

This is the accepted manuscript made available via CHORUS. The article has been published as:

Position-dependent Voronoi probability distribution functions for matter and halos

Drew Jamieson and Marilena Loverde

Phys. Rev. D **103**, 103522 — Published 17 May 2021

DOI: [10.1103/PhysRevD.103.103522](https://doi.org/10.1103/PhysRevD.103.103522)

Position-dependent Voronoi probability distribution functions for matter and halos

Drew Jamieson and Marilena Loverde

*C.N. Yang Institute for Theoretical Physics, Department of Physics &
Astronomy, Stony Brook University, Stony Brook, New York, 11794, USA*

We measure the Voronoi density probability distribution function (PDF) for both dark matter and halos in N-body simulations. For the dark matter, Voronoi densities represent the matter density field smoothed on a uniform mass scale, which approximates the Lagrangian density field. For halos, the Voronoi densities contain information about the local environment of each halo. We measure the halo virial masses, the total amount of dark matter within each halo Voronoi cell, and the halo Voronoi cell volumes, and we show how halo abundances depend on these three quantities. We then study the position-dependent Voronoi density PDF, measured within finite subregions of the Universe, using separate universe simulations. We demonstrate that the spatial variation of the position-dependent PDF is due to large-scale density fluctuations, indicating that the position-dependent PDF is a biased tracer of large-scale structure. We measure this bias for the dark matter, and interpret it as the bias of regions of the Lagrangian density field that are selected based on density. For the halos, this bias can be interpreted as a form of assembly bias. We present the mapping from late-time to early-time Voronoi density for each simulation dark matter particle, which is highly stochastic. We compare the median of this stochastic map with spherical collapse calculations and discuss challenges involved in modeling the evolution of the density field on these scales.

I. INTRODUCTION

The spatial distributions of matter and galaxies contain an abundance of information about the contents, evolution, and origins of the Universe. With upcoming large-scale structure (LSS) surveys [1–4], we will improve upon the already impressive measurements of cosmological parameters from CMB data [5], and possibly detect evidence of new physics beyond Λ CDM, the standard model of cosmology. The physical processes of structure formation involve nonlinear gravitational clustering, which must be theoretically understood and accurately modeled in order to extract optimal and rigorous constraints from observational data. For this reason, the theory and phenomenology of LSS formation continues to be an important area of research in cosmology.

The first step of any LSS analysis is to decide how to smooth the density field. In the Eulerian description, the density field is smoothed on a fixed comoving length scale, so that density fluctuations represent mass fluctuations. In the Lagrangian description, the density field is smoothed into equal-mass fluid cells, which flow under the competing influence of cosmic expansion and gravitational clustering. Since the fluid cells all contain the same amount of mass, the density fluctuations of the Lagrangian field represent volume fluctuations.

These two approaches have complementary strengths and weaknesses. The fixed comoving length scale imposed under Eulerian smoothing allows the level of nonlinearity to be controlled, by averaging out the highly nonlinear small scales. Then the Eulerian density field can be described in the linear and quasilinear regimes using the effective field theory of LSS, which can provide tighter parameter constraints than analyses with standard perturbation theory [6–9]. The trade-

off is that many structural features of the cosmic web, which contain cosmological information, may be washed out by this smoothing. The mass-weighted Lagrangian smoothing provides a more detailed description of the filamental structures in the cosmic web, while losing the ability to reduce contributions from small-scale, nonlinear modes.

The mass-weighted Lagrangian smoothing provides a more detailed description of the filamental structures in the cosmic web [10–13], while losing some control over the level of nonlinearity.

In this work, we study the one-point statistics of the Lagrangian matter density field, smoothed on a uniform mass scale through the use of Voronoi tessellation in N-body simulations. Voronoi tessellation only achieves an approximation of the true Lagrangian field, since it neglects overlap of fluid cells due to multistreaming effects [14]. However, the tessellation allows us to probe the distribution of matter deep into the nonlinear regime, and test our understanding of structure formation on these scales. We also study the one-point statistics of the halo number density field under Voronoi tessellation, which probes the local environment of halos and their nearest neighbors [15].

The position-dependent one-point statistics, measured within finite regions of the Universe, are spatially modulated by the long-wavelength modes of the density field, which couple to small-scale modes due to nonlinear gravitational evolution. This is analogous to the spatial modulations of halo and galaxy abundances, which are parameterized bias models. The concept of cosmic bias can be generalized to any local observable that is affected by the nonlinearity of gravitational clustering. Previously, the generalized bias of the one-point statistics of the Eulerian density field was measured in N-body simulations and accurately modeled [16]. In this work

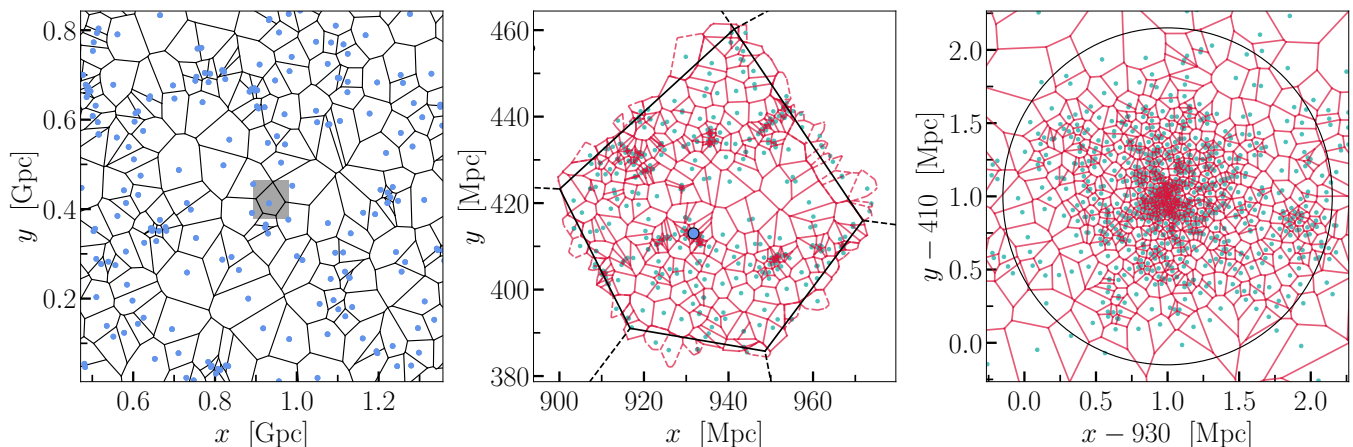


FIG. 1. Left: 2D Voronoi tessellation of simulation halos with mass $M > 10^{13} M_{\odot}$ in a slice of width 0.4 Mpc taken at redshift $z = 0.0$. The halos are indicated by blue dots, and the boundaries of their Voronoi cells are the black lines. The shaded square indicates the area that is shown in the middle plot. Middle: 2D Voronoi tessellation of the dark matter particles within a halo's Voronoi cell. The particles are indicated by green dots and their cell walls are the red lines. The dashed red lines indicate dark matter cells that overlap with the halo's cell while the dark matter particle lies outside of the halo's Voronoi region. The halo has mass $M = 8.7 \times 10^{13} M_{\odot}$ and its position is indicated by the blue circle with radius corresponding to the halo's virial radius $R = 1.1$ Mpc. Right: 2D Voronoi tessellation of the dark matter within the halo's virial radius.

we introduce the generalized bias associated with the one-point statistics of the matter and halo density fields smoothed using Voronoi tessellation.

The outline of this paper is as follows. In Sec. II, we define the Voronoi density PDF and the locally measured, position-dependent Voronoi density PDF. In this section, we also present the dependence of halo abundances on their virial mass, the total mass in their Voronoi cell, and their Voronoi volume. We summarize the details of our N-body simulations in Sec. III. Results for the globally defined Voronoi density PDFs in both dark matter and halos are presented in Sec. IV. We review the separate universe formalism in Sec. V A, and in Sec. V B we present our separate universe determinations of the position-dependent Voronoi PDF for simulation dark matter and halos. We validate our separate universe results using global universe simulations in Sec. V C. In Sec. VI, we discuss the challenges of modeling the matter Voronoi statistics due to stochasticity and nonlinearity in the evolution of the dark matter Voronoi cells, and study the sensitivity of the position-dependent matter PDF to the simulation resolution. In Sec. VII we conclude.

II. VORONOI DENSITIES

LSS surveys provide catalogs of objects that represent the most dense features of the cosmic web. A typical strategy for making cosmological inferences using this data is to construct the number density field of catalog objects and measure its N-point statistics [17]. For recent examples see [18–24]. This involves smoothing the discrete catalog objects over some window function. Typical choices for the smoothing window function

include a spherical top-hat of fixed radius [25–27], or smoothing B-spline interpolation on a comoving Cartesian grid [28, 29]. Both of these involve cells of fixed shape and size, and so they probe the abundance fluctuations of catalog objects and mass fluctuations within each cell. This fixed-volume smoothing gives an estimation of the Eulerian density field.

A complementary approach is to fix the amount of mass, or the number of objects that each cell contains and allow the shapes and sizes of the cells to vary. In the latter case, this gives an equal-mass smoothed estimation of the Lagrangian density field. One way of achieving this is through Voronoi tessellation, which provides an unambiguous assignment of volume and cell shape to each discrete object in a catalog. Voronoi tessellation has been employed in several contexts for cosmological simulations and observational data analysis, including void finders [30, 31], and cosmic web structure analysis [11, 14, 32–35].

