Cosmological *N*-Body Simulations

Edmund Bertschinger and James M. Gelb

Citation: Computers in Physics **5**, 164 (1991); doi: 10.1063/1.4822978
View online: https://doi.org/10.1063/1.4822978
View Table of Contents: https://aip.scitation.org/toc/cip/5/2
Published by the American Institute of Physics

Cosmological *N*-Body Simulations

Edmund Bertschinger and James M. Gelb

Modern simulations use millions of particles to follow thousands of galaxies in a large volume of space

Gravitational N-body simulations are a widely-used theoretical tool in astrophysics and cosmology. Given the dominance of the gravitational force on large scales, the complexity of nonlinear dynamics, and the inability of astrophysicists to perform real experiments, N-body simulations have an important role to play in helping us to understand the formation and evolution of galaxies and larger structures.

The first gravitational N-body simulation was Holmberg's remarkable study of a tidal encounter between galaxies using an analog optical computer: each galaxy was represented by 37 light bulbs, with a photocell and galvanometer measuring the inverse square law.[1] Modern simulations use millions of particles to follow thousands of galaxies in a large volume of space. Cosmologists believe that most—perhaps 95%—of the mass in the universe may be in the form of some unknown and invisible particles collectively called "dark matter." Normal nuclear matter forms the luminous stars produced after matter collapsed into galaxies. The dark matter is believed to have no significant interactions save gravity, and it is thought to dominate the mass everywhere except in the stellar cores of galaxies. To achieve a basic understanding of galaxy formation and clustering, therefore, it may be sufficient only to follow the gravitational interactions of dark matter. The purpose of cosmological N-body simulations is to foster such an understanding.

This article discusses some of the physical and computational aspects of large cosmological N-body simulations that we have performed.

Edmund Bertschinger is an Assistant Professor of Physics at the Massachusetts Institute of Technology. His research interests include cosmology, general relativity and interstellar gas dynamics. James M. Gelb, a Ph.D. student in physics at MIT, has worked extensively in computational physics. His research interests are in cosmology and the numerical modeling of galaxy formation.

Equations of Motion

Cosmological N-body simulations aim to represent the dynamics of galaxies in a large volume (from ~ 10 to 10^3 Mpc across, where 1 Mpc $= 3.26\times 10^6$ light years), whose overall motions are dominated by the mean cosmological expansion, described by a scale factor $a(t)$. In the absence of perturbations, two particles separated initially by $\vec{r}_1$ at time t_1, will, at time t_2, be separated by $\vec{r}_2 = \vec{r}_1 a(t_2)/a(t_1)$. The relative velocity between any pair of particles then obeys Hubble's law $\vec{v} = H\vec{r}$, where $H(t) = \dot{a}/a$ is the Hubble parameter. It is convenient to remove this zeroth-order motion by using a uniformly expanding system of "comoving" coordinates $\vec{x} = \vec{r}/a(t)$. Written in comoving coordinates, Newton's second law for a particle with comoving trajectory $\vec{x}(t)$ is:

$$\frac{d^2\vec{x}}{dt^2} + 2H\frac{d\vec{x}}{dt} = -\vec{\nabla}\phi, \tag{1}$$

where ϕ is the perturbed gravitational potential. This Newtonian equation is correct even in general relativity as long as the gravitational fields are weak and the motions in the comoving system are nonrelativistic.

The gravitational potential ϕ follows from the usual Poisson equation, but with only the spatially varying part of the mass density field, $\delta\rho/\rho \equiv [\rho(\vec{x},t) - \bar{\rho}(t)]/\bar{\rho}(t)$, providing the source:

$$\nabla^2\phi(\vec{x},t) = 4\pi G[\rho(\vec{x},t) - \bar{\rho}(t)] = \tfrac{3}{2}\Omega H^2\frac{\delta\rho}{\rho}. \tag{2}$$

The cosmological density parameter $\Omega(t) = 8\pi G\bar{\rho}/3H^2$ determines the overall dynamics and geometry of the universe: if $\Omega < 1$, the universe will expand forever and space is negatively curved (an "open" universe); if $\Omega > 1$, the universe will recollapse and space is positively curved

(a "closed" universe); if $\Omega = 1$, the universe continually decelerates but expands forever and space is Euclidean (a "flat" universe).

In an expanding universe with a non-exotic equation of state, $\Omega = 1$ is an unstable equilibrium point: $|\Omega - 1|$ increases with time. Most astronomical estimates yield $\Omega \lesssim 0.3$ today[2], but it is possible that not all of the dark matter between galaxies has been accounted for in these estimates. Many cosmologists favor the value $\Omega = 1$, which would be a good approximation at early times in any case. The inflationary universe paradigm[3] predicts that today $|\Omega - 1| < 10^{-4}$. According to this theory, an exotic false vacuum state in the very early universe *accelerated* the expansion, temporarily making $\Omega = 1$ a *stable* equilibrium and driving $\Omega - 1$ toward zero. One of the key goals of cosmological N-body simulations is to aid the determination of Ω by testing which cosmological model—open, closed or flat—gives the best agreement with the patterns of structure observed in the universe.

Initial Conditions

In the early universe, the matter distribution was almost perfectly uniform, with small-amplitude density fluctuations generated by quantum zero-point fluctuations or by some other mechanism. The fluctuations are described by the density power spectrum $P(k)$, defined so that the contribution to the variance of $\delta\rho/\rho$ from plane waves in the wavenumber volume element d^3k at a fixed time is $P(k)d^3k$. For a Gaussian random field, the natural outcome of quantum fluctuations produced during inflation, $P(k)$ provides a complete statistical description. Cosmological N-body simulations require specification of the initial power spectrum (and higher density moments, for non-Gaussian random fields like those arising in theories of cosmic strings or textures) and the cosmological parameters Ω and H. Where distances are related to velocities, we assume in this article that the present value of the Hubble parameter is $H_0 = 50$ km s^{-1} Mpc^{-1}.

The initial positions and velocities of particles in a cosmological N-body simulation can be obtained from the density fluctuation field using a Lagrangian description originated by Zel'dovich.[4] The fluctuations are equivalent to small displacements from a uniform grid of comoving coordinates $\vec{q}$,

$$\vec{x}(\vec{q},t) = \vec{q} + \vec{\psi}(\vec{q},t) \approx \vec{q} + D(t)\vec{\psi}_0(\vec{q}),\ d\vec{x}/dt = \partial\vec{\psi}(\vec{q},t)/\partial t. \quad (3)$$

For linear density perturbations with $|\delta\rho/\rho| \ll 1$, the displacement field separates into a time-dependent growth factor $D(t)$ and a spatially varying piece $\vec{\psi}_0(\vec{q})$. For a flat universe ($\Omega = 1$), $D(t) \propto a(t) \propto t^{2/3}$. The displacement field is related to the density perturbation field $\delta\rho/\rho$ by conservation of mass and circulation:

$$\frac{\delta\rho}{\rho} = -\vec{\nabla}\cdot\vec{\psi}, \quad \vec{\nabla}\times\vec{\psi} = 0. \quad (4)$$

The displacement field is irrotational because any primeval vorticity is reduced by expansion until it is negligible compared with the divergence, which grows by gravitational instability. Equations (4) can be inverted for $\psi(\vec{q},t)$ using Fourier transform techniques.

The most commonly considered theories for $\delta\rho/\rho$ predict a Gaussian random field with a specified power spectrum. A sample of a Gaussian random field is easily generated using independent Gaussian-distributed random numbers for the real and imaginary parts of each Fourier component. Using random fields, we do not expect to match our own universe in detail. Instead, statistical comparisons must be made by averaging diagnostic

