

A COSMOLOGICAL KINETIC THEORY FOR THE EVOLUTION OF COLD DARK MATTER HALOS WITH SUBSTRUCTURE: QUASI-LINEAR THEORY

CHUNG-PEI MA

Department of Astronomy, University of California, 601 Campbell Hall, Berkeley, CA 94720; cpma@astro.berkeley.edu

AND

EDMUND BERTSCHINGER

Department of Physics, Massachusetts Institute of Technology, Room 37-602A, 77 Massachusetts Avenue, Cambridge, MA 02139; edbert@mit.edu

Received 2003 November 26; accepted 2004 April 9

ABSTRACT

We present a kinetic theory for the evolution of the phase-space distribution of dark matter particles in galaxy halos in the presence of a cosmological spectrum of fluctuations. This theory introduces a new way to model the formation and evolution of halos, which traditionally have been investigated by analytic gravitational infall models or numerical N -body methods. Unlike the collisionless Boltzmann equation, our kinetic equation contains nonzero terms on the right-hand side arising from stochastic fluctuations in the gravitational potential due to substructures in the dark matter mass distribution. Using statistics for constrained Gaussian random fields in standard cosmological models, we show that our kinetic equation to second order in perturbation theory is of the Fokker-Planck form, with one scattering term representing drift and the other representing diffusion in velocity space. The drift is radial, and the drift and diffusion coefficients depend only on positions and not velocities; our relaxation process in the quasi-linear regime is therefore different from the standard two-body relaxation. We provide explicit expressions relating these coefficients to the linear power spectrum of mass fluctuations and present results for the currently favored cold dark matter model with a nonzero cosmological constant. Solutions to this kinetic equation will provide a complete description of the cold dark matter spatial and velocity distributions for the average halo during the early phases of galaxy halo formation.

Subject headings: cosmology: theory — dark matter — galaxies: halos — gravitation — large-scale structure of universe

Online material: color figures

1. INTRODUCTION

Most mass in the universe is in the form of dark matter, and most dark matter is gravitationally clustered in the form of halos. Dark matter halos are therefore the building blocks of the universe. Knowledge of the formation and evolution history of dark matter halos is essential for an understanding of the halos' structure and dynamics, as well as the environment that harbors galaxies and clusters. Designs for dark matter search experiments also rely on the predicted spatial and velocity distributions of dark matter particles in our own Galaxy.

According to the theory of structure formation in current cosmological models, dark matter halos arise from tiny primordial fluctuations seeded in the matter and radiation fields in the early universe at, for example, the end of an inflationary era. These small inhomogeneities grow via gravitational instability, with the slightly overdense regions becoming denser and the slightly underdense regions becoming emptier. At a given time, there exists a spectrum of density fluctuations over a wide range of length and mass scales, whose spatial distribution is characterized by the power density in Fourier space $P(k, t)$, the power spectrum. Dark matter halos therefore rarely form and evolve in isolation, growing instead by frequent accretion of smaller mass halos and occasional major mergers with other halos of comparable mass.

The earliest theoretical insight into cosmological halo formation came from the spherical radial infall model, which considers an isolated spherical overdense volume and describes the secondary infall of bound but initially expanding

collisionless matter onto this region (Gunn & Gott 1972). If the initial density perturbation spectrum is scale-free and the velocity distribution follows the Hubble law, the infall solution is found to approach a self-similar form in an Einstein–de Sitter universe. The resulting density profile asymptotically approaches $\rho \propto r^{-\alpha}$ with $\alpha \geq 2$ (Fillmore & Goldreich 1984; Bertschinger 1985). The exact value of α depends on the index of the initial perturbation spectrum $P(k) \propto k^n$: it is isothermal ($\alpha = 2$) for $n \leq -1$ and for $n > -1$ steepens to $\alpha \approx 3(n+3)/(n+4)$ (Hoffman & Shaham 1985) or $3(n+3)/(n+5)$ (Syer & White 1998; Nusser & Sheth 1999). The collapse of halos therefore retains memory of the initial conditions in the radial infall model.

By contrast, modern cosmological N -body simulations find dark matter halos over a wide range of mass scales to follow a density profile that is nearly independent of the initial power spectrum and is shallower near the halo centers (with $\alpha \sim 1$) than that predicted by the radial infall model (Navarro et al. 1996, 1997; Moore et al. 1998). Furthermore, the dark matter in the simulated halos is not entirely smoothly distributed spatially; instead, at least 10% of the halo mass is in the form of hundreds to thousands of smaller, dense satellite halos of varying mass (Ghigna et al. 1998; Klypin et al. 1999; Moore et al. 1999). Thus, even though the analytic similarity solutions derived from one-dimensional infall onto a smooth overdensity have provided valuable insight into the evolution of self-gravitating systems, realistic halo formation is clearly more complicated. Numerical simulation has been the tool most frequently used in cosmology to study structure

formation. It has provided powerful insight and occasionally surprising results (e.g., universal density profile) unexplainable by the simulations themselves. Our understanding of structure formation can certainly be enhanced if a complementary tool is developed. Complementary approaches are also especially needed at this time to help us better understand and perhaps resolve the “cuspy halo” and “dwarf satellite” problems of the cold dark matter (CDM) model.

In this paper we offer a new perspective on the subject of dark matter halo formation by proposing an approach different from both the radial infall model and N -body simulations. We focus on the phase-space distribution of dark matter in galaxy halos instead of on the motion of a mass shell (as in infall models) or the orbit of a simulation particle (as in N -body). We attempt to understand the evolution of the phase-space distribution with a statistical description based on kinetic theory. Our method is to be contrasted with the standard method where the individual dark matter particles are described by the collisionless Boltzmann equation; N -body simulations are in essence a Monte Carlo method for solving this equation. Instead, we consider the phase space of dark matter particles in an average halo and develop a cosmological kinetic theory for halo evolution in the presence of a spectrum of density fluctuations.

The essential features of our approach are represented by the fluctuation-dissipation theorem for time-dependent stochastic processes. To elaborate on this concept, we recall that stochastic modeling was first developed to explain the Brownian motion of pollen particles suspended in a fluid (see § 5.3 for a more detailed discussion). This phenomenon is now understood to arise from a fluctuating force on a Brownian particle due to many rapid collisions with the molecules in the surrounding fluid. These fluctuations cause successive small changes in the particle velocity, hence to a random walk. The consequence of many such random walks is diffusion, a form of dissipation.

Because of the large number of molecules, it is impossible to solve the coupled equations of motion for the trajectories of all molecules and the pollen particle in the fluid. Fortunately, a stochastic treatment suffices. One mathematical prescription is provided by the Langevin equation (Langevin 1908), which models the forces arising from the fluctuations as a stochastic term that is added to all other deterministic forces in the equation of motion for a particle. An alternative prescription is given by the Fokker-Planck equation (Fokker 1914; Planck 1917), which is an equation of motion for the probability density or distribution function of particles undergoing random walks. In this case, the collective changes in the velocities due to the fluctuations in the system are represented by a drift coefficient and a diffusion coefficient arising from the collision terms in the collisional Boltzmann equation.

Stochastic phenomena occur beyond molecular scales. In astrophysics, the most studied example is the evolution of dense stellar systems such as globular clusters (Chandrasekhar 1943; Spitzer 1987; Binney & Tremaine 1988). A star in a globular cluster experiences a fluctuating force from many gravitational two-body encounters and suffers a succession of small velocity changes. The fluctuations in these systems arise from granularity in the mass distribution at the locations of the stars. The Fokker-Planck approximation is valid for these weak encounters. We explore this example in greater detail and contrast it with our model of galaxy halo evolution in § 5.4.

The key feature common to all of these systems is the presence of stochastic fluctuations that result in dissipation.

In the cosmological setting addressed by this paper, the fluctuations arise from cosmological perturbations produced in the early universe. At early times, these perturbations are Gaussian and fully characterized by their power spectrum; at late times, the substructures in halos seen in recent simulations are the nonlinear counterparts of this spectrum of early ripples. These fluctuations scatter particles. If one considers an ensemble of halos, the masses and locations of the substructures (or their Gaussian predecessors) are random variables. The substructures thereby introduce “noise” into the evolution of dark matter halos. As we show in this paper, the result is dissipation in the form of drift and velocity space diffusion.

Recent work has explored the use of the Fokker-Planck equation in cosmological problems. Evans & Collett (1997), for example, searched for steady state solutions of the Fokker-Planck equation assuming that the fluctuations arise from two-body encounters. This paper also assumed a simple isotropic phase-space distribution dependent only on the total energy and a radial power-law form for the potential and density. Under these assumptions, Evans & Collett (1997) found that $\rho \propto r^{-4/3}$ was a steady state solution under the effects of encounters. There was no dynamics in this work, however: the system was assumed to be in collisionless equilibrium and the nature of the clumps producing the collisions was left arbitrary. It also remains to be determined whether galaxy halos, either dark matter or stellar, actually evolve to $\alpha = 4/3$ cusp. Weinberg (2001a, 2001b) examined the dynamics of halos due to stochastic fluctuations from three different noise processes for the Fokker-Planck scattering term: the merging clumps were assumed to fly by with constant velocity, to follow decaying orbits due to dynamical friction, or to orbit inside halos. The results of this study were promising and supported the notion that the density profiles of halos were affected by mergers and substructures. The fluctuations in this work, however, were constructed by hand instead of being self-consistently calculated, and the results about a central cusp were inconclusive.

Our objective here is to construct a kinetic equation in which the drift and diffusion terms are derived from realistic density fluctuations in current cosmological models. We also retain the full phase-space information by, in effect, using $f(E, J, v_r, t)$ without averaging over the angular momentum or radial velocity. We also do not assume collisionless equilibrium; instead, we base our approach on exact dynamics.

The organization of this paper is as follows. In § 2 we derive the kinetic equation relevant for the evolution of an average halo, starting with the collisionless Boltzmann equation for individual dark matter particles. This procedure gives rise to a collision term on the right-hand side of the kinetic equation. This term represents a correlated force density and arises from fluctuations in the ensemble-averaged gravitational potential due to clustering of the matter distribution.

In § 3 we introduce the concepts and physical variables needed to apply this equation to cosmological problems. The key step is in relating the phase-space density $f(\mathbf{r}, \mathbf{v}, t)$ to the probability densities of two standard variables in cosmology: the mass density ρ and velocity \mathbf{v} . We then focus on the quasi-linear regime in which the fluctuations about the mean density and Hubble flow of an isotropic and homogeneous universe are small. This focus allows us to relate ρ and \mathbf{v} to the density and displacement fields δ and ψ , which are Gaussian random fields in standard cosmological models.

We leave the most mathematical part of the paper to Appendix A, where we describe how to use the statistics of

constrained Gaussian random fields to express the probability densities of § 3 in terms of the (constrained) means and covariances of the density and displacement fields. The idea of a constrained field is central to this study because our aim is to construct a kinetic theory for dark matter particles that will later reside in a collapsed halo rather than at any random location in the universe. The relevant perturbations are for a density field constrained to have certain properties, e.g., a specified height or gradient at some point in space. The spectrum for such constrained fields is related but not identical to the familiar linear power spectrum $P(k)$ for unconstrained Gaussian density fluctuations.

In § 4 we put the final expression for the correlated force derived in Appendix A back into the kinetic equation and show that the resulting equation is of the Fokker-Planck form where one term represents a drift force and the other term represents diffusion in velocity space. We work out these diffusion coefficients for the currently favored cold dark matter plus a cosmological constant (Λ CDM) model and provide explicit expressions for these coefficients in terms of the linear matter fluctuation power spectrum. We also present the results for the coefficients computed for this model.

In § 5 we discuss further the physical meaning of the results in § 4. We interpret the diffusion terms by taking velocity moments of the kinetic equation. To place our cosmological kinetic equation in a broader perspective, we discuss in more detail the classical examples of Brownian motion and globular clusters and point out the similarities and differences among the three cases. In § 6 we provide a summary and conclusions. In Appendix B we investigate the relaxation to stationary solutions of our cosmological Fokker-Planck equation.

2. DERIVATION OF THE KINETIC EQUATION

We derive the kinetic equation for galaxy halo evolution following the methods presented by Bertschinger (1993) for the derivation of the BBGKY hierarchy. The starting point is the one-particle phase-space density for dark matter particles in a single halo, the Klimontovich (or Klimontovich-Dupree) density (Klimontovich 1967; Dupree 1967)

$$f_{\mathbf{K}}(\mathbf{r}, \mathbf{v}, t) = m \sum_i \delta_{\mathbf{D}}[\mathbf{r} - \mathbf{r}_i(t)] \delta_{\mathbf{D}}[\mathbf{v} - \mathbf{v}_i(t)], \quad (1)$$

where $\delta_{\mathbf{D}}$ is the Dirac delta function and the subscript K denotes the use of the Klimontovich density. We suppose that each halo is made of equal-mass particles of mass m , for example, the elementary particles of nonbaryonic dark matter models. For convenience, we use velocity rather than momentum and choose units so that $f d^3v$ is a mass density. We use proper coordinates except in § 3.2 and Appendix A, where we switch to the appropriate cosmological (comoving) coordinates.

The Klimontovich density is simply a way of casting the exact trajectories of all N particles in a system into the form of a phase-space density. The number of particles in $d^3r d^3v$ about (\mathbf{r}, \mathbf{v}) is $f_{\mathbf{K}} d^3r d^3v$, which equals either 0 or 1. The Klimontovich density consists of delta functions because, when one is dealing with a single system (as opposed to an ensemble), every particle has a well-defined position and velocity. The reason for introducing this quantity is that it gives a particularly clean way of going from exact N -body dynamics to a statistical description by taking expectation values of $f_{\mathbf{K}}$ or products of more than one $f_{\mathbf{K}}$ over an ensemble of N -body

systems. This approach to kinetic theory is different from the usual method that begins with the N -particle distribution for an ensemble.

The Klimontovich density obeys the Klimontovich-Dupree equation

$$\frac{\partial f_{\mathbf{K}}}{\partial t} + \mathbf{v} \cdot \frac{\partial f_{\mathbf{K}}}{\partial \mathbf{r}} + \mathbf{g}_{\mathbf{K}} \cdot \frac{\partial f_{\mathbf{K}}}{\partial \mathbf{v}} = 0, \quad (2)$$

where

$$\mathbf{g}_{\mathbf{K}}(\mathbf{r}, t) = -Gm \sum_i \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3} = -G \int d^6w' f_{\mathbf{K}}(\mathbf{w}', t) \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}. \quad (3)$$

For notational convenience we have grouped together all six phase-space coordinates into $\mathbf{w} \equiv \{\mathbf{r}, \mathbf{v}\}$. Equation (2) is exact, as may be verified by substituting equation (1) and using the equations of motion for individual particles,

$$\frac{d\mathbf{r}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{v}}{dt} = \mathbf{g}_{\mathbf{K}}. \quad (4)$$

In practice, the gravitational force may be softened by modifying equation (3) to reduce artificial two-body relaxation, as is done in N -body simulations. Equations (1) and (2) are entirely equivalent to equation (4) for one system. The Klimontovich phase-space density retains *all* information about a particular halo because it specifies the trajectories of all particles. Equations (1)–(4) are formally equivalent to a perfect N -body simulation. That is not what we want. Rather than giving a perfect description of a single halo, we average over halos to obtain a statistical description of halo evolution. While it is impossible to completely describe a single galactic halo of dark matter particles, by using statistical mechanics methods we *can* describe the average of infinitely many halos!

Before proceeding further, we make one important modification to the usual procedure in kinetic theory. Equation (4) assumes that we work in an inertial frame. This is satisfactory for an isolated halo but not for a halo formed in a cosmological context where neighbors cause it to orbit in the net gravitational field produced by all matter. The center of mass of a halo typically moves much farther in a Hubble time than the size of the halo itself, even in an inertial frame in which the halo begins at rest. For example, for the CDM family of models, the displacement of a mass element from its initial position (in comoving coordinates) is typically several megaparsecs today. If we statistically average over an ensemble of halos, each of which wanders this far in a Hubble time, the phase-space density will be artificially smeared out, making the resulting average “halo” much larger than any realistic halo. We avoid this by doing the same thing that simulators do when they construct halo profiles: we work in the center-of-mass frame of each halo.

Indeed, at each moment the halo center of mass must be at rest in our reference frame. This frame is noninertial because of the gravitational forces acting on the halo. Thus, we must modify the kinetic equation to apply to an accelerating frame. Fortunately, the accelerations are small and the frame is extended only a few megaparsecs, so we do not have to worry about relativistic effects. We simply include the inertial (“fictitious”) forces associated with working in an accelerating frame.

We choose to define $\mathbf{r} = 0$ as the center of mass of each halo. Transforming equation (4) to the center-of-mass frame

is very simple: we replace \mathbf{g} by the *tidal* gravitational field relative to the center of each halo, $\mathbf{g}(\mathbf{r}) - \mathbf{g}(0)$, and then subtract the center-of-mass velocity from the initial velocity of each particle. Equation (2) is unchanged except that \mathbf{g} is replaced by

$$\begin{aligned} \mathbf{g}_{KT}(\mathbf{r}, t) &= -Gm \sum_i \left(\frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3} + \frac{\mathbf{r}_i}{|\mathbf{r}_i|^3} \right) \\ &= -G \int d^6 w' f_K(\mathbf{w}', t) \left(\frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} + \frac{\mathbf{r}'}{|\mathbf{r}'|^3} \right). \end{aligned} \quad (5)$$

The subscript T is a reminder that this is the tidal gravitational field.

Now we return to the usual procedure for deriving a kinetic equation. The Klimontovich density is the phase-space density for a single realization of a halo. We might imagine performing many N -body simulations of different halos, each with slightly different initial conditions. If we average the Klimontovich density over this ensemble of halos, the result (at least in the limit where the size of the ensemble becomes infinite) is no longer a sum of delta functions but instead is a smooth *one-particle distribution function*:

$$f(\mathbf{w}, t) \equiv \langle f_K(\mathbf{w}, t) \rangle, \quad (6)$$

where angle brackets denote the ensemble average. Recall that for notational convenience we have grouped all six phase-space variables into \mathbf{w} .

Our goal is to derive a tractable kinetic equation for f . The obvious next step is to take the ensemble average of equation (2). To proceed, however, we will need the *two-particle distribution function* $f_2(\mathbf{w}_1, \mathbf{w}_2, t)$, which follows from averaging the product of Klimontovich densities at two points:

$$\langle f_K(\mathbf{w}_1, t) f_K(\mathbf{w}_2, t) \rangle \equiv \delta_D(\mathbf{w}_1 - \mathbf{w}_2) f(\mathbf{w}_1, t) + f_2(\mathbf{w}_1, \mathbf{w}_2, t). \quad (7)$$

The delta-function term $\delta_D(\mathbf{w}_1 - \mathbf{w}_2) \equiv \delta_D(\mathbf{r}_1 - \mathbf{r}_2) \delta_D(\mathbf{v}_1 - \mathbf{v}_2)$ corresponds to the case in which the two points lie on top of one particle. Two distinct particles give the f_2 term, which may always be written as a product term plus a two-particle correlation:

$$f_2(\mathbf{w}_1, \mathbf{w}_2, t) \equiv f(\mathbf{w}_1, t) f(\mathbf{w}_2, t) + f_{2c}(\mathbf{w}_1, \mathbf{w}_2, t). \quad (8)$$

This equation defines the two-point correlation function in phase space, f_{2c} .