A. Tessellation

Voronoi tessellation is a method for partitioning a region of space around a catalog of discrete, point-like objects [36, 37]. Each object is assigned a Voronoi cell such that the space within an object's cell is nearer to it than to any other object in the catalog. A Voronoi cell is bounded by a set of planes, or faces, that form a polyhedron completely enclosing the cell's volume. The

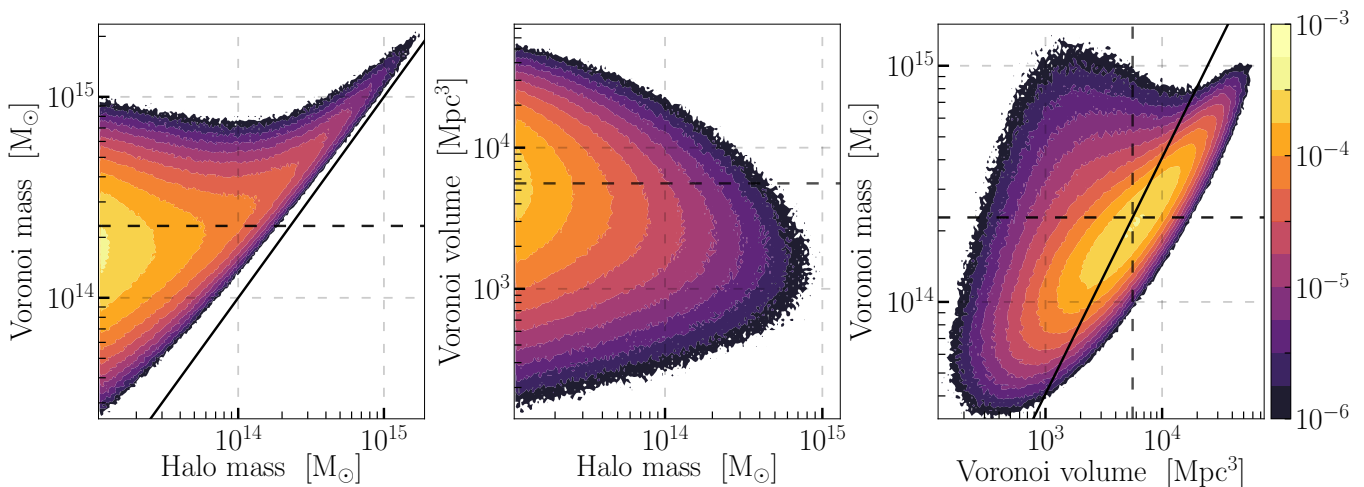


FIG. 2. Left: Number density of halos measured in a 2D histogram of the halo virial mass and the total mass within a halo's Voronoi cell, the color map indicates the value of $\partial^2 \bar{n} / \partial \log M / \partial \log M_V$, where $\bar{n}$ is the halo number density in units of $\text{Mpc}^{-3} h^3$, M is the halo virial mass, and M_V is the total mass within the halo's Voronoi cell. The black line corresponds to $M = M_V$, when all mass within a halo's Voronoi cell equals the mass inside its virial radius. Middle: Number density of halos measured in a 2D histogram of the halo virial mass and the halo Voronoi volume, $\partial^2 \bar{n} / \partial \log M / \partial \log \mathcal{V}$, where $\mathcal{V}$ is the halo's Voronoi volume. Right: Number density of halos measured in a 2D histogram of Voronoi volume and total mass within a halo's cell, $\partial^2 \bar{n} / \partial \log \mathcal{V} / \partial \log M_V$. The dashed lines correspond to the mean values of Voronoi volume and total cell mass. The solid line represents mean density cells, with $M_V = \bar{\rho} \mathcal{V}$. The histograms are binned uniformly in log of mass and log of volume, using 200 bins within the ranges of the plot axes, averaged over 20 simulations in $1 \text{ Gpc } h^{-1}$ boxes. The halo catalogs were taken at redshift $z = 0.0$ with a lower virial mass threshold $M > 10^{13} M_\odot$, and had an average of 5.2×10^5 halos per simulation.

faces meet at edges, three of which meet at a vertex¹. When two cells share a common face, they are said to be neighbors. A cell's shape and size encode information about its distribution of immediate neighbors. There is an enormous amount of information within a Voronoi tessellation, including the statistical distributions of volumes, faces, edges, vertices, and spatial correlations among all of these quantities. In this work, we will focus exclusively on the cell volumes as an estimator of density.

Two dimensional examples of Voronoi tessellations from a slice through one of our N-body simulations are shown in Fig. 1. The plot on the left is a tessellation around a catalog of halos with mass greater than $10^{13} M_\odot$. One of the halos is highlighted by a gray square, and its Voronoi cell is plotted in greater detail in the middle panel, along with the tessellation of matter particles from the N-body simulation. The halo's Voronoi cell is much larger than its virial radius, and contains complex substructures of clustered matter. The plot on the right shows the Voronoi tessellation of the dark matter particles within the halo's virial radius.

The additional mass in the halo's cell is clustered, and some of the overdensities may be unresolved halos with masses below the threshold of our catalog. By lowering

the mass threshold, some of these low-mass objects will be included in the catalog, and the Voronoi cell of the original halo will be significantly altered, dividing into smaller cells around the newly resolved halos. The halo's Voronoi volume is also not equal to the sum of volumes of all the dark matter particles within its cell. Some contributions to the volume come from particles that lie outside the cell, and some particles within the cell have volumes that extend to its exterior.

The halo Voronoi properties are plotted as 2D histograms in Fig. 2. The plot on the left is the differential, *joint halo mass function*²,

$$\frac{\partial^2 \bar{n}(M, M_V)}{\partial \log M \partial \log M_V}, \quad (1)$$

where $\bar{n}(M, M_V)$ is the cumulative halo number density, for halos with virial mass greater than M , and total mass within their Voronoi cells greater than M_V . The middle plot is the same, but for Voronoi volume $\mathcal{V}$ instead of total cell mass,

$$\frac{\partial^2 \bar{n}(M, \mathcal{V})}{\partial \log M \partial \log \mathcal{V}}. \quad (2)$$

¹In three dimensions it is exceedingly rare for more than three edges to meet at a single vertex, even computationally within floating point accuracy.

²We use the term *joint* here to indicate that the function is related to the joint, bivariate, probability density of finding a halo with two distinct properties.

The right plot is,

$$\frac{\partial^2 \bar{n}(M_V, \mathcal{V})}{\partial \log M_V \partial \log \mathcal{V}}. \quad (3)$$

The color bar on Fig. 2 shows the values of differential number density in units of $\text{Mpc}^{-3} h^3$.

The low-mass halos have a wide range of masses within their Voronoi cells, and a corresponding wide range of cell volumes. This demonstrates that low-mass halos are found in diverse environments, from strongly clustered, small volume, low-mass cells, to strongly anti-clustered, large volume, high-mass cells. The lowest mass halos, with $M \sim 10^{13}$, are very weakly clustered, in the sense that their distribution is peaked near the mean volume per halo ($5.7 \times 10^3 \text{ Mpc}^3$) in the catalog, which is expected for a Poisson sample of points. For high-mass halos with $M > 10^{14} M_\odot$, the total mass in a halo's Voronoi cell is much more narrowly peaked near and above the virial mass, which bounds the total mass from below. These halos also have a narrower distribution of Voronoi volumes, which peaks 0.2–0.5 times below the mean volume per halo, indicating their strong clustering.

In the right-most plot of Fig. 2, the solid line corresponds to $M_V = \bar{\rho} \mathcal{V}$, indicating halos with total Voronoi mass density within their cells equal to $\bar{\rho}$, the mean density of the Universe. Points above (below) this line are overdense (underdense) cells. For unclustered, Poisson distributed matter, the histogram would be symmetric around this line, whereas the peak of the distribution follows a curve that is tilted downwards with respect to this line. Note that the axes scales are logarithmic, and a symmetric scatter around this line is not a symmetric scatter if linear axes scales were used instead. On the other hand, for Poisson distributed halos, the distribution of volumes is much narrower and the distribution of total cell masses is wider, which would make the contours tilt upwards, with a higher slope than the mean density line. The shapes and orientations of these contours reflect the gravitational clustering of both matter and halos.

B. Global one-point statistics

Consider a dark matter particle of mass M_p in an N-body simulation, and let this particle's comoving Voronoi volume be $\mathcal{V}$. We define its Voronoi density to be,

$$\rho_V = \frac{M_p}{\mathcal{V}}. \quad (4)$$

Rather than defining density contrasts with respect to the mean Voronoi density, we define,

$$\delta_V = \frac{\rho_V}{\bar{\rho}} - 1, \quad (5)$$

where $\bar{\rho}$ is the usual mean comoving Eulerian matter density. In a simulation box with sides of length L_{box}

and N_p particles, $\bar{\rho} = N_p M_p L_{\text{box}}^{-3}$. The Voronoi density contrast defined in this way is the fractional difference between the mean volume per particle and the volume of an individual particle's Voronoi cell. At early enough times every particle, to a good approximation, has the same Voronoi volume. Then the Voronoi density contrast can be interpreted as the fractional amount of change in an individual particle's Voronoi volume since early times.

The mean Voronoi mass density is not equal to the mean mass density, $\bar{\rho}$, but rather,

$$\bar{V} = \frac{M_p}{\bar{\rho}}. \quad (6)$$

The Voronoi density contrasts can be expressed in terms of the mean Voronoi volume, $\bar{V}$, as

$$\delta_V = \frac{\bar{V}}{\mathcal{V}} - 1. \quad (7)$$

The Voronoi density contrasts do not average to zero, but their volume weighted average vanishes,

$$\langle \delta_V \mathcal{V} \rangle = 0. \quad (8)$$

We denote the probability of finding a cell with Voronoi density between $1 + \delta_V$ and $1 + \delta_V + d\delta_V$ as

$$\mathcal{P}(1 + \delta_V) d\delta_V. \quad (9)$$

We refer to $\mathcal{P}(1 + \delta_V)$ as the Voronoi density PDF. We will now consider how the shape of this distribution, measured in a finite subregion of the Universe, differs from its shape measured globally.

