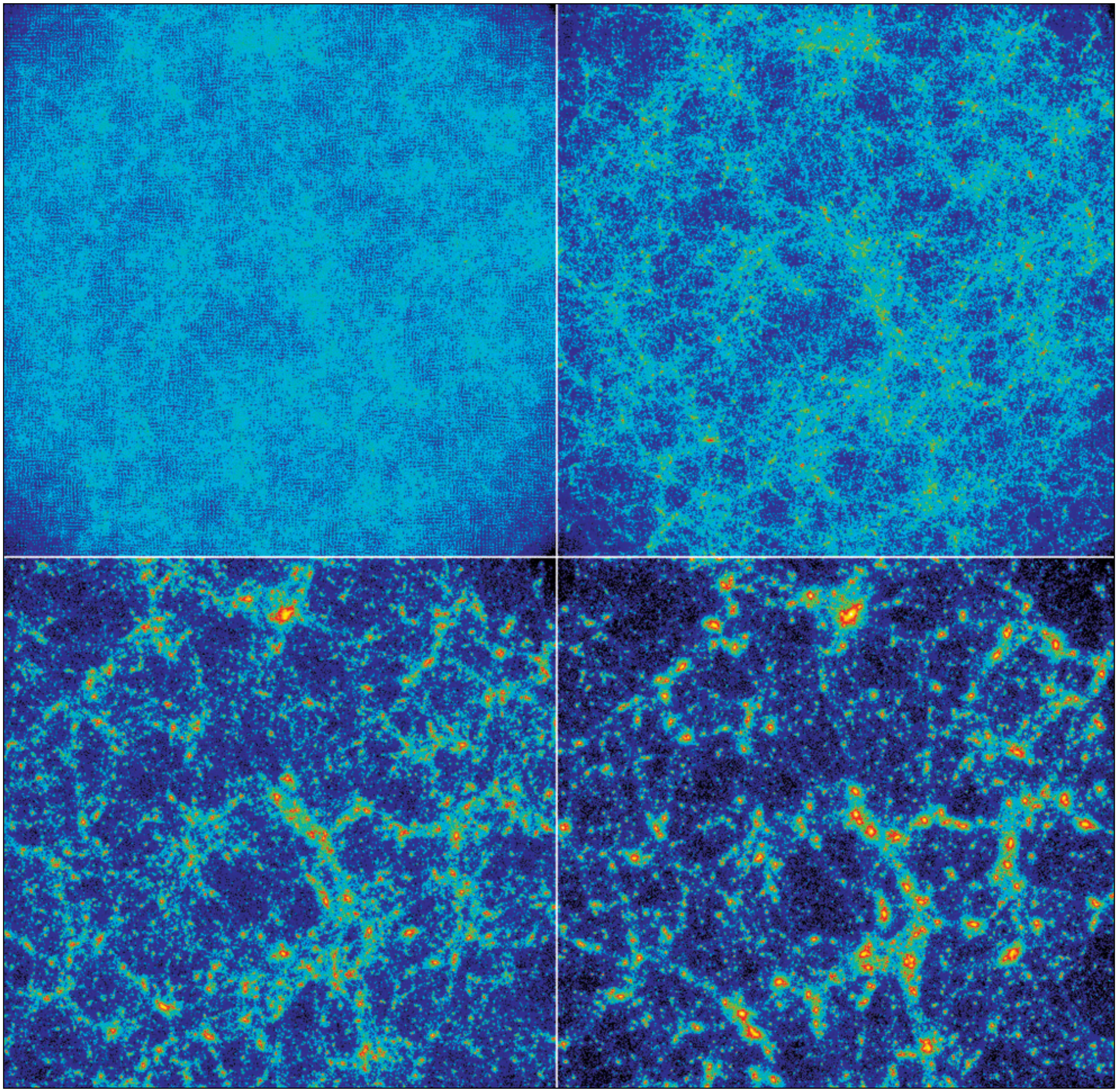


Figure 7. The Evolution of Structure in a Simulation

These four frames show consecutive stages of the particle distribution in one of our simulations of the matter distribution. The simulation starts a few tens of million years after the Big Bang ($z = 64$), shortly before density fluctuations began to grow nonlinearly. The first frame is at $z = 10$; the last is at the present. The color of each pixel indicates the logarithm of the particle density along the line of sight through the computational volume; blue indicates the lowest density, then cyan, green, red, yellow, and white. As the simulation progresses, the mass becomes more clumped, eventually forming structures qualitatively similar to those observed. The region of space shown expands as the universe expands; it is about 200 megaparsecs across in the final frame.

