Mon. Not. R. Astron. Soc. **000**, 1–16 (2003) Printed 9 June 2018 (MN LATEX style file v2.2)

Phase-Space Structure of Dark-Matter Haloes: Scale-Invariant PDF Driven by Substructure

Itai Arad[1], Avishai Dekel[2], Anatoly Klypin[3]
[1]*Institute of Astronomy, Madingley Road, Cambridge CB3 OHA, UK, arad@ast.cam.ac.uk*
[2]*Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel, dekel@phys.huji.ac.il*
[3]*Astronomy Department, New Mexico State University, Box 30001, Dept. 4500, Las Cruces, NM 88003, USA, aklypin@nmsu.edu*

9 June 2018

ABSTRACT

We present a method for computing the 6-dimensional coarse-grained phase-space density $f(\mathbf{x},\mathbf{v})$ in an N-body system, and derive its distribution function $v(f)$. The method is based on Delaunay tessellation, where $v(f)$ is obtained with an effective fixed smoothing window over a wide f range. The errors are estimated, and $v(f)$ is found to be insensitive to the sampling resolution or the simulation technique. We find that in gravitationally relaxed haloes built by hierarchical clustering, $v(f)$ is well approximated by a robust power law, $v(f) \propto f^{-2.5\pm0.05}$, over more than 4 decades in f, from its virial level to the numerical resolution limit. This is tested to be valid in the ΛCDM cosmology for haloes with masses $10^9 - 10^{15} M_\odot$, indicating insensitivity to the slope of the initial fluctuation power spectrum. By mapping the phase-space density in position space, we find that the high-f end of $v(f)$ is dominated by the "cold" subhaloes rather than the parent-halo central region and its global spherical profile. The value of f in subhaloes near the virial radius is typically > 100 times higher than its value at the halo centre, and it decreases gradually from outside in toward its value at the halo centre. This seems to reflect phase mixing due to mergers and tidal effects involving puffing up and heating. The phase-space density can thus provide a sensitive tool for studying the evolution of subhaloes during the hierarchical buildup of haloes. It remains to be understood why the evolved substructure adds up to the actual universal power law of $v(f) \propto f^{-2.5}$. It seems that this behaviour results from the hierarchical clustering process and is not a general result of violent relaxation.

Key words: cosmology: theory — dark matter — galaxies: dwarfs — galaxies: formation — galaxies: haloes — galaxies: interactions — galaxies: kinematics and dynamics

1 INTRODUCTION

The standard paradigm assumes that dark-matter haloes are the basic entities in which luminous galaxies form and live. The haloes dominate the gravitational potential over a wide range of radii and they have a crucial role in determining the galaxy properties. While many of the systematic features of halo structure and kinematics have been revealed by N-body simulations, the origin of these features is still an open theoretical puzzle, despite the fact that they are due to simple Newtonian gravity.

The halo *density profile* $\rho(r)$ is an example for such a puzzle. It is found in the simulations to have a robust non-power-law shape (which we refer to in general as "NFW"), with a log slope -3 near the virial radius, flattening gradually toward -1 at about 1% of the virial radius, and perhaps flattening further at smaller radii. (e.g. Navarro, Frenk & White 1997; Moore et al 1998; Klypin et al 2001; Power et al 2003; Hayashi et al 2004, and references therein). This density profile is insensitive (or at most weakly sensitive) to parameters of the cosmological model and the initial fluctuation power spectrum (e.g., Cole & Lacey 1996; Navarro, Frenk & White 1997; Subramanian, Cen & Ostriker 2000; Ricotti 2003; Colín et al 2003; Navarro et al 2004) indicating that its origin is due to a robust relaxation process rather than specific initial conditions. In particular, violent relaxation (Lynden-Bell 1967) may be involved in shaping up the density profile. In addition, secondary infall may be important in the outer regions (Gunn and Gott 1972; Dekel, Kowitt & Shaham 1981; Fillmore and Goldreich 1984; Hoffman & Shaham 1985; White & Zaritsky 1992) while some argue that resonances may have a role in the inner regions (Weinberg & Katz 2002; but see Valenzuela & Klypin 2003). Nevertheless, there is no clear understanding for why the haloes actually pick up the particular density profile they have.

The properties of the *velocity dispersion* tensor is another theoretical puzzle. The haloes tend to be rather round, with a velocity dispersion profile that is slightly rising at small radii and slightly falling at large radii but is rather flat overall (Huss et al 1999a,b).

The profile of the anisotropy parameter $\beta(r)$, which is a measure of radial versus tangential velocities, indicates near isotropy at small radii, which gradually develops into more radial orbits at large radii (Cole & Lacey 1996; Huss et al 1999a,b; Colín et al 2000a). For a spherical system in equilibrium, the $\sigma(r)$ and $\beta(r)$ are related to $\rho(r)$ via the Jeans equation, but it is not at all clear why the haloes in the simulations choose the characteristic $\sigma(r)$ or $\beta(r)$ that they actually do.

An interesting attempt to address the origin of the halo profile has been made by Taylor & Navarro (2001), who measured a poor-man phase-space density profile by $f_{\rm TN}(r) = \rho(r)/\sigma(r)^3$, and found that it displays an approximate power-law behaviour, $f_{\rm TN} \propto r^{-1.87}$, over more than two decades in r. Using the spherisymmetric Jeans equation, they showed that this power law permits a whole family of density profiles, whose limiting case is a profile similar to NFW, which asymptotically approaches a slope -0.75 as $r \to 0$. The general power-law shape of $f_{\rm TN}(r)$ is confirmed in the simulated haloes described below. This scale-free behaviour of $f_{\rm TN}(r)$ is intriguing, and it motivates further studies of halo structure by means of phase-space density.

Simulations of the ΛCDM cosmology also reveal a roughly self-similar *hierarchical clustering* process, where smaller building blocks accrete and merge into bigger haloes. At every moment, every halo contains a substructure of subhaloes on top of a smooth halo component that has been tidally stripped from an earlier generation of substructure. Some of the important dynamical processes involved in this hierarchical halo buildup are understood qualitatively. These include, for example, the dynamical friction which governs the decay of the satellite orbits, the tidal stripping of subhaloes due to the host halo potential, and the mergers and flyby interactions of the subhaloes among themselves. However, a complete theoretical understanding of how these processes work in detail, and how they combine to produce the halo structure and kinematics, is lacking.

Attempts have been made to explain an inner density cusp using toy models of dynamical stripping and tidal effects during the halo buildup by mergers (Syer & White 1998; Nusser & Sheth 1999; Dekel, Devor & Hetzroni 2003; Dekel et al. 2003). However, a similar halo density profile seems to be produced also in some of the simulations where substructure has been artificially suppressed (Moore et al. 1999b; Huss et al 1999a; Avila-Reese et al. 2001; Bullock, Kravtsov & Colin 2002; Alvarez, Shapiro & Martel 2002), indicating that the process responsible for the origin of this density profile might be a somewhat more general feature of gravity and not unique to the merger scenario.

The issue of halo substructure has become timely both because of its relevance to observations and its implications on other major issues in galaxy formation. Tidal tails and streams associated with dwarf satellite galaxies have been observed in the haloes of the Milky Way and M31 (Ibata et al 1994, 2001a,b), and they start to allow detailed modelling of the halo history through the satellite orbits. Gravitational-lensing observations provide preliminary indications for the presence of substructure in haloes at the high level predicted by the dissipationless ΛCDM scenario (e.g. Dalal & Kochanek 2002) In contrast, the observed number density of dwarf galaxies seems to be significantly lower, thus posing a "missing dwarf problem" Klypin et al. (1999b); Moore et al (1999a). Also, the "angular-momentum problem" of disk galaxies (e.g. Navarro & Steinmetz 2000; Bullock et al. 2001b) is probably associated with the evolution of substructure in haloes (Maller & Dekel 2002; Maller, Dekel & Somerville 2002). While the dwarf and angular-momentum problems necessarily involve baryonic processes, understanding the gravitational evolution of substructure is clearly a key for solving them.

In order to better understand the origin the various aspects of halo structure and buildup mentioned above, we make here a first attempt at addressing directly and in some detail the *phase-space* structure of dark-matter haloes. The fundamental quantity in the dynamical evolution of gravitating systems is the full, six-dimensional, coarse-grained, phase-space density $f(\mathbf{x},\mathbf{v})$, which intimately relates to the underlying Vlasov equation and lies behind any relaxation process that may give rise to the virialized halo structure (Binney & Tremaine 1987, chapter 4).

Ideally, one would have liked to compute it free of assumptions concerning spherical symmetry, isotropy, or any kind of equilibrium. However, computing densities in a six-dimensional space is a non-trivial challenge which requires simulations of a very broad dynamical range. The state-of-the-art N-body simulations, with more than million particles per halo, allow an attempt of this sort for the first time. We describe below a successful algorithm for measuring $f(\mathbf{x},\mathbf{v})$, and study its relevant properties including the associated systematic and random uncertainties. We then apply this algorithm to simulated virialized haloes in the ΛCDM cosmology.

We report in this paper two surprising new results. First, we discover that the volume distribution function of the phase-space density, $v(f)$, displays a universal scale-invariant *power-law* shape, valid in all virialized haloes that form by hierarchical clustering. Second, we realise that this power law is not directly related to the overall density profile, but is rather driven by the halo *substructure*. This implies that the phase-space density provides a useful tool for studying the hierarchical buildup of dark-matter haloes and the evolution of substructure in them.

In §2 we introduce $f(\mathbf{x},\mathbf{v})$ and $v(f)$. In §3 we describe the method of computing f and $v(f)$ from an N-body halo, and summarise its properties and the associated errors, which are addressed in more detail in Appendix A. In §4 we present the universal power-law shape of $v(f)$ based on several different simulations, and demonstrate its robustness to the mass scale and simulation technique. In §5 we display maps of phase-space density and show that the high-f contributions to $v(f)$ come from substructures within the parent halo. In §6 we summarise and discuss our results and future work.

2 THE PDF OF PHASE-SPACE DENSITY: $v(f)$

This is a more technical and elaborate introductory section, aimed at introducing the concepts and nomenclature relevant for the analysis of this paper.

2.1 Definitions

The state of a collisionless system is completely determined by the fine-grained phase-space density function $f(\mathbf{x},\mathbf{v},t)$, which measures the mass contained in an infinitesimal phase-space patch of volume $d\mathbf{x}d\mathbf{v}$, located at $(\mathbf{x},\mathbf{v})$. The evolution of $f(\mathbf{x},\mathbf{v},t)$ is governed by the Vlasov equation,

$$\partial_t f + \mathbf{v}\cdot\nabla_{\mathbf{x}} f - \nabla_{\mathbf{x}}\Phi\cdot\nabla_{\mathbf{v}} f = 0\ , \qquad (1)$$

with $\Phi(\mathbf{x})$ the gravitational potential, related self-consistently to $f(\mathbf{x},\mathbf{v})$ by the Poisson's integral

$$\Phi(\mathbf{x}) = -G\int d\mathbf{x}'d\mathbf{v}\,\frac{f(\mathbf{x}',\mathbf{v})}{|\mathbf{x}-\mathbf{x}'|}\ . \qquad (2)$$

It is therefore sensible to assume that a true understanding of the nature of self-gravitating collisionless systems must involve $f(\mathbf{x},\mathbf{v})$ as a primary ingredient. However, being a function of six variables, $f(\mathbf{x},\mathbf{v})$ is hard to deal with computationally. It turns out that there is a much simpler function, $v(f)$, which is intimately related to $f(\mathbf{x},\mathbf{v})$, but is much easier to handle both analytically and numerically.

Before defining $v(f)$, it is worth recalling that the Vlasov equation preserves the phase-space density along particle trajectories: It implies eq. (1) implies that

$$\frac{d}{dt}f[\mathbf{x}(t),\mathbf{v}(t)] = 0\ , \tag{3}$$

where $[\mathbf{x}(t),\mathbf{v}(t)]$ is the trajectory of a phase-space element. This conservation of phase-space density along trajectories can also be viewed as a conservation of phase-space volume. In order to illustrate what this means, consider any smooth function $C(x)$, and define the integral

$$I(t) \equiv \int d\mathbf{x}d\mathbf{v}\, C[f(\mathbf{x},\mathbf{v},t)]\ . \tag{4}$$

Using the Vlasov equation we find that

$$\frac{dI}{dt} = \int d\mathbf{x}d\mathbf{v}\, C'(f)\partial_t f \tag{5}$$

$$= -\int d\mathbf{x}d\mathbf{v}\, C'(f)\big[\mathbf{v}\cdot\nabla_{\mathbf{x}}f - \nabla_{\mathbf{x}}\Phi\cdot\nabla_{\mathbf{v}}f\big] \tag{6}$$

$$= -\int d\mathbf{x}d\mathbf{v}\, \big[\mathbf{v}\cdot\nabla_{\mathbf{x}}C(f) - \nabla_{\mathbf{x}}\Phi\cdot\nabla_{\mathbf{v}}C(f)\big] = 0\ . \tag{7}$$

The last equality follows by integrating the first term over $\mathbf{x}$ and the second term over $\mathbf{v}$, and assuming that $f\to 0$ as $\mathbf{x},\mathbf{v}\to\infty$. This implies that I is conserved under the dynamics.