The averaging we perform is similar to the averaging used in defining ordinary clustering correlation functions in cosmology. The differences here are that we are not averaging over arbitrary systems but only over those that lead to formation of a halo at $\mathbf{r} = 0$ and that we retain velocity information. Had we averaged over all space and integrated over velocities, we would have gotten $\int d^3 v f = \bar{\rho}$ and $\int d^3 v_1 d^3 v_2 f_{2c} = \bar{\rho}^2 \xi$, where $\bar{\rho}$ is the cosmic mean mass density and ξ is the usual spatial two-point correlation function.

Now, by taking the ensemble average of equation (2), we arrive at a kinetic equation for the evolution of the average halo phase-space density (Bertschinger 1996):

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{g}_T \cdot \frac{\partial f}{\partial \mathbf{v}} = - \frac{\partial}{\partial \mathbf{v}} \cdot \mathbf{F}, \quad (9)$$

where we define the ensemble-averaged tidal field

$$\mathbf{g}_T(\mathbf{r}, t) \equiv \langle \mathbf{g}_{KT}(\mathbf{r}, t) \rangle = -G \int d^6 w' f(\mathbf{w}', t) \left(\frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} + \frac{\mathbf{r}'}{|\mathbf{r}'|^3} \right) \quad (10)$$

and the correlated force density

$$\begin{aligned} \mathbf{F}(\mathbf{w}, t) &\equiv \text{Cov}[\mathbf{g}_{KT}(\mathbf{r}, t), f_K(\mathbf{w}, t)] \\ &= -G \int d^6 w' f_{2c}(\mathbf{w}, \mathbf{w}', t) \left(\frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} + \frac{\mathbf{r}'}{|\mathbf{r}'|^3} \right). \end{aligned} \quad (11)$$

The covariance of two random variables is denoted $\text{Cov}[A, B] \equiv \langle (A - \langle A \rangle)(B - \langle B \rangle) \rangle = \langle AB \rangle - \langle A \rangle \langle B \rangle$. The product of \mathbf{g}_{KT} and f_K involves a product of two Klimontovich densities, for which we have used equations (7) and (8). The one-particle discreteness term in equation (7) makes no contribution because the particle has no self-force.

Equation (9) is the first BBGKY hierarchy equation (Ichimaru 1973; Davis & Peebles 1977; Peebles 1980). In cosmological applications this equation is usually derived starting from the Liouville equation for the N -particle distribution function. In order to clarify the averaging that is done here, we have started instead from the one-particle Klimontovich density. The result, equation (9), is an evolution equation that looks very much like equation (2). By averaging over halos, however, we have introduced a correlation integral term on the right-hand side. The *average* halo therefore does not evolve according to the Vlasov equation; instead, it evolves according to a kinetic equation with an effective collision term. Equations (9)–(11) are exact. No approximation has been made yet.

Equation (9) is a continuity equation for density in phase space. Without the right-hand side, it says that phase-space density is conserved along trajectories given by equation (4) modified for the tidal force. The right-hand side gives a correction due to the fact that fluctuations about the ensemble average lead to correlated forces. As we will see, the effects of these fluctuations lead to dissipation. Because the correction term is a divergence, the mass density $\rho(\mathbf{r}, t) = \int f d^3 v$ is not dissipated but instead is conserved.

In most applications of kinetic theory, the right-hand side of equation (9) arises from particle collisions. In the usual treatment of a gas of particles with short-range forces, one assumes $f_{2c} = 0$ always except during particle collisions. This is the assumption of molecular chaos that Boltzmann introduced to derive his celebrated kinetic equation. Our situation, however, is different: we cannot neglect f_{2c} because it can be much larger than the single-particle term in equation (8) as a result of strong gravitational clustering.

Heuristically, f_{2c} describes the substructure within a galaxy halo at the two-point level. (Higher order correlation functions would be needed for a complete description.) Current cosmological models and observations imply that the initial density field has fluctuations that cause the formation of many small halos that subsequently merge into present-day halos. The lumpiness of the matter distribution represents a fluctuation about the ensemble-average density field. Through the correlated force density \mathbf{F} , these fluctuations cause changes in the energy and angular momentum of individual particle orbits that are important for the evolution of the one-particle distribution function.

Equation (9) is useful only if we find an expression for f_{2c} . One way to proceed would be to derive a kinetic equation for f_{2c} by taking the ensemble average of equation (2) multiplied by the Klimontovich density. This leads to the second BBGKY equation, which depends on the three-point correlation in phase space, which obeys the third BBGKY equation, and so on. Thus, we either have an infinite chain of coupled equations of increasing dimensionality, or we must close the system. Davis & Peebles (1977) handled this problem in their study of gravitational clustering by writing the three-point function in terms of products of two-point functions. We proceed differently, by seeking to close the hierarchy at the first level, equation (9).

To summarize this section, we have derived a formally exact evolution equation for the average halo. The one-particle phase-space density is conserved up to a fluctuating force term arising from the two-point correlation function in phase space. Physically this term represents the effects of substructure within and around halos, which scatter particles to new orbits. In the next two sections we show that we can in fact express f_{2c} in terms of the one-particle phase-space density f using a relation that is exact to second order in cosmological perturbation theory. This relation will enable us to close the first BBGKY hierarchy.

3. DISTRIBUTION FUNCTIONS FROM PROBABILITY THEORY

Solving our evolution equation in equations (9)–(11) requires specifying the phase-space density f at an initial time and the two-particle correlation f_{2c} at all times. In this section we show how to compute these quantities from the probability distributions of mass density and velocity at one and two points in space.

3.1. Probability Density versus Phase-Space Distribution Function

We wish to derive expressions that would relate $f(\mathbf{w}_1, t)$ and $f_{2c}(\mathbf{w}_1, \mathbf{w}_2, t)$ to probability distributions of mass density and velocity because the latter quantities are naturally specified by a cosmological model for structure formation. The method is to examine the mass and momentum in a small volume of space and to construct ensemble averages of the Klimontovich density using the probability distribution of density and velocity.

We use the symbol $p(x, y, \dots)$ to refer to the probability density with respect to its arguments, where the joint probability distribution for (x, y, \dots) is $dP(x, y, \dots) = p(x, y, \dots) dx dy \dots$. We assume that these probability distributions are always normalized to unit total probability. The variables relevant for this paper are the densities ρ and velocities \mathbf{v} at two points $\mathbf{r}_1, \mathbf{r}_2$ and at one time t : $\rho_1 \equiv \rho(\mathbf{r}_1, t)$, $\mathbf{v}_1 \equiv \mathbf{v}(\mathbf{r}_1, t)$, $\rho_2 \equiv \rho(\mathbf{r}_2, t)$, and $\mathbf{v}_2 \equiv \mathbf{v}(\mathbf{r}_2, t)$. Initially we assume that the velocity field in a given realization is single valued in space. We use the following probability distributions:

$$\begin{aligned} dP(\rho_1) &= p(\rho_1) d\rho_1, \\ dP(\rho_1, \mathbf{v}_1) &= p(\rho_1, \mathbf{v}_1) d\rho_1 d^3v_1, \\ dP(\rho_1, \mathbf{v}_1, \rho_2) &= p(\rho_1, \mathbf{v}_1, \rho_2) d\rho_1 d^3v_1 d\rho_2, \\ dP(\rho_1, \mathbf{v}_1, \rho_2, \mathbf{v}_2) &= p(\rho_1, \mathbf{v}_1, \rho_2, \mathbf{v}_2) d\rho_1 d^3v_1 d\rho_2 d^3v_2, \end{aligned} \quad (12)$$

where the different probability densities are related to one another by

$$\begin{aligned} p(\rho_1, \mathbf{v}_1, \rho_2) &= \int d^3v_2 p(\rho_1, \mathbf{v}_1, \rho_2, \mathbf{v}_2), \\ p(\rho_1, \mathbf{v}_1) &= \int d\rho_2 p(\rho_1, \mathbf{v}_1, \rho_2), \\ p(\rho_1) &= \int d^3v_1 p(\rho_1, \mathbf{v}_1), \\ 1 &= \int d\rho_1 p(\rho_1). \end{aligned} \quad (13)$$

These probability distributions depend on time, but we suppress the t -dependence for clarity.

For a small volume in phase space, $f(\mathbf{w}_1, t) d^3r_1 d^3v_1$ equals the mean mass contained in $d^3r_1 d^3v_1$. The same quantity can be calculated by averaging $\rho_1 d^3r_1$ over density at fixed velocity using $dP(\rho_1, \mathbf{v}_1)$, i.e., $\int \rho_1 d^3r_1 dP(\rho_1, \mathbf{v}_1)$. Equating the results gives

$$\begin{aligned} f(\mathbf{w}_1, t) &= \int_0^\infty d\rho_1 \rho_1 p(\rho_1, \mathbf{v}_1) = \int_0^\infty d\rho_1 \rho_1 p(\rho_1 | \mathbf{v}_1) p(\mathbf{v}_1) \\ &= \langle \rho_1 | \mathbf{v}_1 \rangle p(\mathbf{v}_1). \end{aligned} \quad (14)$$

We have used the conditional probability density $p(\rho_1 | \mathbf{v}_1)$ to define the conditional mean $\langle \rho_1 | \mathbf{v}_1 \rangle$. Note that the probability densities depend on (\mathbf{r}_1, t) through $\rho_1 = \rho(\mathbf{r}_1, t)$ and $\mathbf{v}_1 = \mathbf{v}(\mathbf{r}_1, t)$, so equation (14) yields the desired one-particle phase-space dependence on $\mathbf{w}_1 = \{\mathbf{r}_1, \mathbf{v}_1\}$.

This argument is easily extended to the two-point phase-space density, using the joint probability distribution of density and velocity at two points. As equation (11) shows, we need $f_{2c}(\mathbf{w}_1, \mathbf{w}_2, t)$, which depends on \mathbf{v}_1 but not \mathbf{v}_2 ; i.e., we need the velocity at only one point. Recalling that $f_{2c}(\mathbf{w}_1, \mathbf{w}_2, t) = f_2(\mathbf{w}_1, \mathbf{w}_2, t) - f(\mathbf{w}_1, t)f(\mathbf{w}_2, t)$, we obtain

$$\begin{aligned} f_{2c}(\mathbf{w}_1, \mathbf{w}_2, t) &\equiv \int d^3v_2 f_{2c}(\mathbf{w}_1, \mathbf{w}_2, t) \\ &= \int_0^\infty d\rho_1 \rho_1 \int_0^\infty d\rho_2 \rho_2 [p(\rho_1, \mathbf{v}_1, \rho_2) \\ &\quad - p(\rho_1, \mathbf{v}_1)p(\rho_2)] \\ &= \int_0^\infty d\rho_1 \rho_1 \int_0^\infty d\rho_2 \rho_2 [p(\rho_1, \rho_2 | \mathbf{v}_1) \\ &\quad - p(\rho_1 | \mathbf{v}_1)p(\rho_2)] p(\mathbf{v}_1) \\ &= [\langle \rho_1 \rho_2 | \mathbf{v}_1 \rangle - \langle \rho_1 | \mathbf{v}_1 \rangle \langle \rho_2 \rangle] p(\mathbf{v}_1). \end{aligned} \quad (15)$$

Although we have derived equations (14) and (15) assuming a single-valued velocity field in each realization, they are also valid when there is a distribution of velocities at each point. One simply interprets ρ_1 as the total mass density while \mathbf{v}_1 is a single-particle velocity. Therefore, these equations are fully general.

3.2. Cosmological Variables

At high redshift, before dark matter halos formed, the density and velocity fields were only slightly perturbed from a uniformly expanding cosmological model. It is thought that the fluctuations of density and velocity responsible for all cosmic structure were Gaussian random fields produced during the inflationary era of the very early universe. These fluctuations provide us with both the initial conditions and the random

element leading to a probabilistic description for halo formation. Our goal now is to obtain expressions for equations (14) and (15) at early times while the density fluctuations are small.

Before proceeding further, we review the description of density and velocity fields at high redshift in an almost uniformly expanding cosmological model. First, the mean expansion is described by the cosmic expansion scale factor $a(t)$, which is related to the redshift by $1+z=1/a(t)$. The Hubble expansion rate is $H(t)=\dot{a}/a$. The matter contributes a fraction $\Omega_m(t)$ to the closure density. We account for the mean expansion by defining new coordinates conventionally called comoving coordinates: $\mathbf{x} \equiv \mathbf{r}/a(t)$.

Adopting the standard cosmological model, we assume that the dark matter is cold. This assumption means that the velocity field is single valued in space at early times before halos form. We can then write the density field ρ and velocity field \mathbf{v} in terms of perturbations δ and $\boldsymbol{\psi}$ around a uniformly expanding background:

$$\rho(\mathbf{r}, t) = \bar{\rho}(t)[1 + \delta(\mathbf{x}, t)], \quad \mathbf{v}(\mathbf{r}, t) = H(t)[\mathbf{r} + b(t)\boldsymbol{\psi}(\mathbf{x}, t)]. \quad (16)$$

Here $b(t) = a(d \ln D / d \ln a) \approx a[\Omega_m(t)]^{0.6}$ is the growth rate for density fluctuations with time dependence $\delta(\mathbf{x}, t) \propto D(a)$. We emphasize that equation (16) does not assume that the density fluctuations are small; it only assumes that the velocity field is single valued. We are implicitly assuming that all of the matter may be described as a single cold noninteracting fluid (i.e., we assume that baryons move like CDM over the length scales of interest).

The computation of the functions $a(t)$, $H(t)$, $\bar{\rho}(t)$, $D(t)$, and $b(t)$ is standard in cosmology and is not described here. At early times when $\delta^2 \ll 1$, mass conservation relates the density and velocity perturbation fields by $\delta = -(\partial/\partial \mathbf{x}) \cdot \boldsymbol{\psi}$, or

$$\boldsymbol{\psi}(\mathbf{x}, t) = - \int \frac{d^3 \mathbf{x}'}{4\pi} \frac{(\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} \delta(\mathbf{x}', t). \quad (17)$$

Now we make a crucial step by changing variables from $(\mathbf{r}, \mathbf{v}, \rho)$ used previously to $(\mathbf{x}, \boldsymbol{\psi}, \delta)$. The reason for doing so is that the description in comoving coordinates \mathbf{x} and perturbation variables $\boldsymbol{\psi}$ and δ is much simpler in the small perturbation limit. Following the notation in § 3.1, we write $\delta_1 \equiv \delta(\mathbf{x}_1, t)$, $\delta_2 \equiv \delta(\mathbf{x}_2, t)$, and $\boldsymbol{\psi}_1 \equiv \boldsymbol{\psi}(\mathbf{x}_1, t)$. These variables have normalized joint probability densities $p(\delta_1)$, $p(\delta_1, \boldsymbol{\psi}_1)$, and $p(\delta_1, \boldsymbol{\psi}_1, \delta_2)$ in analogy with equations (12) and (13).

To change variables from (\mathbf{v}, ρ) to $(\boldsymbol{\psi}, \delta)$, we use $d\rho = \bar{\rho} d\delta$, $d^3 \mathbf{v} = (Hb)^3 d^3 \boldsymbol{\psi}$, and $d\rho d^3 \mathbf{v} p(\rho, \mathbf{v}) = d\delta d^3 \boldsymbol{\psi} p(\delta, \boldsymbol{\psi})$. Equation (14) then becomes

$$f(\mathbf{w}_1, t) = \bar{\rho}(1 + \langle \delta_1 | \boldsymbol{\psi}_1 \rangle) p(\boldsymbol{\psi}_1) (Hb)^{-3}, \quad (18)$$

where

$$\langle \delta_1 | \boldsymbol{\psi}_1 \rangle \equiv \int d\delta_1 p(\delta_1 | \boldsymbol{\psi}_1) \delta_1, \quad p(\delta_1 | \boldsymbol{\psi}_1) = \frac{p(\delta_1, \boldsymbol{\psi}_1)}{p(\boldsymbol{\psi}_1)}, \quad (19)$$

and $p(\delta_1 | \boldsymbol{\psi}_1)$ is the conditional distribution of δ_1 . Similarly, equation (15) becomes

$$f_{2c}(\mathbf{w}_1, \mathbf{x}_2, t) = \bar{\rho}^2 (\langle \delta_2 | \boldsymbol{\psi}_1 \rangle - \langle \delta_2 \rangle + \langle \delta_1 \delta_2 | \boldsymbol{\psi}_1 \rangle - \langle \delta_1 | \boldsymbol{\psi}_1 \rangle \langle \delta_2 \rangle) p(\boldsymbol{\psi}_1) (Hb)^{-3}. \quad (20)$$

Although the derivation of equations (18) and (20) assumed that the velocity field is single valued, they are valid in general provided that one interprets δ_1 and δ_2 as total mass density perturbations (not necessarily small) while $\boldsymbol{\psi}_1$ is proportional to the single-particle velocity. Equations (18) and (20) are equivalent to equations (14) and (15) expressed in different variables. Their utility becomes evident in Appendix A, where we evaluate the probability distributions of density and velocity in cosmological models.

All that remains is to calculate the expectation values in equations (18) and (20) and to substitute the latter into equation (11). This calculation is somewhat complicated and is presented in Appendix A. Part of the complication arises because of an important detail we have so far neglected: halos form preferentially in regions that are initially overdense. Our calculations assume that a halo forms at $\mathbf{r} = 0$. Consequently, we should not evaluate the expectation values for arbitrary points in space but rather for those that satisfy appropriate constraints.

Strictly speaking, it is impossible to specify in advance the regions that will form halos in a cosmological model with random initial fluctuations because the dynamics of gravitational clustering is too complicated. A simple model, however, has been found to give reasonable agreement with numerical simulations: halos form at maxima of the smoothed initial density field (Bardeen et al. 1986). These constraints are found to modify the expectation values in equations (18) and (20). To emphasize the application of constraints, we rewrite these expectation values with a subscript C :

$$f(\mathbf{w}_1, t) = \bar{\rho}(1 + \langle \delta_1 | \boldsymbol{\psi}_1 \rangle_C) p(\boldsymbol{\psi}_1) (Hb)^{-3} \quad (21)$$

and

$$f_{2c}(\mathbf{w}_1, \mathbf{x}_2, t) = \bar{\rho}^2 (\langle \delta_2 | \boldsymbol{\psi}_1 \rangle_C - \langle \delta_2 \rangle_C + \langle \delta_1 \delta_2 | \boldsymbol{\psi}_1 \rangle_C - \langle \delta_1 | \boldsymbol{\psi}_1 \rangle_C \langle \delta_2 \rangle_C) p(\boldsymbol{\psi}_1) (Hb)^{-3}. \quad (22)$$

Equations (21) and (22) are the main results of this section. We use them to specify the initial conditions on f and to determine the correlated force density in equation (11).