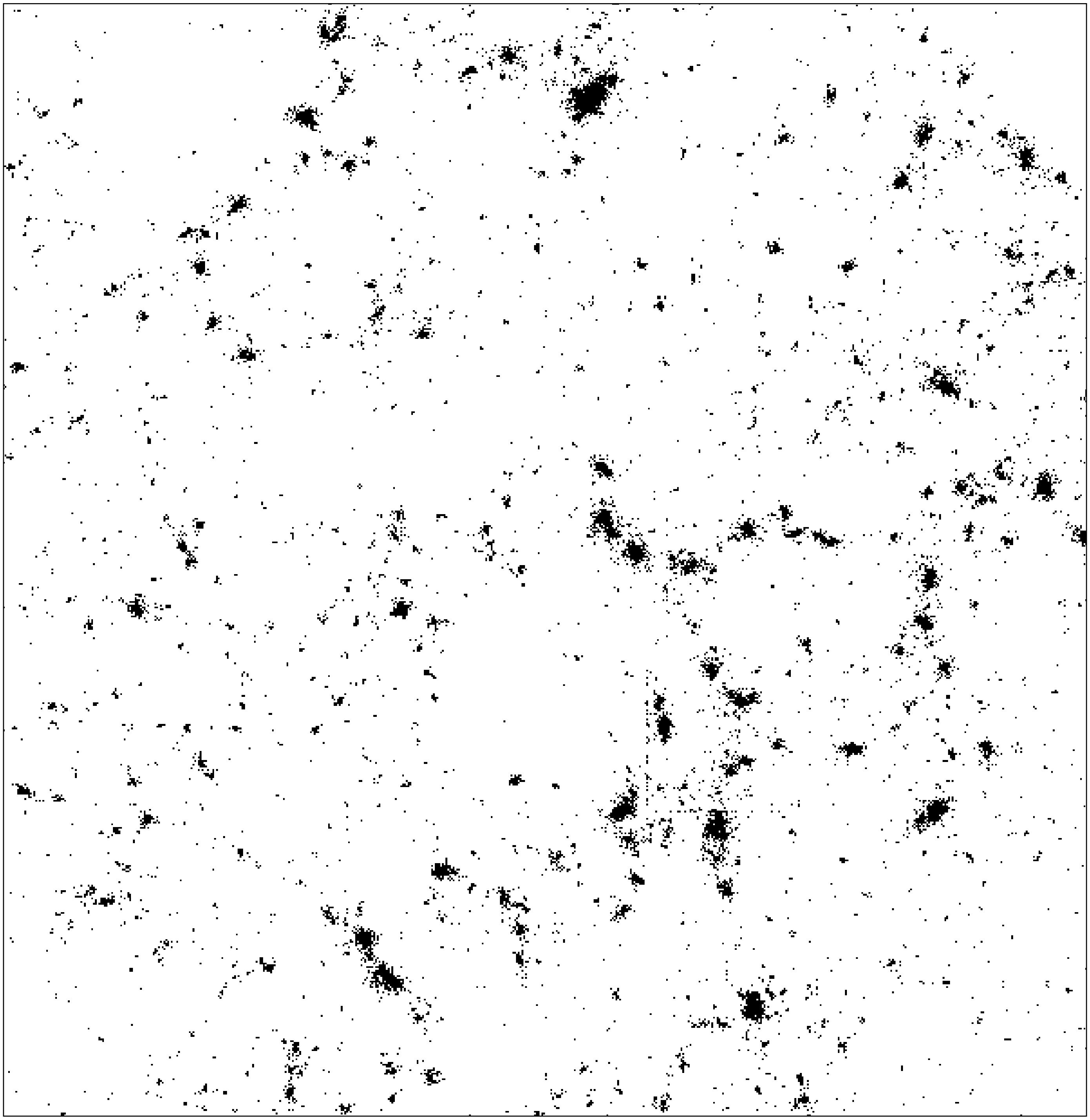


Figure 8. Particles at the End of a Simulation

The figure shows the projected location of particles in a sphere of diameter 250 megaparsecs at the end of a simulation. The distance across the image is about 180 Mpc (part of the sphere is cut off). Only about one in every thirty particles is plotted. The figure shows that the distribution is inhomogeneous on all scales. The square shows the region enlarged in Figure 10. The treecode used for the simulation is described in "A Fast Tree Code for Many-Body Problems," accompanying this article.