Now, if $C(f)$ is the Dirac delta function[1] $\delta(f-f_0)$, then the integral I becomes an f-dependent function, which we refer to as the *PDF* $v(f)$:

$$v(f=f_0) \equiv \int d\mathbf{x}d\mathbf{v}\,\delta\big[f(\mathbf{x},\mathbf{v},t)-f_0\big]\ . \tag{8}$$

From this definition it is clear that $v(f)df$ is the volume of phase-space occupied by phase-space elements whose density lies in the range $(f, f+df)$. The conservation of $v(f)$ implied by Eq. (7) means that this volume is conserved under the dynamics. The fine-grained $v(f)$ [i.e., the $v(f)$ which is calculated using the fine-grained $f(\mathbf{x},\mathbf{v})$] should therefore be viewed as a signature of the system, which remains the same throughout its evolution.

2.2 Coarse-Grained Density and Mixing

One might have thought that the conservation of $v(f)$ poses severe constrains on the evolution of $f(\mathbf{x},\mathbf{v},t)$, but this is not the case due to the effects of *mixing*. In the course of evolution, phase-space patches of high f are stretched and spiral into regions with low f. As the stretching continues, the patches become thiner and thiner, and as a result $f(\mathbf{x},\mathbf{v})$ varies over increasingly smaller scales. Very soon, one can no longer measure $f(\mathbf{x},\mathbf{v})$, but instead measure an average of it over some finite volume. This average is often referred to as the "coarse-grained" phase-space density, as opposed to the original "fine-grained" density $f(\mathbf{x},\mathbf{v})$. We denote the coarse-grained quantities by $\bar{f}$ and $\bar{v}(\bar{f})$, but in subsequent sections we may omit the bar and simply refer to them as f and $v(f)$. The important point to realise is that the coarse-grained $\bar{v}(\bar{f})$ is no longer conserved. For example, given initially a volume V filled with phase-space density f_A and a similar volume filled with density f_B, one ends after mixing with a volume $2V$ filled with an average density $(f_A+f_B)/2$.

When a large fraction of the mass is added to halo, e.g. by collapse or a major merger, rapid global fluctuations of the gravitational potential re-distribute the energies of each phase-space element, and lead to very strong mixing across large scales, resulting in variations in $\bar{f}$. After a few global dynamical times, the potential fluctuations fade away and $\bar{f}$ stabilises. The fine-grained f, on the other hand, continues to spiral and stretch indefinitely, on smaller and smaller scales, which no longer affect $\bar{f}$. After a while, the $\bar{f}$ can be viewed as the "physical" phase-space density of the system, since the microscopic fluctuations of f can no longer be measured, and, in particular, no longer affect the gravitational potential. It is therefore the coarse-grained $\bar{f}$ which, according to Jeans' theorem, can be written as a function of he invariants, the energy E and the angular momentum $\mathbf{L}$. This relaxation process is referred to as *violent relaxation* (Lynden-Bell 1967).

What is the coarse-grained $\bar{v}(\bar{f})$ of a relaxed system, and how is it related to the constant fine-grained $v(f)$ of that system? One may identify the fine-grained $v(f)$ with the coarse-grained $\bar{v}(\bar{f})$ at an early time when the initial f is smooth enough not to vary over scales smaller than the averaging scale associated with $\bar{f}$. Then, the question can be rephrased in terms of the relation between the final $\bar{v}(\bar{f})$ and the early $\bar{v}(\bar{f})$. One obvious constraint on the final $\bar{f}$, as an average of f, is that their maximum values should obey $\bar{f}_{\rm max}\leqslant f_{\rm max}$. In particular, if the early $\bar{v}(\bar{f})$ [namely $v(f)$] vanishes for $\bar{f}>f_{\rm max}$, then so does the final $\bar{v}(\bar{f})$.

The *mixing theorem* (Tremaine, Hénon & Lynden-Bell 1986) specifies several additional constraints on the final $\bar{v}(\bar{f})$, which arise from the fact that it has originated from a given $v(f)$. The *cumulative* distribution function is defined by

$$V(f_0) \equiv \int d\mathbf{x}\mathbf{v}\,\theta\big[f(\mathbf{x},\mathbf{v})-f_0\big] = \int_{f_0}^{\infty} v(f')\,df'\ , \tag{9}$$

where $\theta(\cdot)$ is the Heaviside step function. As a monotonic function, $V(f)$ can be inverted to yield $f(V)$, the *reduced* distribution function. The associated cumulative mass function $M(f)$, which measures the mass in phase-space regions where $f(\mathbf{x},\mathbf{v})<f_0$, is then defined by

$$M(f_0) \equiv \int d\mathbf{x}d\mathbf{v}\, f(\mathbf{x},\mathbf{v})\,\theta\big[f(\mathbf{x},\mathbf{v})-f_0\big] \tag{10}$$

$$= \int_{f_0}^{\infty} f'\,v(f')\,df'\ . \tag{11}$$

Substituting $f=f(V)$ inside $M(f)$, one obtains the function $M(V)$, which can also be written as

$$M(V) = \int_0^V f(V')\,dV'\ . \tag{12}$$

If $M(V)$ and $M'(V)$ denote the functions arising from the density functions $f(\mathbf{x},\mathbf{v})$ and $f'(\mathbf{x},\mathbf{v})$ respectively, then the mixing theorem states that the final (coarse-grained) $\bar{f}'$ is an average of the initial (coarse- or fine-grained) f if and only if $M'(V)\leqslant M(V)$ for every V. This, in turn, implies an implicit constraint on the evolution of the coarse-grained $\bar{v}(\bar{f})$. This is one of the very few rigorous

[1] Strictly speaking, the Dirac delta function is not a smooth function, but it can be approximated by a series of smooth functions, each obeying Eq. (7)

results concerning the evolution of $\bar{v}(\bar{f})$, but its strength is rather limited because it provides only an integro-differential inequality that constrains $\bar{v}(\bar{f})$.

The mixing theorem can be also stated directly in terms of $v(f)$, as shown by Mathur (1988). In his formulation, a necessary condition for the final (coarse-grained) $\bar{v}(\bar{f})$ evolved from the initial $v(f)$ is that[2]

$$\bar{v}(f) = v(f) + \frac{d^2}{df^2}P(f) \ , \tag{13}$$

with $P(f)$ being some continuous function (not to be confused with a probability distribution) such that $P(f) \leqslant 0$ for all f and $P(f) \to 0$ for $f \to \pm\infty$.

In the CDM scenario, the dark matter is initially cold – with vanishing velocity dispersion. This means a one-to-one correspondence between the particle positions and their velocities, $\mathbf{v} = \mathbf{v}(\mathbf{x})$, and the initial (fine-grained) phase-space density can be written as

$$f_0(\mathbf{x},\mathbf{v}) = \rho(\mathbf{x})\delta^3[\mathbf{v}-\mathbf{v}(\mathbf{x})] \ . \tag{14}$$

Thus, in the beginning, the system can be described by the evolution of its density and velocity fields, often approximated by the Zel'dovich approximation. This singular nature of the initial phase-space density, together with its conservation, plays a crucial role in the formation of structures, and in particular in the abundance of dwarf haloes, and in the formation of cusps in dark-matter haloes (e.g. Colín et al 2000b).

Physically, the phase-space density is never really infinite, as it is defined by a finite number of dark-matter particles per given phase-space volume. Numerically, it is the finite mass-resolution of the N-body simulation that prohibits the phase-space density from being infinite. Therefore, the initial phase-space density is expected to be extremely high in regions where $\mathbf{v} \simeq \mathbf{v}(\mathbf{x})$ and vanishingly small elsewhere. Accordingly, $v(f)$ will have contributions from a narrow range of extremely high values of f and from a narrow range near $f = 0$. As the system evolves and mixing takes place, phase-space densities from both regions mix and give rise to intermediate values of f, widening the distribution $v(f)$ under the constraints of the mixing theorem.

2.3 A Relation Between $v(f)$ and $\rho(r)$

In general, different systems may have the same $v(f)$. However, if the system is spherically symmetric and stationary, such that f is a function of the energy alone, $f(\mathbf{x},\mathbf{v}) = f(\epsilon)$, then there is a unique relation between $v(f)$, $f(\epsilon)$ and $\rho(r)$. We shall see in §4 that the $v(f)$ measured in N-body haloes is actually different from the $v(f)$ one would have predicted from the halo $\rho(r)$ using the relation assuming $f(\epsilon)$.

For a spherical system the relation between $\rho(r)$ and $v(f)$ can be derived as follows. Assume that $f(\mathbf{x},\mathbf{v}) = f[\epsilon(\mathbf{x},\mathbf{v})]$, where $\epsilon(\mathbf{x},\mathbf{v}) = v^2/2 + \Phi(r)$, and define the density-of-states function

$$g(\epsilon) \equiv \int d\mathbf{x}d\mathbf{v}\ \delta[\epsilon(\mathbf{x},\mathbf{v}) - \epsilon] \ . \tag{15}$$

The quantity $g(\epsilon)d\epsilon$ measures how much phase-space volume is occupied by phase-space elements with energy in the interval $(\epsilon,\ \epsilon + d\epsilon)$. When the system is spherical, $g(\epsilon)$ can be written as (Binney & Tremaine 1987, Eq. 4-157b),

$$g(\epsilon) = (4\pi)^2 \int_0^{r(\epsilon)} s^2 \sqrt{2\big[\epsilon - \Phi(s)\big]}\, ds \ , \tag{16}$$

with $r(\epsilon)$ the inverse function of the gravitational potential. The overall phase-space volume occupied by energy levels below ϵ is thus given by

$$V = \int_{-\infty}^{\epsilon} g(\epsilon')d\epsilon' \tag{17}$$

$$= \frac{2}{3}(4\pi)^2 \int_0^{r(\epsilon)} s^2 \left\{2\big[\epsilon - \Phi(s)\big]\right\}^{3/2} ds \ . \tag{18}$$

Therefore, if $V(f)$ is the cumulative $v(f)$ defined in Eq. (9), and $f(\mathbf{x},\mathbf{v})$ is a function of the energy alone, we obtain an integro-differential equation for $f(\epsilon)$:

$$V[f(\epsilon)] = \frac{2}{3}(4\pi)^2 \int_0^{r(\epsilon)} s^2 \left\{2\big[\epsilon - \Phi(s)\big]\right\}^{3/2} ds \ . \tag{19}$$

This has to be supplemented by the equation that connects $\Phi(r)$ to $f(\epsilon)$ (Binney & Tremaine 1987, eq. 4-104),

$$-\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\Phi}{dr}\right) = 4\pi G\rho(r) \tag{20}$$

$$= (4\pi)^2 G \int_{\Phi(r)}^{0} f(\epsilon)\sqrt{2(\Phi-\epsilon)}\, d\epsilon \ . \tag{21}$$

In general, these equations can be solved numerically, e.g., using a Picard iteration scheme.

In the asymptotic regime, the above equations can be solved analytically. If $\rho(r) \propto r^{-\alpha}$ as $r \to 0$, then, for $\epsilon \to \Phi(0)$, one can show that $f(\epsilon)$ and $g(\epsilon)$ have the scale-invariant forms

$$f(\epsilon) \propto \big[\epsilon - \phi(0)\big]^{-\frac{6-\alpha}{2(2-\alpha)}} \ , \tag{22}$$

$$g(\epsilon) \propto \big[\epsilon - \phi(0)\big]^{\frac{8-\alpha}{2(2-\alpha)}} \ . \tag{23}$$

Note that the derivative of Eq. (19) with respect to ϵ yields

$$v[f(\epsilon)] = \frac{g(\epsilon)}{f'(\epsilon)} \ . \tag{24}$$

By plugging the scale-invariant solutions of eq. (23) into eq. (24), one finds that for $f \to \infty$, the PDF is also a power law, $v(f) \propto f^{-\beta}$, with

$$\beta = \frac{18-4\alpha}{6-\alpha} \quad \text{or} \quad \alpha = \frac{18-6\beta}{4-\beta} \ . \tag{25}$$

We learn that the density slopes in the range $0 \leqslant \alpha \leqslant 2$, relevant for the inner regions of haloes where f is high, correspond to a narrow range of β values, $3 \geqslant \beta \geqslant 2.5$. In this case of power-law profiles, $\beta = 2.5$ corresponds to the singular isothermal sphere $\alpha = 2$, $\beta = 2.8$ corresponds to an $\alpha = 1$ cusp, and $\beta = 3$ corresponds to a flat core, $\alpha = 0$.

Eq. (25) can also be obtained from simpler dimensional arguments using the virial theorem.

3 MEASURING $v(f)$ IN AN N-BODY SYSTEM

We wish to measure the phase-space density $f(\mathbf{x},\mathbf{v})$ of a system represented by N particles of mass m each. One straightforward approach would have been to divide phase-space into a large number of Cartesian cells of equal volume V each. If cell j contains

[2] This is the original statement in Mathur (1988). However, we believe that by adding the (trivial) condition that $\bar{v}(f) \geqslant 0$ for every f, it becomes also a sufficient condition.