C. Local one-point statistics

In a finite subregion, the local estimation of the Voronoi density PDF will differ from the globally measured PDF. The spatial variations in the locally measured PDF are due to spatial variations in the local matter density. For a subregion of volume V_{su} centered at position $\vec{x}$, a local observer sees a mean density,

$$\bar{\rho}_{\text{su}}(\vec{x}) = \frac{1}{V_{\text{su}}} \int_{V_{\text{su}}} d^3y \rho(\vec{y}). \quad (10)$$

If the region is large enough, the local mean density will differ from the global one by a perturbatively small fluctuation, δ_L ,

$$\bar{\rho}_{\text{su}}(\vec{x}) = \bar{\rho} (1 + \delta_L(\vec{x})). \quad (11)$$

The subscript “su” is meant to indicate “separate universe”. From the point of view of an observer who lives within and only has observational access to V_{su} , they ascribe $\bar{\rho}_{\text{su}}$ as the mean density of the Universe. This observer would therefore disagree about the value of cosmological parameters with a global observer, who has wider observational access to regions outside of V_{su} .

We refer to the observer limited to V_{su} as a *separate universe observer*, and all of the observables measured with respect to their coordinates and cosmology are *separate universe observables*. For example, if δ_L is positive and large enough, then the separate universe observer determines that they live within a closed universe, when in actuality they may only live within a collapsing subregion of an otherwise open universe.

The local expansion history and structure formation occur within the separate universe observer's reference frame as if they live in a Friedmann-Lemaître-Robertson-Walker cosmology, but with parameters shifted from their global values in a way that depends on the evolution of δ_L [38, 39]. This implies that all differences between the separate universe observables and their counterparts in the global cosmology should be characterized by their response to the presence and growth history of δ_L . We refer to this statement as the *separate universe ansatz*.

Voronoi density contrasts measured by a separate universe observer are defined,

$$\delta_{\mathcal{V},\text{su}} = \frac{\bar{V}_{\text{su}}}{\mathcal{V}} - 1, \quad (12)$$

where $\bar{V}_{\text{su}}$ is the local mean Voronoi volume within the subregion, including only particles whose positions lie within V_{su} ³. This definition preserves the volume weighted average of Eq. (8) within the separate universe cosmology. The locally measured Voronoi PDF is the probability of finding a cell within V_{su} that has a density between $1 + \delta_{\mathcal{V},\text{su}}$ and $1 + \delta_{\mathcal{V},\text{su}} + d\delta_{\mathcal{V},\text{su}}$. We denote this,

$$\mathcal{P}_{\text{su}}(1 + \delta_{\mathcal{V},\text{su}} | \vec{x}) d\delta_{\mathcal{V},\text{su}}. \quad (13)$$

The separate universe ansatz states that the spatial modulations of local observables are due to the effects of the large-scale density fluctuations, $\delta_L(\vec{x})$, so that,

$$\mathcal{P}_{\text{su}}(1 + \delta_{\mathcal{V},\text{su}} | \vec{x}) = \mathcal{P}_{\text{su}}(1 + \delta_{\mathcal{V},\text{su}} | \delta_L(\vec{x})). \quad (14)$$

For a large enough subregion, we can expand this around $\delta_L = 0$, keeping only the leading term,

$$\mathcal{P}_{\text{su}}(1 + \delta_{\mathcal{V}} | \delta_L) = \mathcal{P}(1 + \delta_{\mathcal{V}}) + \frac{d\mathcal{P}_{\text{su}}}{d\delta_L} \delta_L + \mathcal{O}(\delta_L^2). \quad (15)$$

On the left-hand side of this equation, $\delta_{\mathcal{V},\text{su}} = \delta_{\mathcal{V}}$ in the limit $\delta_L \rightarrow 0$, so we have dropped that subscript. The higher order terms will also include tidal contributions $\sim \mathcal{O}(\nabla^2 \delta_L)$, which enter at second order. The fractional difference between the local and global PDF gives its linear response to δ_L . This response can be isolated by

taking the logarithm of both sides of Eq. (15), expanding to linear order in δ_L , and then correlating with δ_L ,

$$\frac{d \log \mathcal{P}_{\text{su}}}{d\delta_L} = \frac{\langle \log(\mathcal{P}_{\text{su}}(1 + \delta_{\mathcal{V}} | \vec{x})) \delta_L(\vec{x}') \rangle}{\langle \delta_L(\vec{x}) \delta_L(\vec{x}') \rangle}, \quad (16)$$

which is valid to leading order. We refer to the quantity as the separate universe response of the Voronoi density PDF.

D. Halos

The strength of the correlation between a halo's virial mass and the total mass within its Voronoi cell depends strongly on the halo's virial mass. For low-mass halos, the two masses are very weakly correlated, while for high-mass halos the correlation is strong. Observationally, the amount of mass within a Voronoi cell is not easily known. However, it may be related to the number density of halos through a nonlinear bias model. For these reasons, we measure the halo number density PDF, rather than a halo mass density PDF.

For a single Voronoi cell of volume $\mathcal{V}$, the number density is,

$$n_{\mathcal{V}} = \frac{1}{\mathcal{V}}, \quad (17)$$

and we define the Voronoi density contrast,

$$\delta_{\mathcal{V}} = \frac{n_{\mathcal{V}}}{\bar{n}} - 1, \quad (18)$$

where $\bar{n}$ is the mean Eulerian halo number density. The probability of finding a halo Voronoi cell with density between $1 + \delta_{\mathcal{V}}$ and $1 + \delta_{\mathcal{V}} + d\delta_{\mathcal{V}}$ is denote $\mathcal{P}(1 + \delta_{\mathcal{V}}) d\delta_{\mathcal{V}}$. Note that this notation is the same as for the matter Voronoi PDF, but we will always specify which one we are dealing with to avoid confusion.

In a large, finite subregion of volume V_{su} centered at $\vec{x}$, the local mean number density differs from the global mean,

$$\begin{aligned} \bar{n}_{\text{su}}(M, \vec{x}) &= \bar{n}(M) (1 + \delta_n(M, \vec{x})) \\ &\simeq \bar{n}(M) (1 + b_h(M) \delta_L(\vec{x})), \end{aligned} \quad (19)$$

where, as before, δ_L is the large-scale matter fluctuation within V_{su} , b_h is the linear Eulerian halo bias, and δ_n is the local fluctuation in halo number density. In terms of the local mean number density, the Voronoi density contrasts are defined,

$$\delta_{\mathcal{V},\text{su}} = \frac{n_{\mathcal{V}}}{\bar{n}_{\text{su}}} - 1. \quad (20)$$

We use the same notation for the position-dependent Voronoi PDF and its separate universe response for both halos and dark matter.

Suppose we computed the position-dependent PDFs with respect to the global mean density (i.e. the density

³The total Voronoi volume may not sum to V_{su} due to fluctuations at the boundary, depending on how this is defined. However, if the region is large enough, these fluctuations are negligibly small.

Parameter	Value
Ω_Λ	0.7
Ω_m	0.3
Ω_b	0.05
h	0.7
n_s	0.968
A_s	2.137×10^{-9}
N_p	$(1024)^3$
L_{box}	$1 \text{ Gpc } h^{-1}$
M_p	$1.108 \times 10^{11} M_\odot$

TABLE I. Cosmological and N-body simulation parameters.

contrasts defined in Eqs. (5) and (18), and denote these local PDFs as $\mathcal{P}_{\text{loc}}$. The transformation from $\mathcal{P}_{\text{su}}$ is given by

$$\mathcal{P}_{\text{loc}}(1 + \delta_V | \delta_L) = \mathcal{P}_{\text{su}}(1 + \delta_{V,\text{su}} | \delta_L) \frac{d\delta_{V,\text{su}}}{d\delta_V}. \quad (21)$$

To leading order, the Jacobian is,

$$\begin{aligned} \frac{d\delta_{V,\text{su}}}{d\delta_V} &= \frac{\bar{n}}{\bar{n}_{\text{su}}} \\ &= 1 - b_h \delta_L. \end{aligned} \quad (22)$$

Using Eqs. (20), (22), the leading order expansion of the transformation in Eq. (21) gives

$$\frac{\mathcal{P}_{\text{loc}}(1 + \delta_V | \delta_L)}{\mathcal{P}_{\text{su}}(1 + \delta_V | \delta_L)} = 1 - b_h \delta_L \left(1 + \frac{d \log \mathcal{P}}{d \log(1 + \delta_V)} \right), \quad (23)$$

where we have used,

$$(1 + \delta_V) \frac{d}{d\delta_V} = \frac{d}{d \log(1 + \delta_V)}. \quad (24)$$

Then the linear responses to δ_L are related,

$$\frac{d \log \mathcal{P}_{\text{loc}}}{d\delta_L} = \frac{d \log \mathcal{P}_{\text{su}}}{d\delta_L} - b_h \left(1 + \frac{d \log \mathcal{P}}{d \log(1 + \delta_V)} \right). \quad (25)$$

The second term on the right-hand side can be attributed to the shift in the local mean background density relative to the global mean density. It does not depend on the growth history of δ_L and can be determined directly from the global PDF. The same expression holds for the transformation of the matter PDF, setting the bias equal to one.

III. SIMULATIONS AND ANALYSIS

A. Simulations

We ran three sets of 20 N-body simulations, for a total of 60. One set was run with the global expansion

history from a standard Λ CDM cosmology. The second set, corresponding to an overdense separate universe, was given a slightly modified expansion history, according to Eqs. (32)–(33). The long-wavelength mode δ_L evolved linearly as a growing, adiabatic perturbation normalized to $\delta_L = +0.01$ at the final redshift $z = 0.00$. The third set, corresponding to an underdense separate universe, had $\delta_L = -0.01$ at the final redshift.

All three sets used the same 20 random seeds to generate their initial conditions, which ensures cosmic variance largely cancels at leading order when computing separate universe response observables. The initial conditions were generated using the second order Lagrangian perturbation theory method [40, 41], with input power spectrum from CLASS [42] using the parameters in Table I. Our simulations were run using a modified version of Gadget2 [43], which uses tabulated values of the local separate universe scale factor and Hubble rate, rather than integrating the Friedmann equation to compute the expansion history. Our simulation box size was chosen to be $L_{\text{box}} = 1 \text{ Gpc}/h$ with $N_p = (1024)^3$ dark matter particles. The particle mass was $M_p = 1.108 \times 10^{11} M_\odot$. These simulations were previously used for a study of void bias in the separate universe [44], and the position-dependent Eulerian density PDF [16].