The next step is to evaluate the necessary constrained expectation values using the standard cosmological model. We present the details of this calculation in Appendix A, where we choose to err on the side of simplicity by following Bardeen et al. (1986) and using only the zeroth and first derivatives of the smoothed initial density field $\Delta(\mathbf{x})$ as constraints. Detailed comparison has shown that this simple model does not predict perfectly all halo formation sites (Katz et al. 1993), and more complicated constraints work better (e.g., Monaco et al. 2002). The procedure of Appendix A can presumably be extended to incorporate more complicated constraints.

4. FOKKER-PLANCK EQUATION

After having devoted § 3 and Appendix A to probability and statistics, we are now ready to return to the kinetic equation in § 2 and write down an explicit expression for the correlated force density $\mathbf{F}(\mathbf{w}, t)$ on the right-hand side of the kinetic equation (9). The approach we have followed is exact rather than phenomenological: our starting point was the first BBGKY hierarchy equation, an exact kinetic equation. The only approximation we have made is to assume that the

density and velocity fields are Gaussian random fields. The resulting two-particle correlation function, equation (A22), depends in a simple way on the one-particle distribution $p(\psi_1)$. As we show next, this simple dependence leads to the Fokker-Planck equation. This is a remarkable and nontrivial result. The Fokker-Planck equation is a great simplification of the exact (and intractable) BBGKY hierarchy. Usually the Fokker-Planck equation is *assumed* to hold a priori, and then its diffusion coefficients are estimated. Here, on the other hand, we have *derived* the Fokker-Planck equation. Along with it we obtain the diffusion coefficients in the quasi-linear regime, as follows.

4.1. Diffusion Coefficients

Substituting equation (A22) into equation (11) using equation (15), converting to comoving coordinates, using equation (17) to relate ψ to \mathbf{v} , and using equation (A21), to third order in δ we get

$$\begin{aligned} \mathbf{F}(\mathbf{w}, t) = & -4\pi G\bar{\rho}a \left\{ \left[C(\delta, \psi_0) - C(\delta, \psi) \cdot C^{-1}(\psi, \psi) \right. \right. \\ & \left. \left. \cdot C(\psi, \psi_0) \right] f + (Hb) \left[C(\psi, \psi) - C(\psi_0, \psi) \right. \right. \\ & \left. \left. - \langle \psi \rangle_C \otimes C(\delta, \psi) \right] \cdot \frac{\partial f}{\partial \mathbf{v}} \right\} + O(\delta^4), \end{aligned} \quad (23)$$

where the covariance matrices C of variables δ and ψ are given in equations (A14a)–(A14f) and (A15). Here $\psi \equiv \psi(\mathbf{x}, t)$ and $\psi_0 \equiv \psi(0, t)$. Note that all the terms with subscript 0 arise from our use of the tidal field rather than the total gravitational acceleration.

Substituting equation (23) into equation (9), in proper coordinates our kinetic equation in the quasi-linear regime reduces to the Fokker-Planck equation

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{g}_T \cdot \frac{\partial f}{\partial \mathbf{v}} = -\frac{\partial}{\partial \mathbf{v}} \cdot \left(\mathbf{A}f - \mathbf{D} \cdot \frac{\partial f}{\partial \mathbf{v}} \right), \quad (24)$$

specified by a vector field \mathbf{A} and a tensor field \mathbf{D} . Together, \mathbf{A} and \mathbf{D} are called diffusion coefficients. Heuristically, \mathbf{A} represents a drift force (\mathbf{A} has units of acceleration) while \mathbf{D} represents the diffusive effects of fluctuating forces. The flux density in velocity space is $\mathbf{A}f - \mathbf{D} \cdot \partial f / \partial \mathbf{v}$. Keeping only terms up to second order in the density perturbations, these quantities are given by

$$\begin{aligned} \mathbf{A}(\mathbf{r}, t) & \approx -4\pi G\bar{\rho}a \left[C(\delta, \psi_0) - C(\delta, \psi) \cdot C^{-1}(\psi, \psi) \cdot C(\psi, \psi_0) \right] \\ & = -\text{Cov}(\delta, \mathbf{g}_0 | \psi) = \text{Cov}(\delta, \mathbf{g}_T | \mathbf{v}) \end{aligned} \quad (25)$$

and

$$\mathbf{D}(\mathbf{r}, t) \approx 4\pi G\bar{\rho}a Hb C(\psi - \psi_0, \psi) = \text{Cov}(\mathbf{g}_T, \mathbf{v}). \quad (26)$$

In equations (25) and (26) we have also used the fact that in the quasi-linear regime the peculiar gravity and displacement are related by $\mathbf{g} = 4\pi G\bar{\rho}a\psi$ (Zel'dovich 1970), and the tidal field is $\mathbf{g}_T \equiv \mathbf{g}(\mathbf{r}, t) - \mathbf{g}_0$, where $\mathbf{g}_0 \equiv \mathbf{g}(0, t)$. In writing the second line of equation (25) we have used equation (A7) with $\mathbf{Y}_B = \{\delta, \psi_0\}$ and $\mathbf{Y}_A = \psi - \langle \psi \rangle_C$. Thus, $\text{Cov}(\delta, \mathbf{g}_0 | \mathbf{v})$ should be understood to be subject to the same extremum or other constraints such as the constrained covariance function $C(\delta, \psi)$. The displacement and velocity are related by the

second formula in equation (16). Thus, in linear theory $\mathbf{g} \propto \mathbf{v}$ so that conditioning on \mathbf{v} makes the fluctuation in \mathbf{g} vanish, allowing us to replace $-\text{Cov}(\delta, \mathbf{g}_0 | \psi)$ by $\text{Cov}(\delta, \mathbf{g}_T | \mathbf{v})$ in equation (25). In all cases, the covariances are taken subject to the extremum constraints discussed in Appendix A.

It is worth noting that had we used the total gravitational acceleration $\mathbf{g}(\mathbf{r}, t)$ instead of the tidal field $\mathbf{g}_T(\mathbf{r}, t)$, the drift vector \mathbf{A} would have been zero and $\mathbf{D} = \text{Cov}(\mathbf{g}, \mathbf{v})$ would have been larger. We conclude that tidal effects are crucial for a proper description of halo evolution (Dekel et al. 2003). This is evident in the final expressions for \mathbf{A} and \mathbf{D} . After the lengthy derivations, these expressions are remarkably simple, suggesting that they may have greater validity than the case of small-amplitude Gaussian fluctuations that we have analyzed here.

In going from equations (A21) and (23) to equations (25) and (26) we have dropped the contributions made by $\langle \delta \rangle_C$ and $\langle \psi \rangle_C$ because they do not contribute at second order in δ . At this order, \mathbf{A} and \mathbf{D} are independent of the value of the smoothed density Δ_0 used for the constraint. However, the constraint enters in third order.

Carrying out the derivation to one higher order in δ requires using the following result for Gaussian random variables: $\text{Cov}(AB, C) = \langle A \rangle \text{Cov}(B, C) + \langle B \rangle \text{Cov}(A, C)$. Applying this with $B = (1 + \delta)^{-1} = 1 - \delta + O(\delta^2)$, after some algebra we find

$$\begin{aligned} \mathbf{A}(\mathbf{r}, t) & = \text{Cov}\left(\delta, \frac{\mathbf{g}_T}{1 + \delta} \middle| \mathbf{v}\right) + O(\delta^4), \\ \mathbf{D}(\mathbf{r}, t) & = \text{Cov}\left(\frac{\mathbf{g}_T}{1 + \delta}, \mathbf{v}\right) + O(\delta^4). \end{aligned} \quad (27)$$

Although these results are correct for Gaussian fluctuations to third order in the density perturbations, nonlinear evolution will generate nonvanishing irreducible three-point correlations (i.e., non-Gaussianity) that have been neglected here. For example, nonlinear evolution generates corrections to the Zel'dovich approximation so that velocity and gravity are no longer proportional to each other as we have assumed. For this reason, equation (27) should not be regarded as being more accurate than equations (25) and (26).

Despite its limitations, equation (27) is useful in showing the qualitative effect of applying an initial constraint. Choosing the protohalo to have $\Delta_0 > 0$ means that typically $\delta > 0$ in the protohalo, which decreases \mathbf{A} and \mathbf{D} compared with equations (25) and (26). Conversely, the diffusion coefficients are enhanced in void regions. Note also that the constraint value appears explicitly in the initial conditions for f in the linear regime through the term $\langle \delta \rangle_C$ in equation (A21).

Equations (25) and (26) imply that, in the quasi-linear regime, the drift and diffusivity are independent of the velocity within the halo. (Although \mathbf{A} is equivalent to a covariance at fixed \mathbf{v} , it is independent of the particular value of \mathbf{v} .) This follows because f_{2c} depends on $\psi_1 = (\mathbf{v}_1 - H\mathbf{x}_1)/(Hb)$ only through $p(\psi_1)$ and $\partial p / \partial \psi_1$ in equation (A22). As noted above, at second order in perturbation theory the drift and diffusivity are also independent of the initial overdensity of the peak Δ_0 (but they depend on the smoothing scale used to find extrema). Thus, both the Fokker-Planck equation and the diffusion coefficients appear to be remarkably robust quantities.

To make further use of equations (25) and (26), we need to write out the coefficients explicitly. Using equations

(A14a)–(A14f), (25), and (26), we first factor out the dependence on cosmology by writing

$$\mathbf{A}(r, t) = 4\pi G\bar{\rho}aA_r(r, t)\hat{r}, \quad (28a)$$

$$\mathbf{D}(r, t) = 4\pi G\bar{\rho}aHb[D_r(r, t)\hat{r} \otimes \hat{r} + D_t(r, t)(\mathbf{I} - \hat{r} \otimes \hat{r})], \quad (28b)$$

where D_r and D_t are scaled radial and tangential diffusion coefficients. The scaled variables are

$$A_r(r, t) = -r\eta(r) - \frac{\bar{\eta}(0)}{\sigma_1^2} \frac{d\bar{\xi}(r)}{dr} + \frac{1}{c_r(r)} \frac{d\gamma(r)}{dr} \left\{ \frac{r\bar{\xi}(r)\bar{\eta}(r)}{\sigma_0^2} + \frac{1}{\sigma_1^2} \frac{d\bar{\xi}(r)}{dr} [\bar{\xi}(r) - 2\bar{\eta}(r)] \right\}, \quad (29a)$$

$$D_r(r, t) = c_r(r) - \frac{d\gamma(r)}{dr} = \sigma_\psi^2 - \frac{d\gamma(r)}{dr} - \left[\frac{r\bar{\eta}(r)}{\sigma_0} \right]^2 - \left[\frac{\bar{\xi}(r) - 2\bar{\eta}(r)}{\sigma_1} \right]^2, \quad (29b)$$

$$D_t(r, t) = c_t(r) - \frac{\gamma(r)}{r} = \sigma_\psi^2 - \frac{\gamma(r)}{r} - \left[\frac{\bar{\eta}(r)}{\sigma_1} \right]^2. \quad (29c)$$

Here r is the comoving distance from the center of the halo. The time dependence arises only because the functions appearing in equations (29a)–(29c) depend on the time-dependent power spectrum of density fluctuations. These functions are summarized here for convenience:

$$\eta(r) = \int \frac{d^3k}{(2\pi)^3} P(k) \frac{j_1(kr)}{kr},$$

$$\gamma(r) = \int \frac{d^3k}{(2\pi)^3} P(k) \frac{j_1(kr)}{k^3}, \quad (30a)$$

$$\bar{\xi}(r) = \int \frac{d^3k}{(2\pi)^3} P(k) W_R(k) j_0(kr),$$

$$\bar{\eta}(r) = \int \frac{d^3k}{(2\pi)^3} P(k) W_R(k) \frac{j_1(kr)}{kr}, \quad (30b)$$

$$\sigma_0^2 = \int \frac{d^3k}{(2\pi)^3} P(k) W_R^2(k),$$

$$\sigma_1^2 = \frac{1}{3} \int \frac{d^3k}{(2\pi)^3} k^2 P(k) W_R^2(k),$$

$$\sigma_\psi^2 = \frac{1}{3} \int \frac{d^3k}{(2\pi)^3} k^{-2} P(k), \quad (30c)$$

where we recall that σ_0^2 and σ_1^2 are the covariances of the constraints Δ_0 and $\mathbf{\Delta}_0$ defined in equations (A8a)–(A8c) and $W_R(k)$ is a smoothing window defined in equations (A3) and (A4). We use the Gaussian window $W_R(k) = \exp(-k^2 R^2/2)$.

To gain some insight into the significance of these results, let us consider a protohalo at high redshift. If we suppose that the gravity field is dominated by its small-scale components, we may replace \mathbf{g}_T by \mathbf{g} . If we further ignore the effects of initial constraints, in linear theory $D_r \sim D_t \sim (3/2)H\sigma_v^2$, where σ_v is the one-dimensional velocity dispersion. In one realization, the forces are deterministic, but taken across the ensemble they are random and fluctuating. A diffusivity $D \sim H\sigma_v^2$ implies that these fluctuations cause the typical particle to change its velocity by order itself in about a Hubble time. There is also a net drift force \mathbf{A} with a dissipative timescale of about δ^{-1} dynamical times, where δ is a typical magnitude of

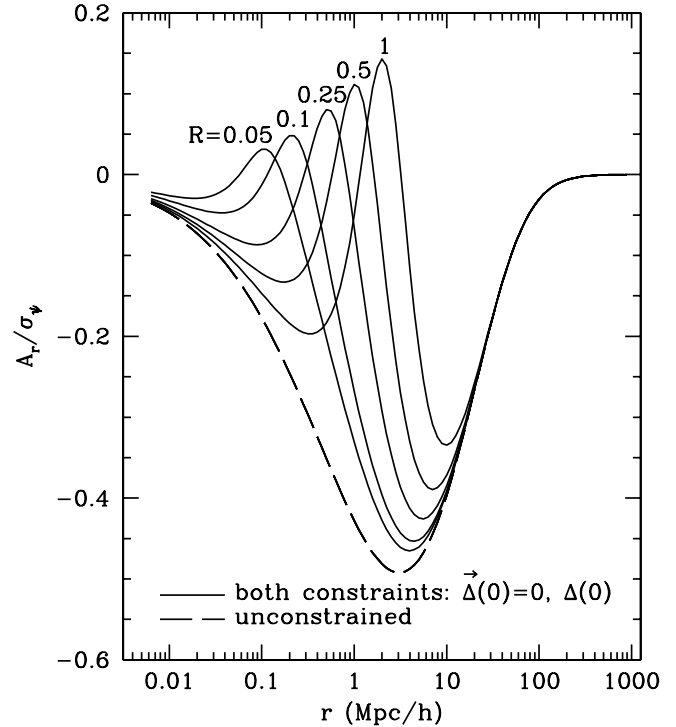


FIG. 1.—Radial drift coefficient A_r plotted as a function of halo radius r for Gaussian smoothing lengths $R = 1, 0.5, 0.25, 0.1,$ and $0.05 h^{-1}$ Mpc. It is computed from eq. (29a) for the Λ CDM model with $(\Omega_m, \Omega_\Lambda, \Omega_b, h) = (0.3, 0.7, 0.05, 0.7)$. For comparison, the dashed curve shows A_r for an unconstrained field. The vertical axis shows the dimensionless A_r/σ_ψ , i.e., the drift normalized by the rms of the peculiar gravitational acceleration.

the density fluctuations. As a result, friction and diffusion will significantly alter halos within a Hubble time of their collapse at high redshift. We conclude that it is not valid to model halos using the spherically symmetric Vlasov equation. This conclusion has also been reached previously by Antonuccio-Delogu & Colafrancesco (1994).

To summarize, equations (24)–(26), as well as their more detailed version, equations (28a)–(30c), are the main results of this subsection. Equation (24) has the standard form of a Fokker-Planck equation. The Fokker-Planck form arises naturally and equations (25)–(26) are exact for Gaussian fluctuations up to second order in perturbation theory. The diffusion coefficients \mathbf{A} and \mathbf{D} are due to fluctuations in the gravitational field over the ensemble of halos. The drift vector \mathbf{A} represents a source of gravitational acceleration in addition to the acceleration \mathbf{g}_T produced by the mean matter distribution. The diffusion term $-\mathbf{D} \cdot \partial f / \partial \mathbf{v}$ causes the temperature (i.e., velocity dispersion) to increase (for a positive-definite \mathbf{D}).

4.2. Results for Λ CDM Model

Having derived analytic expressions for the diffusion coefficients in the previous subsection, we now evaluate them for the currently favored cosmological model at high redshift. Figures 1 and 2 show A_r , D_r , and D_t as a function of radius r (measured from the halo center) for a range of Gaussian smoothing lengths R in $W_R(k)$. We obtain these results by numerically integrating equations (29a)–(30c) with the linear power spectrum $P(k)$ for a flat Λ CDM cosmological model with $(\Omega_m, \Omega_\Lambda, \Omega_b, h) = (0.3, 0.7, 0.05, 0.7)$.

The figures show a number of interesting features. First, the diffusion coefficients are negative over some regions of radius for some parameters. As we show below, \mathbf{A} is the

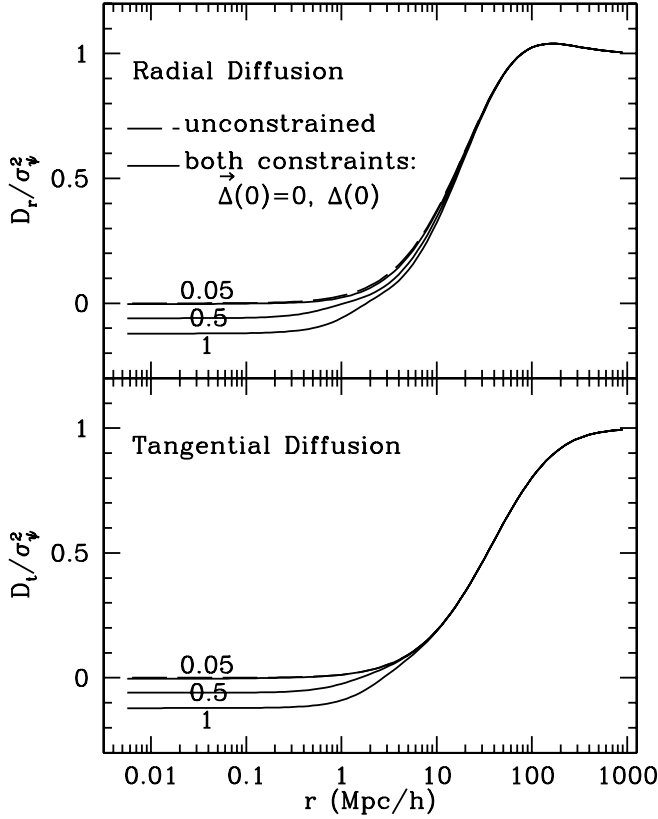


FIG. 2.—Diffusion coefficient D in the radial (r) and tangential (t) directions plotted as a function of halo radius r for Gaussian smoothing lengths $R = 1, 0.5,$ and $0.05 h^{-1}$ Mpc. It is computed from eqs. (29b) and (29c) for the same cosmological model as Fig. 1. The dashed curve in each panel is for an unconstrained field. The vertical axis shows the dimensionless D_{ij}/σ_{ψ}^2 , i.e., the diffusivity normalized by the mean squared peculiar gravitational acceleration.

gravitational acceleration on particles in the halo caused by drift. If the friction were produced by the Chandrasekhar mechanism of tidal wakes, one would expect $A \parallel -v$ (see § 5.4). However, here A is independent of v and the particle mass. The source of friction is not the tidal wake of a flow past a moving massive particle. Instead, it is the fluctuations of the gravitational field of a Gaussian random process, and these fluctuations can lead to a radially outward or inward force. We discuss this further in § 5.