N_j particles, then the average phase-space density in it can be estimated as $\bar{f}_j = N_j m/V$. However, this approach is impractical in a six-dimensional space. A moderate resolution of 100 bins along each axis would require 10^{12} cells, which is beyond the capacity of any present-day computer. Moreover, for the computed $\bar{f}$ to have any statistical meaning, there should be at least 100 particles in each cell (for a relative Poisson error of $\sim 10\%$), which requires a total of 10^{14} particles – about 5 orders of magnitude beyond the number of particles in today's largest simulations.

A possible way to overcome this problem is by using an *adaptive* grid, where the cells vary in size to properly allow high resolution in high-density regions and low resolution in low-density regions. One can simply create an adaptive grid by dividing each Cartesian cell which contains more particles than some prescribed threshold N^* into two or more sub-cells. This division can be done recursively until all cells contain N^* particles or less.

Even more effective would be to vary the shape of the cells as well, allowing them to adapt more efficiently to the geometry of the underlying distribution and thus increase the effective resolution. A particularly robust method of this type is the *Delaunay Tessellation Field Estimator* (DTFE), which has already been implemented in three dimensions in a cosmological context (e.g. Bernardeau & van de Weygaert 1996; Schaap & van de Weygaert 2000, and references therein). We use this method here to measure f in six dimensions.

3.1 Constructing a Delaunay Tessellation

A tessellation is the division of R^d space into a complete covering of mutually disjoint convex polygons. The Delaunay tessellation (Delaunay 1934) is defined for a sample of N points as follows. The *Delaunay cells* that construct the tessellation are the d-dimensional polyhedrons made by connecting every set of $d+1$ points whose circumsphere [the $(d-1)$-dimensional sphere that passes through all of them] does not encompass any other point from the sample. One immediate advantage of the Delaunay tessellation is that it is parameter-free, and it completely adapts itself to the underlying distribution of points. Schaap & van de Weygaert (2000) have demonstrated the superiority of the DTFE over more conventional methods for estimating the density in 3D (methods like cloud-in-cell or the smoothing kernel used in SPH simulations). One may assume that the same holds for the 6D case.

In 2D, the Delaunay cells are triangles, as illustrated in Fig. 1. In 3D, the Delaunay cells are tetrahedrons. In the six-dimensional phase space, the Delaunay cells are six-dimensional polyhedrons, each defined by 7 particles.

Figure 2 shows the Delaunay tessellation of an uneven distribution of points in the 2D plane. It demonstrates the obvious adaptive nature of the tessellation: regions with high density of particles are covered by small triangles, whereas regions with low density are covered with larger triangles.

Constructing the Delaunay tessellation of $N \sim 10^6$ particles in a six-dimensional space is not a straightforward task. Out of the many algorithms that exist in the literature (e.g., Su & Drysdale 1995, and references therein), we followed Bernardeau & van de Weygaert (1996) and van de Weygaert (1994) in picking the algorithm by Tanemura, Ogawa & Ogita (1983). We included some of the programming adjustment by van de Weygaert, e.g., the use of k-trees to speed up searching, and converted the code from three dimensions to six dimensions.

A full account of the algorithm, the numerical implementation and code performance, is provided elsewhere (Arad 2004, in preparation). Here we just mention briefly that for a halo of $\sim 10^6$ particles the C code runs for about a week on a common PC with CPU ∼2GHz and internal RAM ∼1GB. The resulting tessellation is made of $\sim 10^9$ Delaunay cells, where a typical particle is surrounded by $\sim 7,000$ cells involving ~ 200 neighbouring particles.

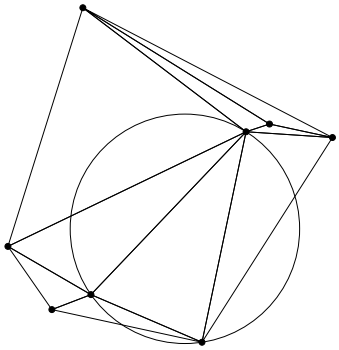

Figure 1. The Delaunay tessellation of 8 points in two dimensions. Shown for example is a circumcircle of one Delaunay cell, which, by construction, does not encompass any other point.

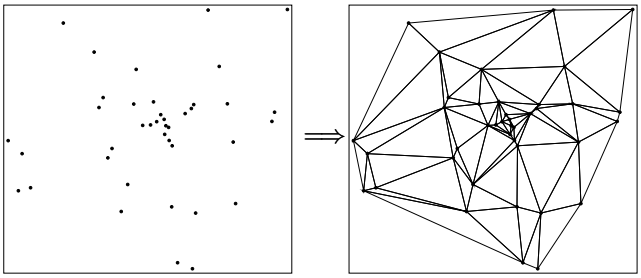

Figure 2. An illustration of the adaptive nature of the Delaunay tessellation. Left: an uneven distribution of points in 2D. Right: the resulting Delaunay tessellation.

3.2 Recovering f and $v(f)$ from the Tessellation

Once the Delaunay tessellation is constructed in phase space, we use it to estimate $f(\mathbf{x},\mathbf{v})$, generalising the method implemented in 3D by Schaap & van de Weygaert (2000). First we estimate f_i for each particle i. We define a *macro cell* by joining all its surrounding Delaunay cells:

$$W_i \equiv \bigcup_\nu D_{\nu_i} \ , \tag{26}$$

where $\{D_j\}$ is the set of all Delaunay cells and $\{D_{\nu_i}\}$ is the subset of cells which contain the particle i as one of their vertices. Figure 3 shows such a macro cell in 2D. The estimated phase-space density at point i should be inversely proportional to the volume of the macro cell, $1/|W_i|$. The proportionality factor must be greater than unity because the W_i cells of different particles partly overlap, and one can show that it should in fact be $d+1$ in order to preserve the total mass of the system. In phase-space we therefore define

$$f_i \equiv 7\frac{m}{|W_i|} \ . \tag{27}$$

In order to estimate f at any general point $(\mathbf{x},\mathbf{y})$, one might have used linear interpolation based on the 7 vertices of the corresponding Delaunay cell D_ν. This means expressing $f(\mathbf{x},\mathbf{v})$ as a linear function of the 6D vector $\omega = (\mathbf{x},\mathbf{v})$, namely $f(\omega) =$

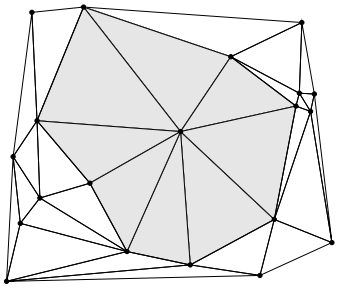

Figure 3. An illustration of a macro cell W_i (grey area), the union of the Delaunay cells that surround the particle i.

$A_\nu \cdot \omega + B_\nu$, with A_ν and B_ν a 6D vector and a scalar whose seven unknown values are found by equating $f(\mathbf{x},\mathbf{v})$ with f_i at the seven vertices of the cell. However, with 10^6 particles involving $\sim 10^9$ Delaunay cells the linear interpolation is CPU-expensive and practically impossible. Instead, we perform a zero-order interpolation where $f(\mathbf{x},\mathbf{v})$ is constant inside each Delaunay cell – the average of the f_i values at the seven vertices of D_ν:

$$f_\nu \equiv \frac{1}{7}\sum_{i_\nu} f_{i_\nu} \ . \tag{28}$$

This has the advantage that any integral of the form

$$I = \int d\mathbf{x}d\mathbf{v}\ \Psi\big[f(\mathbf{x},\mathbf{v})\big] \ , \tag{29}$$

with $\Psi(\cdot)$ some arbitrary function, can be easily estimated by the sum

$$I = \sum_\nu \Psi(f_\nu)|D_\nu| \ . \tag{30}$$

In particular, the desired $v(f)$ can be found by first computing its cumulative counterpart $V(f)$,

$$V(f_0) \quad = \quad \int_{f_0}^{\infty} v(f')\,df' = \int_{f(\mathbf{x},\mathbf{v})>f_0} d\mathbf{x}d\mathbf{v} \tag{31}$$

$$\rightarrow \quad \sum_{f_\nu > f_0} |D_\nu| \ , \tag{32}$$

and then taking its derivative,

$$v(f) = -\frac{dV(f)}{df} \ . \tag{33}$$

In what follows, we often denote the DTFE-measured f and $v(f)$ by $f_{\rm del}$ and $v_{\rm del}$ in order to distinguish them from the exact quantities.

3.3 Error Estimate

When we measure the $v(f)$ of a cosmological N-body system using the Delaunay tessellation, we should expect two types of errors. First are the errors in the underlying f associated with errors in the numerical simulation itself, such as errors due to two-body relaxation effects, force estimation, time integration and so on. A way to estimate these errors is by re-simulating the same system with different codes and with different sets of numerical parameters governing the mass resolution and the force resolution. A systematic testing of this sort will be reported in an associated paper (Arad, Dekel & Stoehr, in preparation). In the current paper, we make a preliminary comparison of several different haloes simulated with different resolutions and with different codes. We find that all the virialized haloes tested recover almost the same shape of $v(f)$.

The second type of errors originate from the fact that we try to estimate a smooth $f(\mathbf{x},\mathbf{v})$ from a finite set of particles using the Delaunay tessellation. Here we may encounter both statistical and systematic errors. Some of these errors would decrease as the number of particles is increased, whereas some are an inherent part of the DTFE independent of the mass resolution. In order to obtain a simple understanding of the nature of the DTFE errors, we first use a simple statistical model based on a Voronoi tessellation, which resembles the DTFE but lends itself more easily to analytical treatment. We then evaluate the DTFE errors empirically by measuring $f_{\rm del}(\mathbf{x},\mathbf{v})$ in synthetic systems were the particles represent a known $f(\mathbf{x},\mathbf{v})$, and compare the results to the error-model predictions. Since this error analysis is somewhat detached from the main focus of this paper, we describe it in detail in Appendix A. Our wisdom regarding the DTFE properties and uncertainties can be summarised in three points as follows:

(i) The measured $f_{\rm del}$ at each particle is chosen, to a good extent, from a probability distribution of the random variable $f_{\rm del}/f_{\rm true}$. In a typical realisation with 10^6 particles, the width of this distribution is about one decade, defining the range of fluctuation of $f_{\rm del}$ about $f_{\rm true}$. As $N\rightarrow\infty$, the shape of the distribution approaches an asymptotic limit $p(f_{\rm del}/f_{\rm true})$ with a *finite* width of about one quarter of a decade, as deduced from the DTFE of a Poisson distribution. Therefore, also in the infinite limit, the DTFE is expected to produce local fluctuations.

(ii) The measured $v_{\rm del}(f)$ can be viewed as a convolution of the true $v(f)$ and a *fixed window function*, which is related to the probability distribution $p(\cdot)$ [Eq. (A5)]:

$$v_{\rm del}(f=f_0) = \int_0^\infty v_{\rm true}(f) f_0^{-1} p(f_0/f)\,df \ . \tag{34}$$

For distributions where $v(f)$ is close to a power-law, the difference between $v_{\rm del}(f)$ and $v_{\rm true}(f)$ is negligible over a large range of scales, as demonstrated in Figs. A2 and A3. This is a very useful feature of the DTFE-measured $v(f)$.

(iii) The relative statistical error in $v_{\rm del}(f)$ is proportional to $1/\sqrt{N}$, and can be approximated by [Eq. (A18)]

$$\Delta(f) = c\left(\frac{m}{f\,\langle V_{\rm del}(f)\rangle}\right)^{1/2} \ , \tag{35}$$

with $V_{\rm del}(f) = \int_f^\infty v_{\rm del}(f')\,df'$ and $m = M/N$. The constant c is of order unity, and can be calibrated by comparing Eq. (35) to the actual error in lower-resolution measurements. In practise, this means that when $N\sim 10^6$ or more, the statistical error in $v_{\rm del}(f)$ is negligible for a very wide range of f. Moreover, in regions where there are large statistical errors, they are likely to be overwhelmed by systematic errors.

4 A UNIVERSAL SCALE-FREE $v(f)$

We have analysed the $v(f)$ of several different haloes, in three different mass ranges, simulated within the ΛCDM cosmology with two different N-body codes.

In order to calculate the $v(f)$ of a given halo, we find the halo centre using a simple max-density algorithm, and extract all particles which lie within a distance R from its centre. The max-density algorithm is based on an iterative counting in cells: at each iteration, space is divided into 8 equal cubic cells and the densest cell is

chosen for the following iteration. The iterative process stops when the densest cell contains no more than 500 particles, and the centre of that cell is defined as the centre of the halo. The radius R is typically set to be $\sim 10-20\%$ larger than the virial radius. We compromise on analysing $\sim 10^6$ particles, such that we explore a significant dynamical range while the computation can be completed in a few days.