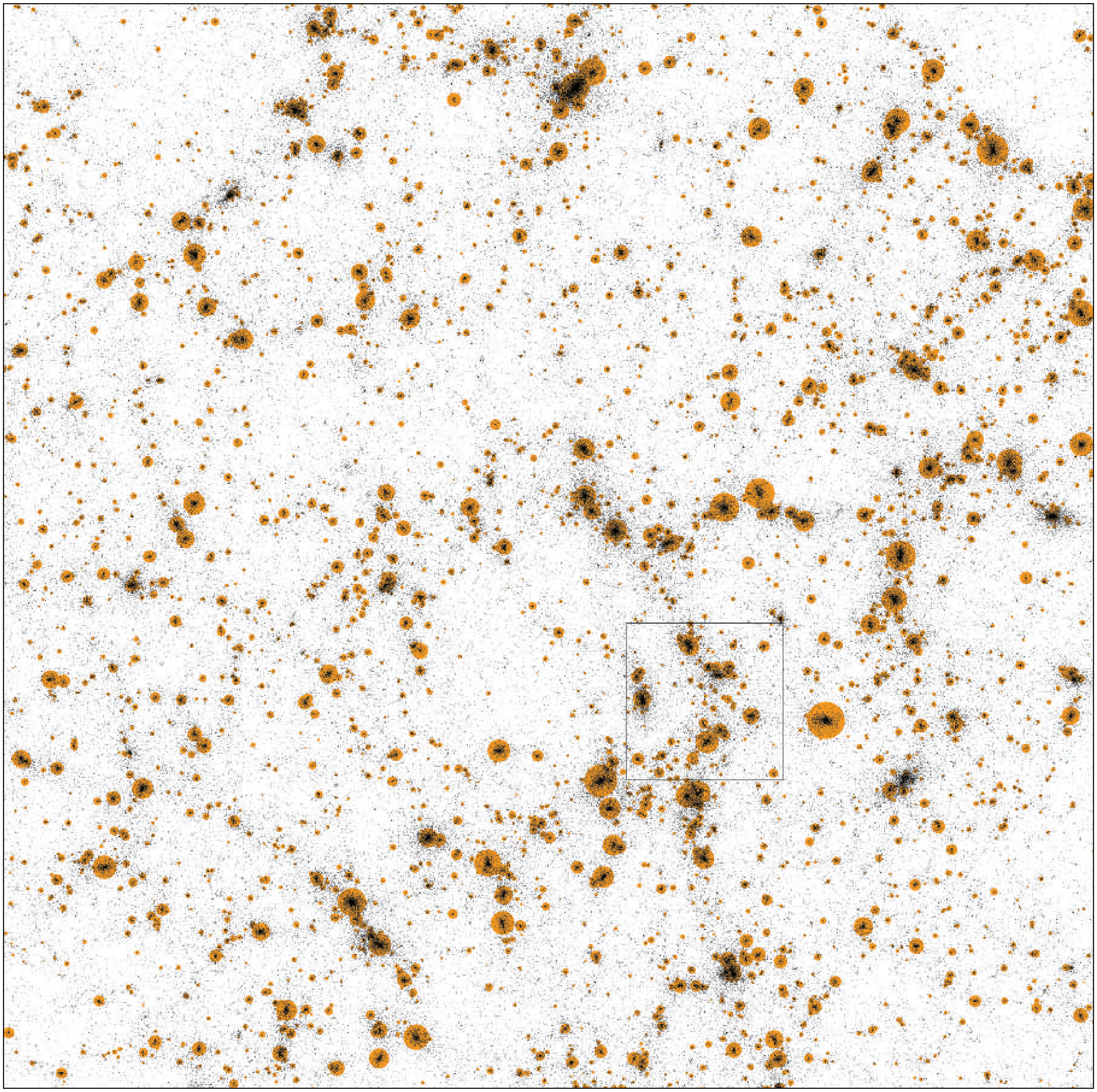


Figure 9. Halos at the End of a Simulation

A plot of the location of "galactic halos" in the system of particles shown in Figure 7, as determined by an algorithm that locates clumps of particles. Halos are defined to be regions whose centers are at least 10,000 times denser than the average density of the universe and contain at least 10 particles. The sizes of the circles are proportional to the masses of the halos. There are about 6000 such halos in this picture. Note that clusters, voids, and sheets (which appear as narrow concentrations on this two-dimensional projection) are present, as in the observations.