More surprising at first is the fact that the diffusivities D_r and D_t can be negative. Negative diffusivity causes the velocity dispersion (or temperature) to decrease, indicating an instability. Such behavior is possible for gravitational systems. In fact, this result is to be expected in the quasi-linear regime of gravitational instability because second-order perturbations enhance gravitational collapse of dark matter halos (Peebles 1980; Jain & Bertschinger 1994). In the strongly nonlinear regime, after halo virialization, we expect the diffusivities to become positive.

To elucidate the origin of the negative diffusivity, we show in Figures 3 and 4 that once the extremum constraint $\Delta_0 \equiv \nabla\Delta(0) = 0$ is removed (by setting $\sigma_1^{-2} = 0$), we obtain $A_r \leq 0$ and $D_r, D_t \geq 0$ for all radii. The extremum constraint is therefore crucial to generating positive A_r and negative diffusivity seen in Figures 1 and 2. This effect is easy to understand: if one stands at a random place in a Gaussian field, nonlinear effects can speed up or slow down gravitational collapse. Gravitational instability, however, speeds up the evolution in the vicinity of a density extremum.

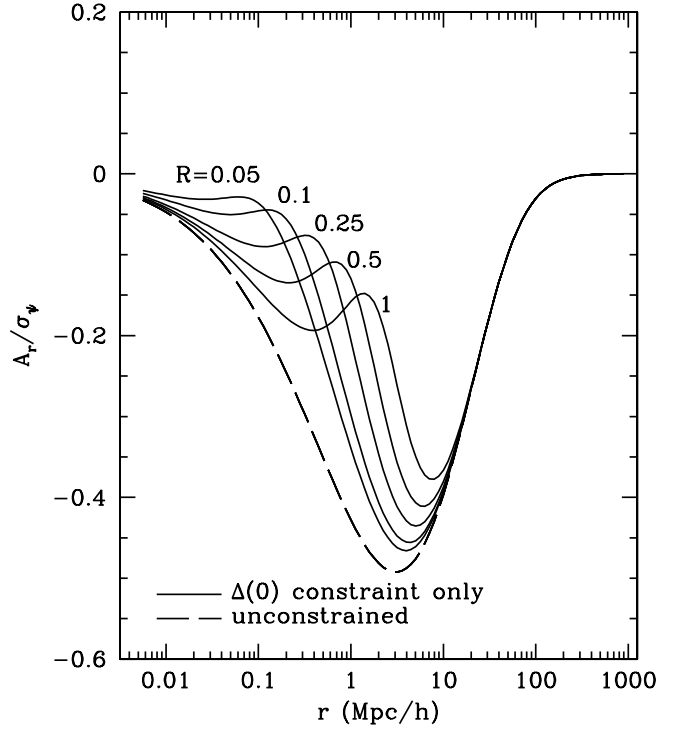


FIG. 3.—Same as Fig. 1, except that no extremum constraint [i.e., $\Delta_0 \equiv \nabla\Delta(x=0) = 0$] is applied.

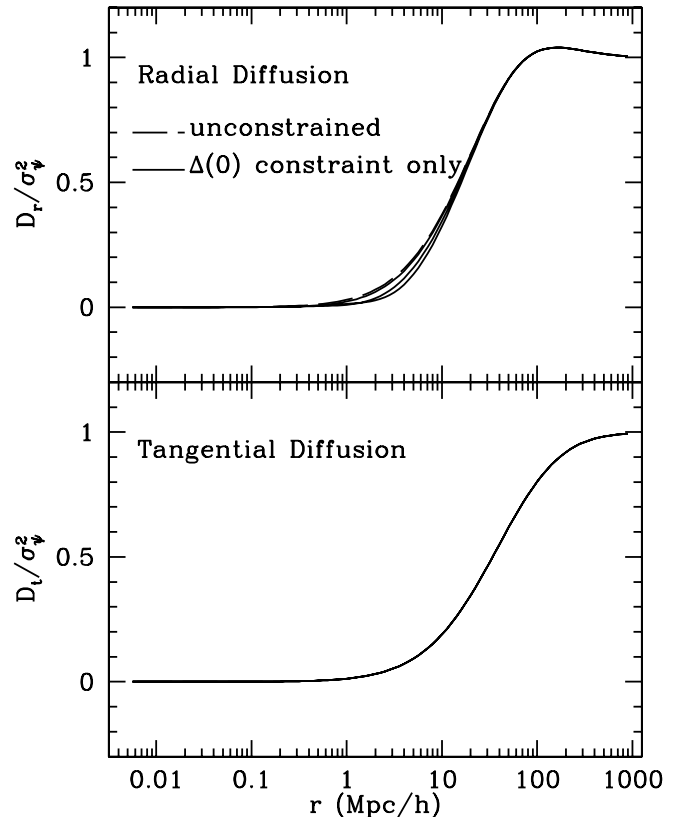


FIG. 4.—Same as Fig. 2, except that no extremum constraint is applied.

Another obvious systematic effect with the diffusion coefficients in Figures 1 and 2 is their dependence on the smoothing radius R used to set the initial constraints. For the power spectrum assumed here, the maximum of A_r occurs at about $2R$. In the limit $R \rightarrow 0$, the variances σ_0^2 and σ_1^2 diverge, causing the constraint terms in the diffusion coefficients (i.e., those terms proportional to σ_0^{-2} and σ_1^{-2}) to vanish.

Figure 2 shows that the diffusivity is isotropic at the halo center but shows radial anisotropy outside the center. At small radii the tangential diffusivity is larger in absolute value than the radial diffusivity. Tangential diffusion gives the infalling matter angular momentum. Our treatment therefore provides a new statistical description of the growth of angular momentum by tidal torques (White 1984) and may enable, in principle, an explanation for the universal angular momentum profile found recently by Bullock et al. (2001).

4.3. Asymptotic Behavior

It is instructive to work out analytically the asymptotic behavior of the functions in equations (29a)–(29c) for small and large r . For small r we find

$$A_r(r, t) = r \left\{ -\frac{1}{3}\sigma_\delta^2 + \alpha\bar{\eta}(0) + \frac{1}{1 - \bar{\eta}^2(0)/\sigma_1^2\sigma_\psi^2} \left[\frac{3\bar{\eta}^2(0)}{\sigma_0^2} - \alpha\bar{\eta}(0) \right] \right\} + O(r^3), \quad (31a)$$

$$D_r(r, t) = -\left[\frac{\bar{\eta}(0)}{\sigma_1} \right]^2 + r^2 \left\{ \frac{1}{10}\sigma_\delta^2 + \frac{3}{5}\alpha\bar{\eta}(0) - \left[\frac{\bar{\eta}(0)}{\sigma_0} \right]^2 \right\} + O(r^4), \quad (31b)$$

$$D_t(r, t) = -\left[\frac{\bar{\eta}(0)}{\sigma_1} \right]^2 + r^2 \left[\frac{1}{30}\sigma_\delta^2 + \frac{1}{5}\alpha\bar{\eta}(0) \right] + O(r^4), \quad (31c)$$

where we have defined

$$\sigma_\delta^2 \equiv \int \frac{d^3k}{(2\pi)^3} P(k), \quad (32a)$$

$$\bar{\xi}(0) = 3\bar{\eta}(0) \equiv \int \frac{d^3k}{(2\pi)^3} P(k) W_R(k), \quad (32b)$$

$$\alpha \equiv \frac{\int d^3k k^2 P(k) W_R(k)}{\int d^3k k^2 P(k) W_R^2(k)} = \frac{1}{3\sigma_1^2} \int \frac{d^3k}{(2\pi)^3} k^2 P(k) W_R(k). \quad (32c)$$

Equations (31a)–(31c) show that if σ_δ^2 is finite, then as $r \rightarrow 0$, A_r drops to zero while D_r and D_t approach a negative constant as discussed above, $-\bar{\eta}^2(0)/\sigma_1^2$. This central value is decreased as the smoothing scale is decreased (i.e., higher σ_0 and σ_1) as shown in Figure 2. As $R \rightarrow 0$, we have $\sigma_0 \rightarrow \sigma_\delta$ and $\sigma_1 \rightarrow \infty$ as long as $d \ln P / d \ln k > -5$ as $k \rightarrow \infty$. In the limit $\sigma_1 \rightarrow \infty$ (no extremum constraint), D_r and D_t are non-negative everywhere (cf. Figs. 2 and 4). Thus, negative diffusivity (instability) is produced by constraining the slope of the density field. Constraints have the opposite effect on A_r : as the comparison of Figures 1 and 3 confirms, the contributions to A_r from constraint terms are positive.

For large r , both $c_r(r)$ and $c_t(r)$ approach σ_ψ^2 , and $\gamma(r) \propto 1/r$ (assuming $P \propto k$ as $k \rightarrow 0$), so

$$D_r(r, t) \rightarrow \sigma_\psi^2, \quad D_t(r, t) \rightarrow \sigma_\psi^2. \quad (33)$$

Figure 2 shows that D_r rises slightly above σ_ψ^2 at a few hundred megaparsecs (for the standard model power spectrum) before reaching σ_ψ^2 . This is because $D_r = c_r - d\gamma/dr$ and $d\gamma/dr$ is negative for this range of r . The drift $A_r(r, t)$ approaches zero at large r . In practice, the large- r limit is unimportant because halo infall occurs only from within about 10 Mpc.

4.4. Power-Law Models

Another way to gain insight into the diffusion coefficients is to examine their behavior for scale-free spectra $P(k) \propto k^n$. Many of the integrals are analytic in this case.

We define a set of functions

$$G_l(a, b) \equiv \int_0^\infty x^{a-1} j_l(bx) \exp(-\frac{1}{2}x^2) dx. \quad (34)$$

The integral converges for $a + l > 0$. Using properties of the spherical Bessel functions $j_l(x)$, we obtain the following useful relations:

$$bG_{l+1}(a, b) = (a + l - 1)G_l(a - 1, b) - G_l(a + 1, b), \quad (35a)$$

$$bG_{l-1}(a, b) = (2 + l - a)G_l(a - 1, b) + G_l(a + 1, b), \quad (35b)$$

$$G_0(a, 0) = 2^{(a-2)/2} \Gamma\left(\frac{a}{2}\right). \quad (35c)$$

We also use the following relation, valid for $a + l > 0$:

$$g_l(a) \equiv \lim_{\sigma \rightarrow \infty} \int_0^\infty x^{a-1} j_l(x) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = 2^{a-2} \sqrt{\pi} \Gamma\left(\frac{a+l}{2}\right) \left[\Gamma\left(\frac{3+l-a}{2}\right) \right]^{-1}. \quad (36)$$

For Gaussian-tapered power-law spectrum $P(k) = 2\pi^2 B k^n \exp(-k^2 R_0^2)$ and Gaussian smoothing $W_R(k) = \exp(-k^2 R^2/2)$, equations (30a)–(30c) give, for $R_0 \rightarrow 0$ and $n > -3$,

$$r\eta(r) = Bg_1(n+2)r^{-(n+2)}, \quad (37a)$$

$$\gamma(r) = Bg_1(n)r^{-n} \quad \text{for } n > -1, \quad (37b)$$

$$\sigma_\psi^2 - \frac{d\gamma}{dr} = B \begin{cases} ng_1(n)r^{-(n+1)} + \frac{1}{6}\Gamma\left(\frac{n+1}{2}\right)R_0^{-(n+1)}, & n \neq -1, \\ -\frac{1}{9} + \frac{1}{6}C + \frac{1}{3}\log\left(\frac{r}{R_0}\right), & n = -1, \end{cases} \quad (37c)$$

$$\sigma_\psi^2 - \frac{\gamma}{r} = B \begin{cases} -g_1(n)r^{-(n+1)} + \frac{1}{6}\Gamma\left(\frac{n+1}{2}\right)R_0^{-(n+1)}, & n \neq -1, \\ -\frac{4}{9} + \frac{1}{6}C + \frac{1}{3}\log\left(\frac{r}{R_0}\right), & n = -1, \end{cases} \quad (37d)$$

$$\bar{\xi}(r) = BG_0(n+3, r/R)R^{-(n+3)}, \quad (37e)$$

$$r\bar{\eta}(r) = BG_1(n+2, r/R)R^{-(n+2)}, \quad (37e)$$

$$\sigma_0^2 = \frac{1}{2}B\Gamma\left(\frac{n+3}{2}\right)R^{-(n+3)},$$

$$\sigma_1^2 = \frac{1}{6}B\Gamma\left(\frac{n+5}{2}\right)R^{-(n+5)}, \quad (37f)$$

$$\sigma_\psi^2 = \frac{1}{6}B\Gamma\left(\frac{n+1}{2}\right)R_0^{-(n+1)} \quad \text{for } n > -1. \quad (37g)$$

Here $C = 0.5772156649 \dots$ is Euler's constant. For a pure power-law spectrum with $n \leq -1$, $\gamma(r)$ and σ_ψ^2 diverge at long

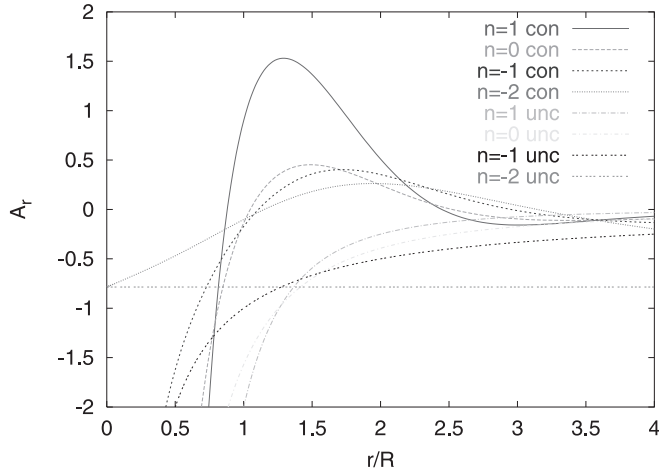


FIG. 5.—Radial drift coefficient $A_r(r)$ for power-law spectra, $P(k) \propto k^n$, plotted as a function of halo radius r scaled to the Gaussian smoothing radius R . Results are shown for four spectral indices, $n = -2, -1, 0$, and 1 , and for constrained and unconstrained fields. The amplitude of A_r corresponds to $P(k) = 2\pi^2 B k^n$ normalized to $B = 1$. [See the electronic edition of the Journal for a color version of this figure.]

wavelength. However, equations (37c) and (37d) are valid for $n > -3$. For $n \leq -1$, $c_r^{-1} d\gamma/dr = 1$ in equation (29a). For $n > -1$, σ_ψ^2 converges at long wavelength but diverges at short wavelength with $\sigma_\psi^2 \propto R_0^{-(n+1)}$ as $R_0 \rightarrow 0$.

The limiting behaviors of $\bar{\xi}(r)$ and $\bar{\eta}(r)$ are given by

$$\bar{\xi}(r) \approx \begin{cases} \frac{1}{2} \Gamma\left(\frac{n+3}{2}\right) \left(\frac{R}{\sqrt{2}}\right)^{-(n+3)} \left[1 - \frac{(n+3)r^2}{6R^2} + O(r^4)\right], & r \ll R, \\ -ng_1(n+2)r^{-(n+3)}, & r \gg R, \end{cases} \quad (38a)$$

$$\bar{\eta}(r) \approx \begin{cases} \frac{1}{6} \Gamma\left(\frac{n+3}{2}\right) \left(\frac{R}{\sqrt{2}}\right)^{-(n+3)} \left[1 - \frac{(n+3)r^2}{10R^2} + O(r^4)\right], & r \ll R, \\ g_1(n+2)r^{-(n+3)}, & r \gg R. \end{cases} \quad (38b)$$

For all r/R , these functions obey the relation $\bar{\xi} + (R^2/r) d\bar{\xi}/dr = -n\bar{\eta}(r)$.

The drift coefficient is plotted in Figure 5 for $n \in \{-2, -1, 0, 1\}$. As expected from the previous section, constraints make A_r more positive. For power-law spectra, however, σ_δ^2 diverges so one cannot use equations (31a)–(31c) to obtain the behavior as $r \rightarrow 0$. The exact results here show that for $-2 < n < 4$, $A_r \rightarrow -\infty$ as $r \rightarrow 0$. For $n < -2$, $A_r \rightarrow 0$ as $r \rightarrow \infty$.

These results show that models with a lot of small-scale power and substructure ($n > -2$ as $k \rightarrow \infty$) have a strong inward drift force, while models that are smoother on small scales ($n < -2$) have negligible drift force as $r \rightarrow 0$. This is easy to understand qualitatively in the context of halo formation. Density fields with a lot of substructure have a non-spherical mass distribution, with mass concentrated into subhalos. For a given halo, the mass between $r/2$ and r , for example, is concentrated into overdense substructures whose gravity field on infalling particles is stronger, on average, than

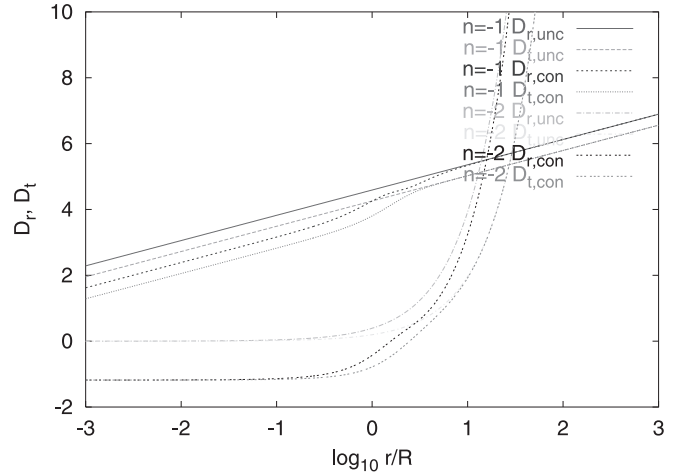


FIG. 6.—Diffusion coefficient D in the radial (r) and tangential (t) directions for power-law spectra, $P(k) \propto k^n$, plotted as a function of halo radius r scaled to the Gaussian smoothing radius R . Results are shown for two spectral indices, $n = -1$ and -2 , and for constrained and unconstrained fields. For $n < -1$, constraints cause the diffusivity to approach a negative constant for $r \ll R$. The amplitude of D corresponds to $P(k) = 2\pi^2 B k^n$ normalized to $B = 1$. [See the electronic edition of the Journal for a color version of this figure.]

if this mass were uniformly distributed over a spherical shell as it is for the average halo. Given a density correlation function that rises with decreasing r , there is a greater concentration of such clumps interior to r than outside of r , leading to a net inward drift force.