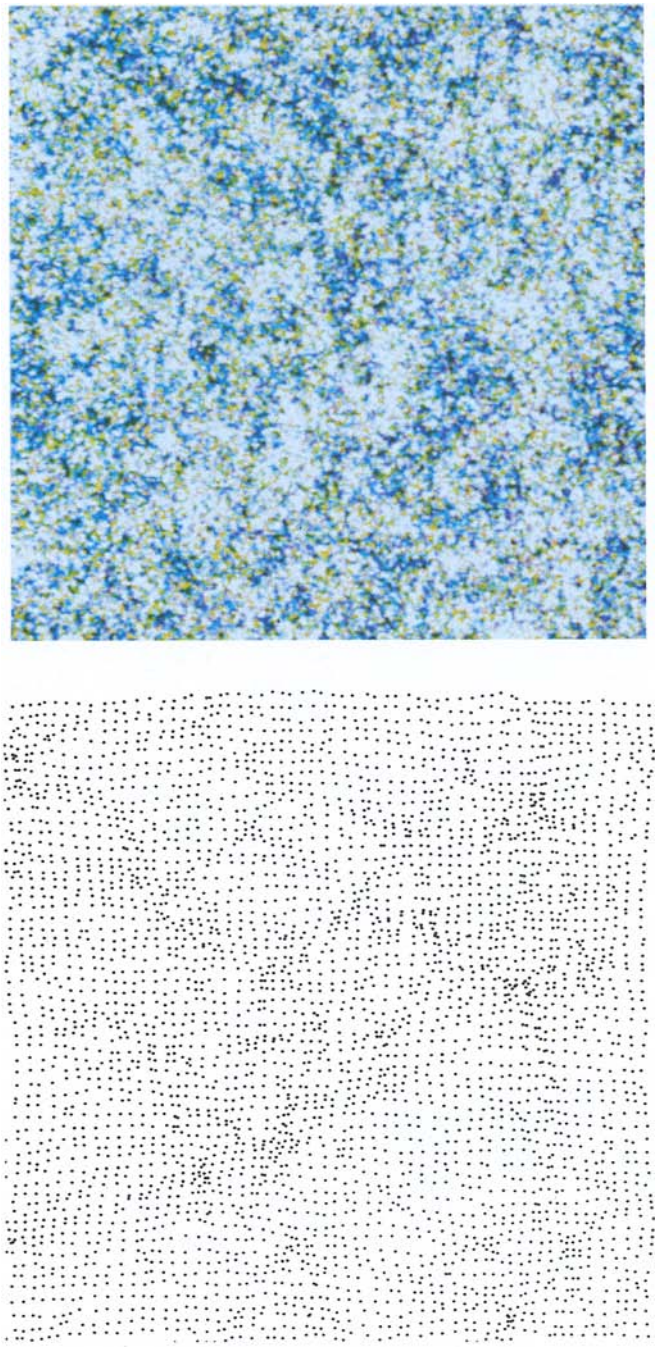

Fig. 1: A Gaussian random field and the corresponding lattice representation used in the initial conditions of an N-body simulation. A three-dimensional random field was first generated on a 256^3 grid; the first plane of the cube is shown as the image on the top. The density was converted into a 256^3 field of lattice displacements. The top plane of this lattice is shown on the bottom, thinned out to 64^2. The color map is a linear stretch from underdense regions (violet/blue) to overdense regions (red/white). Yellow represents the average density.

quantities over large volumes or many simulations performed with different realizations of the random field.

In summary, the procedure for specifying initial conditions is: (1) generate a random density field on a regular cubic lattice; (2) compute the linear displacement field from equations (4) using the Fast Fourier Transform (FFT); and (3) perturb the positions and velocities of lattice points using equation (3).

Figure 1 illustrates the initial density field and the corresponding deformed lattice for one plane of a three-dimensional simulation with 256^3 particles. The initial conditions are Gaussian random noise with power spectrum $P(k) \propto k^{-2}$. Scale-free initial conditions like these are expected to lead to self-similar clustering, the physics of which we will discuss later.

Gravitational instability causes density perturbations to grow in amplitude with time. Once the perturbation amplitude becomes large, linear theory breaks down and a numerical integration of the equations of motion is needed.

Numerical Methods

Numerical methods for cosmological N-body simulations have four desiderata: (1) accurate forces between particles; (2) accurate time integration; (3) large N; and (4) periodic boundary conditions. Periodic boundary conditions allow us to simulate a very large or infinite universe using a finite computer. They are gratis with methods that use FFTs to impose the initial conditions and to evaluate the forces. With periodic boundary conditions, a large fundamental volume is preferred so as to minimize unphysical tides arising from periodic replicas of the structures being simulated. The matter distribution must be discretized into a relatively small number of particles—the largest simulations to date use fewer than $10^{7.3}$ particles, compared with the 10^4 galaxies, 10^{14} stars, 10^{71} atoms, and possibly 10^{87} elementary particles in the volume of a typical cosmological simulation. While we obviously do not need to follow every atom or star in order to get a satisfactory representation of large-scale structure, it is desirable to have at least enough particles to resolve the basic units of cosmological structure—galaxies.

Particle positions and velocities are integrated forward in time using a finite difference approximation to equation (1), which gives one second-order ordinary differential equation for each particle. Leapfrog time integration is generally used because it is second-order accurate while requiring no auxiliary storage; positions and velocities are half a timestep out of phase and time derivatives are replaced by central differences. The time variable is usually rescaled so that equal timesteps will adequately sample the time evolution of clustering. Provided that one takes enough timesteps (typically $\sim 10^3$), accurate time integration is not a problem. This does not mean that all trajectories can be followed with great precision, since orbits in a nonlinear gravitational system are frequently chaotic. However, the smoothed density and velocity fields can be determined with relative precision of a few percent or better.

Obtaining accurate interparticle forces at minimal cost is, perhaps, the most difficult requisite of gravitational N-body simulations and the one to which most attention has been directed. Several algorithms have been devised. The most obvious, and the slowest, is direct summation over all $N(N-1)/2$ pairs. The hierarchical tree code[5] reduces the operations count to $O(N \log N)$ by treating distant clumps of particles as single massive pseudo-particles. The related fast multipole method[6] may reduce the operations count to $O(N)$, but it has not yet been implemented by astrophysicists. With hierarchical methods, accurate forces are readily achievable. Periodic boundary conditions can be added to the direct summation and hierarchical algorithms using Ewald summation to account for the forces from periodic replicas. Tree codes have only recently been applied to cosmology[7], but they are promising because of their speed. Unfortunately, they require a large amount of auxiliary storage. The largest tree code simulations have used one to two million particles.[8,9]

The most popular schemes, which are among the oldest, for computing forces in cosmological simulations rely on FFTs to compute the potential or force on a fixed Cartesian grid from the density evaluated on the grid. Called particle-mesh (PM) by astrophysicists, these techniques are also used in plasma simulations.[10,11] In PM methods, forces are computed in five steps: (1) the density is interpolated from the particles to a grid; (2) the density is Fourier transformed using the FFT algorithm; (3) the Fourier transform of the potential (or force) is computed by multiplication with a suitable Green's function for equation (2) (Fourier convolution); (4) the potential (or force) is returned to the spatial domain using the inverse FFT; and (5) the force is interpolated to the particles from the potential (or force) on the grid.

The main advantage of PM methods is speed. For N particles and N_g grid points, the slowest steps are the FFTs, requiring $O(N_g \log N_g)$ operations. The main disadvantage is that the interparticle force falls below the inverse square law for separations less than about two grid spacings, going to zero at zero separation. In effect, the algorithm treats particles as being fuzzy, nearly round balls of size about 1.4 grid spacings (where our PM force is half the inverse square law value). Some "softening" of the inverse square law force is desirable to avoid collisional effects such as the formation of very tightly bound binaries, which are unphysical in a collisionless simulation because of the large masses of the computational particles (for our largest cold dark matter simulation, $5 \times 10^8\ M_\odot$, where $M_\odot = 2 \times 10^{33}$ g is the mass of the sun) compared with the masses of stars or elementary particles. However, too much force softening is undesirable given the strong clustering caused by gravity. The largest grid used so far in a PM simulation is 384^3, allowing about 270 resolution elements across the length of the simulation volume. Better resolution is needed for many cosmological problems.

The force resolution limitation of the PM algorithm can be removed by supplementing the interparticle forces with a direct sum over pairs separated by $\lesssim 3\Delta x$ (where Δx is the grid spacing), resulting in the particle-particle/particle-mesh algorithm[10,12], or P^3M. The P^3M algorithm has been widely used in cosmological simulations.[13] Its only disadvantage is the fact that it too easily becomes dominated by the direct summation, grinding to a halt even on a supercomputer for very highly clustered mass distributions.