Our halo catalogs were extracted using the friends-of-friends halo finder Rockstar [45], modified so that the density thresholds used for spherical overdensity mass calculations were consistent with the different separate universe cosmologies, and so that the mass assignments were continuous rather than discrete (see [46, 47] for details). We measured the Voronoi number density field in a catalog of halos with lower mass cutoff $M > 10^{13} M_\odot$, corresponding to a minimum of 100 particles per halo.

Our Voronoi tessellations were carried with a C++ code that uses the Voro++ publicly available library [48]. The object-oriented Voro++ library stores particles in container objects, which we defined to be subboxes of our simulations. We included a buffer region in our subbox containers, keeping track of which particles were in the subbox and which were in the buffer. The Voro++ library includes a loop class, which enabled us to easily compute only the Voronoi volumes of particles in the subbox, including ones with neighbors that are in the buffer. If the buffer is taken to be too small, a particle in the subbox may end up with one of the container walls as its neighbor. In this case, the code throws an error and exits, indicating that the size of the buffer region needs to be increased.

The advantage of our code is that it is easily parallelizable, since each subbox container can be analyzed independently, and Voro++ allows for each subbox to be tessellated in parallel one cell at a time. It is also has efficient memory usage, since only particles in the containers that are currently being analyzed are loaded into memory. This allowed us to compute the

Voronoi volumes of all 1024^3 dark matter particles in our simulations in 25 minutes, running on 3 compute nodes with 24 cores per node. We tested our code by comparing to the tessellation from ZOBOV [30], and found that the Voronoi volumes agree within floating point accuracy. Our code uses MPI to distribute processes to each node, and OMP to distribute processes within a node to each core.

B. Kernel density estimator

The Voronoi tessellations around our simulation dark matter particles and halos assign a volume, which we converted to a density, for each point-like tracer. From this list of Voronoi densities, we computed the PDFs using a kernel density estimator (KDE). First, we converted the lists of Voronoi densities into lists $\{\log(1 + \delta_{V,i})\}_{i=1}^N$, and used a Gaussian kernel,

$$\mathcal{P}_{\log}(\log(1 + \delta_V)) = \frac{1}{Nw} \sum_{i=1}^N \exp\left(-\frac{1}{2w^2} \log\left(\frac{1 + \delta_V}{1 + \delta_{V,i}}\right)^2\right). \quad (26)$$

Here, w is the kernel width, which sets the level of smoothing over each data point, and $\mathcal{P}_{\log}$ is the PDF for $\log(1 + \delta_V)$. We then transformed this to $\mathcal{P}(1 + \delta_V)$,

$$\mathcal{P}(1 + \delta_V) = \frac{\mathcal{P}_{\log}(\log(1 + \delta_V))}{1 + \delta_V}. \quad (27)$$

This procedure is justified by the observation that the density PDFs are, to a first approximation, similar to log-normal distributions [49, 50].

The choice of kernel width is important (the kernel function itself is less important), because it sets the level of smoothing, which controls the variance and statistical bias of the estimator. In the asymptotic limit of large sample size, the variance and estimator bias for a Gaussian KDE are [51],

$$\text{Var}(\mathcal{P}_{\log}(x)) \simeq \frac{1}{Nw} \mathcal{P}_{\log}(x), \quad (28)$$

$$\text{Bias}(\mathcal{P}_{\log}(x)) \simeq \frac{w^2}{2} \frac{d^2 \mathcal{P}_{\log}}{dx^2} \Big|_x. \quad (29)$$

where $x = \log(1 + \delta_V)$. If the PDF $\mathcal{P}_{\log}$ were perfectly Gaussian, the optimal width given by minimizing the integral of the squared asymptotic error would be,

$$w_G = \sigma_{\mathcal{P}_{\log}} \left(\frac{4}{3N}\right)^{1/5}, \quad (30)$$

with $\sigma_{\mathcal{P}_{\log}}$ being the standard deviation of $\mathcal{P}_{\log}$, which can be estimated from the sample. For the dark matter, the PDF in the global cosmology prefers a broader width than a Gaussian distribution, so we chose kernel widths $w = 1.5 w_G$. For the halos, the global PDF prefers a

Dark matter						
z	μ	σ	q_1	q_2	q_3	$\delta_V > 0$
0.0	1170	7123	0.834	5.63	132	0.722
0.5	382.9	2565	0.710	2.79	34.3	0.681
1.0	157.4	1185	0.665	1.85	10.9	0.642
49	1.004	0.062	0.960	1.00	1.04	0.505

Halos, $M > 10^{13} M_\odot$						
z	μ	σ	q_1	q_2	q_3	$\delta_V > 0$
0.0	2.403	3.292	0.760	1.42	2.77	0.644
0.5	2.467	3.559	0.750	1.40	2.78	0.638
1.0	2.610	4.316	0.739	1.36	2.78	0.629

TABLE II. Properties of the Voronoi density PDFs. The left column lists the redshift, followed by the distribution mean μ and standard deviation σ . The next three columns list the first, second and third quartiles. That is, q_i is the value of $1 + \delta_V$ for which $i \times 25\%$ of the cells have a lower density. The final column gives the fraction of cells that are overdense.

narrow width of $w = 0.3 w_G$, as we demonstrate in the Appendix to this paper, where we discuss these width selections in greater detail.

Estimating derivatives of PDFs using a KDE permits much more smoothing. Since taking a derivative essentially involves subtracting nearly identical functions, the estimator bias largely cancels in the numerator. This is true for the linear responses of the PDF to δ_L , which are approximated in the separate universe formalism as finite difference derivatives. Minimizing the integral of the asymptotic squared error for the derivative of a Gaussian PDF, the optimal width scales as $N^{1/7}$ rather than $N^{1/5}$. We chose widths of $w_{\text{su}} = 2.5 w_G$ for the separate universe PDFs of the dark matter, and $w_{\text{su}} = 4.5 w_G$ for the halos. We present tests of the effects from varying the separate universe smoothing for the halos in the Appendix. For the dark matter, the effects of varying the smoothing width are smaller, due to the large sample size.

IV. THE GLOBAL VORONOI DENSITY PDF

The mean volume per dark matter particle is constant in N-body simulations. For our simulations, the value is fixed at $\bar{V} = 2.72 \text{ Mpc}^3$. At early times, the distribution of matter cell densities is nearly Gaussian, narrowly peaked around $\rho = \bar{\rho} \equiv M_p/\bar{V}$. The exact distribution cannot be Gaussian, since ρ is bounded from below, however the width of the distribution is sufficiently narrow so that the non-Gaussianity is small. In fact, since we began our simulations at redshift $z = 49$ and used second order Lagrangian perturbation theory to set up the initial conditions, the initial density PDF is better approximated as log-normal. The interplay of cosmic expansion and gravitational collapse concentrate most of the dark matter into overdense regions, producing a

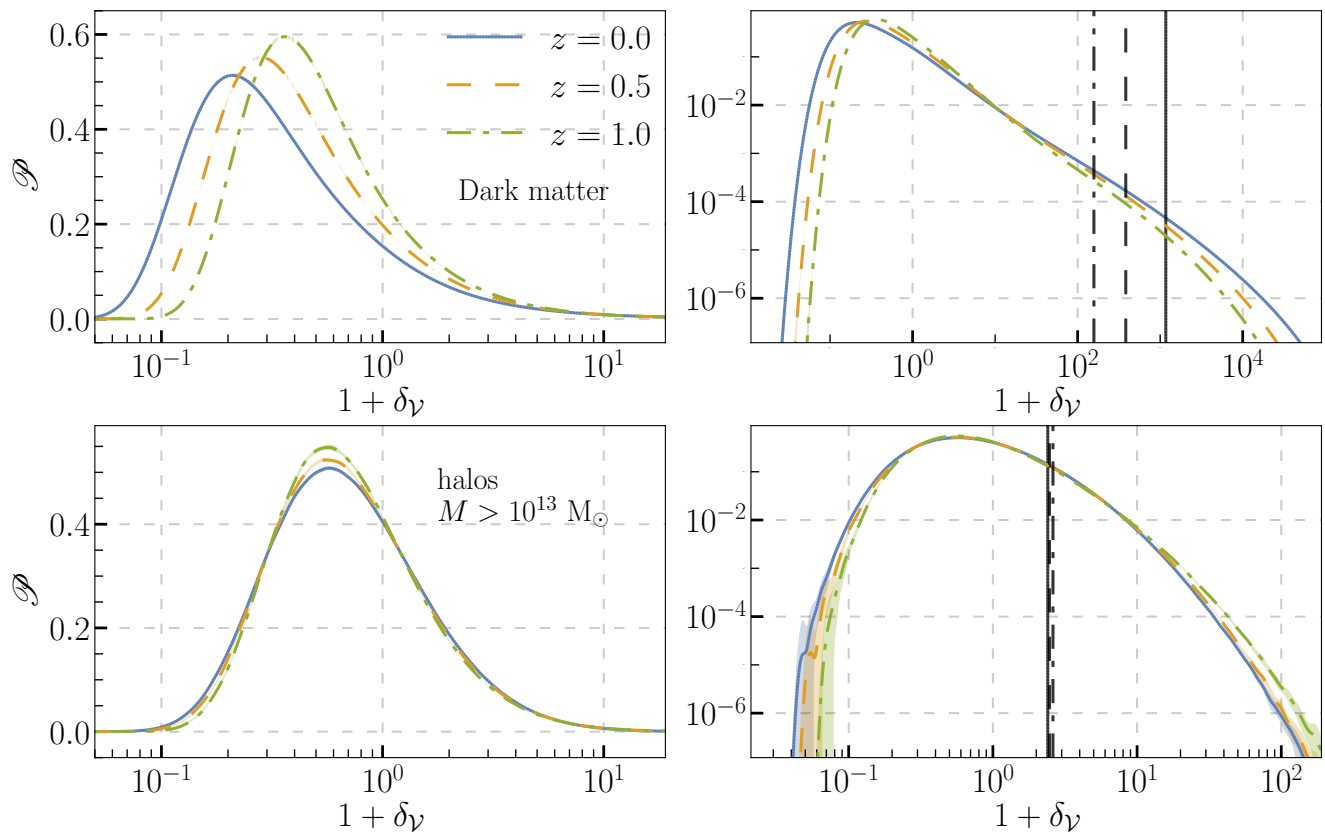


FIG. 3. Top: Dark matter density probability distributions measured with respect to Voronoi tessellations around N-body simulation dark matter particles. The plots on the left have a linear vertical scale to demonstrate the evolution of the distribution near the peak, while the plots on the right have a logarithmic vertical scale to demonstrate the evolution of the tails of the distributions. Bottom: Halo number density probability distributions measured with respect to Voronoi tessellations around halos with masses $M > 10^{13} M_{\odot}$. The legend on the top left plot, indicating the redshift of each distribution, applies to all plots. The black vertical lines indicate the mean of the distribution, with dashes that correspond to redshift according to the legend. The 1σ bootstrap errors are plotted as shaded regions, but are too small to see except in the extreme tails.