To estimate the typical phase-space density at the outer regions of the halo, we define f_{vir} by

$$f_{vir} = \frac{1}{\pi^{3/2}} \frac{\bar{\rho}_{vir}}{\sigma^3_{vir}} , \tag{36}$$

which is the average space-space density that one would measure for a halo of constant density ρ_{vir} and a Maxwellian velocity distribution with a velocity dispersion σ_{vir}. For a given halo in a given cosmology, the mean virial quantities ρ_{vir} and σ_{vir} are defined using the virial theorem and the top-hat model relevant to that cosmology.

As one crude estimate of the upper limit on f for which the measurement of $v(f)$ is reliable, we evaluate in each halo the f value, $f_{20\%}$, below which the *statistical* error in $v(f)$ due to the DTFE procedure is below 20%. For that we use Eq. (35), calibrated by measurements with 10^5 particles. This statistical error is expected to be practically negligible in the range $f_{\rm vir} < f < f_{20\%}$.

4.1 N-body Simulations on Different Scales

The results described in this paper are based on three different cosmological simulations. Two of them used the Adaptive Refinement Tree (ART) code (Kravtsov, Klypin & Khokhlov 1997), and the third used the Tree Particle Mesh (TPM) code (Bode et al 2000; Bode & Ostriker 2003). In all the simulations, the assumed cosmological model is the standard ΛCDM with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$ and $h = 0.7$ today.

The ART simulations were done in periodic boxes of sizes $L = 1\,h^{-1}{\rm Mpc}$ and $L = 25\,h^{-1}{\rm Mpc}$, whereas the TPM simulation was done in a box of $L = 320\,h^{-1}{\rm Mpc}$. We denote these three simulations by $L1$, $L25$ and $L320$ respectively. Three haloes were analysed from each of these simulations, with masses corresponding to to dwarf galaxies ($10^9 - 10^{10} M_\odot$), normal galaxies ($\sim 10^{12} M_\odot$) and clusters ($\sim 10^{15} M_\odot$) respectively. Global properties of of the simulations and the haloes are given in table 4.1. The $v(f)$ curves for these haloes are shown in Fig. 4.

The L1 simulation is from Colín et al (2003). It is analysed at $z = 2.33$ in *physical* coordinates. The three haloes studied, labelled A, B, C, are the largest haloes in the snapshot. Their virial radii refer to a mean overdensity of $\Delta = 183$, as appropriate for the given cosmology and redshift. The L25 simulation is by Klypin et al (2001). The haloes are denoted B, C, D following the notation in the simulation paper. The L320 simulation is by Wambsganss et al (2004); Weller et al (2004). The three haloes studied, labelled A, B, C, are the most massive haloes in the simulation excluding haloes whose real density map shows an ongoing major merger.

In each of the analysed haloes, we find $v(f)$ to be well described by a power-law,

$$v(f) \propto f^{-2.50\pm0.05} , \tag{37}$$

over 3 to 5 decades in f. It is typically valid between slightly above $f_{\rm vir}$ and slightly below $f_{20\%}$. Outside this range, $v(f)$ gradually and systematically deviates downward from the power law. In the low-f regime the deviation is associated with departure from the virial regime, and the high-f deviation is consistent with being due to the limited mass resolution of the specific halo, as seen by the proximity to $f_{20\%}$ and as demonstrated in Appendix A. The high-f deviation from the power-law tends to occur at a smaller f value in L25, and even smaller in L320, due to the fact that $M_{\rm vir}$ is smaller respectively.

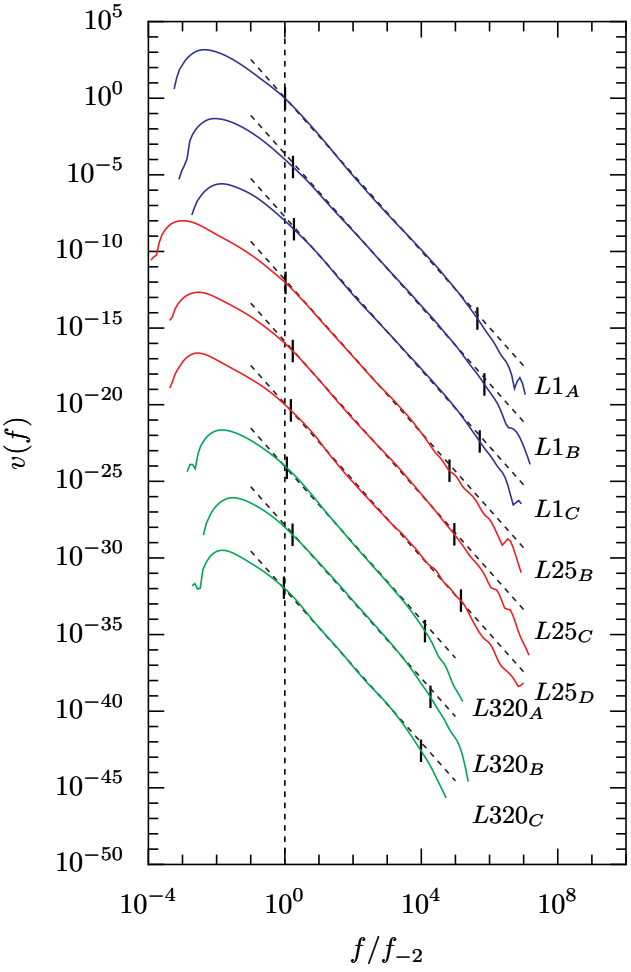

Figure 4. The volume distribution of phase-space density, $v(f)$, for each of the nine haloes analysed in this paper (see Table 4.1). The curves were shifted to coincide at $f = f_{-2}$, where the local log slope of $v(f)$ is -2, and were then shifted vertically by 4 decades relative to each other. A power-law line $v(f) \propto f^{-2.5}$ is shown on top of each curve. Marked on each curve are the virial level $f_{\rm vir}$ and the 20% statistical error limit $f_{20\%}$. The f range is shorter for the L320 haloes because they were sampled with less particles.

There is no evidence for a significant dependence of $v(f)$ on the halo mass. There may be a marginal trend for slight steepening of $v(f)$ as a function of mass, but only from steeper than $f^{-2.45}$ at $\sim 10^9 M_\odot$ to flatter than $f^{-2.55}$ at $\sim 10^{15} M_\odot$. This indicates relative insensitivity to the exact slope of the initial fluctuation power spectrum, which varies across the range from dwarf galaxies to clusters of galaxies. Additionally, the fact that we obtained essentially the same $v(f)$ from simulations using two different numerical codes, indicates that the shape of $v(f)$ is not an artifact of a particular simulation technique.

Table 1. Global properties of the 9 haloes analysed in this paper. The $v(f)$ of each halo was calculated from all the $N_{\rm cut}$ particles that lie within a radius $R_{\rm cut}$. $R_{\rm vir}$ and $M_{\rm vir}$ are the the virial radius and mass. C is the concentration of the halo, calculated from an NFW fit. f_2 is the f level where the logarithmic slope of $v(f)$ is -2. This value, together with $v(f_2)$, are used to scale the $v(f)$ curves of the different haloes in Fig. 4. Finally, "Code", z, $m_{\rm par}$, $r_{\rm force}$ and σ_8 describe the computer code, red-shift, mass-resolution, force-resolution, and normalisation of each simulation.

Halo	$N_{\rm cut}$	$R_{\rm cut}$ [kpc]	$R_{\rm vir}$ [kpc]	$M_{\rm vir}$ [$M_\odot$]	C	$f_2{}^a$	$v(f_2)^b$	Code	z	$m_{\rm par}$ [$M_\odot$]	$r_{\rm force}$ [kpc]	σ_8
$L1_A$	1.6×10^6	23	21	1.1×10^{10}	7.0	4.7×10^8	1.0×10^{-8}	ART	2.33	7.0×10^3	8.7×10^{-2}	0.75
$L1_B$	1.3×10^6	23	20	8.6×10^9	4.3	3.4×10^8	1.3×10^{-8}	ART	2.33	7.0×10^3	8.7×10^{-2}	0.75
$L1_C$	1.1×10^6	20	19	7.8×10^9	7.5	3.5×10^8	1.1×10^{-8}	ART	2.33	7.0×10^3	8.7×10^{-2}	0.75
$L25_B$	1.1×10^6	420	320	1.9×10^{12}	17.4	5.8×10^5	1.3×10^0	ART	0	1.2×10^6	1.4×10^{-1}	0.9
$L25_C$	1.2×10^6	420	330	2.0×10^{12}	12.8	3.4×10^5	4.0×10^0	ART	0	1.2×10^6	1.4×10^{-1}	0.9
$L25_D$	1.4×10^6	420	340	2.2×10^{12}	11.7	3.4×10^5	4.3×10^0	ART	0	1.2×10^6	1.4×10^{-1}	0.9
$L320_A$	4.6×10^5	3,000	2,700	1.1×10^{15}	6.28	8.8×10^2	3.4×10^8	TPM	0	2.6×10^9	4.7×10^0	0.95
$L320_B$	4.5×10^5	3,000	2,700	1.1×10^{15}	5.00	6.0×10^2	6.7×10^8	TPM	0	2.6×10^9	4.7×10^0	0.95
$L320_C$	4.6×10^5	3,000	2,700	1.1×10^{15}	6.43	1.0×10^3	2.4×10^8	TPM	0	2.6×10^9	4.7×10^0	0.95

[a] units: $M_\odot{\rm Mpc}^{-3}{\rm km}^{-3}s^3$
[b] units: $M_\odot^{-1}{\rm Mpc}^6{\rm km}^6s^{-3}$

5 SUBSTRUCTURES

5.1 Clumpiness in Phase-Space Density

Had $f(\mathbf{x},\mathbf{v})$ been a function of the energy alone, and the haloes were completely spherical and isotropic, the power-law $v(f)\propto f^{-2.5}$ would have implied via Eq. (25) that the real-space density profile must also be a power law, in fact an isothermal sphere $\rho(r)\propto r^{-2}$, at least over some finite range in r. This is clearly not the case, as the simulated haloes are well described by a universal average density profile whose local logarithmic slope is varying continuously from -3 at the outer parts to -1 or even flatter in the inner parts (§1). We conclude that f is far from being a function of energy alone, and in particular the system must deviate significantly from spherical symmetry or isotropy. This could be mostly due to the clumpy substructure of the halo, where the surviving subhaloes contribute volume of high phase-space density to $v(f)$, thus making it shallower than expected from a smooth system with an inner density slope flatter than -2.

In order to address the hypothesis that $v(f)$ is driven by substructure, we plot real-density and phase-space density maps of each halo in real-space slices. Fig. 5 and Fig. 6 show such maps for dwarf haloes B and C (all other haloes show a similar qualitative behaviour).

The real-space density ρ of each particle was calculated using a three-dimensional Voronoi tessellation (van de Weygaert 1994), generated using the free `qhull` software package. We chose this technique to estimate the real-space density because it is very similar in its adaptive nature to the Delaunay tessellation technique used to estimate the six-dimensional phase-space density. A brief account of the Voronoi tessellation technique is found in Appendix A, where it is used to estimate the errors in the DTFE method.

The maps were produced in the following way. For each halo, we consider all the particles within an equatorial slice parallel to the xy-plane, whose width is 40% of the virial radius. The slice is divided into $500\times500\times n_z$ equal cubic cells, with n_z set to have the cells cubic. The density (ρ or f) assigned to each cell is the average of the densities of all particle within it. ¿From each group of cells with the same x and y, we plot the one with the highest density.

We see that the real-space density maps are dominated by the familiar relatively smooth trend of density decreasing from the centre outward, with several tight clumps spread throughout the halo. The phase-space density maps, on the other hand, are qualitatively different. While the global trend with radius is much less apparent, the subhaloes become the highest peaks, especially in the outer regions of the halo. For example, the clumps with $f > 10^{12}M_\odot\,{\rm Mpc}^{-3}(km/s)^{-3}$ (yellow-reddish colours) are found everywhere. The very high peaks, with $f > 10^{13}M_\odot\,{\rm Mpc}^{-3}(km/s)^{-3}$ (bright yellow colours), are all found at a considerable distances from the centre. The central peak in f is quite modest in comparison; the elongated structures near the centre of dwarf-halo B are most likely subhaloes in the process of merging.

Fig. 7 highlights the same effect by showing ρ and f associated with a random subset of the N-body particles as a function of their distance r from the halo centre. A large portion of these particles follow the global trend of decreasing density with radius – they could be associated with a smooth-background component, for which f is approximately a function of energy alone. At radii $r > 1\,{\rm kpc}$, the high-f values come in "spikes" corresponding to the subhaloes. While the spikes in ρ reach values comparable to the central peak, the spikes in f could be more than 100 times higher, indicating that the subhaloes are both *compact* and *cold*. We note in dwarf halo B, for example, that all the points where $f\gtrsim10^{12}M_\odot\,{\rm Mpc}^{-3}{\rm km}^{-3}{\rm s}^3$ are in subhaloes.

The other interesting feature of the spikes is the fact that they seem to get lower and broader as they get closer to the halo centre, and that beyond a certain radius (of about $2\,{\rm kpc}$), the spikes completely blend into the smooth background. This indicates that the subhaloes phase-mix and lose their high phase-space densities as they approach the halo centre. This seems to be the natural result of mergers and *tidal effects*, which both puff up the subhaloes and heat them up.