Figure 6 shows the radial and tangential diffusivity for power-law spectra with $n \leq -1$. For $n > -1$ the diffusivity diverges because the velocity dispersion σ_ψ^2 diverges as $k \rightarrow \infty$. For $n = -1$ the divergence is logarithmic, resulting in a logarithmic dependence of D_r and D_t on radius. Figure 6 shows that the effect of the constraints is to increase the diffusivity at radii within a few smoothing lengths of the peak. As we noted above, constraints on the density gradient make the diffusivities negative as $r \rightarrow 0$. For $n \leq -1$ the diffusivities continue to rise with increasing r because σ_ψ^2 diverges as $k \rightarrow 0$. For the physical power spectrum used in Figures 1 and 2, σ_ψ^2 converges at both small and large k , so the asymptotic behavior is different.

5. DISCUSSION OF THE FOKKER-PLANCK EQUATION

The use of the Fokker-Planck equation to describe dark matter clustering is novel. In order to gain insight into what this approach may reveal, in this section we examine the Fokker-Planck equation from several different perspectives.

5.1. Velocity Moments

The Fokker-Planck equation describes relaxation processes in a weakly collisional fluid. In this regard it is similar to the Boltzmann equation, in which the right-hand side of equation (9) is replaced by a two-body collision integral that is bilinear in the one-particle distribution function. The present situation is physically very different, however: “collisions” are not due to individual particles but rather to the gravity field of density fluctuations in a Gaussian random field. Nevertheless, we can borrow one of the techniques used to gain insight into the Boltzmann equation, viz., velocity moments, to obtain a clear physical interpretation of the meaning of the Fokker-Planck equation and the diffusion coefficients \mathbf{A} and \mathbf{D} .

Moment equations are obtained by multiplying equation (24) by powers of v_i and integrating over velocity. The three lowest moments give the mass density ρ , fluid velocity \mathbf{u} , and velocity dispersion tensor θ_{ij} :

$$\rho(\mathbf{r}, t) = \int d^3v f(\mathbf{r}, \mathbf{v}, t), \quad (39a)$$

$$u_i(\mathbf{r}, t) = \rho^{-1} \int d^3v f(\mathbf{r}, \mathbf{v}, t) v_i, \quad (39b)$$

$$\theta_{ij}(\mathbf{r}, t) = \rho^{-1} \int d^3v f(\mathbf{r}, \mathbf{v}, t) (v_i - u_i)(v_j - u_j). \quad (39c)$$

Here we are using physical (proper) spatial coordinates rather than comoving coordinates. The density, velocity, and velocity dispersion tensor obey the following equations:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial r_i} (\rho u_i) = 0, \quad (40a)$$

$$\left(\frac{\partial}{\partial t} + u_k \frac{\partial}{\partial r_k} \right) u_i + \frac{1}{\rho} \frac{\partial}{\partial r_j} (\rho \theta_{ij}) = g_{Ti} + A_i, \quad (40b)$$

$$\left(\frac{\partial}{\partial t} + u_k \frac{\partial}{\partial r_k} \right) \theta_{ij} + \frac{\partial u_i}{\partial r_k} \theta_{jk} + \frac{\partial u_j}{\partial r_k} \theta_{ik} + \frac{1}{\rho} \frac{\partial}{\partial r_k} (\rho \theta_{ijk}) = 2D_{ij}. \quad (40c)$$

The first equation is the usual fluid continuity equation. The second is the usual Euler equation, except that the pressure is generalized to a matrix $\rho \theta_{ij}$ because the spatial stress need not be diagonal for a collisionless fluid. The right-hand side has two force terms: the gravitational tidal field \mathbf{g}_T produced by $\rho(\mathbf{r}, t)$ and the drift acceleration $\mathbf{A}(\mathbf{r}, t)$ caused by the correlated fluctuations of the mass density field. The third equation introduces a tensor θ_{ijk} , which is defined exactly like θ_{ij} except the integrand in equation (39c) acquires an additional factor $(v_k - u_k)$. Equation (40c) is a generalized heat equation, which requires a little more discussion.

The velocity dispersion tensor θ_{ij} generalizes the temperature of an isotropic gas to the case of an anisotropic velocity distribution. For an isotropic gas with $\theta_{ij} = T \delta_{ij}$ (here T has units of v^2) and $D_{ij} = D \delta_{ij}$, this equation takes the more familiar form

$$\frac{3}{2} \frac{dT}{dt} + (\nabla \cdot \mathbf{u})T + \frac{1}{\rho} \nabla \cdot \mathbf{q} = 3D, \quad (41)$$

where $q_i = \frac{1}{2} \rho \sum_{j=1}^3 \theta_{ij}$ is the heat flux. The velocity gradient terms in equation (40c) generalize the $p dV$ term (i.e., the $T \nabla \cdot \mathbf{u}$ term) of equation (41) to an anisotropic velocity distribution. For a perfect gas, the heat flux term is absent because the Maxwellian velocity distribution has vanishing skewness. For an imperfect or collisionless gas, however, the third moment of the velocity distribution generates a generalized heat flux $\rho \theta_{ijk}$. Notice that the heating rate is given not by the heat flux but instead by its divergence. For a weakly imperfect gas, the heat flux can be approximated (e.g., in the Chapman-Enskog expansion approach) by a conductive flux $\mathbf{q} = -\kappa \nabla T$, where κ is the conductivity, but for a collisionless gas, θ_{ijk} cannot be obtained simply from ρ , \mathbf{u} , and θ_{ij} . Thus, equations (40a)–(40c) do not form a closed system. This is precisely why we must use a phase-space description for a

collisionless gas instead of trying to modify the fluid equations. Equations (40a)–(40c) and (41), however, are pedagogically useful in showing us how to interpret \mathbf{A} and \mathbf{D} .

The velocity diffusivity tensor D_{ij} appears as a new term in the anisotropic heat equation, with no familiar counterpart in the dynamics of a collisional gas. Although D_{ij} comes from a *diffusion* (i.e., gradient) term in *phase space*, it appears as a direct *local heating* term in *real space* (i.e., coordinate space \mathbf{r}). Diffusion in velocity space is equivalent to local heating: the temperature increases when particles spread out in velocity at a given point in space. The diffusion occurs in velocity space, not in real space. The presence of this term begs the question, where does the energy come from? From the exact kinetic equation (9), it follows that the heating term is given by the symmetrized first velocity moment of the correlated force density \mathbf{F} produced by fluctuating gravitational fields. From this we see that the heating comes from gravitational energy. For this reason, it is possible for D to be positive or negative: gravitational fluctuations can either locally heat or cool a gas. One should not try to draw conclusions about energy conservation from these results because our calculations are performed in the accelerating frame centered on a halo. However, the equations of motion used in deriving our kinetic theory imply local energy conservation.

5.2. Langevin Equation

The Fokker-Planck equation describes the evolution of the probability distribution for particles undergoing random walks. Instead of describing the system by a distribution function, we can specify equations of motion for individual particle trajectories using the Langevin equation (Langevin 1908). Although this is reversing the logic we used in deriving the Fokker-Planck equation, it is useful for providing both an understanding and a Monte Carlo method for numerically solving the Fokker-Planck equation.

The Langevin equation is a stochastic differential equation for the individual particle trajectories. The Langevin equation corresponding to equation (24) is given by the pair of equations

$$\frac{d\mathbf{r}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{v}}{dt} = \mathbf{g}_T(\mathbf{r}, t) + \mathbf{A}(\mathbf{r}, t) + \mathbf{\Gamma}(\mathbf{r}, t), \quad (42)$$

where $\mathbf{\Gamma}$ is a force due to a zero-mean Gaussian random process with covariance

$$\langle \Gamma_i(\mathbf{r}, t) \Gamma_j(\mathbf{r}, t') \rangle = 2D_{ij}(\mathbf{r}, t) \delta_D(t - t'). \quad (43)$$

The velocity of a particle at time $t + dt$, $\mathbf{v}(t + dt)$, is therefore given by a deterministic term $\mathbf{v}(t) + [\mathbf{g}_T(\mathbf{r}, t) + \mathbf{A}(\mathbf{r}, t)] dt$, plus a random term drawn from a Gaussian distribution with variance $\propto D dt$. Subtle differences exist between the cases in which the variance D is evaluated at $[\mathbf{r}(t), t]$ (“Ito calculus”) versus at $[\mathbf{r}(t + dt/2), t + dt]$ (“Stratonovich calculus”). Details can be found in Risken (1989). Note that the Langevin equation requires $D > 0$.

The interpretation of the Langevin equation is given by the following result (Risken 1989; Gardiner 2002): if a sufficiently large sample of particles is drawn at random at time t_0 from the phase-space distribution $f(\mathbf{r}, \mathbf{v}, t_0)$ and their orbits are integrated using equation (42), then at any later time t their positions and velocities are a sample from the phase-space density $f(\mathbf{r}, \mathbf{v}, t)$ obtained by integrating equation (24).

Equation (42) is similar to the exact equations of motion for individual particles given by equation (4), but there are several important differences. The gravity field \mathbf{g}_K of one realization is replaced by the spherically symmetric gravity field of the ensemble-average density profile. The effects of spatially varying forces in \mathbf{g}_K are represented by the drift term \mathbf{A} and the stochastic acceleration $\mathbf{\Gamma}$. These additional terms have the effect that, at least in the quasi-linear regime (which we assumed in our derivation of the Fokker-Planck equation), the phase-space evolution implied by the Langevin dynamics is equivalent to a Monte Carlo sample from the phase-space density.

This does not mean that we have come full circle. For individual realizations of the ensemble, the density field is clustered and not spherically symmetric and the phase space is six-dimensional. The Langevin equations give orbits in a smooth, spherically symmetric matter distribution. The latter is far easier to model numerically, with much higher resolution possible, than the original dynamics. By taking advantage of the symmetries of the average halo, we have reduced the phase space from six dimensions to three.

5.3. Comparison with Classical Brownian Motion

The Fokker-Planck equation was first written almost a century ago to describe Brownian motion (Fokker 1914; Planck 1917). In this case, a small macroscopic particle (a Brownian particle) undergoes a random walk as a result of its collisions with individual molecules in a liquid. Instead of the phase-space density of molecules, the relevant quantity is the probability density $W(\mathbf{v}, t)$ for the Brownian particle to have velocity \mathbf{v} at time t . Assuming a spatially homogeneous and isotropic medium, this probability distribution obeys the Fokker-Planck equation

$$\frac{\partial W}{\partial t} = \frac{\partial}{\partial \mathbf{v}} \cdot \left(\gamma \mathbf{v} W + \gamma \frac{k_B T}{m} \frac{\partial W}{\partial \mathbf{v}} \right), \quad (44)$$

where γ is a constant (given, for a small spherical body of mass m and radius a immersed in a fluid of viscosity η , by the Stokes formula $\gamma = 6\pi\eta a/m$), T is the temperature, and k_B is Boltzmann's constant. Equation (44) has the same form as equation (24) if we identify $\mathbf{A} = -\gamma\mathbf{v}$ for the drag term and $D_{ij} = \gamma(k_B T/m)\delta_{ij}$ for the diffusion term.

We note that the Brownian motion can be equally described by the Langevin equations in equations (42) and (43) with $\mathbf{g}_T = 0$, $\mathbf{A} = -\gamma\mathbf{v}$, and $D_{ij} = \gamma(k_B T/m)\delta_{ij}$.

Brownian motion describes a very different physical situation than dark matter halo formation. Why are both described by a Fokker-Planck equation? The reason is that both problems involve random walks. Let us examine this more closely. In the case of Brownian motion, the random walks arise from impulsive collisions that cause the Brownian particle's velocity to change significantly on a timescale γ^{-1} . During this time interval, the particle moves a distance $\sim v/\gamma$. After an elapsed time t , the particle undergoes γt steps and therefore moves a distance $\sim (v^2 t/\gamma)^{1/2}$. In thermal equilibrium, $v^2 \sim k_B T/m$ so the particle moves a typical distance $\gamma^{-1}(Dt)^{1/2}$, where $D = \gamma k_B T/m$ is the velocity diffusivity. A random walk in space leads naturally to a diffusion equation for the probability distribution (Risken 1989; Gardiner 2002).

The dark matter case at first appears to be very different. Instead of a large particle being buffeted by small ones, we are studying small particles being buffeted by large-mass fluctuations (i.e., large in mass compared with the dark matter particles themselves,

whose mass never enters our discussion). Moreover, our derivation of the Fokker-Planck equation has assumed that the motion is in the quasi-linear regime described by equation (16) with a universal time dependence for $\psi(\mathbf{x}, t)$. Therefore, each particle has moved in a straight line with (in appropriate units) a constant velocity, which is not a random walk in space. However, *the velocity* of a given particle is a random walk since ψ is obtained by adding up the gravitational accelerations produced by all the mass fluctuations in a Gaussian random field. The net force on a particle is a random walk at fixed time.

Although cosmic Gaussian random fields lead to a Fokker-Planck equation, the diffusion coefficients are very different from Brownian motion. First, the drift acceleration \mathbf{A} is independent of velocity, in contrast with the viscous drag acceleration $-\gamma\mathbf{v}$ for Brownian motion. In the quasi-linear regime of gravitational clustering, the drift is not a friction at all. Instead, it is radially directed (because of spherical symmetry for the average halo), either inward or outward depending on the density correlations of fluctuating substructure. Second, the velocity diffusivity tensor D_{ij} is not proportional to the temperature; the dark matter is treated as being completely cold with vanishing temperature. In both cases the diffusivity is proportional to the mean squared displacement. In the cosmological case this depends simply on the power spectrum of density fluctuations and not on the thermal velocity of the particles.

In the Brownian case \mathbf{A} and \mathbf{D} are both proportional to the Stokes drag coefficient γ . The proportionality between \mathbf{A} and \mathbf{D} is no accident but is a consequence of the fluctuation-dissipation theorem, which states that these coefficients have a proportionality given by the condition of thermal equilibrium. It is instructive to see this in a generalized case of equation (44). Suppose that the Brownian particle is moving in a time-independent potential $\Phi(\mathbf{r})$. The Fokker-Planck equation then becomes

$$\frac{\partial W}{\partial t} + \mathbf{v} \cdot \frac{\partial W}{\partial \mathbf{r}} - \frac{\partial \Phi}{\partial \mathbf{r}} \cdot \frac{\partial W}{\partial \mathbf{v}} = \frac{\partial}{\partial \mathbf{v}} \cdot \left(\gamma \mathbf{v} W + \gamma \frac{k_B T}{m} \frac{\partial W}{\partial \mathbf{v}} \right). \quad (45)$$

The stationary solution (for suitable boundary conditions) is given by the Boltzmann distribution

$$W = N \exp\left(-\frac{E}{k_B T}\right), \quad E = \frac{1}{2} m v^2 + m \Phi(\mathbf{r}), \quad (46)$$

where N is a normalization constant. With this solution not only does the left-hand side of equation (45) vanish, but so does the velocity flux on the right-hand side. Drag and diffusion balance each other in thermal equilibrium. Note that the mean squared velocity at fixed \mathbf{r} equals $3k_B T/m$. For a distribution of masses of Brownian particles, thermal equilibrium results in an equipartition of energy with the heaviest particles moving most slowly. This is a natural consequence of having $(D/\gamma)^{1/2}$ equal the thermal speed.

In thermodynamic systems, the Fokker-Planck equation governs the relaxation to thermal equilibrium. In the cosmological case of quasi-linear clustering considered here, it is not thermal equilibrium but rather the statistics of Gaussian random fields that relate \mathbf{A} and \mathbf{D} . In Appendix B we investigate our cosmological Fokker-Planck equation for the relaxation to stationary solutions.

5.4. Comparison with Globular Cluster Dynamics

Globular cluster evolution is the most studied application of the Fokker-Planck equation in astrophysics. Unlike the

Brownian case, gravity plays the dominant role here. A star in a globular cluster can be approximated as experiencing two types of gravitational forces: a smoothly varying potential $\Phi(\mathbf{r}, t)$ due to the smoothed matter distribution in the system, and a fluctuating force due to many two-body interactions with other stars. The phase-space density of the stars obeys the Fokker-Planck equation (Spitzer 1987; Binney & Tremaine 1988)

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} - \frac{\partial \Phi}{\partial \mathbf{r}} \cdot \frac{\partial f}{\partial \mathbf{v}} = - \frac{\partial}{\partial \mathbf{v}} \cdot [\mathbf{A}(\mathbf{v})f] + \frac{1}{2} \frac{\partial^2}{\partial v_i \partial v_j} [D_{ij}(\mathbf{v})f]. \quad (47)$$

This looks very similar to the cosmological kinetic equation (24) in the quasi-linear regime, aside from a factor of 2 in the diffusivity that is purely a matter of differing conventions, as well as the velocity dependence of the diffusion coefficients. However, the physics of \mathbf{A} and D_{ij} is very different here. The relaxation process in globular cluster dynamics is *two-body relaxation* (Chandrasekhar 1943), i.e., the dissipation arising from the fluctuating forces of discrete Newtonian gravitating point masses. Examining this case will shed light on the relaxation processes that can affect galaxy halo evolution.

The calculation of diffusion coefficients for two-body relaxation is described in § 8.3 of Binney & Tremaine (1988). To understand the results, it is helpful to consider the process of Chandrasekhar dynamical friction acting on a test particle with mass m_t and velocity \mathbf{v}_t moving through a spatially homogeneous sea of background particles of mass m_b and velocity distribution $f(\mathbf{v}_b)$ with mass density $\rho_b \equiv m_b \int f(\mathbf{v}_b) d^3 v_b$. For an isotropic Maxwellian distribution $f(\mathbf{v}_b) \propto \exp(-v_b^2/2\sigma_b^2)$, the test body feels an acceleration

$$\frac{d\mathbf{v}_t}{dt} = - \frac{4\pi G^2 \rho_b (m_t + m_b) \ln \Lambda}{v_t^3} \left[\operatorname{erf}(X) - \frac{2X}{\sqrt{\pi}} e^{-X^2} \right] \mathbf{v}_t, \quad (48)$$

where $X \equiv v_t/(\sqrt{2}\sigma_b)$.