long, high-density tail in the PDF at late times. The dark matter Voronoi PDFs measured for particles with mass $M_p = 1.1 \times 10^{11} M_{\odot}$ are shown in Fig. 3 at redshifts $z = 0.0, 0.5$, and 1.0 . The location of the peak decreases with redshift, and occurs at very low densities near redshift $z = 0.0$. While the distribution peaks at $\delta_v < 0$, the long tail at high densities ensures that a majority of the particles are in overdense cells, with $\delta_v > 0$. The vertical lines in Fig. 3 indicate the mean of each distribution.

Some properties of the dark matter and halo PDFs are summarized in Table II. The mean of the dark matter PDF increases by nearly an order of magnitude between redshifts $z = 1$ and $z = 0$. The mean Voronoi density is 0.77 times the virial density for spherical collapse at redshift $z = 1$ ($148 \bar{\rho}$), while the mean is 3.5 times larger than the virial density at $z = 0$ ($337 \bar{\rho}$). Many of the dark matter particles within the virial radius of a halo have cell densities that are lower than the virial density, and while the individual Voronoi cells are typically far from spherical, we will demonstrate in Sec. VI that the

virial density for spherical collapse still appears as an important scale in the dark matter Voronoi PDF.

The number of halos is not constant over time, so the mean Voronoi volume per halo also evolves with redshift. The halo Voronoi density PDF captures information about the local environment of halos and their clustering with each other. Unlike the spatial correlation function, which is computed using the distances of each halo from all other halos in the catalog, the Voronoi PDF characterizes only the nearest neighbors of each halo. The halo Voronoi density PDFs are shown in Fig. 3 for halos with $M > 10^{13} M_{\odot}$. The peak of the distribution evolves considerably less than that of the dark matter distribution. The clustering of halos is made evident by the broadening of the distribution with time, as more collapsed objects are concentrated, with smaller cells, into high-density regions, leaving halos with larger Voronoi cells in lower density regions. However, the clustering also causes mergers, which combine high-density cells into single, somewhat larger cells for high-mass halos. This can be seen from the decrease in the

high-density tail of the halo PDF with time.

V. THE LOCAL VORONOI DENSITY PDF

A. Separate universe formalism

The dynamics of gravitational clustering is nonlinear, so the long-wavelength perturbations of the density field couple to, and effect the dynamics of structure formation on smaller scales. In the separate universe formalism, these effects are computed by defining a local cosmology in which the long-wavelength density perturbations are absorbed into the background, homogeneous density,

$$\bar{\rho}_{\text{su}} = \bar{\rho}(1 + \delta_L). \quad (31)$$

The locally defined mean density, $\bar{\rho}_{\text{su}}$, yields a locally defined cosmology that differs from the global one due to the presence of δ_L , the large-scale fluctuations. The local scale factor and local Hubble rate can be computed to linear order in δ_L [38] as

$$a_{\text{su}} \simeq a \left(1 - \frac{1}{3} \delta_L \right), \quad (32)$$

$$H_{\text{su}} \simeq H \left(1 - \frac{1}{3} \delta'_L \right), \quad (33)$$

where $'$ indicates $d/d \log a$.

Since δ_L includes only linear modes, its evolution can be computed in the global cosmology, using perturbation theory. From the above equations, we can define a pair of cosmologies, one corresponding to an overdense region ($\delta_L > 0$), and one corresponding to an underdense region ($\delta_L < 0$). By running pairs of N-body simulations with the corresponding overdense and underdense expansion histories, the linear response of an observables in the separate universe $\mathcal{O}_{\text{su}}(\delta_L)$ can be determined by taking the finite difference derivative with respect to δ_L ,

$$R_{\mathcal{O}} = \left. \frac{d \log \mathcal{O}_{\text{su}}}{d \delta_L} \right|_{\delta_L=0} \quad (34)$$

$$\simeq \frac{\mathcal{O}_{\text{su}}(+\delta_L) - \mathcal{O}_{\text{su}}(-\delta_L)}{2\delta_L \mathcal{O}}, \quad (35)$$

where $\mathcal{O}$ is the observable in the global universe. If we use the same set of initial conditions in each separate universe pair, we also benefit from cosmic variance cancellation in determinations of $R_{\mathcal{O}}$.

The separate universe response of the halo mass function is the linear halo bias [46, 52]. Similarly, the linear void bias is given by the separate universe response of the void size function [44, 53]. Higher order terms in the bias expansion can also be measured using separate universe techniques. A generalized bias can be associated to any observable, $\mathcal{O}$, that is affected by long-wavelength density perturbations. In a previous paper, we studied the generalized bias associated with the Eulerian matter

density PDF using the separate universe method [16]. In the following subsections we present the generalized bias of the Voronoi density PDF in dark matter and in halos.

B. The Voronoi density PDF in the separate universe

The separate universe response of the Voronoi density PDF is defined as

$$R_{\mathcal{P}}(1 + \delta_V) = \frac{\mathcal{P}_{\text{su}}(1 + \delta_V | +\delta_L) - \mathcal{P}_{\text{su}}(1 + \delta_V | -\delta_L)}{2 \mathcal{P}(1 + \delta_V) \delta_L}, \quad (36)$$

where $\mathcal{P}_{\text{su}}(1 + \delta_V | \pm \delta_L)$ is the local Voronoi density PDF measured in an overdense or underdense separate universe cosmology. Measurements of $R_{\mathcal{P}}$ from our simulations are shown in Fig. 4 for both the dark matter particles and the halos.

The matter PDF response has an intuitive density dependence. The response is positive for low-density cells ($\delta_V < -0.5$) and high-density cells ($\delta_V > 10$), indicating that the tails of the PDF are enhanced in an overdense region, and reduced in an underdense region. The response is negative in cells near the background density ($-0.5 < \delta_V < 10$), indicating the distribution's peak is enhanced in an underdense region, and reduced in an overdense region. This density dependence reflects gravitational clustering, which concentrates most of the matter into high-density cells, leaving behind a few isolated, low-density cells. Clearly, the response is dynamical in nature, depending on the details and evolution of δ_L , which is sensitive to cosmology.

The valley of the PDF response near its minimum is narrower at earlier redshifts, although it is not significantly deeper at $z = 1.0$ than it is at $z = 0.0$. On the other hand, the positive responses are stronger at earlier times. This reflects the relative rarity of extreme density regions at early times compared to later times. This is analogous to the increase in halo bias at fixed mass with increasing redshift.

Halos have a similar PDF response shape, although with less pronounced features. The response is positive for overdense cells with $0 < \delta_V < 10$ and for underdense cells with $\delta_V < -0.8$ at redshift $z = 1$. These bounds evolve toward lower densities with decreasing redshift. The negative response regime spans a more underdense range compared with the matter response, especially at lower redshift. This is due to the lack of evolution in the position of the halo distribution's peak, which can be seen in Fig. 3. An overdense separate universe has enhanced clustering, and can be thought of as being dynamically more advanced in time than the mean universe. For the matter PDF, the peak migrates deeper into the underdense region with time as the overall distribution broadens. Both of these effects contribute to the negative response near $\delta_V \simeq 0$ for the matter. For the halos, the peak does not significantly change position over time, and

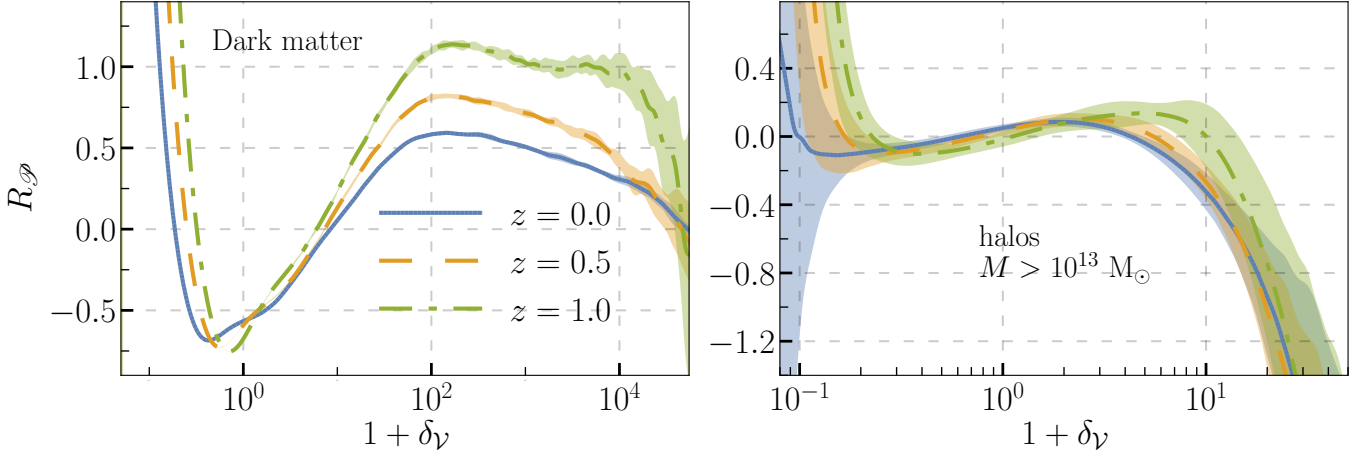


FIG. 4. Separate universe responses of the Voronoi density PDF of dark matter (left) and halos (right), at redshifts $z = 0.0$, 0.5 , and 1.0 , measured in N-body simulations. The smoothing scale for the dark matter is given by the particle mass $M_p = 1.1 \times 10^{11} M_\odot$. For the halos, a lower mass cutoff of $10^{13} M_\odot$ was used to construct the catalogs. The shaded regions correspond to 1σ bootstrap errors.

so the response is negative at lower densities, although this depends on the halo mass threshold.