5.2 Toy Model: Adding Up Small Haloes

As a first attempt at trying to understand how the power law $v(f)\propto f^{-2.5}$ in a halo of mass M may originate from its substructure, we

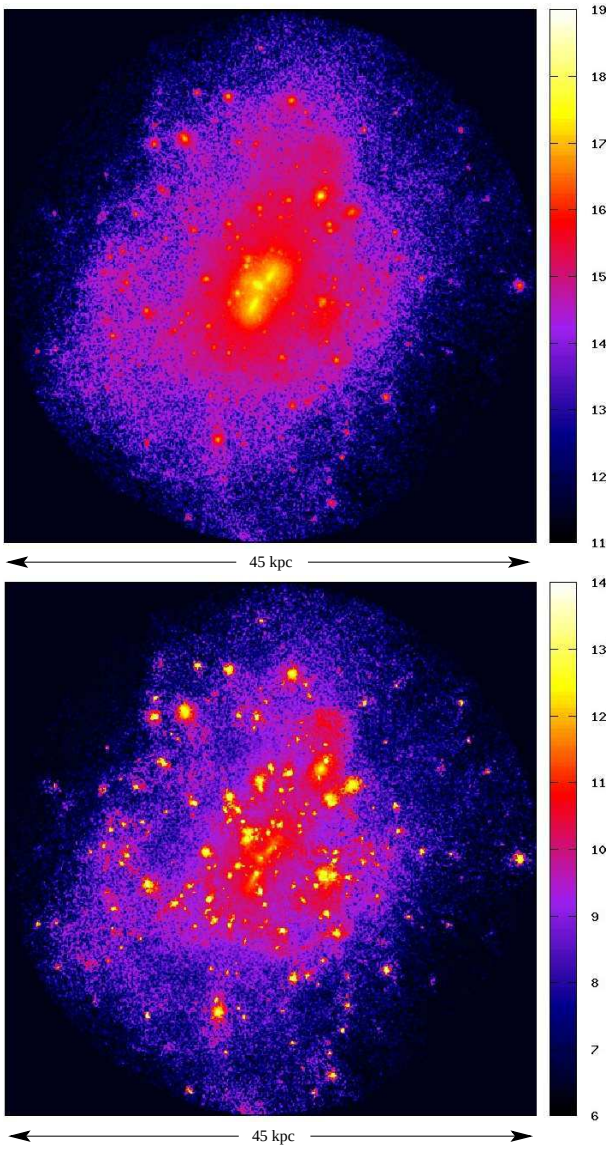

Figure 5. Density maps of dwarf halo $L1_B$ in a slice of thickness $0.4R_{\rm vir}$. Top: real-space density. Bottom: phase-space density. The units in the colour key are $\log(\rho/[M_\odot Mpc^{-3}])$ and $\log(f/[M_\odot Mpc^{-3}{\rm km}^{-3}s^3])$ respectively. The very-high f values are found inside clumps which are typically far away from the halo centre.

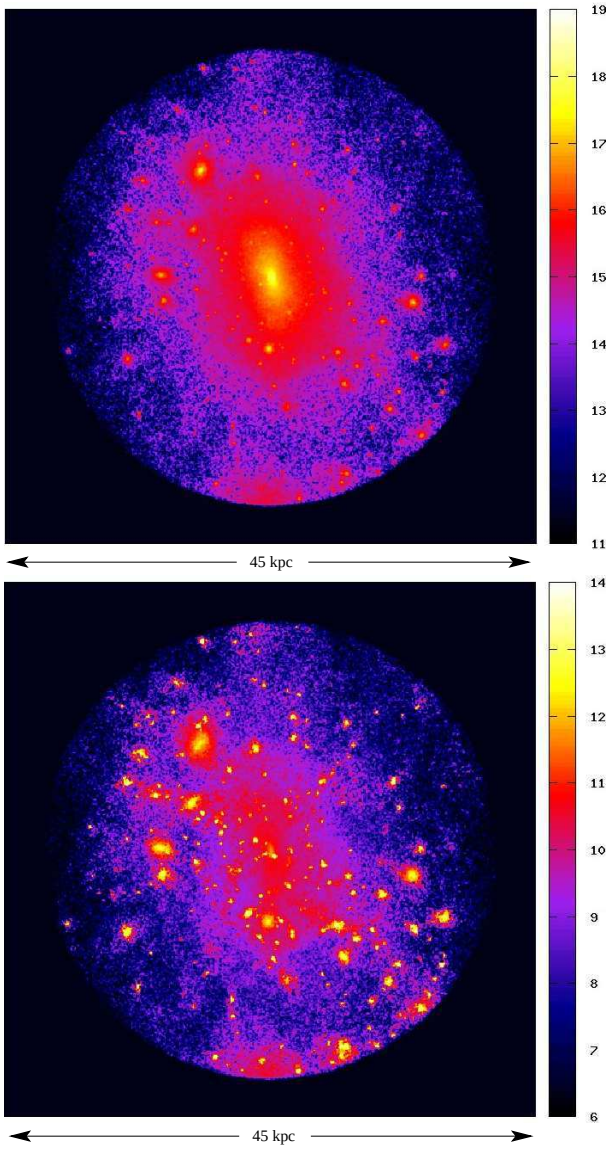

Figure 6. Density maps of dwarf halo $L1_C$. See Fig. 5.

simply add up the typical contributions from the general population of haloes of different masses m smaller than M, as predicted in the ΛCDM cosmology. Based on cosmological N-body simulations (Moore et al 1999a; Ghigna el al 2000; De Lucia et al. 2004), and in accordance with the Press-Schechter approximation (Press & Schechter 1974) and its extensions (e.g., Lacey & Cole 1993; Sheth 2003), the mass function of small-mass haloes (not necessarily subhaloes) can be approximated by $dn/dm \propto m^{-\gamma}$, with $\gamma \simeq 1.8-2.0$. Additionally, we assume that the average density profiles of haloes of different masses have the same functional form and are simply scaled versions of each other. This is established by n-body simulations for the case of *isolated* haloes, but is less clear when one considers subhaloes (e.g., Hayashi et at 2003). Nevertheless, we adopt this assumption as our first crude toy model.

As a first approximation we assume that the haloes all form at the same time in an Einstein-deSitter cosmology, so they have the same characteristic real-space density $\rho_m = \rho$ (a constant factor times the universal density), and therefore their typical radii scale like $r_m \propto m^{1/3}$. Based on the virial theorem, the velocity dispersions then scale like $\sigma_m \propto m^{1/3}$. Therefore, the typical phase-space volume of a halo of mass m scales like $V_m \propto r_m^3\sigma_m^3 \propto m^2$, and its typical phase-space density is $f_m \propto m/V_m \propto m^{-1}$. If we denote by $\tilde{v}(f)$ the universal, un-scaled, normalised, dimensionless probability distribution function relevant for all the haloes, then the volume distribution function $v_m(f)$ of a halo of mass m, defined such that $v_m(f)df$ is a volume, is given by

$$v_m(f) = \frac{V_m}{f_m}\tilde{v}\Big(\frac{f}{f_m}\Big) = m^3\tilde{v}(mf)\ . \tag{38}$$

Then the total contribution to $v(f)$ of halo M from the population of smaller haloes $m < \mu M$ is

$$v(f) = \int_0^{\mu M}\frac{dn}{dm}v_m(f)\,dm = \int_0^{\mu M} m^{3-\lambda}\tilde{v}(mf)\,dm\ , \tag{39}$$

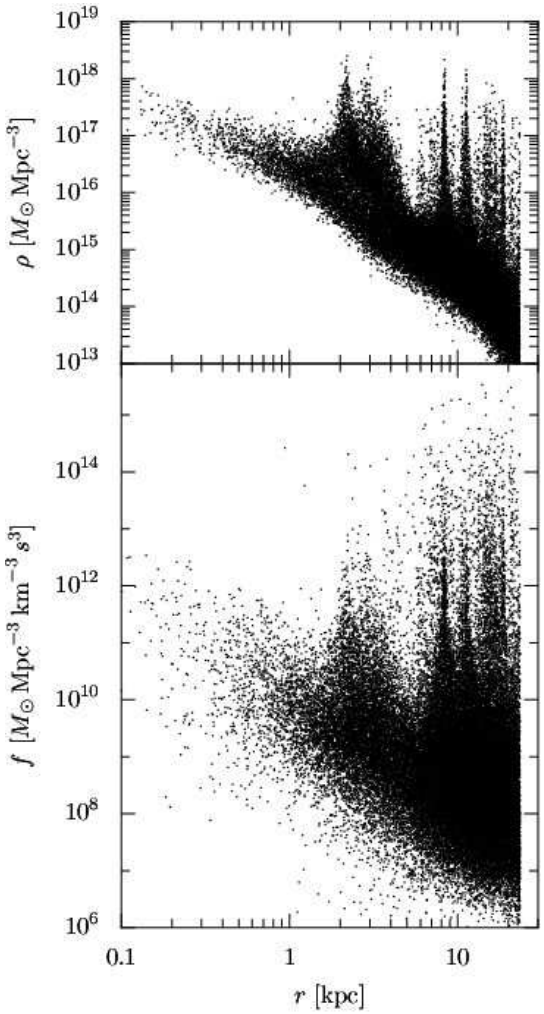

Figure 7. Densities as a function of radius for dwarf halo $L1_B$, using a random set of 4% of the particles. Top: real-space density. Bottom: phase-space density. The background particles define a general trend of decreasing density with radius, while the spikes correspond to subhaloes. The phase-space density spikes are higher than the central peak because the subhaloes are cold. They become shorter and broader at smaller radii, indicating heating and puffing up by tidal effects and mergers.

where μM is the mass of the largest subhalo ($\mu < 0.5$). Changing variables $m \to mf$, we finally get

$$v(f) = f^{-(4-\gamma)} \int_0^{\mu M f} s^{3-\gamma} \tilde{v}(s)\, ds\ . \tag{40}$$

Thus, $v(f)$ is a multiple of a power-law $f^{-(4-\gamma)}$ and a monotonically increasing function of f; it therefore has to be shallower than the power law $f^{-(4-\gamma)}$. For $\gamma \geqslant 1.8$, this means that $v(f)$ is shallower than $f^{-2.2}$, which is significantly shallower than the measured $f^{-2.5\pm0.05}$. This idealised toy model does not seem to work.

This calculation can be generalised to the case where haloes of smaller masses form first, as implied from the slope of the initial fluctuation power spectrum P_k. For an Einstein-deSitter cosmology, the formation time scales with m such that $\rho_m \propto m^{-\nu}$, with $\nu = (n+3)/2$, and n is the local power index, $n = d\ln P_k/d\ln k$, at the relevant effective scale for masses $< \mu M$. Then, in analogy to the calculation in the previous paragraph, we have $\sigma_m \propto m^{(2-\nu)/6}$, $V_m \propto m^{(4+\nu)/2}$, $f_m \propto m^{-(2+\nu)/2}$, and we finally obtain that $v(f)$ must be flatter than the power law

$$v(f) \propto f^{-2-2(2-\gamma)/(2+\nu)}\ . \tag{41}$$

For $n = -3$, namely $\nu = 0$, the asymptotic index for small haloes where all haloes form at the same epoch, we recover the earlier result that $v(f)$ is flatter than $f^{-(4-\gamma)}$. For ΛCDM on galactic scales, the effective power index is $n \leqslant -2.3$, so $\nu \leqslant 0.35$, and for $\gamma \geqslant 1.8$ we obtain that $v(f)$ is flatter than $f^{-2.17}$. We see that the time dependence makes only a little difference.

The total $v(f)$ of the halo should be the sum of the background contribution and the subhalo contribution, but at the high-f range we expect $v(f)$ to be dominated by the contribution of subhaloes, as seen earlier in this section. The above toy model thus predicts that $v(f)$ should be flatter than $f^{-2.2}$, in conflict with the measured $v(f) \propto f^{-2.5}$.

We conclude that a halo is not simply an ensemble of clumps drawn from the general population of smaller haloes. The subhaloes may have a different mass function, their shape properties may vary differently with mass, and they both could vary with distance r from the host-halo centre. If one keeps the scaling relation with $\nu = 0$ and ignores any variation with r, the required subhalo mass function for matching the measured $v(f) \propto f^{-2.5}$ is with $\gamma = 1.5$ (compared to $\gamma = 1.8$). Indeed, a flattening of the subhalo mass function could be a natural result of the inevitable dynamical evolution of the subhaloes in the potential well of their host halo. The phase mixing due to tidal effects, including total disruption, is likely to be more efficient in less massive subhaloes, thus flattening the mass function. Also, dynamical friction is more efficient in making more massive subhaloes sink into the halo, thus making the mass function flatter in the inner parts. However, recent simulations indicate that the subhalo mass function is not flatter than $\gamma \simeq 1.7$ (De Lucia et al. 2004), indicating that the tidal effects on the inner structure of subhaloes must also have an important role. These are matters for more detailed future studies, but the failure of the idealised toy model analysed above to reproduce the magic power law $v(f) \propto f^{-2.5}$ indicates that the phase-space density is likely to provide a useful tool for studying the dynamical evolution of subhaloes in host haloes.