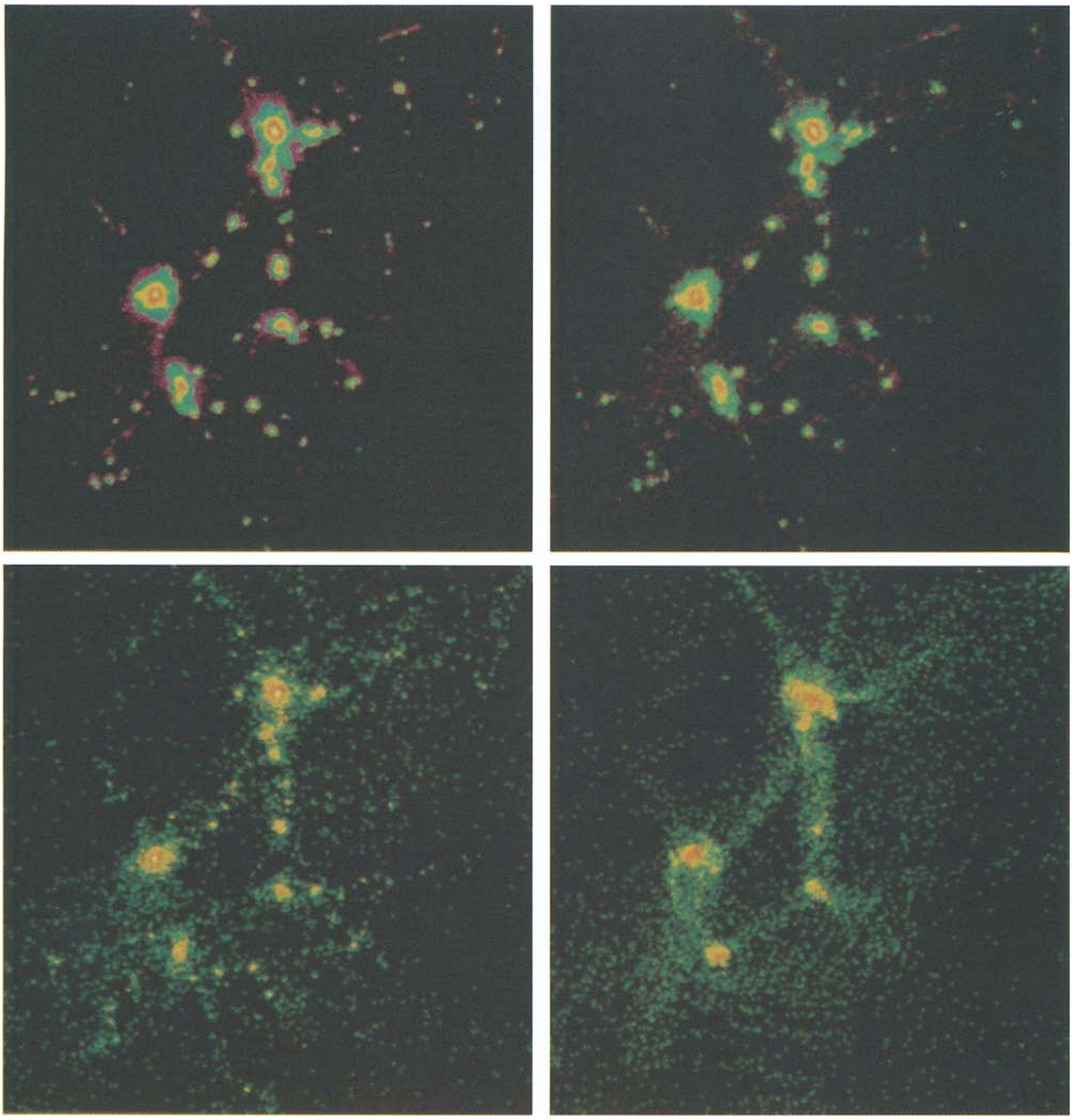

Fig. 2: Comparison of logarithmic mass density images from a $25.6^2 \times 6.4$ Mpc^3 subvolume of 4 cold dark matter N-body simulations in a 51.2 Mpc cube. The 4 panels show results from simulations starting from equivalent initial conditions but having different mass and force resolution. Top left: $N=256^3$, 384^3 mesh force. Top right: $N=128^3$, 256^3 mesh force. Bottom left: $N=64^3$, P^{3}M forces. Bottom right: $N=64^3$, 64^3 mesh force. The graininess is due to shot noise for finite N.

A solution to the slowing-down problem of the P^{3}M algorithm was introduced recently by Couchman.[14] His mesh-refined P^{3}M algorithm uses subgrids to speed up the direct summation of nearby neighbors by shifting some of the burden to high-resolution FFTs. This scheme may well be faster than a hierarchical tree code for highly clustered distributions. Besides being fast, this algorithm allows an accurate fit to an arbitrary force law with maximum rms errors as small as one percent for separations of about $2\Delta x$.

We have implemented a modified version of Couchman's algorithm, improving its efficiency in the way that the subgrid replaces direct summation. Even with a partial implementation—only one level of subgrid and no direct summation below this—we gained a dramatic speedup over the old P^{3}M algorithm. In a simulation of self-similar clustering with $\Omega = 1$, $N = 256^3$ particles, and a 384^3 PM grid, whose initial conditions were partially depicted in Fig. 1, we followed about 1.1 million of the particles (in 5% of the volume) with P^{3}M with 1024 force resolution elements across the cube. We began the simulation using the direct summation algorithm, which ground to a halt with the cpu time per timestep increasing as a^4. Changing horses midstream, we switched to Couchman's algorithm,

with an immediate four-fold speedup. Even more happily, this algorithm slowed little as clustering increased, enabling us to expand by another factor of 2.2, for a total expansion factor of more than 30. With 1200 timesteps, we conserved energy to better than 0.25%. At the end of the simulation, the standard P^3M direct summation algorithm would have been a factor of 100 slower per timestep. Even so, we still spent 70% of the simulation cpu time computing short-range forces for only 6.5% of the particles. Further improvements in Couchman's algorithm should result in additional speedup by a factor of two or more.

Memory, Vectorization and Parallelism

All of our simulations involving two million particles or more were run on an IBM 3090-600J supercomputer at the Cornell National Supercomputer Facility. The IBM 3090 has 512 MB of fast memory and allows up to 999 MB of virtual memory. Storing the positions and velocities for 256^3 particles requires 384 MB. Our PM force calculation uses two grids (one for density and one for one component of the force), requiring an additional 434 MB for 384^3 grid points; storage of the Fourier transforms as real numbers saves almost a factor of two here. With other auxiliary data, our total memory requirement for a PM simulation of this size is nearly 950 MB. In order to add short-range forces for the P^3M algorithm, we borrow storage from the large density and force arrays, which are otherwise unused during this stage of the computation.

Vector and parallel operations are additional considerations for large-scale supercomputer applications. The IBM 3090 is a vector/coarse-grained parallel machine with a typical vector speedup, in our applications, of a factor of two to four. All loops vectorize, with the exception of the interpolation of the mass to the density grid, which cannot be vectorized because it is a multiple scatter operation to each grid point, resulting in memory conflicts. The interpolation of the forces back to the particles does, however, vectorize.

For simulations like ours, requiring hundreds of supercomputer hours, parallelism is an attractive feature, allowing a significant reduction of wall clock time. The IBM 3090-600J has six processors but we use only four in our large simulations because we are limited by the large rate of virtual memory paging. A 384^3 complex FFT requires about 65 cpu seconds and achieves $\sim 300\%$ performance on four processors in our fully-loaded environment. An additional 15 seconds of cpu time is required to precondition our data to save the factor of two in storage for real transforms. We parallelized the direct summation part of our calculations by simultaneously working on four independent subvolumes. This scheme achieves up to 350% parallelism but averages only about 200% because processors that finish early sit idle. This factor can be improved by using more subvolumes and by efficiently queuing them to the processors. With large amounts of input/output for producing images and some analysis each timestep, and with our jobs paging at up to the machine limit, the total wall clock time with four processors is 1.4 times less than the cpu time. Since the single-processor performance for our large jobs is only 50–60%, parallelism still leads to significant improvements.

For our largest N-body simulation performed on the IBM 3090, with $N = 256^3$, $N_g = 384^3$ and P^3M applied to 1.1 million particles, the PM computation requires about 730 cpu seconds per timestep. The mesh-refined P^3M

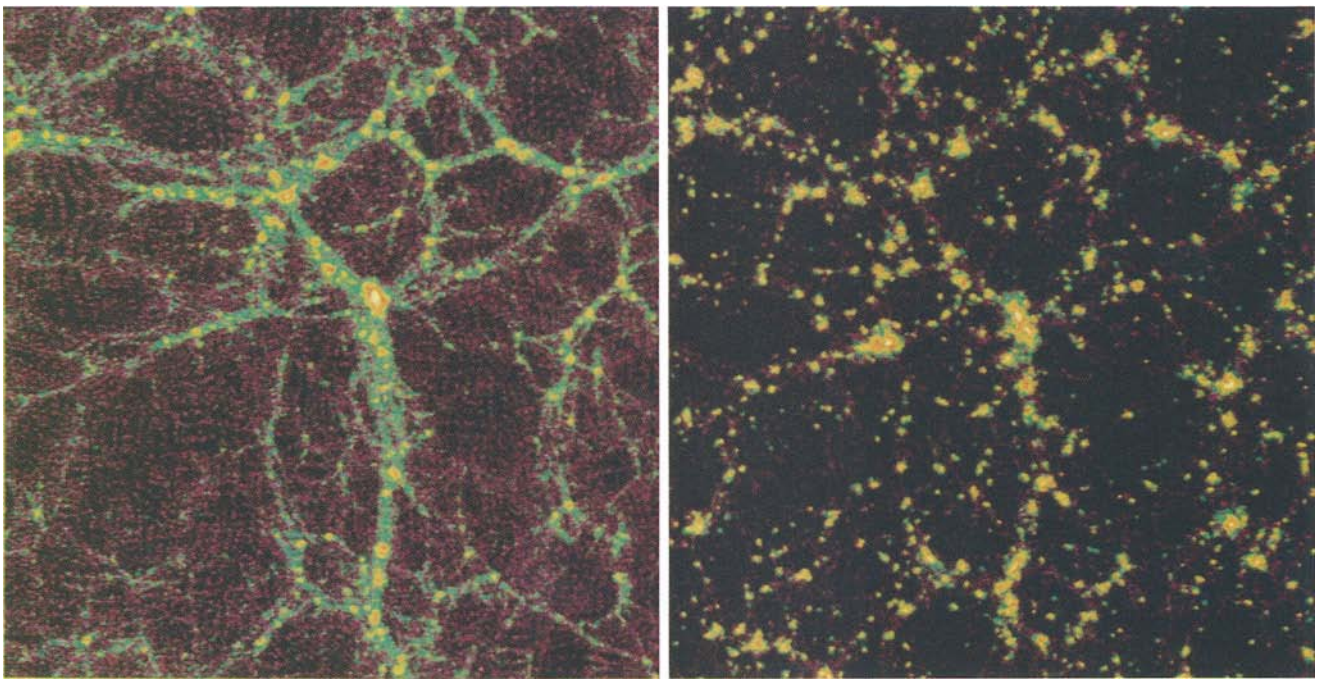

Fig. 3: *Logarithmic* density images from slices of thickness one-eighth of the cube. Results are from evolved N-body simulations with 128^3 particles. Left: scale-free initial conditions with increased amplitude on large scales. Right: white noise initial conditions.

short-range force computation requires at most 1400 cpu seconds per timestep per million particles.