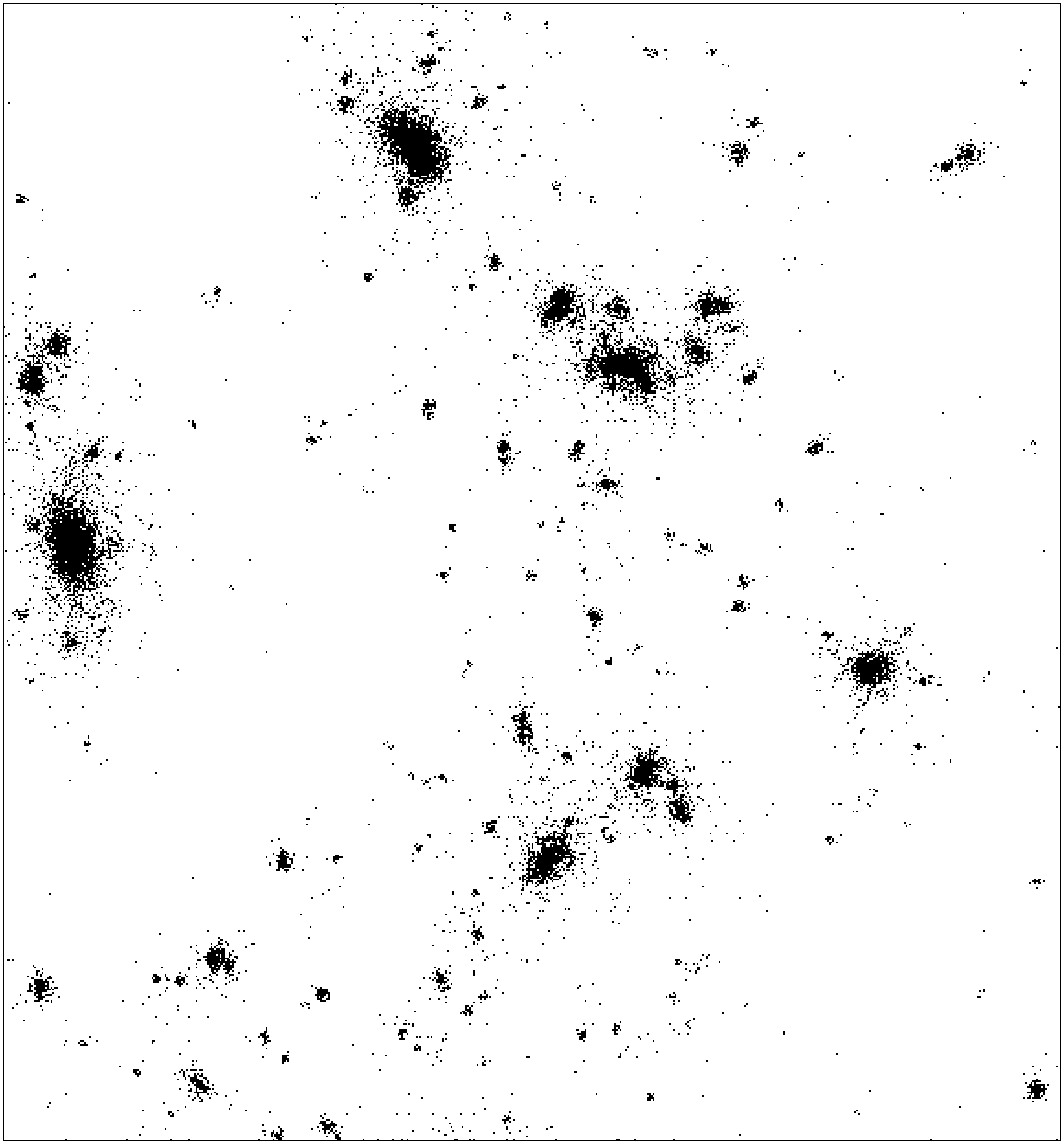


Figure 10. Detailed Particle Distribution at the End of a Simulation
A high-resolution plot of every particle (about 250,000) in one fiftieth of the total system (indicated by the small square in Figure 8). The circles represent galactic halos as identified in Figure 9.