For the halos, the response at high Voronoi density ($\delta_V > 10$) becomes negative, indicating that there are more extremely high-density cells in an underdense compared to an overdense separate universe. This may seem counter intuitive, but it is simply due to the merger of halos. Extremely high-density neighboring cells coalesce and becomes one larger cell as halos merge. Since the overdense cosmology can be thought of as advanced in time with respect to clustering, more of these mergers have occurred at a fixed global time compared with the underdense separate universe, so the response is negative. This interpretation is consistent with the redshift evolution of the global PDF in Fig. 3, which shows the high-density tail decreasing with time.

Since both $\mathcal{P}$ and $\mathcal{P}_{\text{su}}$ are normalized, this implies,

$$\int_{-1}^{\infty} d\delta_V R_{\mathcal{P}}(1 + \delta_V) \mathcal{P}(1 + \delta_V) = 0. \quad (37)$$

The orthogonality of the PDF and its linear response allows us to further interpret the physical meaning of the PDF response. To see this, we first note that we can recast the halo Voronoi density PDF as a *joint halo mass function*,

$$n(M, 1 + \delta_V) = \bar{n}(M) \mathcal{P}(1 + \delta_V), \quad (38)$$

where $\bar{n}(M)$ is the cumulative halo mass function. Note that $\mathcal{P}$ also depends on the lower mass threshold since, if we vary the threshold, we change the halo catalog and must redo the tessellation. The linear response of this *joint halo mass function* is,

$$\frac{d \log n_{\text{su}}(M, 1 + \delta_V)}{d\delta_L} = b_h(M) + R_{\mathcal{P}}(1 + \delta_V). \quad (39)$$

The left-hand side of this expression is the generalized bias of the conditional halo mass function, for halos selected based both on their Voronoi density and a lower mass threshold. Halos of the same mass but different Voronoi cell size have different biases, so this generalized bias can be interpreted as a particular kind of assembly bias. From Eq. (37), we find

$$\int_{-1}^{\infty} d\delta_V \frac{d \log n_{\text{su}}(M, 1 + \delta_V)}{d\delta_L} \frac{n(M, 1 + \delta_V)}{\bar{n}(M)} = b_h(M). \quad (40)$$

The generalized bias of the *joint halo mass function* decomposes the total halo bias into contributions from halos of different Voronoi density. The interpretation of the dark matter PDF response is the same, except in that case the bias is unity, and the dark matter PDF response gives the bias of equal-mass regions of the density field selected based on their Voronoi density.

C. Validation in the global universe

The separate universe ansatz claims that the spatial modulations in local, position-dependent observables measured in finite subregions are due to the effects of long-wavelength modes on gravitational clustering. To test this, we measured the local Voronoi density PDF in subboxes of global universe simulations, and cross-correlated the fluctuations of the local PDF with the matter field.

First, we constructed a mesh of 32^3 overlapping cubic subboxes with sides of length $125 \text{ Mpc}/h$. In each subbox, we load the N-body particles from one of our global universe simulations with positions that lie within the subbox, along with their Voronoi densities. We

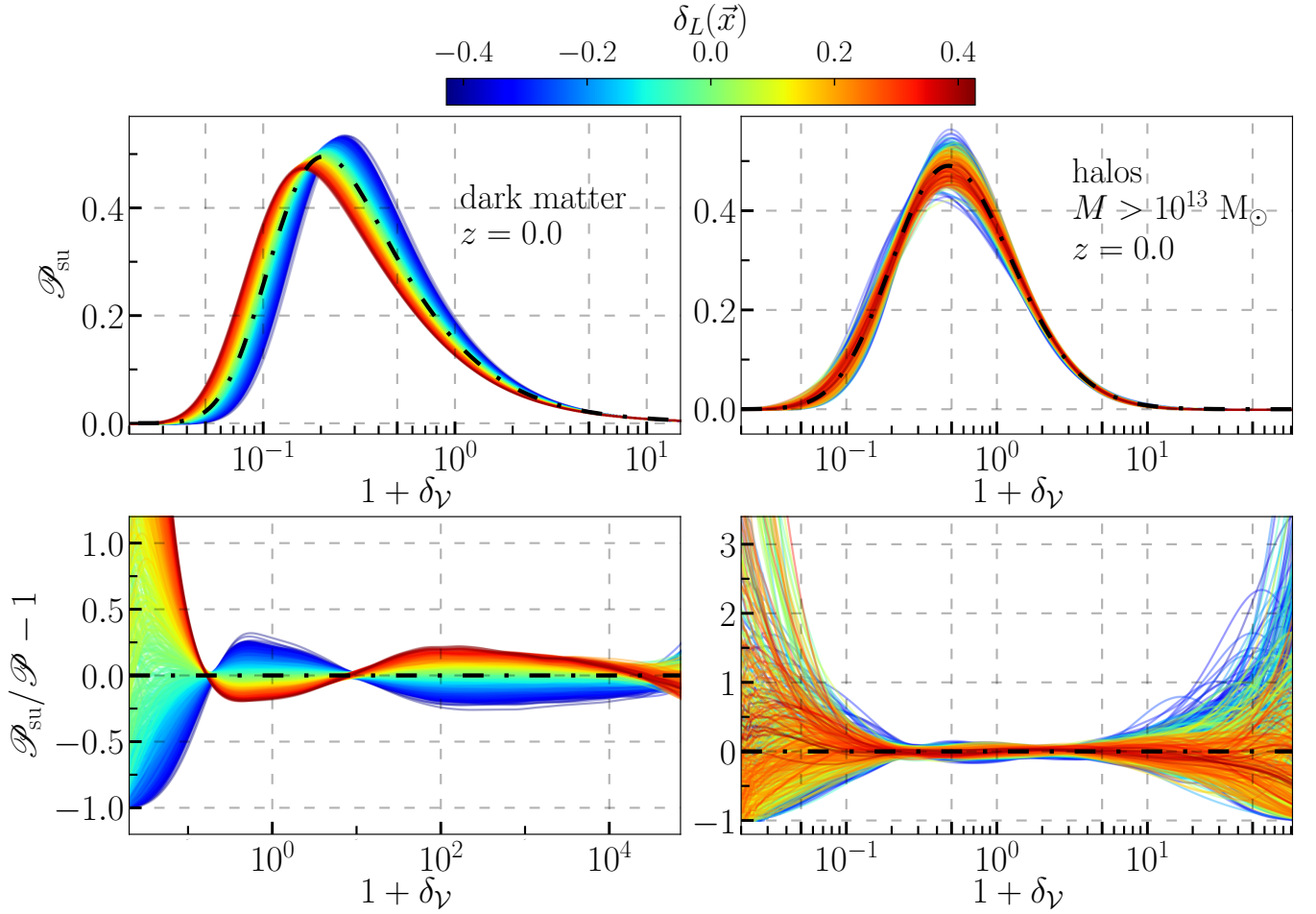


FIG. 5. Local Voronoi density PDFs for the matter (left) and halos (right) at redshift $z = 0.0$, measured in subboxes with sides lengths of $125 \text{ Mpc}/h$ within a global universe simulation. The color of each curve corresponds to the value of the large-scale density fluctuation within each subbox.

computed the local PDF $\mathcal{P}_{\text{su}}(1 + \delta_{\mathcal{V}} | \vec{x})$, and the local matter fluctuation, $\delta_L(\vec{x})$, where $\vec{x}$ is the position of the center of the subbox. Next, we computed the local PDF fluctuation,

$$\frac{\mathcal{P}_{\text{su}}(1 + \delta_{\mathcal{V}} | \vec{x})}{\mathcal{P}(1 + \delta_{\mathcal{V}})} - 1 \simeq \log \left(\frac{\mathcal{P}_{\text{su}}(1 + \delta_{\mathcal{V}} | \vec{x})}{\mathcal{P}(1 + \delta_{\mathcal{V}})} \right). \quad (41)$$

In practice we computed the fractional difference on the left-hand side of the above expression, but we use the logarithmic notation on the right-hand side here for brevity. They are equivalent to leading order in δ_L .

We used FFTW3 [54] to compute the matter modes, $\delta_L(\vec{k})$, and the modes of the PDF fluctuations, $\log(\mathcal{P}_{\text{su}}(1 + \delta_{\mathcal{V}} | \vec{k}))$. According to the separate universe ansatz, and from Eqs. (16) and (36),

$$R_{\mathcal{P}}(1 + \delta_{\mathcal{V}}) = \lim_{k \rightarrow 0} \frac{\langle \log(\mathcal{P}_{\text{su}}(1 + \delta_{\mathcal{V}} | \vec{k})) \delta_L(\vec{k}') \rangle}{\langle \delta_L(\vec{k}) \delta_L(\vec{k}') \rangle}. \quad (42)$$

The left-hand side is the separate universe response of the PDF, and we refer to the right-hand side as the clustering bias of the local Voronoi density PDF. To extract this from our simulation data, we first compute $R_{\mathcal{P}}(1 + \delta_{\mathcal{V}}, k)$, which is the clustering bias in the above expression prior to taking the $k \rightarrow 0$ limit. We then fit a linear polynomial in k^2 ,

$$R_{\mathcal{P}}(1 + \delta_{\mathcal{V}} | k) = b_1(1 + \delta_{\mathcal{V}}) + b_2(1 + \delta_{\mathcal{V}}) k^2. \quad (43)$$

and take the k -independent term $b_1(1 + \delta_{\mathcal{V}})$ as the linear clustering bias.