6 DISCUSSION AND CONCLUSION

Using Delaunay tessellation, we developed a method for measuring the 6-dimensional coarse-grained phase-space density $f(\mathbf{x},\mathbf{v})$ in N-body systems. We focused, in particular, on measuring the phase-space volume distribution function, $v(f)$. We applied this technique to several simulated haloes of $\sim 10^6$ particles, formed by hierarchical clustering in the standard ΛCDM scenario, and obtained two striking new results.

First, $v(f)$ is well described by a power law, $v(f) \propto f^{-2.5\pm0.05}$, over 3 to 5 decades in f. The power-law regime starts at an f value which corresponds to the characteristic size of the virialized halo. It ends at an f value which is determined by the dynamical resolution limit of the specific simulation. Therefore, the true power-law range may extend to $f \to \infty$. This power law seems to be insensitive to the halo mass in the range $10^9 - 10^{15} M_\odot$, indicating insensitivity to the exact slope of the fluctuation power spectrum, as long as the haloes are built by hierarchical merging of clumps bottom up.

Second, this power-law originates from substructures within the halo rather then the overall trend with radius. The substructure

completely dominates the high-f parts of the $v(f)$ distribution. The infalling clumps seem to phase-mix — by puffing up, heating and stripping — as their orbits decay from the virial radius inwards toward the halo centre and they melt into the halo smooth background.

Our first worry is that these results could be numerical artifacts, or severely contaminated by such. Based on our error analysis and tests with mock datasets, we argue that the $v(f)$ measured by the DTFE algorithm genuinely reflects the true phase-space properties of the given N-body system over a broad range of f. The question is whether the phase mixing suffered by the subclumps as they approach the halo centre might be an artifact of numerical effects such as two-body relaxation, leading to underestimated inner densities and/or overestimated internal velocities. A similar effect has been pointed out using a one-dimensional toy model (Binney 2003). The apparent agreement between simulations run with different codes and different resolutions is encouraging. In order to specifically address the effect of few-body relaxation, we intend to run twice a simulation of the same halo with the same number of particles but with a different force resolution (ongoing work with F. Stoehr).

Assuming that the simulations genuinely reflect the true physical behaviour under the Vlasov equation, the origin of the robust power-law shape of $v(f)$ from the merging substructure becomes a very interesting theoretical issue. As demonstrated in §5, a simple model using the mass function and the scaled profiles of the general halo population in the ΛCDM scenario does not reproduce the correct power law. This, and the apparent trend of the f spikes with radius, indicate that the structural and kinematic evolution of the subhaloes in the parent halo are important. Studies of tidal heating and stripping may be found useful in this modelling.

It would be interesting to follow the phase-space evolution and the contribution to the overall $v(f)$ by a single, highly resolved subhalo, or many of those, as they orbit within the parent halo and approach its centre. This may help us understand the nature of the interaction between the parent halo and its subhaloes, and the origin of the $v(f)$ power law (ongoing works with E. Hayashi and with B. Moore).

Another more general but speculative possibility is that the $f^{-2.5}$ power law represents some sort of a cascade of relaxation processed in phase-space, in which high phase-space densities turn into lower (coarse-grained) densities through the process of mixing. In general, the fact that our findings are expressed in terms of the fundamental concept of phase-space density should make them more directly accessible to analytical treatment. In this respect, it may prove beneficial to investigate more closely the time evolution of the $v(f)$ of a cosmological halo and its components. This may shed light on the connection between the $v(f)$ power-law behaviour and the relaxation processes within the halo.

We saw that the power-law behavior of $v(f)$ is limited to the virial regime. It would be interesting to learn how this shape evolves in time as the halo virializes. A preliminary study (to be concluded and reported in another paper) indicates that in the intermediate-f regime the $v(f)$ of a pre-virialized system is significantly flatter than $f^{-2.5}$, while in the high-f regime it drops in a much steeper way. The $f^{-2.5}$ behaviour seems to be a feature unique to virialized systems.

We learnt that in the haloes that are built by hierarchical clustering, the power-law behaviour $v(f) \propto f^{-2.5}$ reflects the halo substructure. It would be interesting to find out whether this power-law behaviour actually requires substructure, or it is a more general phenomenon of virialized gravitating systems, valid independently of substructure. One way to answer this question would be to analyse simulated haloes in which all fluctuations of wavelengths smaller than the halo scale were removed, resulting in a smooth halo formed by monolithic collapse, with no apparent substructure in the final configuration. As described in §1, such haloes are known to still have NFW-like density profiles in real space, and one wonders whether they also have the magic power-law $v(f)$. There are preliminary indications for a steeper $v(f)$ in this case (Arad, Dekel & Moore, in preparation). If confirmed, it would indicate that the $f^{-2.5}$ behaviour, while insensitive to the exact slope of the initial power spectrum, is unique to the hierarchical clustering process, and is not a general result of violent relaxation.

Our current results are just first hints from what seems to be a promising rich new tool for analysing the dynamics and structure of virialized gravitating systems. The analysis could become even more interesting when applied to haloes including the associated gaseous and stellar components.

ACKNOWLEDGMENTS

We thank Paul Bode for making his earlier simulations available for analysis. We acknowledge related ongoing collaborations with Stefan Gottloeber, Eric Hayashi & Julio Navarro, Ben Moore and Felix Stoehr. We are grateful for stimulating discussions with Stefan Colombi, Donald Lynden-Bell, Gary Mamon, Jerry Ostriker and Simon White. Part of this work has been done at the IAP in Paris. AK acknowledges support of NASA and ASF grants to NMSU. Our simulations were performed at the National Energy Research Scientific Computing Centre (NERSC). IA is a Marie Curie Postdoctoral Fellow. AD is a Miller Visiting Professor at UC Berkeley. This research has been supported by the Israel Science Foundation grant 213/02, by the German-Israel Science Foundation grant I-629-62.14/1999, and by NASA ATP grant NAG5-8218.

REFERENCES

Alvarez M.A., Shapiro P.R., Martel H., 2002, AAS, 200, 4103
Avila-Reese V., Colmn P., Valenzuela O., D'Onghia E., Firmani C., 2001, ApJ, 559, 516
Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693
Binney, J., Tremaine S., 1987, Galactic Dynamics (Princeton Univ Press)
Binney J., 2003, astro-ph/0311155, submitted to MNRAS.
Bode P., Ostriker J. P., Xu. G., 2000, ApJS, 128, 561
Bode P., Ostriker J. P., 2003, ApJS, 141, 1
Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001b, ApJ, 555, 240
Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001a, MNRAS, 321, 559
Bullock J. S., Kravtsov A. V., Colin P., 2002, ApJL, 564, 1
Cole S., Lacey C., 1996, MNRAS, 281, 716
Colín P., Klypin A., Kravtsov A. V., 2000, ApJ, 539, 561
Colín P., Avila-Reese V., Valenzuela O., 2000, ApJ, 542, 622
Colín P., Klypin A., Valenzuela O., Gottlöber S., 2003, astro-ph/0308348, submitted to ApJ.
Dalal N., Kochanek C. S., (2002), ApJ, 572, 25
Dekel A., Kowitt M., Shaham J., 1981, ApJ, 250, 567.
Dekel A., Devor J., Hetzroni G., 2003, MNRAS, 341, 326.
Dekel A., Arad I., Devor J., Birnboim Y., 2003, ApJ, 588, 680

Delaunay B. V., 1934, Bull. Acad. Sci (VII) Classe Sci. Mat., 793
De Lucia G., Kauffmann G., Springel V., White S. D. M., Lanzoni B., Stoehr F., Tormen G., Yoshida N., 2004, MNRAS, 348, 333
Fillmore J. A., Goldreich P., 1984, ApJ, 281, 1
Gunn J. E, Gott J. R., 1972, ApJ, 176, 1
Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 2000, ApJ, 544, 616
Hayashi E., Navarro J. F., Taylor J., Stadel J., Quinn T., 2003, ApJ, 584, 541
Hayashi E., Navarro J. F., Power C., Jenkins A., Frenk C. S., White D. M., Springel V., Stadel J., Quinn T. R., MNRAS, submitted (astro-ph/0310576)
Hockney R. W., Eastwood J. W., 1981, Computer simulations using particles (New York: McGraw-Hill)
Hoffman Y., Shaham J., 1985, ApJ, 297, 16
Huss A., Jain B., Steinmetz M., 1999, ApJ, 517, 64
Huss A., Jain B., Steinmetz M., 1999, MNRAS, 308, 1011
Ibata R., Gilmore G., Irwin M., 1994, Nature, 370, 194
Ibata R., Irwin M., Lewis G. F., Stolte A., 2001, ApJ, 547, 133
Ibata R., Irwin M., Lewis G. F., Ferguson A., Tanvir N., 2001, Nature, 412, 49
Klypin A., Gottloeber S., Kravtsov A. V., Khokhlov A. M., 1999a, ApJ, 516, 530
Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999b, ApJ, 522, 82
Klypin A., Kravtsov A. V., Bullock J. S., Primack J. R., 2001, ApJ, 554, 903
Kravtsov A. V., Klypin A., Khokhlov A. M., 1997, ApJS, 111, 73
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lynden-Bell D., 1967, MNRAS, 136, 101
Maller A. H., Dekel A., 2002, MNRAS, 335, 487
Maller A. H., Dekel A., Somerville, R. S., 2002, MNRAS, 329, 423
Mathur S. D., 1988, MNRAS, 231, 367
Moore B., Governato F., Quinn T., Stadel J., Lake G., 1998, ApJ, 499, L5
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19
Moore B., Quinn T., Governato F., Stadel J., Lake, G., 1999b, MNRAS, 310, 1147
Navarro J. F., Frenk C. S., White S. D. M, 1995, MNRAS, 275, 720
Navarro J. F., Frenk C. S., White S. D. M, 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M, 1997, ApJ, 490, 493
Navarro J. F., Steinmetz M., 2000, ApJ, 538, 477
Navarro J. F., Hayashi E., Power C., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T. R., MNRAS, submitted (astro-ph/0311231)
Nusser A., Sheth R., 1999, MNRAS, 303, 685
Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T. R., MNRAS, 338, 14
Press W. H., Schechter P., 1974, ApJ, 187, 425
Ricotti M., 2003, MNRAS, 334, 1237
Schaap W., van de Weygaert R., 2000, A&A, 363, L29
Sheth R. K., 2003, MNRAS, 345, 2000
Su P., Drysdale R. L. S., 1995, Proceedings of the eleventh annual symposium on Computational geometry, p 61
Subramanian K., Cen R., Ostriker J. P., ApJ, 538, 528
Syer D., White S. D. M., 1998, MNRAS, 293, 337
Tanemura M., Ogawa T., Ogita N., 1983, Journal of Computational Physics, 51, 191
Taylor J. E., Navarro J. F., 2001, ApJ, 563, 483
Tremaine S., Hénon M., Lynden-Bell D., (1986), MNRAS, 219, 285
van de Weygaert R., 2000, A&A, 363, L29.
Valenzuela O, Klypin A., 2003, MNRAS345, 406
Wambsganss J., Bode P., Ostriker J. P., sumbitted to ApJ, astro-ph/0306088
Weinberg M. D., Katz N., 2002, ApJ, 580, 627
Weller J., Bode P., Ostriker J. P., in preperation.
White S. D. M., Zaritsky D., 1992, ApJ, 394, 1

APPENDIX A: ERRORS IN $v_{\rm del}(f)$

The "true", underlying coarse-grained $f(\mathbf{x},\mathbf{v})$ is what one would have measured using a fixed smoothing window in phase space (e.g. counts in fixed cells) and an arbitrarily large number of particles. Instead, the Delaunay Tessellation Field Estimator uses adaptive cells in order to deal with the mass resolution limitations. Therefore, the relation between the measured $v_{\rm del}(f)$ and the underlying $v_{\rm true}(f)$ is not trivial. Both statistical and systematic errors might influence our results. We estimate these errors empirically below, but we start with an approximate model that provides a simple understanding of the origin of the errors.

A1 A Voronoi Model: Fixed Smoothing $v(f)$

A technique similar to the Delaunay tessellation, but somewhat simpler to interpret, is the Voronoi tessellation (van de Weygaert 1994). For each particle, the Voronoi cell is defined as the region of phase space in which every point is closer to that particle than to any other particle. In this case, N particles define exactly N Voronoi cells which cover all of phase space with no overlaps. If V_i is the Voronoi cell of particle i, then a natural mass-preserving way of estimating the phase-space density inside that cell is by $f_i \equiv m_i/|V_i|$. We denote the quantities measured this way by $f_{\rm vor}$ and $v_{\rm vor}(f)$.

Much like the Delaunay tessellation, the Voronoi tessellation is an adaptive grid that enables one to estimate $f(\mathbf{x},\mathbf{v})$ even in the presence of a relatively small number of particles. We use the Delaunay method in our main analysis because it is somewhat more accurate (Schaap & van de Weygaert 2000), and is easier to calculate numerically. However, the similar Voronoi method provides a simple way of learning about the properties of the measured $v(f)$ and understanding the uncertainties associated with such a measurement. The empirical tests of the Delaunay measurements below demonstrate the relevance of the wisdom gained by analysing the Voronoi model.