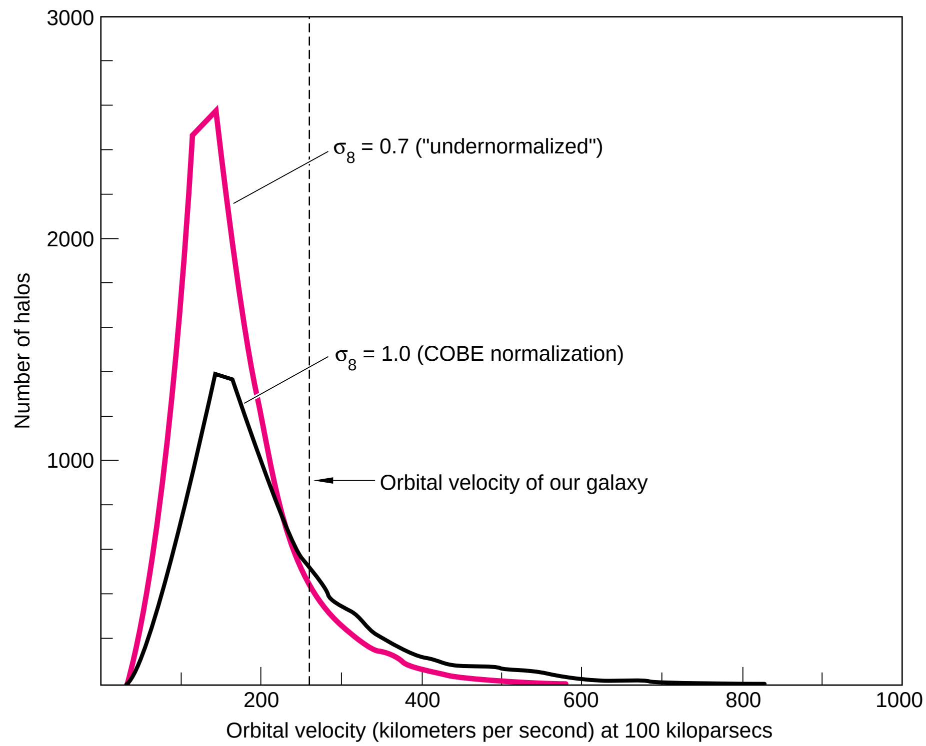


Figure 11. Distribution of Halo Number as a Function of Inferred Orbital Velocities of Stars in Halos

The graph shows the number of halos in our simulation versus the inferred orbital velocity, v_{circ}, of a star around the center of each halo at a radius of 100 kiloparsecs. The radius of 100 kiloparsecs was chosen because orbital velocities of stars around galactic centers are usually independent of distance at such large distances (as in Figure 4). The velocity associated with each halo is inferred from the mass distribution within the halo. The most abundant halos produced in our simulation are those with the lowest masses that the simulation can resolve, and thus they usually have low orbital velocities. Therefore the peaks of the curves shown are artifacts of the simulation, not characteristics to be compared with observations. We show results from runs of the simulation using two different normalizations of the density-fluctuation spectrum: $\sigma_8 \approx 1.0$ (black curves), the naive prediction from COBE measurements, and a lower normalization, $\sigma_8 \approx 0.7$ (red curves), which, as argued in the text, may be a much better estimate. The vertical line at 260 kilometers per second indicates the velocity of stars orbiting the center of our galaxy; the highest velocities observed in other galaxies are about 600 kilometers per second. Thus the rotational velocities predicted by our simulation are on the right order of magnitude, indicating that the local mass distribution within halos is about right. The lower normalization gives better agreement with observations since it does not lead to orbital velocities in excess of 600 kilometers per second.

obvious disagreement with observation, as can be shown by computing halo-halo correlation functions as well as other measures of the halo distribution, and comparing them with similar quantities computed on the basis of observational data. To evaluate the distribution of mass within each halo and check whether it is compatible with the observations, we used the positions of the tens or hundreds of particles in each halo to determine for each halo $M(<r)$, the mass within a radius r as a function of r. Then, using the previously discussed formula for circular velocity, $v_{circ} = \sqrt{GM(<r)/r}$, we determined the orbital velocities that stars would have around the center of the halo. The mass distributions of our halos do indeed look like those inferred from the observations of spiral galaxies: $M(<r)$ in our simulated halos is roughly proportional to r at large r, so orbital velocities of stars would be roughly independent of distance (see Figure 4). The average orbital velocity, or equivalently the average mass, of the halos also seems to be in general agreement with observation. Figure 11 is a plot of the number of prospective galactic halos versus the inferred orbital velocity of a star at a distance of 100 kiloparsecs from the center of the halo. The two curves in the figure are derived from simulations with different normalizations of the density-fluctuation spectrum, one specified by the COBE value of σ_8 and the other by a somewhat lower value. The peaks in both curves, at a velocity of 150 kilometers per second, are probably an artifact of our resolution—we cannot resolve halos less massive than about one tenth of the mass of our galaxy. A more important feature is that relatively few orbital velocities are much higher than the orbital velocity of stars in the Milky Way around the galactic center, about 260 kilometers per second. In previous simulations, low resolution had caused halos to merge into overly large halos, so the inferred orbital velocities were considerably higher than those observed.

Apart from that "overmerging" problem, CDM cosmology never had much difficulty accounting for spatial structure on megaparsec scales and smaller. The model ran into serious trouble only when it had to account simultaneously for both the spatial distribution and relative motions of galactic halos. Figure 12 shows results from our simulations for the root-mean-square pairwise radial velocities of halos as a function of the distance separating the halos. (The root-mean-square pairwise radial veloc-