A sample of local Voronoi PDFs is shown in Fig. 5, with colors corresponding to the local large-scale density fluctuation. The underdense matter PDFs are clearly narrower, more sharply peaked, and with peak positions nearer to $\delta_{\mathcal{V}} = 0$. The PDFs in overdense subboxes are broader and have their peaks at lower densities. The shapes of the fractional differences between the local and global PDFs, shown in the bottom panel, reproduce the shape of the separate universe response. This is also true for the halos, although this is not as easy to see since the

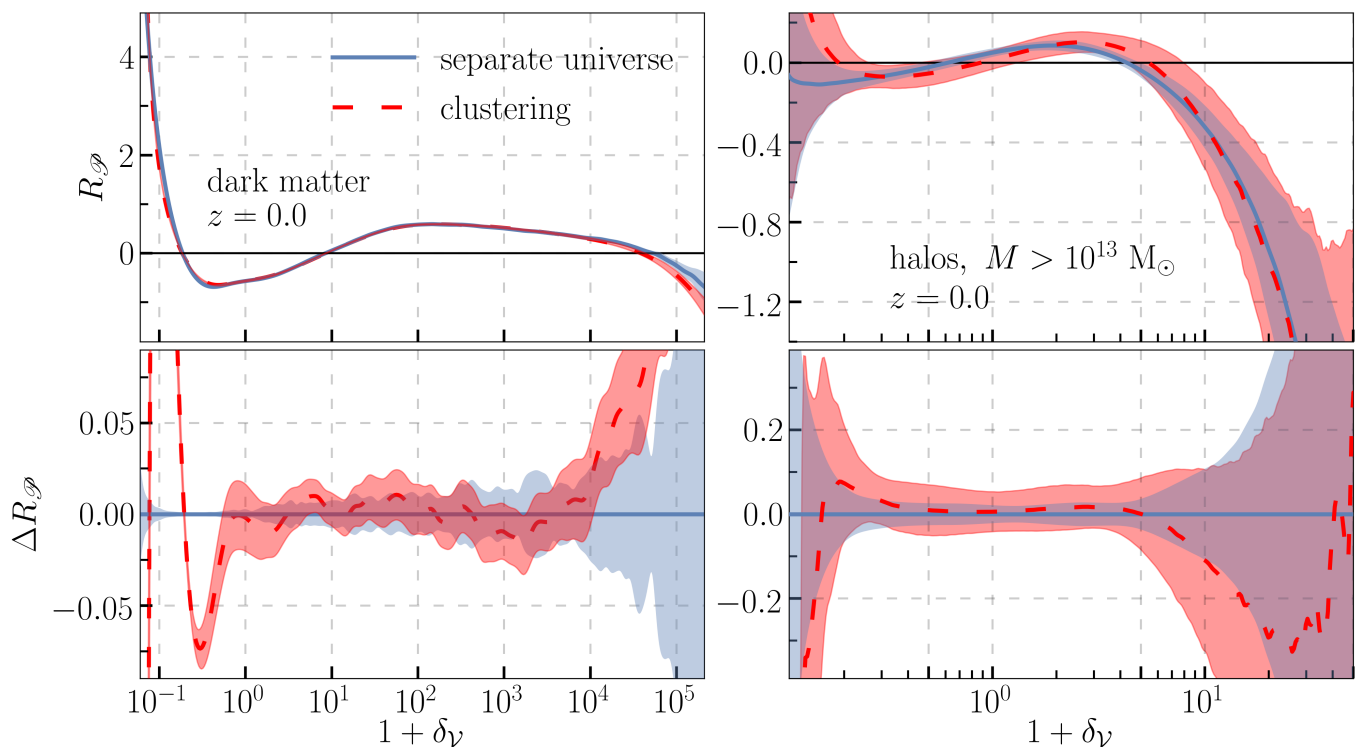


FIG. 6. Comparison between the separate universe response and the clustering bias of the Voronoi density PDFs for matter (left) and halos (right) at redshift $z = 0.0$. The difference between the separate universe responses and clustering biases are shown in the bottom panels.

response is weaker in this case, and the local PDFs are noisier due to the lower sample size.

The comparison between the PDF clustering bias and the separate universe PDF response is shown in Fig. 6. For the matter PDF, the agreement is excellent for $\delta_V > -0.5$, until around $\delta_V \simeq 2 \times 10^4$, spanning nearly five orders of magnitude of Voronoi cell densities. This clearly demonstrates that the separate universe ansatz is quite accurate over an enormous range of densities. Interestingly, the PDF response provides a test that distinguishes between overdense and underdense cells, unlike previous work that has tested the separate universe ansatz using the power spectrum [55, 56]. Since the power spectrum is quadratic, overdensities and underdensities are treated as equivalent. Our results for the matter Voronoi density PDF response represent a validation of the separate universe method across the widest range of densities to date.

The disagreement at extremely high densities is unsurprising. On the one hand, these densities are so far out into the PDF’s tail that their abundances are considerably low. Given the difference in sample sizes between the separate universe simulations and the global universe subboxes, the corresponding PDF tails must incur different statistical biases from the KDE. On the other hand, particles with the smallest Voronoi cells are subject to force softening. We set the softening

length to 20 kpc in our N-body simulations. A spherical cell with radius equal to this softening length has a density corresponding to $1 + \delta_V = 2.8 \times 10^4$. Note, we have not consistently adjusted the comoving softening length in the separate universe simulations to match the physical softening length in the global universe, so it is unsurprising to find discrepancies at these extreme densities.

For underdense cells with $\delta_V < -0.5$, the clustering bias and separate universe response disagree. This is also not surprising, since the low-density tail of the PDF drops quickly, with the very steep slope, and there is a comparatively smaller sample size of extremely underdense cells. Although the statistical bias incurred from the KDE smoothing on $\mathcal{P}_{\log}(\log(1 + \delta_V))$ is proportional to its curvature (or second derivative), transforming to $\mathcal{P}(1 + \delta_V)$ yields a contribution to the statistical bias that is proportional to the PDF’s slope. We have optimized the KDE widths based on the integrated squared error, which must be suboptimal for the steep, low-density part of the PDF. It may be possible to improve on the KDE using a more sophisticated smoothing, such as choosing local widths as a function of $1 + \delta_V$, but we leave this for future work.

In Fig. 6, we have also shown the comparison between the halo PDF clustering bias and the separate universe response. The agreement is good across the entire range

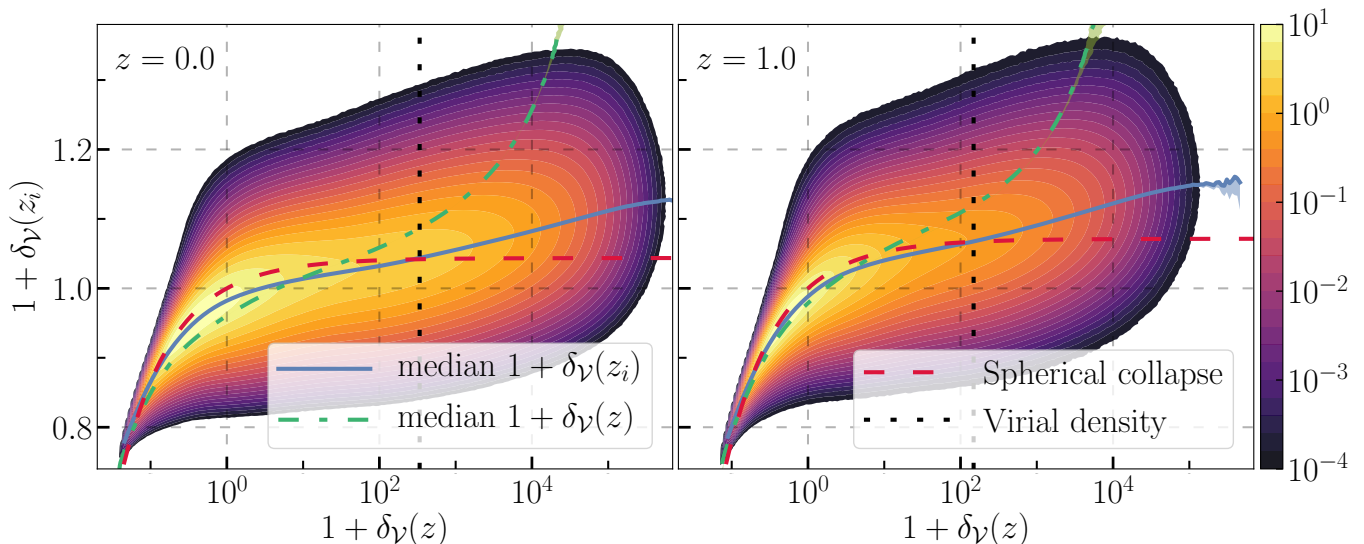


FIG. 7. 2D contour plots of the initial and final Voronoi densities of the dark matter particles at final redshifts $z = 0.0$ (left) and $z = 1.0$ (right). The contours and color bar scale correspond to the normalized bivariate PDF, $\mathcal{P}(1 + \delta_V(z_i), 1 + \delta_V(z_f))$, estimated from a 2D histogram with 200 uniformly spaced bins along the vertical axis, and 200 logarithmically spaced bins along the horizontal axis. The solid blue curves indicate the median initial density for a fixed final density, while the dash-dotted, light green curves indicate the median final density at fixed initial density. The red dashed lines are the EdS spherical collapse map. The dotted black vertical lines corresponds to the virial density.

of densities, spanning three orders of magnitude. The variance of the clustering bias is larger than that of the separate universe response. This is consistent with previous results for halo bias, void bias, and the Eulerian PDF response [16, 44, 46]. The separate universe method also yields a lower variance for the matter PDF response at densities where $\delta_V < 10^2$. At higher densities, the clustering method and separate universe method yield similar variances.