Dynamic Range

Gravitationally-induced cosmological structure spans an enormous range of mass and length scales, from dwarf galaxies of mass $\lesssim 10^9 M_\odot$ and size $\lesssim 10^{-3}$ Mpc, to galaxy superclusters of mass $\gtrsim 10^{16} M_\odot$ and size $\gtrsim 100$ Mpc. Galaxy redshift surveys, whose results often challenge cosmological theories[15,16], record structure over scales ranging from less than 0.1 Mpc up to 300 Mpc. No computer simulation has come close to having the dynamic range needed to follow all of this structure. Existing simulations suffer from limited dynamic range in both mass (too few particles) and distance (small volume and large force softening distance). The pioneering work of Davis et al., for example, which set the amplitude of the power spectrum in the cold dark matter theory and demonstrated that the galaxy clustering pattern could be significantly biased compared with the overall dark matter distribution, was based on simulations with only $10^{4.5}$ particles.[17] In order to fit several galaxy clusters into their simulation volumes (cubes of side 130 Mpc), they used particles of mass $5\times 10^{12} M_\odot$, ten times the mass of a typical galaxy, and the force softening length was 0.6 Mpc, larger than the core of a cluster of galaxies. One must worry whether the conclusions drawn from simulations with such limited dynamic range would change with improved resolution.

Figure 2 illustrates the effects of varying force and mass resolution. The $N = 64^3$ PM and P^3M simulations represent the state of the art as it was about 8 and 4 years ago, respectively. These simulations have too little dynamic range in both mass and length to resolve galaxies well; the poor force resolution of the small PM simulation is particularly harmful. It is clear that accurate results demand high dynamic range for both mass and force resolution, i.e. large particle number N and a force softening distance significantly less than the mean interparticle spacing. In this volume, $N = 128^3$ particles with high force resolution from P^3M should give accurate results on galaxy scales. Such a simulation has not been performed yet, but is not beyond the present state of the art. The largest PM simulation, with $N = 256^3$ particles and a 384^3 grid, has good mass resolution—our own Milky Way would be resolved into about 10^3 particles—but suffers from inadequate force resolution, with a force softening distance of about 0.18 Mpc, compared with a distance of 0.06 Mpc for the Milky Way's closest companions, the Magellanic Clouds.

We could improve the force resolution by using a smaller cube, so that the PM grid spacing is a smaller physical distance. However, such a solution would be unacceptable because nonlinear structures would span the entire cube at a relatively early time, with important nonlinear contributions from longer wavelengths being neglected. This is already a concern in a cube of side 51.2 Mpc. A finer PM grid is possible, but 512^3, or 128 Mwords (512 MB for 4-byte words), is about the largest possible with present supercomputers. Achieving factors of three or more reduction in the force softening length really requires using P^3M or a tree code.

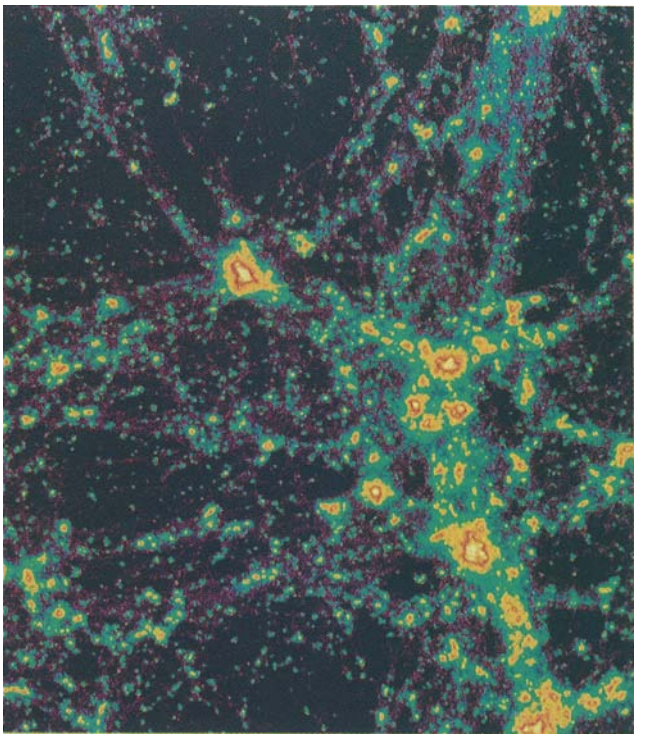

Fig. 4: Projection of all the mass in a highly-evolved simulation with 256^3 particles beginning from scale-free initial conditions like those used for the left-hand part of Fig. 3. Colors show the logarithm of the projected density.

Self-Similar Hierarchical Clustering

Do we really need enormous dynamic range? How much resolution is necessary for accurate results? Answering these questions is not easy because, for a complex nonlinear system of millions of interacting mass points, there exist few checks of accuracy besides global energy, momentum and angular momentum conservation and repeatability with simulations of higher resolution. As Fig. 2 showed, repeatability is an important test. However, dynamic range cannot be extended ad infinitum. Or can it?

There does exist a class of cosmological systems for which the limitations of finite dynamic range can be diagnosed and for which statistical results can be extrapolated to infinite resolution: hierarchical clustering models in a flat ($\Omega = 1$) universe with scale-free initial conditions. Gravity, obeying the inverse square law, has no preferred scale. A power-law power spectrum of initial linear density fluctuations, $P_i(k) \propto k^n$, introduces no scales. Power spectra with $n \leqslant -3$ are unphysical because they lead to divergences for long wavelengths, which are inconsistent with the isotropy of the microwave background radiation and the large-scale uniformity of the galaxy distribution. For $n > -3$, the density fluctuations, with variance $P_i(k)d^3k$, are linear (small amplitude) for long wavelengths, but are nonlinear at sufficiently short wavelengths (large wavenumber k). Thus, one physical length scale is introduced by the critical wavenumber $k_{nl}(t)$ separating these two regimes. The time-dependence of k_{nl} is obtained by setting the mean square density perturbation to unity using linear perturbation theory: $D^2(t)k_{nl}^3 P_i(k_{nl}) = 1 \Rightarrow k_{nl}(t) \propto t^{-2\alpha/3}, \alpha = 2/(3+n)$. For $k \ll k_{nl}(t)$, linear perturbation theory is valid and it

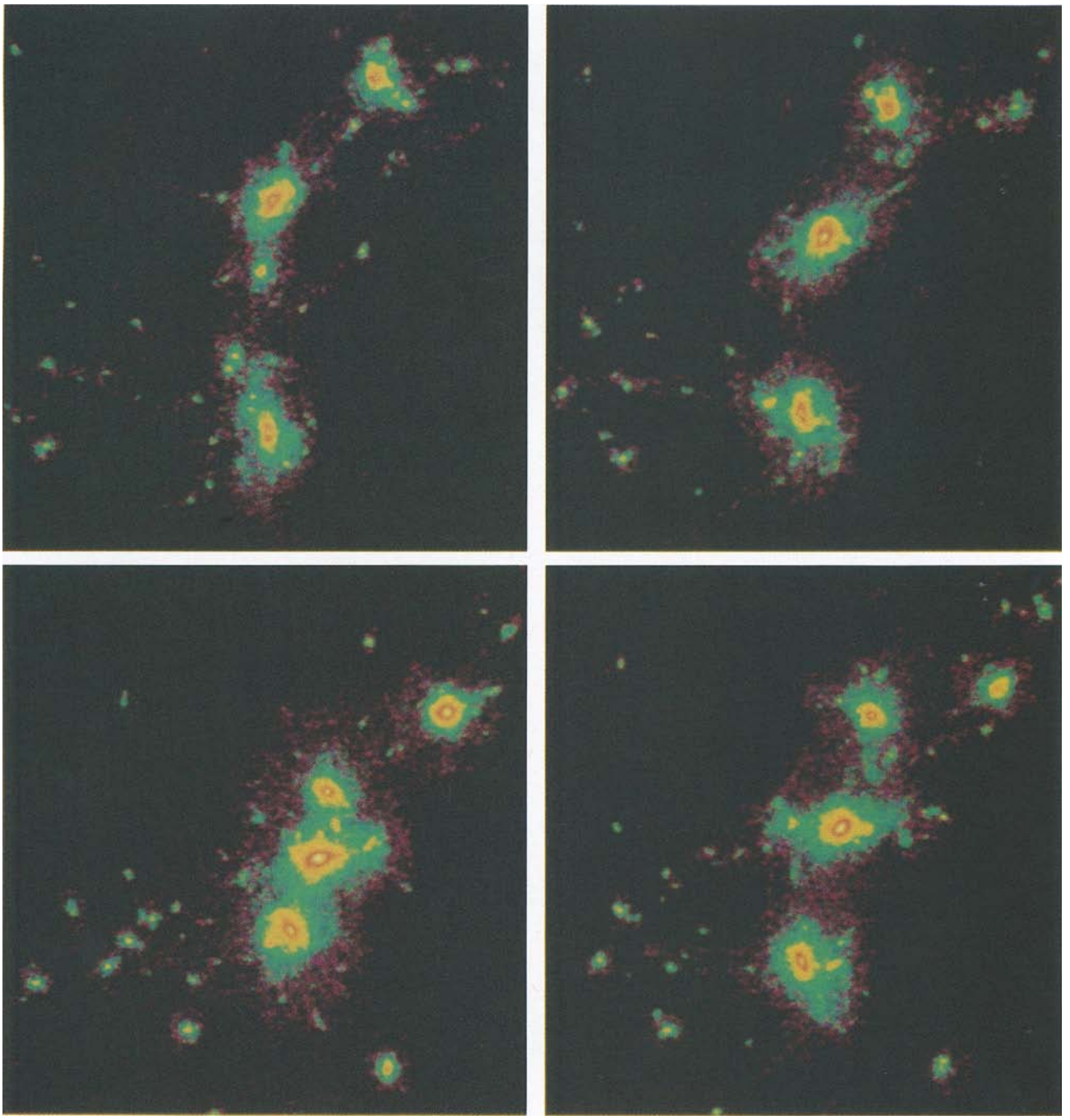

Fig. 5: Time evolution of mass density from a $(1/6)^2 \times (1/20)$ subvolume of the simulation shown in Fig. 4. The P^{3}M algorithm was used to obtain high force resolution: the interparticle force follows the inverse square law down to the smallest scales resolvable in these images. The mosaic shows time evolution proceeding clockwise from the upper left. During the period shown, the universe expanded by a factor 1.7. The lower left image should be magnified by this factor for comparison with the upper left image in proper coordinates.