To understand the errors in the Voronoi density estimation, let us start with the trivial case where all of infinite phase-space is uniformly filled with phase-space density f_0, which is represented by an infinite number of particles with mass m. A volume V of phase-space would then contain on average a finite number of particles, $f_0 V/m$. The Voronoi estimate of f_i for each particle would fluctuate about f_0 due to the discreteness of the particle distribution. Since there is no typical scale in the problem, and each cell always contains one particle, the fluctuations $\delta f/f$ per Voronoi cell would remain at the same level even if one increases the average number density of particles while decreasing the mass of each particle in proportion, keeping f_0 the same. Therefore, the probability that the Voronoi estimated f would lie in the interval $f \rightarrow f + df$ may

be written in terms of a universal probability distribution function $p_\infty(f/f_0)\,d(f/f_0)$.

When we consider a finite system in a box of volume V, with a finite number of particles N inside it, we may expect the probability distribution $p_N(\cdot)$ to deviate from its asymptotic form $p_\infty(\cdot)$. As N decreases, we expect $p_N(\cdot)$ to widen due to the increasing effect of the boundaries. Next, examine a system with a non-uniform phase-space density, such as a cosmological halo. If the number of particles that represent this f is sufficiently large, we may approximate every region in phase-space as being locally uniform, and estimate its Voronoi f using the asymptotic $p_\infty(\cdot)$. This assumption is expected to break down in regions with very high phase-space density, where the sampling may become poor and insufficient, or in regions where f has large gradients over small scales. Nevertheless, lets assume for the moment that there exists an effective $p(\cdot)$ [not necessarily $p_\infty(\cdot)$] which properly approximates the fluctuation distribution of the Voronoi f for all particles.

This assumption allows us to calculate the expectation value of $v_{\rm vor}(f)$ for a finite system with a given $f(\mathbf{x},\mathbf{v})$. This is done by first calculating the average of $V_{\rm vor}(f)$, the cumulative version of $v_{\rm vor}(f)$ defined in (9), and then differentiate it to obtain the average of $v_{\rm vor}(f)$. Assuming that $f(\mathbf{x},\mathbf{v})$ is realised by $N \gg 1$ particles, we divide phase space into a large number of cells ω_α, which are small enough to guarantee that (a) each cell is very unlikely to contain more than one of the N particles, and (b) the value of $f(\mathbf{x},\mathbf{v})$ is approximately constant within every cell.

For each cell ω_α, we calculate $\langle V_\alpha(f)\rangle$, the average of the contribution of this cell to $V_{\rm vor}(f)$. The contribution $V_\alpha(f)$ would be non-zero only when there is a particle in the cell ω_α, a particle whose assigned Voronoi estimated density is $f_{\rm vor}$. If f_α is the true phase-space density in that the cell, then, according to our assumption, $f_{\rm vor}$ would be chosen at random out of the probability distribution $p(f_{\rm vor}/f_\alpha)\,d(f_{\rm vor}/f_\alpha)$. Once $f_{\rm vor}$ is chosen, $V_\alpha(f)$ is given by

$$V_\alpha(f) = \begin{cases} m/f_{\rm vor} & , \quad f < f_{\rm vor} \\ 0 & , \quad f > f_{\rm vor} \end{cases} \; . \tag{A1}$$

Using our assumption (a) above, the probability of ω_α to host one particle is $P_\alpha = NM^{-1}\int_{\omega_\alpha} f(\mathbf{x},\mathbf{v})\,d\mathbf{x}d\mathbf{v}$, with $M = Nm$ the total mass of the system. Therefore,

$$\langle V_\alpha(f)\rangle = P_\alpha \times \int_f^\infty p(f_{\rm vor}/f_\alpha) d(f_{\rm vor}/f_\alpha)\, m/f_{\rm vor} \tag{A2}$$

$$= \int_{\omega_\alpha} d\mathbf{x}d\mathbf{v} \int_f^\infty p[f_{\rm vor}/f(\mathbf{x},\mathbf{v})] f_{vor}^{-1}\, df_{\rm vor} \; , \tag{A3}$$

and so,

$$\langle V_{\rm vor}(f)\rangle = \sum_\alpha \langle V_\alpha(f)\rangle$$

$$= \int d\mathbf{x}d\mathbf{v} \int_f^\infty p[f_{\rm vor}/f(\mathbf{x},\mathbf{v})] f_{vor}^{-1}\, df_{\rm vor}$$

$$= \int_0^\infty df'\, v(f') \int_f^\infty p(f_{\rm vor}/f') f_{vor}^{-1}\, df_{\rm vor} \; . \tag{A4}$$

In the last equality, we have used the exact $v(f)$ to replace the six-dimensional phase-space integration. By differentiating Eq. (A4) with respect to f_0, we finally obtain the desired expectation value:

$$\langle v_{\rm vor}(f)\rangle = \int_0^\infty v(f') f'^{-1} p(f/f')\, df' \; . \tag{A5}$$

We see that the measured $\langle v_{\rm vor}(f)\rangle$ is a convolution of the exact $v(f)$ with a *fixed window function* $p(f/f_{f_{\rm true}})$. The narrower $p(\cdot)$ is, the closer $\langle v_{\rm vor}(f)\rangle$ would be to the true $v(f)$. However, as argued above, even when $N \to \infty$ the window $p(\cdot)$ does not approach a Dirac delta function; it rather converges to some finite-width distribution $p_\infty(\cdot)$. Therefore, even with an infinite resolution the Voronoi tessellation would not produce the exact $v(f)$; it converges to a convolution of it with a fixed window $p_\infty(\cdot)$.

This convolution would not affect the shape of the measured $\langle v_{\rm vor}(f)\rangle$, and would preserve the true shape of $v(f)$, provided that $v(f)$ does not vary drastically over f scales which are smaller than the width of the window. In particular, when $v(f)$ is a power law, the Voronoi algorithm would recover the same power law for $\langle v_{\rm vor}(f_0)\rangle$.

A2 Empirical Testing with Mock Systems

For an empirical study of the errors in the DTFE-measured $v(f)$, and for testing how well the Voronoi model approximates these errors, we have performed a series of $v_{\rm del}(f)$ measurements on systems with known phase-space densities of the form

$$f(\mathbf{x},\mathbf{v}) = \rho(x)\left[2\pi\sigma^2(x)\right]^{-3/2} e^{-v^2/2\sigma^2(x)} \; , \tag{A6}$$

corresponding to a spherical system in real space with a Maxwellian velocity dispersion. We have examined six such systems with three different density profiles parametrised by α,

$$\rho_\alpha(x) = \frac{e^{-x/5}}{x^\alpha(1+x)^{3-\alpha}} \; , \quad \alpha = 0,\ 0.5,\ 1.0 \; , \tag{A7}$$

and the following two types of dispersion profiles:

$$\sigma_{\rm v}(x) = \left[\frac{M(x)}{x}\right]^{1/2} \; , \tag{A8}$$

$$\sigma_{\rm c}(x) = 0.1 \; . \tag{A9}$$

The subscripts "v" or "c" denote a varying dispersion profile versus a constant one respectively. The $v(f)$ for such systems is

$$v(f) = \frac{(4\pi)^2}{f} \int_0^{x(f)} x^2\sigma^3(x) \sqrt{2\log\frac{f(x)}{f}}\, dx \; , \tag{A10}$$

with

$$f(x) \equiv \frac{\rho(x)}{\left[2\pi\sigma^2(x)\right]^{3/2}} \; , \tag{A11}$$

and $x(f)$ its inverse.

Figure A1 shows the cumulative distributions of $f_{\rm del}/f_{\rm true}$ in different bins (f_j, f_{j+1}) of $f_{\rm true}$, for the $\alpha = 1$ $\sigma_{\rm v}$ system and the $\alpha = 1$ $\sigma_{\rm c}$ system. Both systems were realised using 10^6 particles. In both cases we have also plotted the cumulative distribution $f_{\rm del}/f_{\rm true}$ for a homogeneous Poisson distribution, realised in a six-dimensional cubic box with 10^6 particles. We have verified that this distribution is essentially unchanged when the calculation is done with 10^5 particles. Therefore, it should be regarded as good approximation to the asymptotic limit we would have reached in the different bins, had we used an infinite number of particles. The other four systems give essentially the same results.

In the 1-v system, five bins were defined by $f_j = 10^{-7}, 10^{-5}, 10^{-3}, 10^{-1}, 10, 10^3$. We see that the shape of the distribution, corresponding to the width of the differential distribution, is very much independent of $f_{\rm true}$. For all bins, the full width is less than a 1.5 decades. On the other hand, there does seem to be a systematic shift of the median toward larger values of $f_{\rm del}/f_{\rm true}$

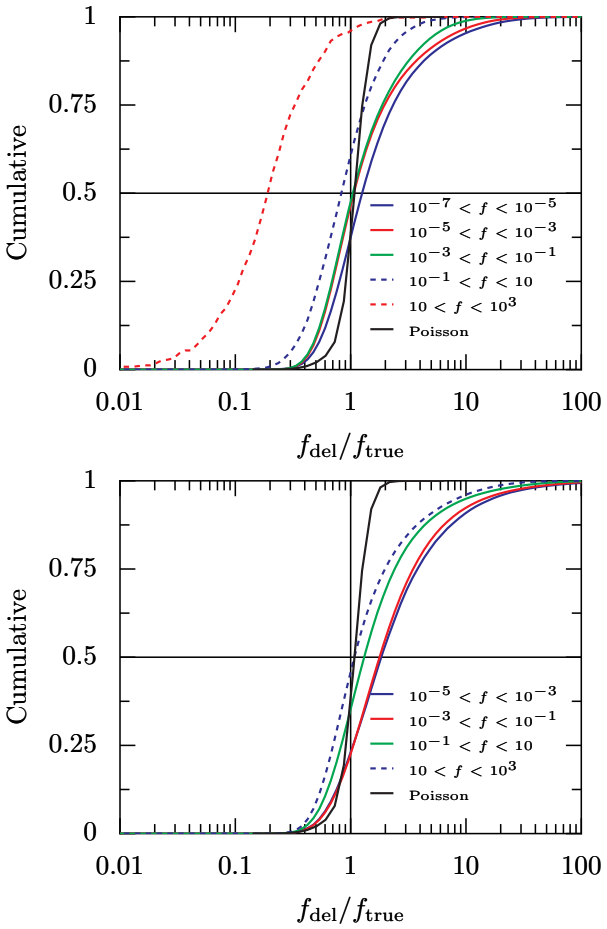

Figure A1. The cumulative distribution of $f_{\rm del}/f_{\rm true}$ as measured in different bins of $f_{\rm true}$ for two mock systems described by Eq. (A6) with 10^6 particles. Top: an NFW $\rho(x)$ given by Eq. (A7) ($\alpha = 1.0$) with a varying $\sigma_{\rm v}^2(x) = M(x)/x$. Bottom: same $\rho(x)$ but with a constant $\sigma_{\rm c}(x) = 0.1$. The solid black line represents the cumulative distribution of $f_{\rm del}/f_{\rm true}$ for a homogeneous Poisson distribution, which is realised in a six-dimensional cube using 10^6 particles. This should be regarded as the asymptotic limit with an infinite resolution.

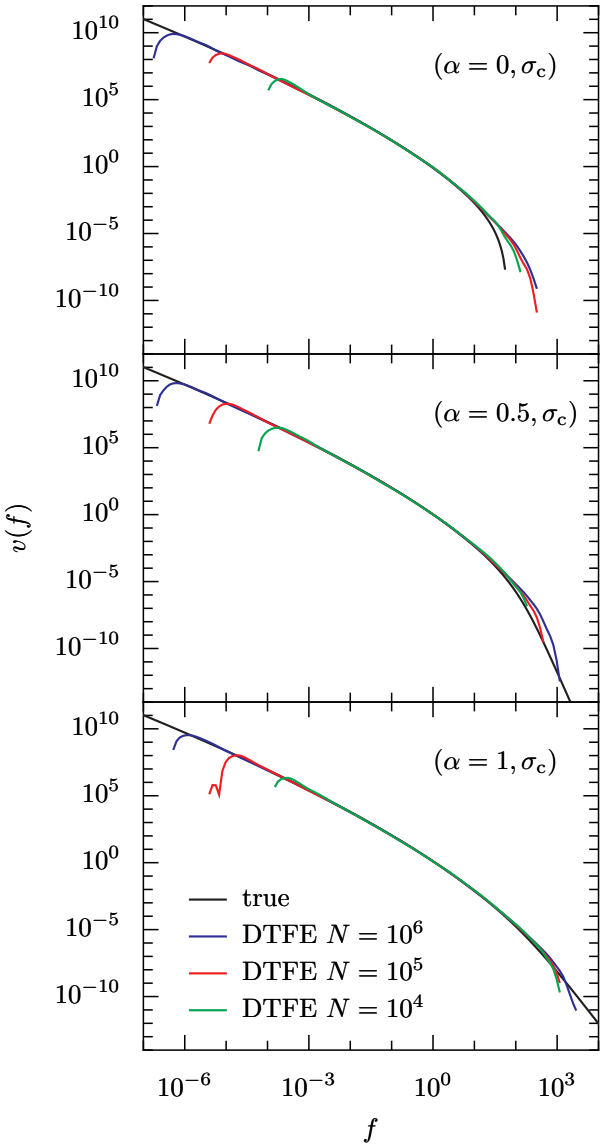

Figure A2. The measured $v_{\rm del}(f)$ versus the true $v(f)$ for the three $\sigma_{\rm c}$ mock systems, each performed with three different resolutions.

for smaller values of $f_{\rm true}$. The large shift of the highest-f bin ($10 < f < 10^3$) toward small $f_{\rm del}/f_{\rm true}$ is a result of the very low number of particles in that bin, less than 500, which is insufficient for representing such high phase-space densities.