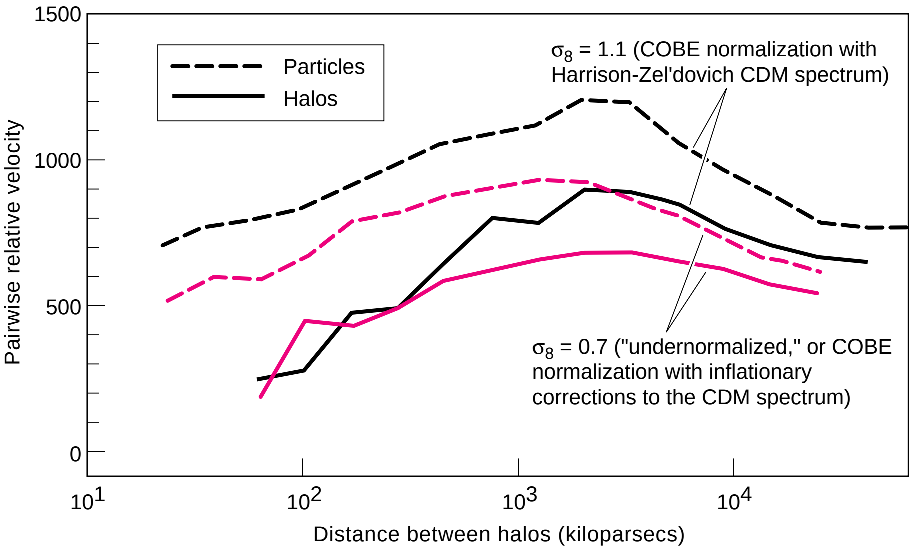


Figure 12. Pairwise Radial Velocities of Halos

The solid lines indicate the root-mean-square value of the pairwise radial velocity—the differences of the velocities of two halos along the line connecting them, excluding the Hubble expansion—as a function of separation of the halos. The dashed lines show the relative velocities of particles in the simulation rather than those of halos. As in Figure 11, the simulation was run twice, using the COBE normalization (black curves) and a lower normalization (red curves). The data from the lower-normalization run are in the same range as the observed pairwise radial velocities of real galaxies, provided the same statistical procedures are applied to the simulated and observed data (see discussion in the main text).

ity is closely related to σ_v, the standard deviation of pairwise radial velocities. The difference in our simulation is about 20 percent.) Again the two sets of results derive from the different normalizations used in the simulations. For the lower normalization the σ_v of halos separated by 1 to 2 megaparsecs is on the order of 600 kilometers per second. Figure 13 shows the distribution of pairwise radial velocities for halos separated by 1 to 2 megaparsecs.

The value of σ_v of 600 kilometers per second obtained from our high-resolution simulation is less than half of the value indicated by the old "one particle per galaxy" CDM simulations, but still appears to be too large compared with the usually quoted observational value—340 ± 40 kilometers per second on megaparsec scales. At first we thought that this discrepancy ruled out the CDM scenario. However, having re-examined observed and simulated relative velocities more carefully, we have concluded that the quoted value for σ_v is not a very reliable diagnostic for evaluating the global dynamics of simulations. In particular, the values deduced can vary by nearly an order of magnitude from dense clusters to underdense "backwaters" like our own local group of galaxies. Similarly, large variations in σ_v (factors of 2 to 5) are seen among fragments of our simulated universes containing roughly 1400 galaxies. These fragments are comparable in size to the regions of the universe usually surveyed to measure σ_v.

The usually quoted value of σ_v was obtained by M. Davis and P. J. E. Peebles from their analysis of the sample of about 1200 galaxies in the northern sky compiled by the Harvard-Smithsonian Center for Astrophysics. This sample contains the galaxies in the Virgo cluster, which move at high speeds and tend to affect relative velocities to an extent disproportionate to their small numbers. In addition, overall infall into the Virgo cluster results in significant infall velocities quite far from its core. Davis and Peebles employed a special procedure to compensate for some of those effects. When we reanalyzed the observations without such special treatment, we found that the standard deviation of the pairwise radial velocities was around 500 kilometers per second, in close agreement with our simulations, especially those with the lower normalizations. Thus, contrary to recent popular prejudice, the observations of pairwise radial velocities do not seem to rule out CDM cosmology.

Our results for the relative velocities of halos and the orbital velocities within halos agree particularly well with observations when we take the value of σ_8 to be 0.7, rather than 1.0 to 1.3 as has been inferred from the COBE measurements of temperature variations. We would argue that the agreement is not pure happenstance, but rather that the value close to 0.7 is the relevant value for galaxy formation. First, the density-fluctuation power spectrum resulting from inflation is proportional not to k but to $k^{1-\epsilon}$, where $\epsilon = 0.02$ to 0.05. This small difference is not important at the long wavelengths (small k) measured by COBE, but it does affect the extrapolation of the COBE results to the shorter-wavelength (larger-k) modes relevant to galaxy formation, as it reduces the predicted amplitude of modes on megaparsec scales. Second, COBE infers differences in temperature from differences in radiation intensity, but those differences can be caused by gravity waves as well as by density variations. Gravity waves do not develop into astronomical structures. Therefore the part of the COBE temperature-fluctuation data due to gravity waves must be estimated and subtracted to obtain the part due to density fluctuations.