implies that the time-dependent power spectrum is $P(k,t) = D^2(t)P_i(k)$. For $k \gg k_{nl}(t)$, a nonlinear approach is needed.

Because there is only one physical scale, given by $k_{nl}(t)$, the evolution of clustering in these models is *self-similar*. Under self-similar evolution, all statistical quantities depending on wavenumber, like the power spectrum, must depend on wavenumber only through the dimensionless combination k/k_{nl}. Moreover, the asymptotic behavior of the power spectrum at high wavenumber can be determined[18] from the argument that gravitationally bound structures should be in virial equilibrium. This argument yields the prediction $P(k,t) \propto t^{2\alpha} (k/k_{nl})^{\gamma-3}$, $\gamma = 3(3+n)/(5+n)$. Note that the self-similarity refers to the evolution in *time and space*: the similarity variable is $k/k_{nl} \propto kt^{2\alpha/3}$. Unlike fractals, the models are *not* self-similar in *space* at one moment of time.

Before testing the self-similar prediction using computer simulations, it is helpful to get a visual idea of hierarchical clustering in these models. Fig. 3 compares the density distributions from PM simulations (128^3 particles, 256^3 grid points) with $n = -2$ and $n = 0$ (white noise) initial conditions. The same random number seed has been used in both cases so that the large-

scale structure is similar. Both simulations have been evolved until the nonlinear wavelength is one-eighth of the length of the cube. The evolved white noise has significantly more power on small scales than the other simulation: the matter distribution is very clumpy. The evolved $n = -2$ simulation has prominent filaments and sheets connecting the clumps. These linear and two-dimensional structures are remnants of the larger amplitude, longer wavelength waves present in the initial conditions.

Thin, two-dimensional slices give a limited view of the structure existing in a large N-body simulation. Displaying all of the structure at once in a comprehensible way is difficult because of the large dynamic range in mass and length; standard volume-rendering algorithms fare poorly with our data. Fig. 4 is a full projection of the mass density in the cube for one of our large simulations (256^3 particles, 384^3 grid points) with $n = -2$ initial conditions. At the time shown here the simulation has been evolved 940 timesteps and the nonlinear wavelength has grown to 40% of the size of the cube. Regions of density less than the cosmic mean are black; white regions are at least 125 times the mean density, in projection. The peak three-dimensional density exceeds 10^4 times the mean. The smallest visible clumps contain about 100 particles; the largest contains about 10^6.

Although the simulations with $n = -2$ initial conditions have prominent filamentary structure at low density, they also have a broad mass spectrum of clumps similar to galaxy halos. Fig. 5 shows the time evolution in a small comoving volume in the region of our large simulation where P^3M was used to obtain high force resolution. Close inspection shows that there are several small clumps orbiting in the larger halos. The satellites gradually spiral into the massive cores as a result of dynamical friction—the drag on an object caused by the gravitational wake created as the object travels though the background medium.

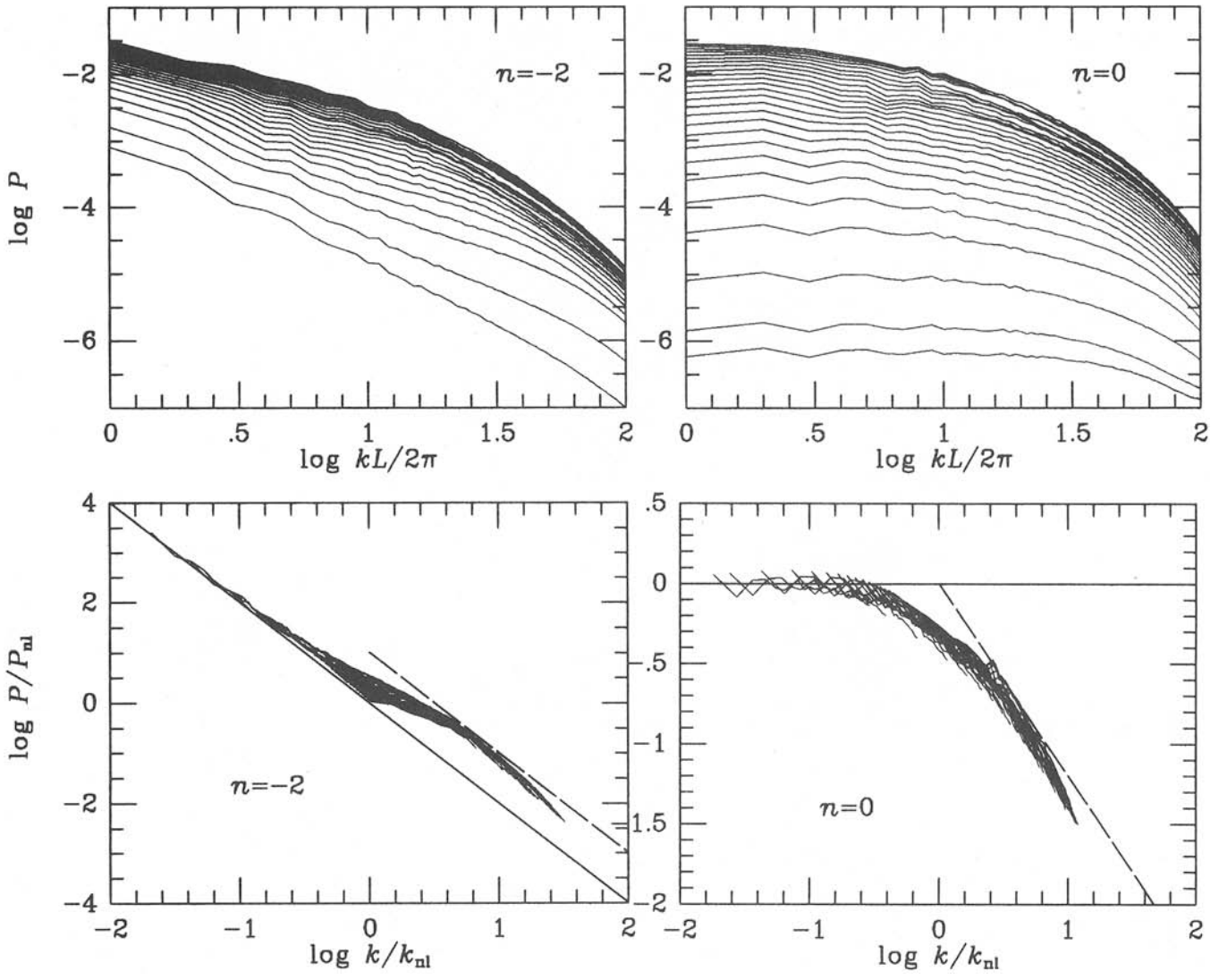

Fig. 6: Power spectra for 2 different simulations starting from scale-free initial conditions $P(k) \propto k^n$. The top set shows the evolution of the power spectrum in comoving coordinates. In the bottom set, the wavenumber and power spectrum have been scaled with time according to the prediction of the similarity solutions, with the longest wave and wavelengths shorter than 4 times the mean interparticle spacing excluded. The solid straight line shows the linear power spectrum and the dashed line shows the limiting nonlinear slope expected at high wavenumbers (the offset is arbitrary).

We now test whether the numerically-computed evolution is self-similar, as is predicted theoretically. Fig. 6 presents the test using power spectra from our largest $n = -2$ simulation and from the smaller white noise ($n = 0$) simulation. Aside from departures at the smallest and largest wavenumbers, due to the neglect of waves longer than the cube and to force softening, respectively, the expected self-similarity is confirmed remarkably well. Our results show the best numerical evidence to date for the asymptotic behavior of the power spectrum at high wavenumbers.