Now treat each star within the globular cluster one at a time as a test particle, with all the other stars serving as background particles. If the velocity distribution is Maxwellian and spatial gradients are ignored, dynamical friction will contribute $\mathbf{A} = -d\mathbf{v}/dt$ given by equation (48) with \mathbf{v}_t replaced by \mathbf{v} . More generally, for a spatially homogeneous distribution on scales larger than the interparticle spacing, two-body relaxation gives (Binney & Tremaine 1988)

$$\mathbf{A} = 4\pi G^2 m_b (m_t + m_b) \ln \Lambda \frac{\partial}{\partial \mathbf{v}} h(\mathbf{v}), \quad (49a)$$

$$D_{ij} = 4\pi G^2 m_b^2 \ln \Lambda \frac{\partial^2}{\partial v_i \partial v_j} g(\mathbf{v}), \quad (49b)$$

where $\ln \Lambda$ is the Coulomb logarithm, and the Rosenbluth potentials h and g are given by (Rosenbluth et al. 1957)

$$h(\mathbf{v}) \equiv \int \frac{f(\mathbf{v}_b) d^3 v_b}{|\mathbf{v} - \mathbf{v}_b|}, \quad (50a)$$

$$g(\mathbf{v}) \equiv \int f(\mathbf{v}_b) |\mathbf{v} - \mathbf{v}_b| d^3 v_b. \quad (50b)$$

For the isotropic Maxwellian distribution $f(\mathbf{v}_b) \propto \exp(-v_b^2/2\sigma_b^2)$, $\mathbf{A} = -\gamma(\mathbf{v})\mathbf{v}$ is a drag force and the needed coefficients are

$$\gamma = \frac{4\pi G^2 \rho_b (m_t + m_b) \ln \Lambda}{v^3} \left[\operatorname{erf}(X) - \frac{2X}{\sqrt{\pi}} e^{-X^2} \right], \quad (51a)$$

$$D_{\parallel} = 8\pi G^2 \rho_b m_b \frac{\sigma_b^2}{v^3} \ln \Lambda \left[\operatorname{erf}(X) - \frac{2X}{\sqrt{\pi}} e^{-X^2} \right], \quad (51b)$$

$$D_{\perp} = 8\pi G^2 \rho_b m_b \frac{1}{v} \ln \Lambda \left[\left(1 - \frac{1}{2X^2} \right) \operatorname{erf}(X) + \frac{1}{\sqrt{\pi} X} e^{-X^2} \right], \quad (51c)$$

where X is defined as before (with $v_t \rightarrow v$) and D_{ij} is decomposed into two components:

$$D_{ij} = \frac{v_i v_j}{v^2} D_{\parallel} + \frac{1}{2} \left(\delta_{ij} - \frac{v_i v_j}{v^2} \right) D_{\perp}. \quad (52)$$

We note that the drift \mathbf{A} contains two terms: the first term is proportional to the test particle mass m_t , while the second is independent of m_t . The first term represents a ‘‘polarization cloud’’ effect, which arises from the fact that a test particle moving relative to a background deflects more upstream than downstream background particles (Gilbert 1968; Mulder 1983). This process results in a density gradient and hence a drag force on the test particle, whose amplitude increases linearly with m_t since the gravitational deflection is proportional to m_t . In contrast, the second term in \mathbf{A} and the diffusivity tensor \mathbf{D} in equations (51a)–(51c) are both $\propto \rho_b m_b$ and independent of m_t . These two terms are due to fluctuations in the gravitational potential caused by granularity in the background particle distribution.

The derivation of the Fokker-Planck diffusion coefficients for two-body relaxation is completely different from our derivation based on quasi-linear cosmological density fluctuations. In both cases, fluctuating forces lead to random walks in velocity space and hence to the Fokker-Planck equation. The nature of the fluctuating forces and their consequences, however, are very different in the two cases.

Our cosmological Fokker-Planck derivation is exact (in second-order cosmological perturbation theory); it yields no drag, and the resulting drift and diffusion coefficients depend on position but are independent of velocity, as shown in equations (28a)–(30c). In contrast, the two-body relaxation calculation is based on a Taylor series expansion of the phenomenological (not exact) master equation; it yields no radial drift, and the drag and diffusion coefficients depend on velocity but are (in the usual derivation) independent of position (Rosenbluth et al. 1957). Furthermore, the two-body relaxation rates are proportional to G^2 , while the rates we compute are proportional to G from equations (28a) and (28b). In the cosmology case, the phase-space correlations are built into the initial power spectrum of density fluctuations, whose amplitude is given independently of G . With two-body relaxation, however, correlations are gravitationally induced by fluctuations instead of being present ab initio; hence, the rates pick up another factor of G .

The differences in the \mathbf{A} term (radial drift vs. drag) have an important consequence for equipartition of energy. As shown previously in the discussion of Brownian motion, drag and diffusion work together to drive a system to thermal equilibrium in which the mean squared velocity is proportional to

kT/m . Binney & Tremaine (1988, p. 513) show that this also happens with two-body relaxation. Because equilibrium systems with only gravitational forces generally have negative specific heat, there is no stable thermal equilibrium. Nonetheless, two-body relaxation tends to drive the velocity distribution toward the Maxwellian form, with equipartition of different mass species.

In the quasi-linear cosmological case, the drift and diffusion coefficients are independent of both velocity and particle mass. Therefore, the equilibrium velocity distribution cannot depend on particle mass: there is no equipartition of energy. This result is similar to violent relaxation, a process arising during the initial collapse and virialization of a dark matter halo when the gravitational potential is rapidly varying in time (Lynden-Bell 1967). Our relaxation process, however, arises in the quasi-linear regime and is present even if $\partial\phi/\partial t = 0$ as it is (at fixed comoving position) if $\Omega_m = 1$. Moreover, violent relaxation (as well as its partner, phase mixing) is generally understood as arising from the collisionless dynamics represented by the left-hand side of equation (47), while both two-body relaxation and our cosmological relaxation process are represented by the ‘‘collision’’ terms on the right-hand side.

Why did our cosmological calculation not yield a drag term (and a corresponding diffusion term linked to it by the fluctuation-dissipation theorem)? We suspect that it is because our calculation was limited to small-amplitude perturbations about a homogeneous and isotropic expanding cosmological model. The induced effects of substructure would only come in at higher order in perturbation theory. In the nonlinear regime we expect two types of drift terms to be present: $\mathbf{A} = \alpha\hat{r} - \gamma\mathbf{v}$, where α is the radial drift and γ is the drag coefficient. The Chandrasekhar calculation suggests that the drag and its accompanying diffusivity will depend on both position (through ρ_b) and velocity. The radial drift is absent in the Chandrasekhar calculation because the background there is assumed to be spatially homogeneous; with an inhomogeneous distribution in a virialized halo there should be a radial drift (and corresponding diffusivity) that could depend on both position and velocity. In the nonlinear regime drag should arise from dynamical friction just as with the globular clusters, except that the ‘‘background’’ particles whose discreteness causes the m_b terms in equations (51a)–(51c) are now the sub-halos and substructures that rain on a halo. Because we have not yet extended our derivation to the fully nonlinear regime, these expectations, while plausible, have yet to be demonstrated.

6. SUMMARY AND CONCLUSIONS

In this paper we have developed a cosmological kinetic theory, valid to second order in perturbation theory, to describe the evolution of the phase-space distribution of dark matter particles in galaxy halos. This theory introduces a new way to model the early phases of galaxy halo formation, which has traditionally been studied by analytic infall models or numerical N -body methods.

The key physical ingredients behind our kinetic description are stochastic fluctuations and dissipation caused by substructures arising from a spectrum of cosmological density perturbations. The kinetic equation that we have obtained at the end of our derivation, equations (24)–(26), has the standard form of a Fokker-Planck equation. The diffusion coefficients \mathbf{A} and \mathbf{D} represent acceleration due to a drift force and velocity space diffusion, respectively. To second order in perturbation theory, these coefficients are related to various covariance matrices of the cosmological density and

velocity fields given by equations (25) and (26). These expressions can in turn be written in terms of the familiar linear power spectrum $P(k)$ of matter fluctuations, as shown in equations (29a)–(30c).

The results for \mathbf{A} and \mathbf{D} from second-order perturbation theory are shown in Figures 1–4 for the currently favored Λ CDM model and in Figures 5 and 6 for various scale-free models with power law $P(k) \propto k^n$. Our results indicate that dissipative processes are important during the initial collapse of a dark matter halo, and it is not valid to model halos using the spherically symmetric Vlasov (collisionless Boltzmann) equation. Furthermore, we find that the diffusivity is initially negative close to the center of a halo, indicating a thermodynamic instability. The source of this instability is gravity: perturbations enhance gravitational collapse of dark matter halos.

We emphasize that our derivation leading to the Fokker-Planck equation given by equations (24)–(26) is exact to second order in cosmological perturbation theory. We do not follow the frequently used approach (e.g., Binney & Tremaine 1988) that begins by expanding the collisional terms of the nonexact master equation in an infinite series of changes in phase-space coordinates (i.e., $\Delta\mathbf{w}$) and arrives at the Fokker-Planck equation by assuming weak encounters (i.e., small $|\Delta\mathbf{w}|$) and truncating the series after the second-order terms. Instead, our starting point is equation (9), the first BBGKY hierarchy equation for a smooth one-particle distribution function. This is an exact kinetic equation; solving it requires specifying the one-particle distribution at an initial time and the two-particle correlation function at all times. We have shown in § 3 and Appendix A that if the cosmological density field δ and displacement (or velocity) field ψ are Gaussian random fields, the resulting two-particle correlation function depends on the one-particle distribution in a simple way given by equation (A22). Equation (A22) is exact to second order in perturbation theory and provides the closure relation for the BBGKY hierarchy in equation (9). Combining equation (A22) with equation (9) results in equation (24), which is a Fokker-Planck equation.

The cosmological dissipative processes identified in this paper are quite different from the standard two-body relaxation, which also leads to a Fokker-Planck equation. We find that there is no dynamical friction (i.e., terms proportional to $-\mathbf{v}$ in the diffusion coefficients) in second-order cosmological perturbation theory. Instead, there is a radial drift toward (or away from) the center of a halo as a result of the clustering of matter within the halo. The usual treatment of two-body relaxation has no such drift term because it assumes the medium to be homogeneous. Thus, we have identified a new relaxation process that must affect the phase-space structure of dark matter halos.

Although our derivation is exact to second order in the density fluctuations, it is valid only when the fluctuations are small. We are therefore describing only the early stages of dark matter halo formation. The expressions we obtain for the diffusion coefficients are plainly invalid in the nonlinear regime, where we expect dynamical friction to be present, as well as radial drift. In a later paper we will investigate the nonlinear generalizations of the exact second-order results derived in this paper.

The Fokker-Planck description should still apply in the nonlinear regime provided that the fluctuating force on a particle can be modeled as a Markov process consisting of a random walk of many steps. We conjecture that this description will be approximately valid when the matter distribution is modeled as a set of clumps (i.e., the halo model) that

scatter individual dark matter particles away from the orbits they would have in a smooth, spherical potential. The Fokker-Planck description should apply to not only the initial collapse and virialization of a dark matter halo containing substructure but also the subsequent evolution under minor mergers. Our inability to solve exactly the nonlinear dynamics, however, means that the diffusion coefficients will have to be calibrated using the nonlinear halo model or numerical N -body simulations.

Calibrating the diffusion coefficients is not equivalent to reproducing the N -body simulations. N -body simulations provide a Monte Carlo solution of the BBGKY hierarchy, which is much more general than the Fokker-Planck equation. If the simulation results are describable by Fokker-Planck evolution, this already represents a significant new result. Moreover, even if functional forms for the radial dependence of the diffusion coefficients have to be calibrated with simulations, the solution of the Fokker-Planck equation contains far more information because it gives the complete phase-space density distribution, not simply radial profiles.

Another important step will therefore be to actually solve the Fokker-Planck equation. In deriving it we have identified and quantified the relaxation processes affecting halo formation and evolution, but this is only the first step. In Appendix B we presented solutions in unrealistically simple cases just to highlight the roles played by radial drift, drag, and diffusion. An outstanding question is whether these processes actually are rapid enough to erase the memory of cosmological initial

conditions sufficiently so that dark matter halos relax to a universal profile. Once we have a nonlinear model for the diffusion coefficients, we will address this important question.

The work presented in this paper has a number of implications. First, we have highlighted the importance of the phase-space density for understanding dark matter dynamics and have developed a method for calculating its evolution during the early phases of structure formation. Second, we have identified a dissipative process that may erase memories of initial conditions during the early phases of galaxy halo evolution. This erasure is necessary for universal halo density profiles, although further investigation is needed to quantify how this dissipative process affects the density profiles. Third, the statistical description of the potential fluctuations caused by substructure that we have developed can also be added to investigate related problems such as the heating of galactic disks (Benson et al. 2004) and the merging of central black holes when galaxy halos merge (Kauffmann & Haehnelt 2000; Haehnelt & Kauffmann 2002; Hughes & Blandford 2003).

We thank Jon Arons, Avishai Dekel, and Martin Weinberg for useful conversations. The Aspen Center for Physics provided a stimulating environment where a portion of the research was carried out. C.-P. M. is partially supported by an Alfred P. Sloan Fellowship, a Cottrell Scholars Award from the Research Corporation, and NASA grant NAG5-12173.

APPENDIX A

STATISTICS OF CONSTRAINED GAUSSIAN RANDOM FIELDS

In this appendix we derive expressions for the probability distributions of displacement and density perturbation, $p(\psi_1)$ and $p(\delta_2, \psi_1)$, that are needed in equations (21) and (22). We then use them to evaluate equations (21) and (22) at high redshift when the matter distribution was described by a Gaussian random field of small-amplitude density perturbations. We use comoving spatial coordinates $\mathbf{x} = \mathbf{r}/a(t)$.

In accordance with the standard cosmological model, we assume that $\delta(\mathbf{x}, t)$ is a Gaussian random field with power spectrum $P(k, t)$ and that the displacement and density are related by equation (17). These are the appropriate assumptions at high redshift in standard models of structure formation. Strictly speaking, at high redshift, when linear theory applies, the evolution of the distribution function is trivial because the density and velocity fields evolve by spatially homogeneous amplification. Here we use the statistics of Gaussian random fields to compute quantities driving the lowest order corrections to linear evolution. We find that the fluctuating force is formally second order in perturbation theory.

A Gaussian random field is described most simply by its Fourier transform, which we define by

$$\delta(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \delta(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}}. \quad (\text{A1})$$

We suppress the time dependence for convenience. Before we apply a constraint, both $\delta(\mathbf{x})$ and $\delta(\mathbf{k})$ have zero mean. The covariance of $\delta(\mathbf{k})$ gives the power spectrum,

$$\langle \delta(\mathbf{k}_1) \delta(\mathbf{k}_2) \rangle = (2\pi)^3 P(k_1) \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2), \quad (\text{A2})$$

where δ_{D} is the Dirac delta function. In the linear regime, $\psi(\mathbf{x})$ is a linear functional of $\delta(\mathbf{x})$. The multivariate distribution of any linear functional of a Gaussian random field is itself Gaussian (i.e., multivariate normal) and is therefore described completely by its mean and covariance matrix. This remains true even for a constrained Gaussian random field (Bardeen et al. 1986).

In our case, we wish to constrain the initial density field to represent a protohalo. The constraints are applied to the smoothed density field

$$\Delta(\mathbf{x}) = \int d^3x' W_{\text{R}}(\mathbf{x} - \mathbf{x}') \delta(\mathbf{x}') = \int \frac{d^3k}{(2\pi)^3} W_{\text{R}}(\mathbf{k}) \delta(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}}, \quad (\text{A3})$$

where

$$W_R(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} W_R(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \quad (\text{A4})$$

is a smoothing window with unit volume integral, $\int d^3x W_R(\mathbf{x}) = 1$. The subscript R indicates the smoothing length. We assume that the smoothing window is spherically symmetric so that $W_R(k)$ is real and depends only on the magnitude of the wavevector. Throughout the paper we will use the Gaussian window $W_R(x) = (2\pi R^2)^{-3/2} \exp(-x^2/2R^2)$ and $W_R(k) = \exp(-k^2 R^2/2)$.

Unfortunately, there is no analytic theory to tell us where halos will form. However, there are a variety of plausible approaches. For example, following Bardeen et al. (1986), we might constrain Δ to have a maximum of specified height, orientation, and shape at $\mathbf{x} = 0$. This peak constraint involves 10 variables (Δ and the components of its first and second derivatives with respect to each of the coordinates), presenting a formidable challenge to evaluating the covariance matrix of the four variables (δ_2, ψ_1). The peak constraint is appealing but is difficult to calculate and in practice has not been found to predict well the actual formation sites of dark matter halos (Katz et al. 1993).

Instead of trying to impose complicated constraints for halo formation (e.g., Monaco et al. 2002), we err on the side of simplicity by choosing to use only the zeroth and first derivatives of the smoothed initial density field $\Delta(\mathbf{x})$. We require that $\mathbf{x} = 0$ be an *extremum* of the smoothed density field by requiring $\nabla\Delta(0) = 0$. Of course, an extremum may be a maximum, a minimum, or a saddle point. We therefore keep track also of the smoothed density $\Delta_0 \equiv \Delta(0)$. Provided that Δ_0 is greater than 2σ , the extrema of $\Delta(\mathbf{x})$ are close to the peaks (Bardeen et al. 1986).

How will we know if our results depend strongly on the choice of constraint? We will do this in two ways. First, we will vary the height of the extremum, Δ_0 . Second, we will drop the gradient constraint and use the single constraint Δ_0 . While these tests will not prove that we have a good model for the sites of halo formation, they do provide control cases for us to assess the sensitivity of our results to the details of the initial constraints.