These two effects are interrelated and for small ϵ have similar magnitudes. Taken together, they imply that the primordial density perturbations at short wavelengths (large values of k) have less power than that usually in-

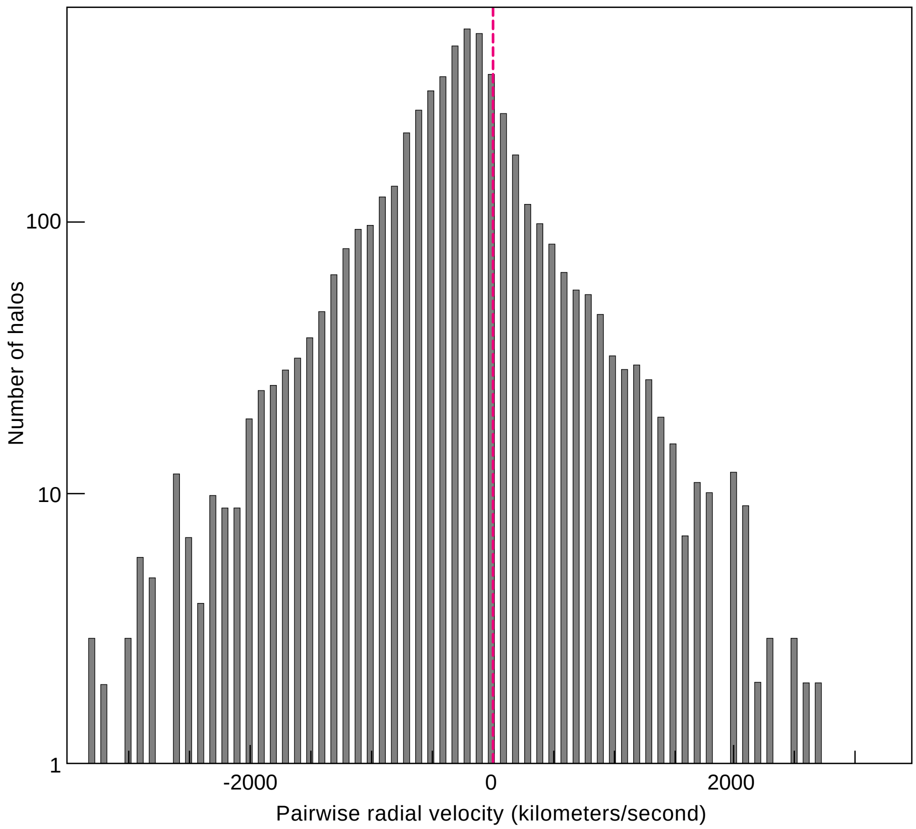


Figure 13. Distribution of Pairwise Relative Velocities of Halos at 1–2 Megaparsec Separations

The graph shows the number of pairs of halos in which the two halos are separated by 1–2 megaparsecs versus the pairwise relative velocities of the two halos in the pair. These results are from a simulation with $\sigma_8 \approx 0.7$. The center of the distribution is not at 0 but at -280 km/s; the negative average velocity indicates a net infall, or movement of halos toward each other, that is, toward local mass concentrations. This infall, due to gravitational attraction, is similar to observed movements of real galaxies. The local increases in density as nearby halos move toward each other are examples of the gravitational growth of inhomogeneity. Note that the graph appears as a triangle on this log plot, so the distribution of velocities is exponential. Therefore σ_v at this separation is simply proportional to the slopes of the sides of the triangle.

ferred from the COBE data; extrapolating that result to the scale of our simulations implies that σ_8 is about 20 to 40 percent less than the COBE data would seem to imply. Thus the agreement between our "undernormalized" simulations and the observational data might not be fortuitous; rather it might be telling us that inflation really did happen and that the resulting primordial fluctuation spectrum was only approximately of the Harrison-Zel'dovich form. If so (and admittedly we are getting carried away with optimism here), our simulations would provide the first concrete example in which detailed effects of inflation on the primordial fluctuation spectrum have to be taken into account to reconcile theoretical predictions with observations!