Our demonstration of self-similar hierarchical clustering has several important corollaries. First, the results provide a useful check on the correctness of the computer simulations. Second, in realistic models, physical processes taking place during the first million years introduce scales that break any primeval scale-invariance such as that predicted in the inflationary paradigm. However, the

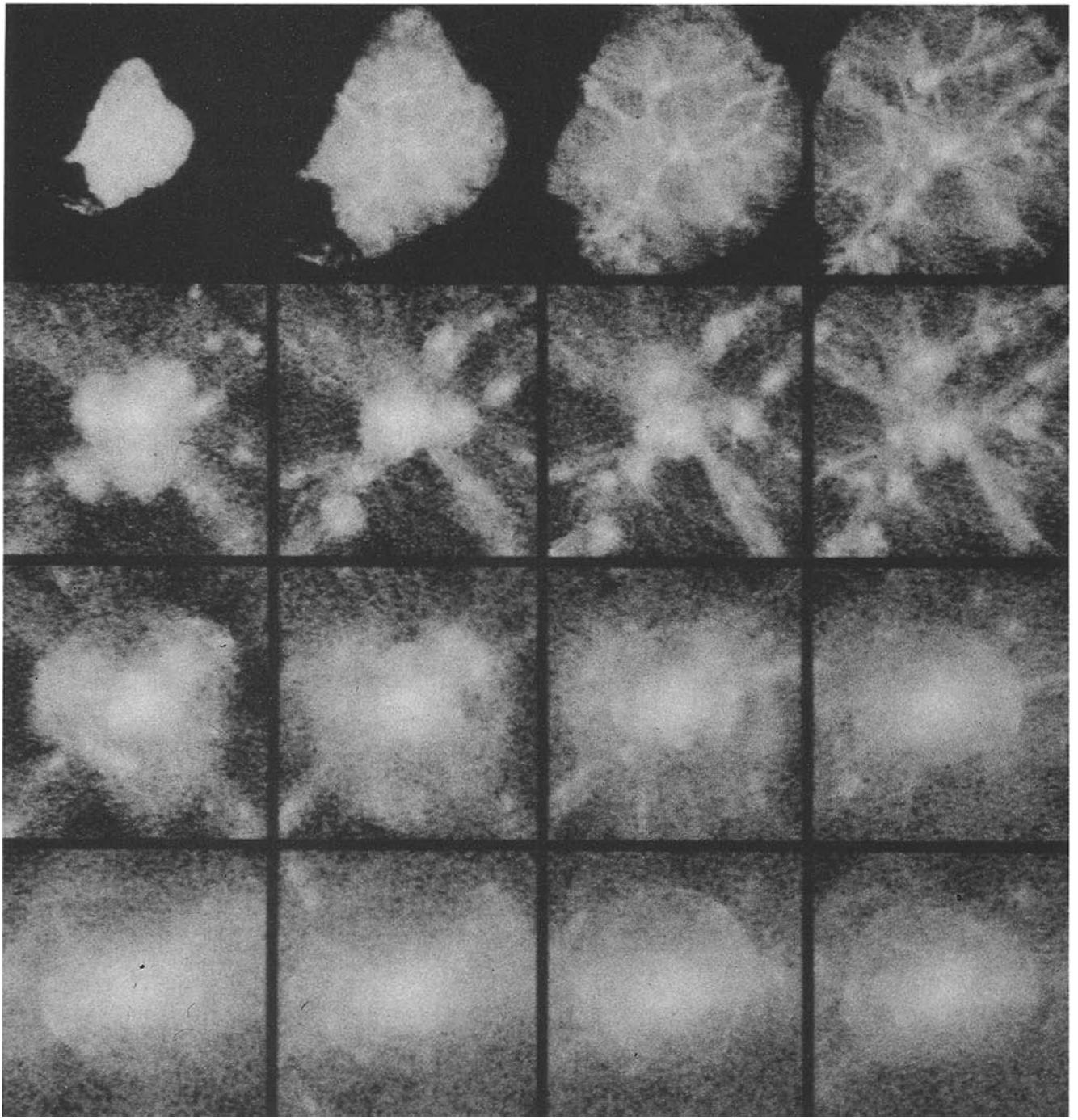

Fig. 7: The formation of a shell galaxy in the cold dark matter model. The 16 panels show the time evolution (beginning in the upper left, with time increasing from left to right and vice versa along alternate rows) in a cube of fixed proper length 2.5 Mpc (a 20× magnification at the end of the simulation). The same particles are followed throughout; the particles that end up in the galaxy are initially in a region shaped like Maine in this projection. The sharp-edged arcs apparent in the bottom two rows are caustics from streams of particles with high phase space density. A black and white image is preferred to color because of the low contrast of the shells.

predicted variation of the slope of the linear power spectrum is gradual enough for the similarity solutions to provide useful approximations over broad ranges of length scales. For example, the cold dark matter spectrum has $n \approx -2$ on the galactic scales and $n \approx 0$ on the scale of clusters of galaxies. Thus, the two panels of Fig. 3 may be viewed as a high-resolution view of the formation of a small group of galaxies (left) and a slice of large-scale structure (right).

Third, using the asymptotic behavior at long and short wavelengths for the similarity solutions, it is possible to extrapolate the finite-resolution numerical results to the limit of infinite dynamic range. This procedure is suggested by the results appearing in the bottom part of Fig. 6. One may take the upper envelope of the scaled power spectra obtained from the simulations, and extend the spectra to small wavenumbers using the solid lines and to high wavenumbers using the dashed lines. In this way we have obtained expressions for $P(k,t)$ for $n = -2$ and $n = 0$ that we estimate to be everywhere accurate to better than 5%. These formulae effectively extrapolate the simulation results for the power spectrum to the limit of infinite volume and infinitesimal force softening distance. Similar extrapolations may be performed for other quantities.

Finally, from the departures of the numerical solutions from self-similarity, we can deduce the limitations imposed by finite mass and force resolution and a finite box size. These limitations allow us to estimate the domain of validity of numerical results obtained for other models. From Fig. 6, for example, we learn that the PM force softening reduces the power for wavelengths shorter than about $8\Delta x$ even though departures from the inverse square law are negligible for separations greater than $3\Delta x$. We also learn that the power is underestimated for long wavelengths after the nonlinear wavelength has grown to one-quarter the side of the cube. When applied to the cold dark matter power spectrum, this result implies that a cube of side 51.2 Mpc is too small even for accurate computation of the galaxy two-point correlation function at late times. A cube twice as large (with the particle mass increased by a factor of eight) would be large enough.

A Case Study of Galaxy Formation

The most successful theory for the formation of galaxies and cosmological structure on intermediate scales (up to about 20 Mpc) combines an exotic (non-nuclear) dark matter component with negligible thermal velocities, generically called cold dark matter, with the prediction from cosmic inflation that $\Omega = 1$. This theory leads to hierarchical clustering of matter like that discussed above, except that the linear power spectrum does not have a constant logarithmic slope because of physical processes that took place before galaxy formation. The cold dark matter theory has one important free parameter: the amplitude of the power spectrum. For an $\Omega = 1$ universe, uncertainty in the initial amplitude of density fluctuations translates into uncertainty about when to stop the computer simulation to identify the results with the present day. In practice, this determination is made by comparing the clustering and motions of simulated galaxies with real galaxies. A realistic comparison requires that the simulated galaxies be well resolved in mass and length and that the volume be large enough to include all relevant contributions to the power spectrum.

We have performed the largest cold dark matter simulation to date, using 256^3 particles each of mass about 10^{-3} of the mass of the Milky Way. This simulation allows us to follow the process of galaxy formation, or at least the assembly of the dark matter halo of a galaxy, with the best mass resolution ever achieved. Unfortunately,

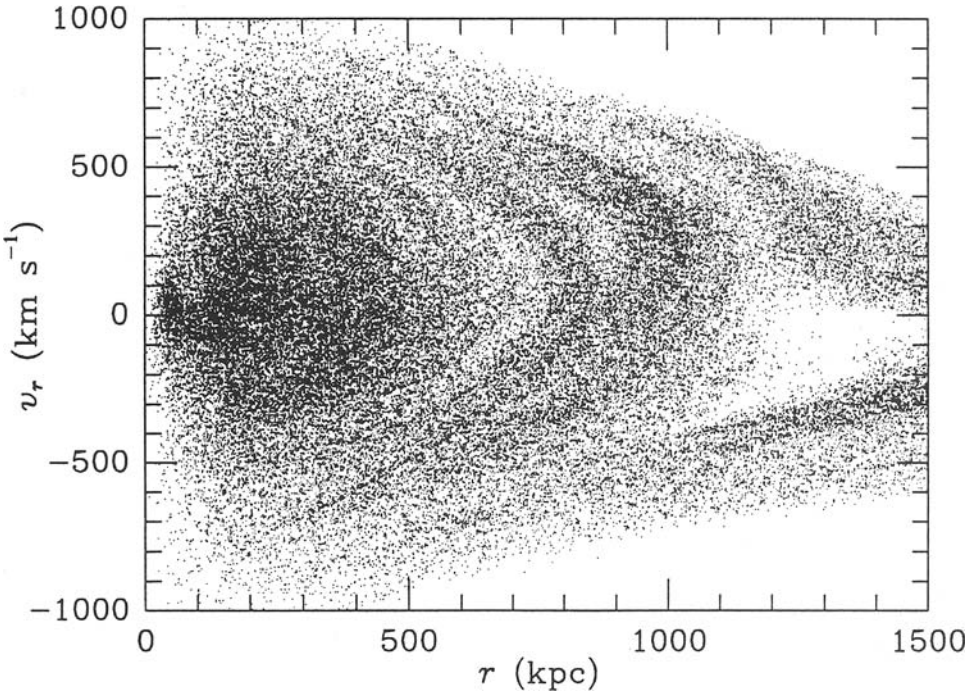

Fig. 8: A phase space diagram for the shell galaxy of Fig. 7.

even this simulation suffers from limited dynamic range, as already noted—the force softening distance is rather large (0.18 Mpc) and the cube is too small (51.2 Mpc) for reliable determination of the correlation function. Despite these limitations, our simulation has excellent resolution for following the formation of a massive galaxy halo.