Now we must calculate the mean and covariances of (δ_2, ψ_1) subject to constraints $(\Delta_0, \mathbf{\Delta}_0)$, where

$$\Delta_0 \equiv \Delta(\mathbf{x} = 0), \quad \mathbf{\Delta}_0 \equiv \nabla\Delta(\mathbf{x} = 0). \quad (\text{A5})$$

The method for this calculation is based on the theorem presented in Appendix D of Bardeen et al. (1986), which states that if \mathbf{Y}_A and \mathbf{Y}_B are zero-mean Gaussian variables (more generally, vectors of any length), then the conditional distribution for \mathbf{Y}_B given \mathbf{Y}_A , $p(\mathbf{Y}_B|\mathbf{Y}_A) = p(\mathbf{Y}_A, \mathbf{Y}_B)/p(\mathbf{Y}_A)$, is Gaussian with mean

$$\langle \mathbf{Y}_B | \mathbf{Y}_A \rangle \equiv \int d\mathbf{Y}_B p(\mathbf{Y}_B | \mathbf{Y}_A) \mathbf{Y}_B = \langle \mathbf{Y}_B \otimes \mathbf{Y}_A \rangle \langle \mathbf{Y}_A \otimes \mathbf{Y}_A \rangle^{-1} \mathbf{Y}_A \quad (\text{A6})$$

and covariance matrix

$$C(\mathbf{Y}_B, \mathbf{Y}_B) \equiv \int d\mathbf{Y}_B p(\mathbf{Y}_B | \mathbf{Y}_A) \Delta \mathbf{Y}_B \otimes \Delta \mathbf{Y}_B = \langle \mathbf{Y}_B \otimes \mathbf{Y}_B \rangle - \langle \mathbf{Y}_B \otimes \mathbf{Y}_A \rangle \langle \mathbf{Y}_A \otimes \mathbf{Y}_A \rangle^{-1} \langle \mathbf{Y}_A \otimes \mathbf{Y}_B \rangle, \quad (\text{A7})$$

where $\Delta \mathbf{Y}_B = \mathbf{Y}_B - \langle \mathbf{Y}_B | \mathbf{Y}_A \rangle$. The symbol \otimes denotes a tensor product and angle brackets denote mean values. Thus, $\langle \mathbf{Y}_A \otimes \mathbf{Y}_A \rangle$ is the covariance matrix of the constraints, while $C(\mathbf{Y}_B, \mathbf{Y}_B) = \langle \Delta \mathbf{Y}_B \otimes \Delta \mathbf{Y}_B \rangle_C = \langle \Delta \mathbf{Y}_B \otimes \Delta \mathbf{Y}_B | \mathbf{Y}_A \rangle$ is the covariance matrix of \mathbf{Y}_B subject to the constraints. (The symbol C , used either as a function or as a subscript, implies that a constraint is applied.) The juxtaposition of two matrices implies matrix multiplication.

A1. COVARIANCE OF CONSTRAINTS AND VARIABLES

To apply these results, first we compute the covariance matrix of constraints $\mathbf{Y}_A = (\Delta_0, \mathbf{\Delta}_0)$. Using equations (A1)–(A4), we get the covariances

$$\langle \Delta_0^2 \rangle = \sigma_0^2 \equiv \int \frac{d^3k}{(2\pi)^3} P(k) W_R^2(k), \quad (\text{A8a})$$

$$\langle \Delta_0 \mathbf{\Delta}_0 \rangle = 0, \quad (\text{A8b})$$

$$\langle \mathbf{\Delta}_0 \otimes \mathbf{\Delta}_0 \rangle = \sigma_1^2 \mathbf{I}, \quad \sigma_1^2 \equiv \frac{1}{3} \int \frac{d^3k}{(2\pi)^3} k^2 P(k) W_R^2(k), \quad (\text{A8c})$$

where \mathbf{I} is the unit tensor.

Next we consider the covariance matrix of the variables $\mathbf{Y}_B = (\delta_1, \delta_2, \psi_1)$. Since $\delta = -(\partial/\partial\mathbf{x}) \cdot \boldsymbol{\psi}$, the displacement field $\boldsymbol{\psi}$ has an additional factor $i\mathbf{k}/k^2$ in the integrand of equation (A1). For the off-diagonal elements of the covariance matrix of \mathbf{Y}_B we obtain

$$\langle \delta_1 \delta_2 \rangle = \xi(r) \equiv \int \frac{d^3k}{(2\pi)^3} P(k) j_0(kr), \quad \mathbf{r} \equiv \mathbf{x}_1 - \mathbf{x}_2, \quad (\text{A9a})$$

$$\langle \delta_1 \psi_1 \rangle = 0, \quad (\text{A9b})$$

$$\langle \delta_2 \psi_1 \rangle = -\eta(r)\mathbf{r}, \quad \eta(r) \equiv \int \frac{d^3k}{(2\pi)^3} P(k) \frac{j_1(kr)}{kr}, \quad (\text{A9c})$$

where the spherical Bessel functions are $j_0(x) = x^{-1} \sin x$ and $j_1(x) = x^{-2}(\sin x - x \cos x) = -dj_0/dx$. The diagonal variances are

$$\langle \delta_1^2 \rangle = \langle \delta_2^2 \rangle = \xi(0), \quad (\text{A10a})$$

$$\langle \boldsymbol{\psi}_1 \otimes \boldsymbol{\psi}_1 \rangle = \sigma_\psi^2 \mathbf{I}, \quad \sigma_\psi^2 = \frac{1}{3} \int \frac{d^3k}{(2\pi)^3} k^{-2} P(k). \quad (\text{A10b})$$

Note that in equations (A9a)–(A9c) (and in the following) we are using \mathbf{r} as a comoving displacement vector and not as the radius vector in proper coordinates. Proper coordinates are not used in this appendix.

A2. CROSS COVARIANCE BETWEEN VARIABLES AND CONSTRAINTS

Finally, we have the cross covariances between our primary variables and the constraints, i.e., the components of $\langle \mathbf{Y}_B \otimes \mathbf{Y}_A \rangle$. These follow from equations (A1)–(A3):

$$\langle \delta(\mathbf{x}) \Delta_0 \rangle = \bar{\xi}(r) \equiv \int \frac{d^3k}{(2\pi)^3} P(k) W_R(k) j_0(kr), \quad r = |\mathbf{x}|, \quad (\text{A11a})$$

$$\langle \boldsymbol{\psi}(\mathbf{x}) \Delta_0 \rangle = -\bar{\eta}(r)\mathbf{x}, \quad \bar{\eta}(r) \equiv \int \frac{d^3k}{(2\pi)^3} P(k) W_R(k) \frac{j_1(kr)}{kr}, \quad (\text{A11b})$$

$$\langle \delta(\mathbf{x}) \boldsymbol{\Delta}_0 \rangle = -\frac{d\bar{\xi}}{dr} \hat{\mathbf{r}}, \quad \hat{\mathbf{r}} = \mathbf{x}/r, \quad (\text{A11c})$$

$$\langle \boldsymbol{\psi}(\mathbf{x}) \otimes \boldsymbol{\Delta}_0 \rangle = r \frac{d\bar{\eta}}{dr} \hat{\mathbf{r}} \otimes \hat{\mathbf{r}} + \bar{\eta} \mathbf{I}. \quad (\text{A11d})$$

To arrive at the last expression above, we have used the identity

$$\int \frac{d\Omega}{4\pi} e^{-i\mathbf{k} \cdot \mathbf{x}} \hat{\mathbf{n}} \otimes \hat{\mathbf{n}} = r \frac{\partial}{\partial r} \left[\frac{j_1(kr)}{kr} \right] \hat{\mathbf{r}} \otimes \hat{\mathbf{r}} + \frac{j_1(kr)}{kr} \mathbf{I}, \quad (\text{A12})$$

where $\mathbf{k} = k\hat{\mathbf{n}}$, $\mathbf{x} = r\hat{\mathbf{r}}$, and \mathbf{I} is the unit tensor.

A3. MEAN OF VARIABLES SUBJECT TO CONSTRAINTS

Using equations (A6) and (A8a)–(A11d), we get the mean values subject to the extremum constraints ($\Delta_0, \boldsymbol{\Delta}_0 = 0$),

$$\langle \delta_1 \rangle_C \equiv \langle \delta_1 | \Delta_0, \boldsymbol{\Delta}_0 \rangle = \frac{\Delta_0}{\sigma_0^2} \bar{\xi}(r_1), \quad r_1 = |\mathbf{x}_1|, \quad (\text{A13a})$$

$$\langle \delta_2 \rangle_C \equiv \langle \delta_2 | \Delta_0, \boldsymbol{\Delta}_0 \rangle = \frac{\Delta_0}{\sigma_0^2} \bar{\xi}(r_2), \quad r_2 = |\mathbf{x}_2|, \quad (\text{A13b})$$

$$\langle \boldsymbol{\psi}_1 \rangle_C \equiv \langle \boldsymbol{\psi}_1 | \Delta_0, \boldsymbol{\Delta}_0 \rangle = -\frac{\Delta_0}{\sigma_0^2} \bar{\eta}(r_1) \mathbf{x}_1. \quad (\text{A13c})$$

The subscript C denotes “subject to constraints.”

Equations (A13a)–(A13c) would be unchanged had we dropped the extremum condition and imposed instead the single constraint $\Delta(0) = \Delta_0$; the constraint $\boldsymbol{\Delta}_0 = 0$ has no effect on the mean values. This follows because $\langle \Delta_0 \boldsymbol{\Delta}_0 \rangle = 0$ so that the only way that $\boldsymbol{\Delta}_0$ can enter the mean values is through terms proportional to $\boldsymbol{\Delta}_0/\sigma_1^2$, which vanish when the extremum constraint is imposed.

A4. COVARIANCE OF VARIABLES SUBJECT TO CONSTRAINTS

We now compute the covariance matrix of our variables $\mathbf{Y}_B = (\delta_1, \delta_2, \boldsymbol{\psi}_1)$ subject to the extremum constraints $\mathbf{Y}_A = (\Delta_0, \mathbf{\Delta}_0)$. The covariances follow from equations (A7)–(A11d). Recalling the definition $C(\mathbf{Y}_B, \mathbf{Y}_B) \equiv \langle \Delta \mathbf{Y}_B \otimes \Delta \mathbf{Y}_B | \mathbf{Y}_A \rangle$, we obtain

$$C(\delta_1, \delta_2) = \xi(r) - \frac{\bar{\xi}_1 \bar{\xi}_2}{\sigma_0^2} - \frac{1}{\sigma_1^2} \frac{d\bar{\xi}}{dr_1} \frac{d\bar{\xi}}{dr_2} \hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2, \quad \mathbf{r} \equiv \mathbf{x}_1 - \mathbf{x}_2, \quad \hat{\mathbf{r}}_1 \equiv \frac{\mathbf{x}_1}{r_1}, \quad \hat{\mathbf{r}}_2 \equiv \frac{\mathbf{x}_2}{r_2}, \quad (\text{A14a})$$

$$C(\delta_2, \boldsymbol{\psi}_1) = -\eta(r)\mathbf{r} + \frac{\bar{\xi}_2 \bar{\eta}_1 \mathbf{x}_1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \frac{d\bar{\xi}}{dr_2} [(\bar{\xi}_1 - 3\bar{\eta}_1)(\hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2)\hat{\mathbf{r}}_1 + \bar{\eta}_1 \hat{\mathbf{r}}_2], \quad (\text{A14b})$$

$$C(\delta_1, \boldsymbol{\psi}_1) = \frac{\bar{\xi}_1 \bar{\eta}_1 \mathbf{x}_1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \frac{d\bar{\xi}}{dr_1} (\bar{\xi}_1 - 2\bar{\eta}_1)\hat{\mathbf{x}}_1, \quad (\text{A14c})$$

$$C(\boldsymbol{\psi}_1, \boldsymbol{\psi}_1) = c_r \hat{\mathbf{r}}_1 \otimes \hat{\mathbf{r}}_1 + c_t (\mathbf{I} - \hat{\mathbf{r}}_1 \otimes \hat{\mathbf{r}}_1), \quad (\text{A14d})$$

$$c_r(r_1) \equiv \sigma_\psi^2 - \left(\frac{r_1 \bar{\eta}_1}{\sigma_0} \right)^2 - \left(\frac{\bar{\xi}_1 - 2\bar{\eta}_1}{\sigma_1} \right)^2, \quad (\text{A14e})$$

$$c_t(r_1) \equiv \sigma_\psi^2 - \left(\frac{\bar{\eta}_1}{\sigma_1} \right)^2, \quad (\text{A14f})$$

where $\bar{\xi}_i \equiv \xi(r_i)$ and $\bar{\eta}_i \equiv \eta(r_i)$, and we have used the relation $r_1(d\bar{\eta}_1/dr_1) = \bar{\xi}_1 - 3\bar{\eta}_1$. Note that $C(\delta_i, \boldsymbol{\psi}_1)$ is a vector and $C(\boldsymbol{\psi}_1, \boldsymbol{\psi}_1)$ is a second-rank tensor. Later we will need the tensor $C^{-1}(\boldsymbol{\psi}_1, \boldsymbol{\psi}_1)$, which is the matrix inverse of $C(\boldsymbol{\psi}_1, \boldsymbol{\psi}_1)$.

The role of constraints is easy to see in equations (A14a)–(A14f). The constraint on Δ_0 is responsible for the terms proportional to σ_0^{-2} , while the constraint on $\mathbf{\Delta}_0$ gives rise to the terms proportional to σ_1^{-2} . Therefore, the constraints are easily removed by setting $\sigma_1^{-2} \rightarrow 0$ (to eliminate the constraint on $\mathbf{\Delta}_0$) or $\sigma_0^{-2} \rightarrow 0$ (to eliminate the constraint on Δ_0). As expected, in the absence of any constraints, the covariance matrix recovers the same expressions as the unconstrained covariance in equations (A9a)–(A10b).

We will need one more covariance, $C(\boldsymbol{\psi}_1, \boldsymbol{\psi}_0)$ subject to the extremum constraints where $\boldsymbol{\psi}_0 \equiv \boldsymbol{\psi}(0)$. Using equations (A7), (A11a)–(A11d), and (A12), we get

$$C(\boldsymbol{\psi}_1, \boldsymbol{\psi}_0) = \frac{d\gamma}{dr_1} \hat{\mathbf{r}}_1 \otimes \hat{\mathbf{r}}_1 + \frac{\gamma}{r_1} (\mathbf{I} - \hat{\mathbf{r}}_1 \otimes \hat{\mathbf{r}}_1), \quad \gamma(r_1) \equiv \int \frac{d^3k}{(2\pi)^3} P(k) k^{-3} j_1(kr_1). \quad (\text{A15})$$

A5. PROBABILITY DISTRIBUTIONS

Now we are ready to calculate the probability distribution functions and means of various quantities needed in § 3. Specifically, we need $p(\boldsymbol{\psi}_1)$, $\langle \delta_1 | \boldsymbol{\psi}_1 \rangle_C$, $\langle \delta_2 | \boldsymbol{\psi}_1 \rangle_C$, $\langle \delta_2 \rangle_C$, and $\langle \delta_1 \delta_2 | \boldsymbol{\psi}_1 \rangle_C$ that appear in the key equations (21) and (22) for $f(\mathbf{w}_1, t)$ and $f_{2c}(\mathbf{w}_1, \mathbf{x}_2, t)$. Equation (A13b) already gives $\langle \delta_2 \rangle_C$, but we need to calculate the other quantities.

First, the probability density distribution for the displacement $p(\boldsymbol{\psi}_1)$ (subject to the extremum constraints) is a three-dimensional Gaussian whose mean and covariance are given above, implying

$$p(\boldsymbol{\psi}_1) = (2\pi)^{-3/2} (c_r c_t^2)^{-1/2} \exp \left[-\frac{1}{2} (\boldsymbol{\psi}_1 - \langle \boldsymbol{\psi}_1 \rangle_C) \cdot C^{-1}(\boldsymbol{\psi}_1, \boldsymbol{\psi}_1) \cdot (\boldsymbol{\psi}_1 - \langle \boldsymbol{\psi}_1 \rangle_C) \right]. \quad (\text{A16})$$

Note that the argument of the exponential is a scalar: the dot denotes a contraction (dot product).

The joint distribution of displacement and density (subject to the constraints) is a little more complicated to work out since this requires inverting the 4×4 covariance matrix of $(\delta_2, \boldsymbol{\psi}_1)$. The distribution factors into conditional and marginal distributions, $p(\delta_2, \boldsymbol{\psi}_1) = p(\boldsymbol{\psi}_1) p(\delta_2 | \boldsymbol{\psi}_1)$. Straightforward algebra gives the conditional distribution,

$$p(\delta_2 | \boldsymbol{\psi}_1) = (2\pi Q)^{-1/2} \exp \left\{ -\frac{1}{2Q} [\delta_2 - \langle \delta_2 \rangle_C - C(\delta_2, \boldsymbol{\psi}_1) \cdot C^{-1}(\boldsymbol{\psi}_1, \boldsymbol{\psi}_1) \cdot (\boldsymbol{\psi}_1 - \langle \boldsymbol{\psi}_1 \rangle_C)]^2 \right\}, \quad (\text{A17})$$

where

$$Q \equiv C(\delta_2, \delta_2) - C(\delta_2, \boldsymbol{\psi}_1) \cdot C^{-1}(\boldsymbol{\psi}_1, \boldsymbol{\psi}_1) \cdot C(\boldsymbol{\psi}_1, \delta_2). \quad (\text{A18})$$

Note that the second term in Q (and the similar term in eq. [A17]) is a scalar. Writing it out with indices and using the summation convention, $C(\delta_2, \delta_2) - Q = C_i(\delta_2, \boldsymbol{\psi}_1) C_j^{-1}(\boldsymbol{\psi}_1, \boldsymbol{\psi}_1) C_j(\boldsymbol{\psi}_1, \delta_2)$.

Equations (A17) and (A18) could also have been derived by noting that $\langle \mathbf{Y}_B | \mathbf{Y}_A \rangle_C = \langle \delta_2 | \psi_1 \rangle_C - \langle \delta_2 \rangle_C$ for $\mathbf{Y}_A = \psi_1 - \langle \psi_1 \rangle_C$ and $\mathbf{Y}_B = \delta_2 - \langle \delta_2 \rangle_C$. Then using equation (A6) and relations $\langle \mathbf{Y}_B \mathbf{Y}_A \rangle_C = \langle (\delta_2 - \langle \delta_2 \rangle_C)(\psi_1 - \langle \psi_1 \rangle_C) \rangle_C = C(\delta_2, \psi_1)$ and $\langle \mathbf{Y}_A \mathbf{Y}_A \rangle_C = C(\psi_1, \psi_1)$, we obtain

$$\begin{aligned} \langle \delta_2 | \psi_1 \rangle_C &= \langle \delta_2 \rangle_C + C(\delta_2, \psi_1) \cdot C^{-1}(\psi_1, \psi_1) \cdot (\psi_1 - \langle \psi_1 \rangle_C) \\ &= \langle \delta_2 \rangle_C - C(\delta_2, \psi_1) \cdot \frac{\partial}{\partial \psi_1} \log p(\psi_1), \end{aligned} \quad (\text{A19})$$

which is identical to the mean in equation (A17). The constrained mean $\langle \delta_1 | \psi_1 \rangle_C$ follows simply by replacing $\langle \delta_2 \rangle_C$ with $\langle \delta_1 \rangle_C$ and $C(\delta_2, \psi_1)$ with $C(\delta_1, \psi_1)$.