The bias toward larger phase-space densities in the low-$f_{f_{\rm true}}$ bins may be attributed to boundary effects: while the total mass of the N particles in the realisation is equal to the total mass one would obtain from the exact $f(\mathbf{x}, \mathbf{v})$ integrated over the infinite phase-space, the total phase-space volume used by the DTFE to estimate $f(\mathbf{x}, \mathbf{v})$ is finite. It is the smallest possible convex polygon containing all N particles. Therefore, we may expect an overestimate of f, which would be more pronounced near the boundaries. Nevertheless, as we shall see by comparing $v_{\rm del}(f)$ to $v_{\rm true}(f)$, on scales of a few-decades, this bias is rather meaningless.

The $f_{\rm del}/f_{\rm true}$ distributions of the 1-c system are essentially the same as the ones in the 1-v system. Here 4 bins were defined by $f_i = 10^{-5}, 10^{-3}, 10^{-1}, 10^{1}, 10^{3}$. The overall shape of the plots changes very little from bin to bin, and its width is about a decade and a half. Additionally there exists the bias towards larger $f_{\rm del}/f_{\rm true}$ ratios as $f_{\rm true}$ gets smaller.

To see how well the DTFE reconstructs the true $v(f)$, we have measured the $v_{\rm del}(f)$ of the six mock systems using 10^4, 10^5 and 10^6 particles for each system. Figures A2, A3 present the results of these numerical experiments. We see that with the highest resolution, of 10^6 particles, the recovery is excellent over a range of almost 7 decades in f. Systematic deviations begin at the high-f end and the low-f end. At both ends, the deviations appear at about one to two decades inward to the highest and lowest values of $f_{\rm del}$ in that realisation. When the resolution is decreased, the f-range where $v_{\rm del}(f)$ closely matches $v_{\rm true}(f)$ narrows gradually.

As argued above, the systematic deviations at the low-f end are probably a result of the finite phase-space volume occupied by the particles. As the number of particles is increased, a larger portion of phase-space is covered, enabling the reconstruction of lower values of $f_{\rm del}$.

The high-f systematic deviations can be qualitatively understood using the Voronoi model and its convolution formula (A5). Since the DTFE uses a finite number of particles to recover the f

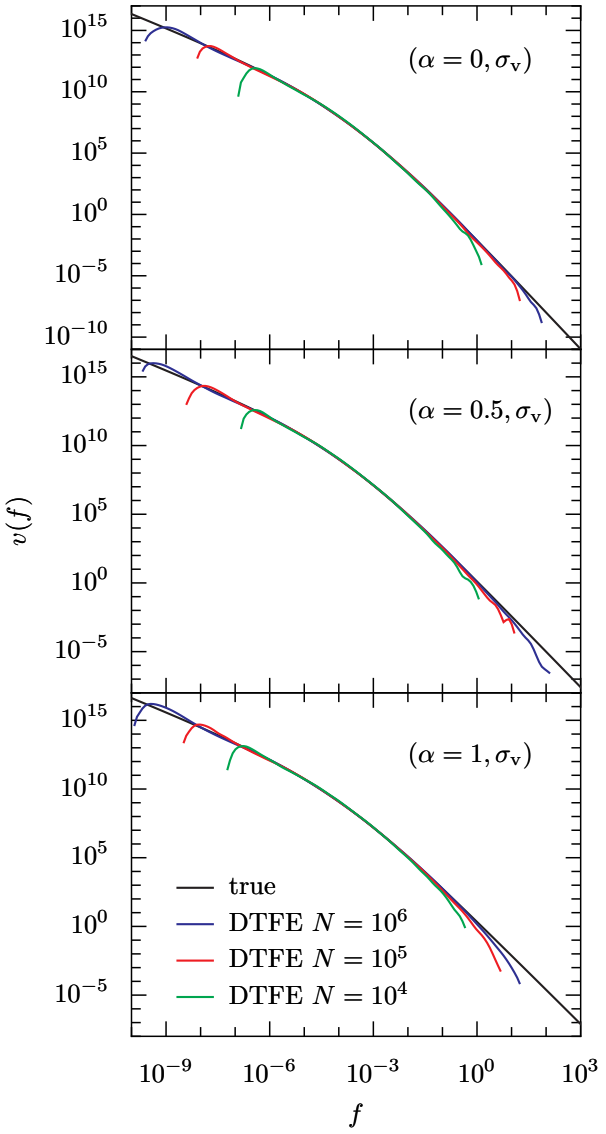

Figure A3. Same as Fig. A2 but for the σ_v mock systems.

field, there must be an upper cutoff, f_1, for the spectrum of f values that the DTFE can produce in principle. As far as the DTFE is concerned, the true $v(f)$ has an effective cutoff at f_1. If we plug this "true" truncated $v(f)$ into the convolution formula (A5), assume a distribution function $p(\cdot)$ with a width of about 1 decade, and set $f_1 \simeq 50$, we recover the qualitative behaviour of the measured $v_{\rm del}(f)$ for the $N = 10^6$ mock samples with σ_v and the three values of α. This simple picture is, however, an oversimplification because the high-f region is also effected by statistical fluctuations due to the low number of particles there.

The convolution formula can also explain the *overestimation* by the $v_{\rm del}(f)$ in the high-f regions of the mock systems with σ_c. In these systems, the transition between the low-f power-law and the steep high-f decline is rather sharp; it occurs over a scale comparable to the width of the window function $p(\cdot)$. As a result, the $v_{\rm del}(f)$ measurement at a high-f includes contributions from the higher $v(f)$ values at lower f. This becomes very apparent in the case with $\alpha = 0$ and σ_c, in which f has an upper bound of $f \simeq 63$. While the true $v(f)$ vanishes for all f values larger than this limit, the DTFE-measured $v(f)$ vanishes only at an f value that is an order-of-magnitude larger, due to the 1-decade width of the assumed probability distribution function $p(\cdot)$.

A3 Statistical Errors in $v_{\rm del}(f)$.

The Voronoi model can also be used to estimate the *statistical* errors in $v_{\rm vor}(f)$. Strictly speaking, $v_{\rm vor}(f)$ is an ill-defined random variable, as it is measured by differentiating $V_{\rm vor}(f)$, which is a super-position of Heaviside step functions, and as such $v_{\rm vor}(f)$ is a sum of Dirac delta functions. Much like a white-noise process, its variance is infinite. In practise, however, we always compute $v_{\rm vor}(f)$ by differentiating a *smoothed* version of $V_{\rm vor}(f)$ (using a spline, for example). Therefore, we may expect the statistical error in $v_{\rm vor}(f)$ to be comparable to the statistical error in $V_{\rm vor}(f)$. The latter can be estimated in a way similar to how we estimated the average of $V_{\rm vor}(f)$.

To calculate $\left\langle [\Delta V_{\rm vor}(f)]^2 \right\rangle = \left\langle V_{vor}^2(f) \right\rangle - \langle V_{\rm vor}(f) \rangle^2$, we can use the definition of $V_\alpha(f)$ to write

$$\left\langle V_{vor}^2(f) \right\rangle = \sum_{\alpha \neq \beta} \langle V_\alpha(f) V_\beta(f) \rangle + \sum_\alpha \left\langle V_\alpha^2(f) \right\rangle \ . \tag{A12}$$

Assuming that $V_\alpha(f)$ is independent of $V_\beta(f)$, the cross terms would cancel out from $\left\langle [\Delta V_{\rm vor}(f)]^2 \right\rangle$, leaving us with the upper limit

$$\left\langle [\Delta V_{\rm vor}(f)]^2 \right\rangle \leqslant \sum_\alpha \left\langle V_\alpha^2(f) \right\rangle \ . \tag{A13}$$

Using arguments similar to those used for calculating $\langle V_\alpha(f) \rangle$, one can show that

$$\left\langle V_\alpha^2(f) \right\rangle = m \int_{\omega_\alpha} d\mathbf{x} d\mathbf{v} \int_f^\infty p[f_{\rm vor}/f(\mathbf{x},\mathbf{v})] f_{vor}^{-2} \, df_{\rm vor} \tag{A14}$$

$$\leqslant \frac{m}{f} \int_{\omega_\alpha} d\mathbf{x} d\mathbf{v} \int_f^\infty p[f_{\rm vor}/f(\mathbf{x},\mathbf{v})] f_{vor}^{-1} \, df_{\rm vor} \tag{A15}$$

$$= \frac{m}{f} \langle V_\alpha(f) \rangle \ . \tag{A16}$$

Therefore,

$$\left\langle [\Delta V_{\rm vor}(f)]^2 \right\rangle \leqslant \frac{m}{f} \langle V_{\rm vor}(f) \rangle \ , \tag{A17}$$

and the relative error $\Delta(f)$ is given by

$$\Delta(f) \leqslant \left(\frac{m}{f \langle V_{\rm vor}(f) \rangle} \right)^{1/2} \ . \tag{A18}$$

Plugging $m = M/N$ into the formula above, where M is the total mass of the system and N the total number of particles, we recover the common large-numbers limit $\Delta(f) \propto 1/\sqrt{N}$.

To check how good this estimate is for the DTFE-measured $v(f)$, we have measured the $v(f)$ of 100 realisations of the $\alpha = 1$ σ_v mock system with $N = 10^3$ and $N = 10^4$ particles, and 30 realisations with $N = 10^5$ particles. From these measurements we computed the true relative error in $v(f)$ [with respect to the average of the DTFE-measured $v(f)$, not with respect to the exact $v(f)$], and compared it to the prediction of Eq. (A18). Figure A4 shows the comparison for the three resolutions. We see that Eq. (A18) performs well as an upper bound for the statistical errors, except for the low-f region. In that region the cumulative $V(f)$ approaches a constant as $f \to 0$, due to the finite phase-space volume of the halo. This introduces fluctuations to $v(f)$ as a result of the numerical differentiation in Eq. (33).

However, it is interesting to notice that whenever the statistical

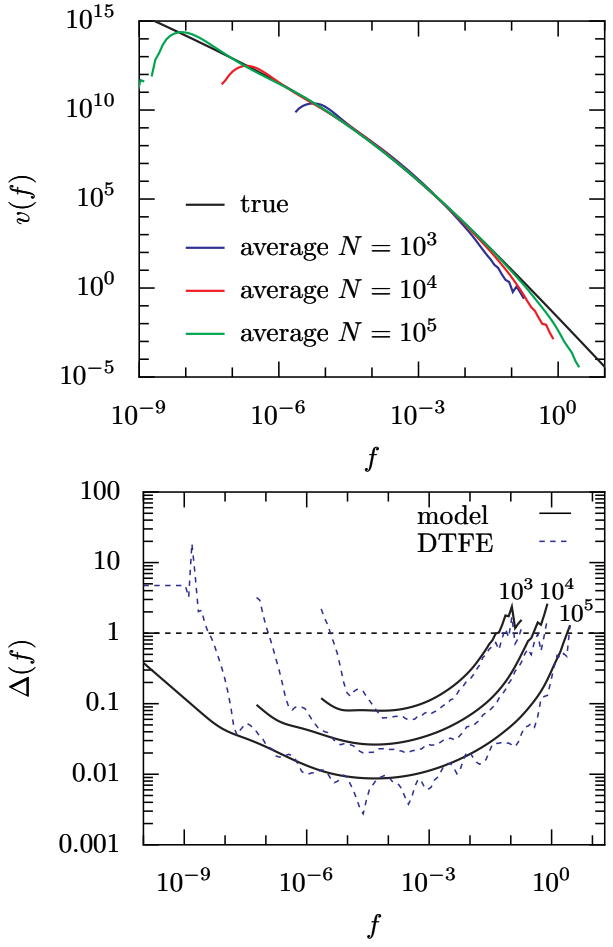

Figure A4. The average (top) and relative error (bottom) in $v(f)$ as measured in the mock system ($\alpha = 1$, $\sigma_{\rm v}$) sampled with three different resolutions. The relative errors are compared to the analytic prediction of Eq. (A18).

errors in $v_{\rm del}(f)$ become important, they are overwhelmed by the low-f or high-f systematic errors. In that respect, the statistical errors in $v_{\rm del}(f)$ are of no big relevance.