Figure 7 shows the formation of an object that caught our attention as we scanned through the volume of our simulation. This beautiful galaxy has shells similar to those seen around some elliptical galaxies. Although it has been recognized for several years that such shells are produced by galaxy mergers[19], ours is the first cosmological simulation showing their formation. Hints of shells appear in our smaller simulations, but only with tens of thousands of particles (there are a little more than 10^5 particles in this object) do they stand out so clearly. While this shell galaxy may be the most beautiful in our simulation, we noticed dozens more. The numbers are roughly consistent with the observation that about 15% of all ellipticals have shells, or 2–3% of all galaxies. We suspect that the formation of so many shell systems may require a power spectrum of just the right shape because the rate of merging of sub-galactic clumps depends strongly on the power spectrum shape. In future simulations, we plan to test this hypothesis to see if we can use shell galaxy observations to constrain the primeval power spectrum.

Figure 8 can help us to understand the formation of the shells. This phase space diagram plots the radial velocity of particles (relative to the center-of-mass velocity of the galaxy) versus their distance from the center, for half of the particles, at a time corresponding to the second from last image in Fig. 7. Several well-defined particle streams are seen. These streams are produced by the clumps that fall into the big galaxy in the middle. As a satellite falls in (with $v_r < 0$), it is tidally disrupted into a stream of debris having a range of orbital radii. When each particle reaches the origin, it emerges on the other side with $r > 0$ and $v_r > 0$. Caustics arising from the accumulation of matter at one radius (but with a range of velocities) occur at $r = 1100$, 1000, 800, 700 and 500 kpc, to name a few. The caustics appear on alternate sides of the galaxy; they are interleaved in radius rather than being complete spherical shells.

Identifying Galaxy Halos

From our maps one can see that nonlinear gravitational instability gathers matter into dense clumps. We would like to identify these clumps with the dark matter halos of galaxies. While it is possible that more than one luminous galaxy forms and survives in a common dark matter halo, it is implausible that galaxies could have formed in regions where the dark matter has failed to collapse, unless nongravitational forces are much more significant than we believe. As a working hypothesis, we identify galaxy halos in a one-to-one manner with maxima of the dark matter density distribution. Since real galaxies are composed of luminous matter that, as a result of gaseous dissipation, has contracted significantly with respect to the dark matter, it is dangerous to identify galaxies themselves with the dark matter maxima. Nevertheless, it is useful to assume a simple relation between luminous galaxies and dark halos so that we can make predictions for the galaxy distribution based on simulations.

A standard method for identifying density peaks in particle distributions is friends-of-friends cluster analysis. All pairs of particles separated by less than some amount (typically 0.1 or 0.2 times the mean interparticle spacing) are linked recursively. Sets of linked particles define clusters; each cluster is identified as a galaxy halo. We have found this algorithm to work poorly when there is a large range of density contrast. A small linking distance is needed to separate distinct halos in high-density regions, with the consequence that halos in low-density regions are undersampled. A large linking distance will correctly assign low-density halos but will merge overlapping halos in dense regions.

The obvious visual distinctiveness of the dark matter clumps formed in the N-body simulations suggests an alternate clustering algorithm based on contour surfaces of the three-dimensional density field. One would like to identify the maxima of the density field with the centers of halos. Each halo should have assigned to it those particles interior to the last closed contour surface surrounding the density maximum, subject to additional requirements such as the particles being gravitationally bound to the clump.

We have devised a simple scheme to find automatically all of the density maxima and, simultaneously, their associated particles. Our mathematical scheme is based on a physical analogy with a highly viscous, perfectly compressible fluid flowing in a fixed gravitational field. Regardless of the initial distribution of the fluid, after it has come to rest the fluid particles will be concentrated entirely at the minima of the gravitational potential. Translating this analogy into a recipe for finding density maxima of a dark matter distribution is straightforward. We perform a pseudo-hydrodynamical calculation, moving the dark matter particles using the equation of motion of a fluid in the limit of large damping:

$$\frac{d\vec{x}}{d\tau} = \vec{\nabla}\frac{\delta\rho}{\rho}, \tag{5}$$

where a fictitious time variable τ has been introduced. The density distribution $\delta\rho/\rho(\vec{x},t)$ is evaluated at a fixed time t, and is unchanged throughout this fluid calculation. The density can be smoothed on a scale appropriate for galaxy halos (about 0.1 Mpc). The density field in equation (5) is evaluated on a fine grid, typically 512^3 or even 1000^3 (covering only a fraction of the total volume at one time in the latter case).

Equation (5) will yield the desired result: particles move normal to the density contour surfaces toward regions of higher density. This equation is easy to model computationally using a finite-difference method; the particles can move independently with different timesteps and the scheme is easy to vectorize. The particles arrive exponentially close to the peaks in a finite number of timesteps. Once the particles are sufficiently concentrated into the density peaks (some of which may consist of only one or two particles), it is a relatively simple matter to scoop up the particles in each peak and to record their labels. Their original positions are then restored; the pseudo-hydrodynamical calculation is used only to tag the

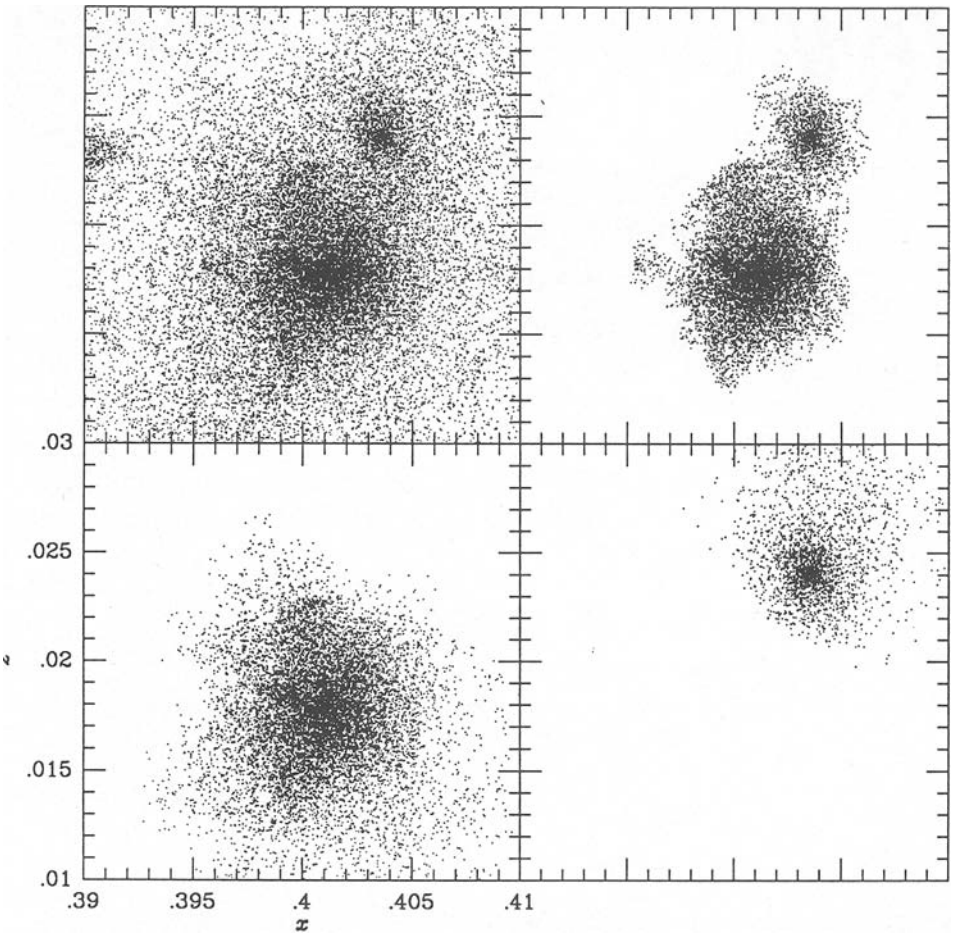

Fig. 9: Particle plots from a small region of the high-force resolution subvolume of the $N=256^3$ simulation shown in Fig. 4. Top left: all 36875 particles in this volume. Top right: the largest single "halo" found in this volume by the friends-of-friends algorithm with a linking distance of 1/2560 (15408 particles). Bottom left: the largest halo found by the density maximum algorithm (16335 particles). Bottom right: the second largest halo found by the density maximum algorithm (2826 particles).

particles associated with each density peak. As a further step, for each peak we find the largest set of particles that are mutually gravitationally bound. This process evaporates many tiny groups (with fewer than 20 or so particles) that arise from Poisson fluctuations in the density. The combined algorithm finds the halos that one would clearly identify in images, and it does even better by applying the gravitational binding criterion and by being objective. Fig. 9 shows an example of the success of our new algorithm in a case where the friend-of-friends algorithm fails by merging obviously distinct halos.