The only remaining constrained mean that we need to evaluate is $\langle \delta_1 \delta_2 | \psi_1 \rangle_C$. Direct evaluation of the conditional distribution would require inverting a 5×5 matrix. It is much simpler to use equations (A6) and (A7) with $\mathbf{Y}_B = \{(\delta_1 - \langle \delta_1 \rangle_C), (\delta_2 - \langle \delta_2 \rangle_C)\}$ and $\mathbf{Y}_A = \psi_1 - \langle \psi_1 \rangle_C$. From the off-diagonal element of $\langle \Delta \mathbf{Y}_B \otimes \Delta \mathbf{Y}_B | \mathbf{Y}_A \rangle$, we get

$$\langle \delta_1 \delta_2 | \psi_1 \rangle_C = \langle \delta_1 \rangle_C \langle \delta_2 \rangle_C + C(\delta_1, \delta_2) - C(\delta_1, \psi_1) \cdot C^{-1}(\psi_1, \psi_1) \cdot C(\psi_1, \delta_2). \quad (\text{A20})$$

At last we are ready to evaluate the one- and two-particle distribution functions for small-amplitude Gaussian fluctuations. Substituting $\langle \delta_1 | \psi_1 \rangle_C$ into equation (21), to second order in δ we obtain

$$f(\mathbf{w}, t) = \bar{\rho}(1 + \langle \delta \rangle_C) p[\psi - C(\delta, \psi)] (Hb)^{-3} + O(\delta^3), \quad (\text{A21})$$

where $\langle \delta \rangle_C$ is given by equation (A13a) with $r_1 = r$ and $C(\delta, \psi)$ is a vector equal to the covariance of the constrained δ and ψ given in equation (A14c). Without an initial constraint we would have $\langle \delta \rangle_C = 0$ and $C(\delta, \psi) = 0$. An initial constraint on the protohalo shifts the mean density and causes the density and velocity to be correlated; hence, it changes the phase-space density. Note that equation (A21) implies that in linear theory the spatial and velocity distributions decouple. As expected, the mean spatial density is $\bar{\rho}(1 + \langle \delta \rangle_C)$. The velocity distribution at each point in the halo is simply the underlying Gaussian of equation (A16) with the mean shifted by $C(\delta, \psi)$. Note that the value of the protohalo smoothed density constraint Δ_0 affects the density profile, $\langle \delta \rangle_C \propto \Delta_0$, but it does not affect the velocity distribution up to second order in perturbation theory. The imposition of the extremum constraint $\Delta_0 = 0$ affects the phase-space density only through the term in $C(\delta, \psi) \propto \sigma_1^{-2}$. Setting $\sigma_1^{-2} \rightarrow 0$ everywhere removes the constraint on Δ_0 .

The two-particle correlation is found by substituting equations (A19) and (A20) into equation (22). The result is

$$f_{2c}(\mathbf{w}_1, \mathbf{x}_2, t) = \bar{\rho}^2 \left\{ C(\delta_1, \delta_2) - C(\delta_1, \psi_1) \cdot C^{-1}(\psi_1, \psi_1) \cdot C(\psi_1, \delta_2) - [C(\delta_2, \psi_1) - C(\delta_1, \psi_1) \langle \delta_2 \rangle_C] \cdot \frac{\partial}{\partial \psi_1} \log p(\psi_1) \right\} p(\psi_1) (Hb)^{-3}. \quad (\text{A22})$$

This equation is the main result of this appendix. It is exact for Gaussian random fields.

APPENDIX B

STATIONARY SOLUTIONS AND RELAXATION TO EQUILIBRIUM

Stationary solutions of the Fokker-Planck equation are relevant for systems that relax on a timescale faster than external changes take place. The stationary solution for classical Brownian motion was given in § 5.3. Stationary solutions may also exist for cosmological systems where the constant-temperature heat bath is replaced by small-scale structure that rains on dark matter halos.

Stationary solutions have $\partial f / \partial t = 0$ and $\mathbf{F} \equiv \mathbf{A}f - \mathbf{D} \cdot \partial f / \partial \mathbf{v} = 0$ so that the flux vanishes in velocity space. We consider spherically symmetric solutions as appropriate for the average halo. From Jeans' Theorem, the distribution function is a function of the integrals of motion (Binney & Tremaine 1988). For spherical systems it is often assumed that stationary solutions have the form $f = f(E, L)$, where

$$E = \frac{1}{2} v^2 + \Phi(\mathbf{r}), \quad L = |\mathbf{r} \times \mathbf{v}|. \quad (\text{B1})$$

(The particle masses are irrelevant, so we use the specific energy and angular momentum.) The collisionless Boltzmann equation gives no constraint on $f(E, L)$. The Fokker-Planck equation, however, constrains f because stationarity requires that $\mathbf{D} \cdot \partial \ln f / \partial \mathbf{v} = \mathbf{A}$. Assuming spherical symmetry, we write

$$\mathbf{A} = \alpha \hat{\mathbf{r}} - \gamma \mathbf{v}, \quad (\text{B2})$$

introducing the radial drift coefficient α and drag coefficient γ . We assume isotropic diffusivity D . For the globular cluster case with Chandrasekhar's dynamical friction, $\alpha = 0$. For the case of quasi-linear cosmological fluctuations arising from Gaussian

random fields, we have seen that $\gamma = 0$. (The cosmological case also allows for different diffusivities in the radial and tangential directions, but that is an unnecessary complication here.)

Requiring the phase-space flux to vanish now gives the following constraints on the distribution function $f(E, L)$:

$$\frac{\partial \ln f}{\partial E} = -\frac{\gamma}{D} + \frac{\alpha}{v_r D}, \quad \frac{r^2}{L} \frac{\partial \ln f}{\partial L} = -\frac{\alpha}{v_r} D. \quad (\text{B3})$$

It is interesting to see what conditions lead to thermal equilibrium with a Maxwellian distribution of velocities, i.e., equation (46) with $f = W$. Equation (B3) shows that the conditions are $\alpha = 0$ and $\gamma/D = m/(k_B T)$. Thus, the drift must be velocity dependent and behave like a drag, $\mathbf{A} \propto -\mathbf{v}$. Friction converts shear into heat in such a way as to drive the system toward thermal equilibrium. Chandrasekhar dynamical friction satisfies these conditions; therefore, two-body relaxation drives the velocity distribution toward the Maxwell-Boltzmann form. Quasi-linear cosmological fluctuations, however, are quite different, having $\gamma = 0$ and $\alpha \neq 0$, implying $\partial \ln f / \partial E = -(r^2/L) \partial \ln f / \partial L = \alpha/(v_r D)$. While these equations do not have simple general solutions, it is clear that cosmological fluctuations do not drive a system toward thermal equilibrium with equipartition of energies. Our cosmological relaxation process is similar to violent relaxation in this respect (Lynden-Bell 1967).

Although the general cosmological case of relaxation in a dark matter halo with $\alpha \neq 0$ is complicated and cannot be fully solved here, there is a special case worthy of closer examination, namely, the evolution of an unconstrained cosmological Gaussian random field, where fluctuations generate peculiar velocities. That is, we drop the constraint of having a halo centered at $r = 0$ and instead consider the velocity distribution of dark matter particles at a randomly selected point in space. To consider the cosmological case, we should use the appropriate comoving variables to factor out the Hubble expansion. In place of \mathbf{r} and \mathbf{v} , we use the following variables:

$$\mathbf{x} = a^{-1} \mathbf{r}, \quad \mathbf{u} = a(\mathbf{v} - H\mathbf{r}), \quad (\text{B4})$$

where $a(t)$ is the cosmic scale factor and $H = d \ln a / dt$. With this change of variables, the Fokker-Planck equation (24) becomes

$$\frac{\partial f}{\partial t} + \frac{\mathbf{u}}{a^2} \cdot \frac{\partial f}{\partial \mathbf{x}} + a \left(\mathbf{g} - \frac{1}{a} \frac{d^2 a}{dt^2} \mathbf{r} \right) \cdot \frac{\partial f}{\partial \mathbf{u}} = a \frac{\partial}{\partial \mathbf{u}} \cdot \left(-A\mathbf{f} + a\mathbf{D} \cdot \frac{\partial f}{\partial \mathbf{u}} \right). \quad (\text{B5})$$

If the velocities are measured in the comoving frame, \mathbf{g}_T is replaced by \mathbf{g} . When using comoving coordinates, the gravitational field is reduced by the Hubble acceleration, $(d^2 a / dt^2) \mathbf{x}$. If the mean density field is homogeneous, then all particles move with the unperturbed Hubble expansion, $\mathbf{r} \propto a(t)$, implying $\mathbf{g} = (d^2 a / dt^2) \mathbf{x}$. Thus, both the velocity and acceleration terms on the left-hand side vanish when $\partial f / \partial \mathbf{x} = 0$. If we apply no initial constraint on the density field, then the one-point distribution function f is a function only of time and peculiar velocity $\mathbf{v} - H\mathbf{r}$. We use the variable \mathbf{u} for the velocity because $\mathbf{v} - H\mathbf{r}$ decreases in proportion to $a^{-1}(t)$ for a freely falling body in an unperturbed homogeneous and isotropic expanding universe. If not for the right-hand side, the solution of equation (B5) for a homogeneous and isotropic model would be $f(\mathbf{x}, \mathbf{u}, t) = f(\mathbf{x} = \mathbf{0}, \mathbf{u}, t = 0)$; i.e., the initial peculiar velocity distribution would simply redshift with cosmic expansion and remain spatially homogeneous.

The Fokker-Planck equation does not, however, describe an unperturbed homogeneous and isotropic universe. Rather, it describes the statistical average of an ensemble of universes with fluctuations. In the case of unconstrained initial conditions, $f = f(u, t)$ and equation (B5) reduces to

$$\frac{1}{a} \frac{\partial f}{\partial t} = \frac{1}{u^2} \frac{\partial}{\partial u} \left[u^2 \left(\alpha_u f + aD \frac{\partial f}{\partial u} \right) \right], \quad (\text{B6})$$

where $\alpha_u = \alpha_u(u, t)$ and $D = D(u, t)$ in general. We have used statistical isotropy to require $\mathbf{A} = -\alpha_u \hat{\mathbf{u}}$, where $\hat{\mathbf{u}}$ is a unit vector in the direction of \mathbf{u} . In second-order perturbation theory (§ 4.1), $\alpha_u = 0$ (no drag) and $D = 4\pi G \bar{\rho} a H b \sigma_v^2$ (assuming that we measure velocities in the comoving frame). The Fokker-Planck equation then takes the form of a diffusion equation with D independent of u , and the general solution is

$$f(\mathbf{u}, t) = \int \exp \left(-\frac{|\mathbf{u} - \mathbf{u}'|^2}{4 \int_0^t a^2 D dt} \right) \frac{f(\mathbf{u}, 0) d^3 u'}{(4\pi \int_0^t a^2 D dt)^{3/2}}. \quad (\text{B7})$$

The effect of fluctuations is simply to spread the initial velocity distribution (if $D > 0$) while retaining the Gaussian form. Physically, the gravitational fluctuations of substructure cause local heating.

Contrary to equation (B7), cosmological simulations show that the distribution of peculiar velocities in the nonlinear regime is closer to exponential rather than Gaussian, at least in the tails (Sheth & Diaferio 2001). It is interesting to ask whether the Fokker-Planck equation can explain such a result. With the drift term retained, equation (B6) has the stationary solution

$$f(\mathbf{u}, t) = \exp \left[-\int^u \frac{\alpha_u(u, t) du}{a(t)D(u, t)} \right]. \quad (\text{B8})$$

If α_u/D is positive and independent of u , the result is indeed an exponential distribution of peculiar velocities u . In second-order perturbation theory we found that $\alpha_u = 0$ and D is independent of u ; however, Chandrasekhar's dynamical friction gives $A/D \propto u$.

Evidently, a third case is needed to explain the exponential velocity distribution: there must be a drag such that α_u/D is independent of u , quite unlike dynamical friction from two-body relaxation. Determining whether such a drag term arises from nonlinear gravitational clustering is a subject for further work.

As one final example, we consider the effects of both drag and radial drift in a simplified model illustrating the effects of collisions on a system for which the relaxation time is much longer than the dynamical time (e.g., the crossing time for a particle in a halo). Under these conditions, the system at all times is nearly in collisionless equilibrium with the phase-space density being a function of the actions alone (and not the angles, using action angle variables). Because the actions are conserved in the absence of collisions, for quasi-static evolution equation (24) simplifies to

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial \mathbf{v}} \cdot \left[(\boldsymbol{\alpha} - \gamma \mathbf{v})f - \mathbf{D} \cdot \frac{\partial f}{\partial \mathbf{v}} \right], \quad (\text{B9})$$

where we have generalized the radial drift to an arbitrary vector field $\boldsymbol{\alpha}$, which we assume to be independent of \mathbf{v} . We can solve this equation exactly for the simple case in which $\boldsymbol{\alpha}$, γ , and \mathbf{D} are all constants. The general solution is the Green's function solution

$$f(\mathbf{r}, \mathbf{v}, t) = \int d^3 v' G(\mathbf{r}, \mathbf{v} - \mathbf{v}', t) f(\mathbf{r}, \mathbf{v}', 0), \quad (\text{B10})$$

with Green's function

$$G(\mathbf{r}, \mathbf{v}, t) = \int \frac{d^3 s}{(2\pi)^3} \exp \left\{ i \mathbf{s} \cdot \left[\mathbf{v} - \frac{1}{\gamma} (1 - e^{-\gamma t}) \boldsymbol{\alpha} \right] - \frac{1}{2\gamma} \mathbf{s} \cdot \mathbf{D} \cdot \mathbf{s} (1 - e^{-2\gamma t}) \right\}. \quad (\text{B11})$$

The parameter γ^{-1} is a relaxation time. Equation (B11) simplifies to a normalized Gaussian in two limits,

$$G(\mathbf{r}, \mathbf{v}, t) = \begin{cases} \mathcal{N}(\boldsymbol{\alpha}t, 2\mathbf{D}t), & \gamma t \ll 1, \\ \mathcal{N}(\boldsymbol{\alpha}/\gamma, \mathbf{D}/\gamma), & \gamma t \gg 1, \end{cases} \quad (\text{B12})$$

where $\mathcal{N}(\mathbf{u}, \mathbf{M})$ is a multivariate normal with mean \mathbf{u} and covariance matrix \mathbf{M} . The solution for $\boldsymbol{\alpha} = \gamma = 0$ was given previously in equation (B7). Now we have a more general solution showing that the equilibrium (for constant diffusion coefficients) is a Maxwellian with velocity dispersion tensor \mathbf{D}/γ just in equation (B8) if $\alpha_u = \gamma u$, but now the drift vector $\boldsymbol{\alpha}$ produces an offset of the mean velocity $\langle \mathbf{v} \rangle = \boldsymbol{\alpha}/\gamma$. Thus, velocity-independent drift $\boldsymbol{\alpha}$ and velocity-dependent drag $-\gamma \mathbf{v}$ play completely different roles in collisional relaxation. In later papers we will investigate these roles for more realistic models of nonlinear halo evolution.

REFERENCES

- Antonuccio-Delogu, V., & Colafrancesco, S. 1994, *ApJ*, 427, 72
 Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, *ApJ*, 304, 15
 Benson, A. J., Lacey, C. G., Frenk, C. S., Baugh, C. M., & Cole, S. 2004, *MNRAS*, in press
 Bertschinger, E. 1985, *ApJS*, 58, 39
 ———. 1993, in *Statistical Description of Transport in Plasma*, Astro-, and Nuclear Physics, ed. J. Misguich, G. Pelletier, & P. Schuck (Commaack: Nova Science), 193
 ———. 1996, in *Cosmology and Large Scale Structure*, ed. R. Schaeffer, J. Silk, M. Spiro, & J. Zinn-Justin (Amsterdam: Elsevier), 273
 Binney, J. J., & Tremaine, S. 1988, *Galactic Dynamics* (Princeton: Princeton Univ. Press)
 Bullock, J. S., Dekel, A., Kolatt, T. S., Kravtsov, A. V., Klypin, A. A., Porciani, C., & Primack, J. R. 2001, *ApJ*, 555, 240
 Chandrasekhar, S. 1943, *ApJ*, 97, 255
 Davis, M., & Peebles, P. J. E. 1977, *ApJS*, 34, 425
 Dekel, A., Devor, J., & Hetzroni, G. 2003, *MNRAS*, 341, 326
 Dupree, T. H. 1967, *Phys. Fluids*, 10, 1049
 Evans, N. W., & Collett, J. L. 1997, *ApJ*, 480, L103
 Fillmore, J. A., & Goldreich, P. 1984, *ApJ*, 281, 1
 Fokker, A. D. 1914, *Ann. Phys.*, 43, 810
 Gardiner, C. W. 2002, *Handbook of Stochastic Methods* (Berlin: Springer)
 Ghigna, S., Moore, B., Governato, F., Lake, G., Quinn, T., & Stadel, J. 1998, *MNRAS*, 300, 146
 Gilbert, I. H. 1968, *ApJ*, 152, 1043
 Gunn, J. E., & Gott, J. R. 1972, *ApJ*, 176, 1
 Haehnelt, M., & Kauffmann, G. 2002, *MNRAS*, 336, L61
 Hoffman, Y., & Shaham, J. 1985, *ApJ*, 297, 16
 Hughes, S. A., & Blandford, R. D. 2003, *ApJ*, 585, L101
 Ichimaru, S. 1973, *Basic Principles of Plasma Physics: A Statistical Approach* (Reading: Addison-Wesley)
 Jain, B., & Bertschinger, E. 1994, *ApJ*, 431, 495
 Katz, N., Quinn, T., & Gelb, J. 1993, *MNRAS*, 265, 689
 Kauffmann, G., & Haehnelt, M. 2000, *MNRAS*, 311, 576
 Klimontovich, Yu. L. 1967, *The Statistical Theory of Non-Equilibrium Processes in a Plasma* (Cambridge: MIT Press)
 Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, *ApJ*, 522, 82
 Langevin, P. 1908, *Comptes rendus*, 146, 530
 Lynden-Bell, D. 1967, *MNRAS*, 136, 101
 Monaco, P., Theuns, T., Taffoni, G., Governato, F., Quinn, T., & Stadel, J. 2002, *ApJ*, 564, 8
 Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., & Tozzi, P. 1999, *ApJ*, 524, L19
 Moore, B., Governato, F., Quinn, T., Stadel, J., & Lake, G. 1998, *ApJ*, 499, L5
 Mulder, W. A. 1983, *A&A*, 117, 9
 Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563
 ———. 1997, *ApJ*, 490, 493
 Nusser, A., & Sheth, R. K. 1999, *MNRAS*, 303, 685
 Peebles, P. J. E. 1980, *The Large Scale Structure of the Universe* (Princeton: Princeton Univ. Press)
 Planck, M. 1917, *Sitzber. Preuf. Akad. Wiss.*, 324
 Risken, H. 1989, *The Fokker-Planck Equation* (Berlin: Springer)
 Rosenbluth, M. N., MacDonald, W. M., & Judd, D. L. 1957, *Phys. Rev.*, 107, 1
 Sheth, R. K., & Diaferio, A. 2001, *MNRAS*, 322, 901
 Spitzer, L. 1987, *Dynamical Evolution of Globular Clusters* (Princeton: Princeton Univ. Press)
 Syer, D., & White, S. D. M. 1998, *MNRAS*, 293, 337
 Weinberg, M. D. 2001a, *MNRAS*, 328, 311
 ———. 2001b, *MNRAS*, 328, 321
 White, S. D. M. 1984, *ApJ*, 286, 38
 Zel'dovich, Ya. B. 1970, *A&A*, 5, 84