Nevertheless, the CDM model is still not "out of hot water." Although the discrepancy of relative velocities of halos was considered the most serious difficulty with the model, there are other tests that deal more directly with the spatial distribution of galaxies. One observational uncertainty that will have a major impact on assessing the validity of cosmological models is the present value of the Hubble parameter, H_0. All of the models that assume that $\Omega = 1$ (as does the standard CDM model) would be seriously endangered if H_0 turned out to be significantly larger than the currently favored value, 50 (km/s)/Mpc. The value of H_0 (in combination with the assumption that the universe is "flat," that is, that $\Omega = 1$) fixes the age of the universe. According to present ideas of stellar evolution, the oldest stars we observe are approximately 14 billion years old, which is about equal to the age of the universe if $H_0 = 50$ (km/s)/Mpc and $\Omega = 1$. If H_0 turns out to be 100 (km/s)/Mpc, those stars would be twice as old as the universe! Such a discrepancy would cause, of course, a major "paradigm shift” in cosmology, as it could be accommodated only by admitting that we are missing some important aspects of stellar evolution (which appears unlikely) or that the universe is open—has $\Omega \approx 0.1$.

Conclusions

The general framework of Big Bang cosmology has been very much strengthened by the COBE results. The cosmic background radiation appears to have a spectrum (that of emission from a perfect black body) and a temperature consistent with the predictions of the Big Bang theory. Furthermore, the COBE observations showed that the form of the primordial perturbations of the density of the universe was consistent with an approximately scale-free spectrum and that their amplitude was sufficiently large that they could act as seeds for the gravitational development of the currently observed density fluctuations, so that no forces other than gravity need to be invoked. The traditional problems with the CDM scenario follow from the *excess* of power on small scales and the resulting excess of the peculiar velocities between the galactic halos (and, therefore, presumably, between galaxies). We believe, however, that the results of our simulations are consistent with the observations when the simulations and observations are analyzed in the same way. Therefore there is no reason yet to abandon the CDM scenario.

Nevertheless, there is an almost embarrassing richness of insufficiently explored alternatives, one of the more popular being hybrid models involving both hot and cold dark matter. Fortunately, all of the alternatives will soon confront even more precise observational data, and their success will be evaluated by means of much more accurate simulations. Indeed, the list of models ruled out by COBE and other observations (with the help of computer simulations) is becoming longer than the list of viable models. “Experimental cosmology,” using parallel supercomputers and sophisticated software, will undoubtedly help bring about continued progress in this exciting and fundamental field. The computer is becoming as important as the telescope in shaping our understanding of the universe. ■

Further Reading

G. Efstathiou, J. R. Bond, and S. D. M. White. 1992. *COBE* background radiation anisotropies and large-scale structure in the Universe. *Monthly Notices of the Royal Astronomical Society* 258: 1P–6P.

Margaret J. Geller and John P. Huchra. 1989. Mapping the universe. *Science* 246: 897–903.

P. J. E. Peebles. 1993. *Principles of Physical Cosmology.* Princeton University Press.

Vera C. Rubin. 1983. The rotation of spiral galaxies. *Science* 220: 1339–1344.

Frank H. Shu. 1982. *The Physical Universe: An Introduction to Astronomy.* University Science Books.

G. F. Smoot *et al.* 1992. Structure in the *COBE* differential microwave radiometer first-year maps. *The Astrophysical Journal* 396: L1–L5.

M. S. Warren, P. J. Quinn, J. K. Salmon, and W. H. Zurek. 1992. Dark halos formed via dissipationless collapse. *Astrophysical Journal* 399: 405–425.

W. H. Zurek, P. J. Quinn, J. K. Salmon, and M. S. Warren. 1993. Large-scale structure after COBE. Los Alamos National Laboratory unclassified release LA-UR-93-2104. *Astrophysical Journal*, in press.

W. H. Zurek, P. J. Quinn, J. K. Salmon, and M. S. Warren. 1993. The second coming of cold dark matter? Los Alamos National Laboratory unclassified release LA-UR-93-4240. In *Proceedings of the Ninth IAP Meeting*, edited by F. Bouchet, in press.

Michael S. Warren (left) received his B.S. in physics, engineering, and applied science from the California Institute of Technology and his M.S. and Ph.D. in physics from the University of California, Santa Barbara. He came to the Laboratory in 1990 as a graduate research assistant in the Theoretical Astrophysics Group and is currently performing post-doctoral work in the Fluid Dynamics Group. In 1992 Warren received the Gordon Bell Prize and the Intel Grand Challenge Computing Award for the work described in “A Fast Tree Code for Many-Body Problems.”

Wojciech H. Zurek was born in Poland and received an M.Sc. from the Technical University of Kraków. He received his Ph.D in Physics from the University of Texas at Austin. Zurek joined the Laboratory’s Theoretical Astrophysics Group as an Oppenheimer Fellow in 1984 and has remained in that group ever since; he is now its leader. He is also an external faculty member of the Santa Fe Institute and an associate of the Cosmology Section of the Canadian Institute for Advanced Research. His research interests include cosmology, relativistic astrophysics, the quantum theory of measurements, the foundations of statistical physics, and the physics of information. In 1988 Zurek was awarded the Los Alamos Fellows Prize for his research on galaxy formation.