Velocity Bias

The cold dark matter theory has, until perhaps very recently, been popular with most theorists because it accounts well for many of the properties of galaxies and clusters of galaxies. Many of the model's successes have appeared to depend on the assumption that the spatial distribution of galaxies is strongly biased relative to that of the dark matter, while the velocity distributions are similar. As a consequence, dynamical measurements of Ω, which depend on both the galaxy clustering and velocities[2], would be biased toward small values like those estimated from galaxy clustering, $\Omega \approx 0.2$. Recently, however, Carlberg, Couchman and Thomas[20] have pointed out that the velocity distribution of galaxies in cosmological N-body simulations may differ significantly from the velocity distribution of all the mass, an effect they call cosmological velocity bias. The effect was not found by previous workers, who either lacked adequate dynamic range to resolve galaxies into more than a few particles, or computed the velocities of galaxies using the most tightly bound particles rather than measuring the center-of-mass velocity of the halo. Carlberg et al. find that simulated galaxies have relative velocities that are about a factor of two smaller than previously estimated. The implication is that previous simulations of the cold dark matter theory assumed an initial fluctuation amplitude that was too small; the simulations should have been evolved longer. A beneficial side-effect would be increased amplitude of large-scale structure, perhaps removing the main defect of the cold dark matter theory.

Figure 10, from our large cold dark matter simulation at the time corresponding to Fig. 8, confirms the existence of a strong velocity bias. Much of the bias arises because massive galaxies in clusters contribute only a small number of galaxy pairs but a large number of particle pairs with large velocities. Because the contribution to the velocity dispersion scales quadratically with the

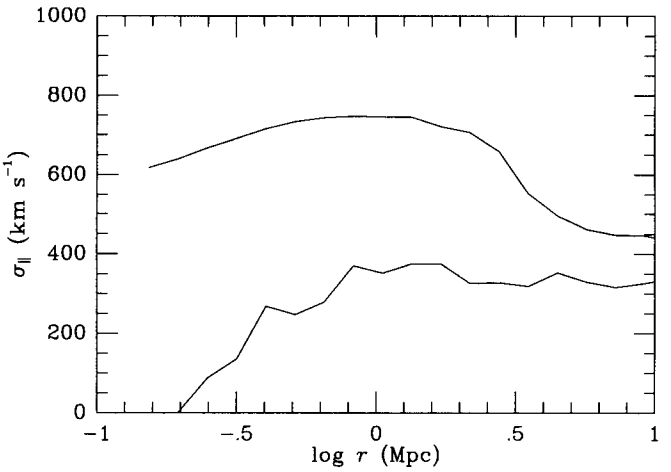

Fig. 10: Relative radial pair velocity dispersions from a large cold dark matter simulation. The top curve is for all particles while the bottom curve uses the center-of-mass velocities of galaxy halos found by the density maximum algorithm; the decline for $r \lesssim 0.5$ Mpc is an artifact of force softening. The relative velocity dispersion of observed galaxies is about 300 km s^{-1}. High resolution is needed to detect this strong velocity bias.

number of objects, there is a strong enhancement of the particle-particle velocity dispersion in the clusters. Dynamical friction also slows down the galaxies relative to the background dark matter in clusters. Because dynamical estimates of Ω scale quadratically with velocity, these results imply that measurements using the relative motions of galaxies, assuming that mass and galaxies have the same clustering and motions, would underestimate Ω by about a factor of four in this theory. Uncorrected estimates using real galaxies yield $\Omega \approx 0.2$. Therefore, the low apparent Ω may not exclude the inflationary paradigm, which predicts $\Omega = 1$. Much more work needs to be done comparing simulations with actual measurements before firm conclusions are reached.

Conclusions

Recent technological developments in both supercomputer hardware and software have made possible dramatic increases in the size and resolution of cosmological N-body simulations. Our largest current simulations, with 256^3 particles, have comparable force resolution but 512 times as many particles as the state-of-the-art $N = 32^3$ simulations of five years ago. If large N is the first priority, particle-mesh codes can easily follow more than 10^7 particles on today's large supercomputers; this should increase to 10^8 in several years, once the accessible memory of supercomputers reaches 10 Gigabytes.

Large N is not enough; concomitantly high spatial resolution demands good force resolution, with the inverse square law being modeled accurately to scales as small as 10^{-3} of the size of the volume simulated. This resolution can be readily achieved using the hierarchical tree or particle-particle/particle-mesh algorithms. Recently, Couchman[14] has described a major improvement in the latter algorithm that stops it from grinding to a halt for highly clustered particle distributions. We have implemented a modified version of Couchman's algorithm on an IBM 3090-600J supercomputer. With it, we have followed more than 10^6 particles with high force resolu-

tion in a simulation with 16.8 million particles in total. Further improvements should increase the size of high-force resolution simulations to more than 10^7 particles within a few years.

Our numerical simulations confirm the self-similarity of gravitational clustering starting from scale-free initial conditions. Although such initial conditions are unrealistic for structure formation in our own universe, our results have several important applications and implications. First, the similarity solutions should provide useful approximations over broad ranges of length scales for models with more realistic initial conditions. Second, departures from self-similarity, arising from the limited dynamic range of the simulations, can be easily identified, providing an important estimate of the limitations of other simulations with similar dynamic range but with non-scale-free initial conditions. Third, because the asymptotic behavior of the similarity solutions is seen despite the limited dynamic range, the results can be extrapolated to effectively infinite resolution.

Regardless of the power spectrum of initial density fluctuations, gravitational instability causes the dark matter to coalesce into dense clumps looking very much like galaxy halos. Proper identification and analysis of these clumps is crucial for correctly assessing the viability of galaxy formation theories. With a large simulation (with 256^3 particles and 384^3 force grid points) of the cold dark matter theory, we have demonstrated the formation of galaxies with shells like those seen around many ellipticals. We have also confirmed that galaxies in this theory move more slowly relative to each other than do the individual dark matter particles. These results bias estimates of the mean cosmological density, perhaps reconciling the inflationary universe paradigm with observations. The bias also implies that the amplitude of primeval fluctuations needed for successful galaxy formation is larger than previously estimated, improving the viability of the cold dark matter theory by increasing the amplitude of large-scale structure. It remains to be seen whether this theory can account for recent observations of large-scale structure.

Despite the major improvements in resolution that have been achieved, cosmological N-body simulations still lack adequate dynamic range for addressing many basic theoretical questions, especially those relating to the large-scale structure of the universe. We believe, however, that computer simulations of size and resolution comparable with observational surveys of galaxies will be made within a few years. When this happens, computer simulation will have become an equal partner of cosmological theory and observation.

Acknowledgements: This research was conducted using the Cornell National Supercomputer Facility, a resource of the Center for Theory and Simulation in Science and Engineering at Cornell University, which receives major funding from the National Science Foundation and IBM Corporation, with additional support from New York State and members of its Corporate Research Institute. We appreciate the programming assistance of CNSF consultant Paul Schwarz. This work was supported by NSF grant AST90-01762. This paper was submitted to the 1990 IBM Supercomputer Competition. ■

References

1. E. Holmberg, *Astrophys. J.*, **94**, 385 (1941).
2. P. J. E. Peebles, *Nature*, **321**, 27 (1986).
3. A. H. Guth, *Phys. Rev.*, **D23**, 347 (1981).
4. Ya. B. Zel'dovich, *Astron. Astrophys.*, **5**, 84 (1970).
5. J. Barnes and P. Hut, *Nature*, **324**, 446 (1986).
6. L. Greengard, *Computers in Phys.*, **4**, 142 (1990).
7. F. R. Bouchet and L. Hernquist, *Astrophys. J. Suppl.*, **68**, 521 (1988).
8. Y. Suto, M. Itoh and S. Inagaki, *Astrophys. J.*, **350**, 492 (1990).
9. M. S. Warren, W. H. Zurek, P. J. Quinn and J. K. Salmon, in *After the First Three Minutes*, ed. S. Holt, V. Trimble and C. Bennett (New York: American Institute of Physics, 1991).
10. R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (New York: McGraw-Hill, 1981).
11. J. M. Dawson, *Rev. Mod. Phys.*, **55**, 403 (1983).
12. G. Efstathiou, M. Davis, C. S. Frenk and S. D. M. White, *Astrophys. J. Suppl.*, **57**, 241 (1985).
13. S. D. M. White, in *The Early Universe*, ed. W. G. Unruh and G. W. Semenoff (Dordrecht: Reidel, 1988), p. 239.
14. H. M. P. Couchman, *Astrophys. J. (Letters)*, **368**, L23 (1991).
15. M. J. Geller and J. P. Huchra, *Science*, **246**, 897 (1989).
16. W. Saunders, C. Frenk, M. Rowan-Robinson, G. Efstathiou, A. Lawrence, N. Kaiser, R. Ellis, J. Crawford, X.-Y. Xia and I. Parry, *Nature*, **349**, 32 (1991).
17. M. Davis, G. Efstathiou, C. S. Frenk and S. D. M. White, *Astrophys. J.*, **292**, 371 (1985).
18. M. Davis and P. J. E. Peebles, *Astrophys. J. Suppl.*, **34**, 425 (1977).
19. P. J. Quinn, *Astrophys. J.*, **279**, 596 (1984).
20. R. G. Carlberg, H. M. P. Couchman and P. Thomas, *Astrophys. J. (Letters)*, **352**, L29 (1990).