VI. OUTLOOK ON MODELING

In a previous paper, we demonstrated how the Eulerian PDF response of the smoothed matter field could be modeled accurately using only the linear power spectrum and simple, spherical collapse calculations [16]. For the Voronoi density PDF responses, the situation is much more difficult. For the halos, a typical approach might be to invoke the halo model. The large-scale correlations between halos are described using a linearly biased matter power spectrum (the two halo term), and small-scale correlations are described in terms of a radial halo profile (the one halo term). From Fig. 1, it is clear that the Voronoi tessellation probes the intermediate regime, precisely outside of a halo's interior, but nearer to the halo than its nearest neighbors. Voronoi statistics may offer a way of testing and improving halo model calculations, precisely by providing information about these intermediate scales, but this is outside the scope of this work.

A. Stochasticity and nonlinearity

For the matter PDF response, the two major challenges to providing an accurate model are due to the extreme nonlinearity of the scales it probes, and the stochasticity of Voronoi cell evolution on these scales. The typical approach to modeling a density PDF of the smoothed matter field relies on the fact that, at early enough times, the density statistics are very nearly Gaussian. Suppose we can provide a deterministic map from the late-time densities $\delta_V(t_f)$ back to the early-time densities $\delta(t_i)$,

$$F(1 + \delta_V(t_f)) = \frac{\delta_V(t_i)}{\sigma}, \quad (44)$$

where σ is the standard deviation of the early fluctuations. The quantity on the right-hand side of the above equation is drawn from a unit normal distribution. Then $\mathcal{P}(1 + \delta_V(t_f))$ is obtained by multiplying this unit normal distribution by the Jacobian of the mapping provided by F [57, 58],

$$\mathcal{P}(1 + \delta_V) = \frac{1}{\sqrt{2\pi}} \frac{dF}{d\delta_V} \exp\left(-\frac{F^2}{2}\right). \quad (45)$$

As a first approximation, one might use the analytic solution for spherical collapse (expansion) of overdensities (underdensities) in the Einstein-de Sitter cosmology. This mapping is plotted in Fig. 7 as the red dashed curves. While this approximation is an excellent starting point for modeling the Eulerian PDF, it cannot work for the Lagrangian PDF. For the Eulerian densities,

the variance σ^2 is evaluated at the Lagrangian length scale, which encompasses the late-time mass within the smoothing window. For high-density perturbations, the length scale is large, and σ is small on these large scales, so F can take arbitrarily large values.

For the Voronoi density PDF, δ_V already corresponds to the density measured on a fixed mass scale (the particle mass), so σ is constant. The spherical collapse map, however, has a maximum $\delta_V(t_i)$ corresponding to the critical density for collapse, so F is bounded from above in this case, and cannot be described by a normal distribution. The problem can be seen in Fig. 7, where a 2D contour plot of the map from late-time to early-time Voronoi densities is plotted for redshifts $z = 0.0$ and $z = 1.0$. We measured that Voronoi densities in our simulation initial conditions at redshift $z = 49$, and matched them with the late-time Voronoi densities using the unique particle ID numbers.

Clearly the mapping is highly stochastic. Also, many particles are above the critical density for spherical collapse, which corresponds to the horizontal asymptote of the spherical collapse curve. While these high-density particles do end up within collapsed objects (i.e. halos), their collapse is eventually halted by virialization. The late-time virial densities are shown in Fig. 7 as the black dashed vertical lines.

The dot-dashed curves of Fig. 7 are the median late-time densities at fixed early-time density. This quantity indicates what kind of environment a particle is likely to end up in if you know its early-time cell size. The solid blue curve shows the median early-time density at fixed late-time density. This quantity indicates which density a particle is likely to have started at if you know the size of its Voronoi cell at late times. Not only are the early densities are nearly Gaussian overall, but the distributions along vertical slices through the contour plots are also nearly Gaussian and thus symmetric about their peaks, so their mean, median, and mode coincide. The same is not true of horizontal slices through the contour plots. Along this direction, the mean curve is significantly to the left (towards late-time underdensity) of the median, and the mode is even further to the left, although we have only shown median curve.

Interestingly, the spherical collapse curve roughly approximates the median early-time density curve below the virial density. This approximation is better at earlier redshifts. The crossing point of these two curves occurs near the virial density, which evolves with redshift. The crossing between the median and spherical collapse maps occurs at Voronoi densities $1 + \delta_V = 278, 181, \text{ and } 140$ at redshifts $z = 1.0, 0.5, \text{ and } 0.0$ respectively. The virial densities at these redshifts are $337\bar{\rho}, 203\bar{\rho}, \text{ and } 148\bar{\rho}$ respectively, so the crossing is nearer to the virial density at earlier redshifts. Above the virial density, the two curves diverge as the median early-time densities increase exponentially, and the spherical map asymptotes to the critical density for collapse. Note that many of the cells within the virial radius of a halo have Voronoi densities

below the virial density of spherical collapse.

At these extremely nonlinear scales, Lagrangian perturbation theory has been shown not to offer significant improvement with regards to this stochastic mapping [33]. One way of introducing stochasticity into the spherical collapse calculation is to generalize it to ellipsoidal collapse [58]. This is an attractive approach, because the distribution of ellipsoidal regions is known analytically in a random Gaussian field [59]. However, since the dark matter cells undergo dramatic flows into halos and their surrounding environments, and out of voids, a more detailed study of the dark matter velocity field is likely required to physically motivate an accurate model [60], and we leave this for future work.

B. Downsampling validation

Since each Voronoi cell contains precisely one dark matter particle, each cell is approaching the limit of resolution achieved in the simulation. A reasonable concern is that the Voronoi statistics we measured in the dark matter could be dominated by finite-resolution artifacts. If that were the case, then we would have measured something that reveals more about the N-body simulations than it reveals about the physical matter density field that the simulations are intended to model.

To verify that this is not the case, and to test how well converged the matter PDFs are, we ran a single set of separate universe simulations with twice the number of particles, $N_p = 2 \times (1024)^3$, and randomly downsampled to 1024^3 particles. We downsampled on the particle ID numbers, so that the same particles were removed from the overdense, underdense, and global universe simulations. We then measured the separate universe response of the matter Voronoi density PDF from the higher resolution simulations.

The results are plotted in Fig. 8. For cell densities in the range $-0.8 < \delta_V < 10^4$, differences between the lower resolution and higher resolution PDFs are at most 5%, and the differences are significantly lower between $0 < \delta_V < 10^3$. While the differences are clearly smooth, and thus systematic, they are a subdominant contribution to the overall shape of the distribution. The bootstrap errors of the PDF are about 0.5% on average, and as large as 1% for $\delta_V > -0.9$. In the tails of the PDF, where $\delta_V < -0.5$ or $\delta_V > 500$, the resolution effects are larger than the bootstrap errors. At intermediate densities where $-0.5 < \delta_V < 500$, the resolution effects are smaller than or comparable to the bootstrap errors.

For the separate universe response, the agreement between the different simulation resolutions is excellent for $\delta_V > 0$. The oscillations in their difference indicate that the dominant contribution is due to statistical uncertainty. There also seems to be a small systematic error, causing the response from the lower resolution simulations to be 1–2% above the higher resolution response. For underdense cells, the response discrepancy

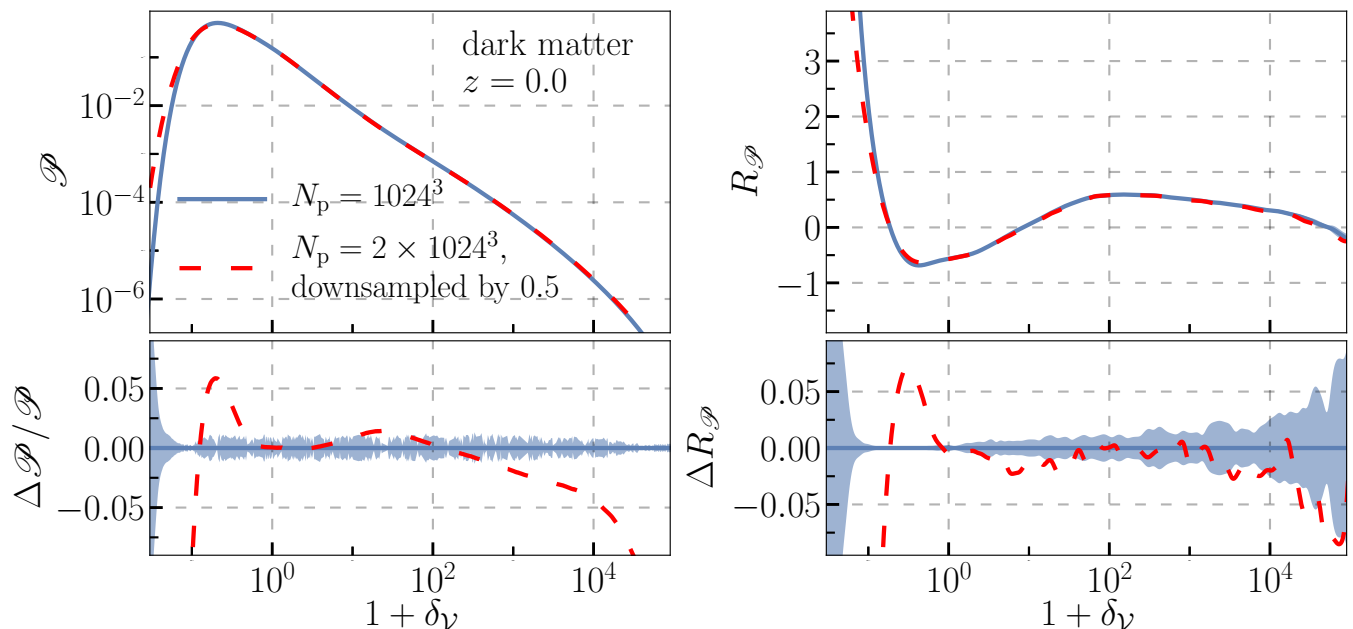


FIG. 8. Resolution convergence test for the PDF (left) and its separate universe response (right). The fractional difference between the lower resolution and higher resolution PDFs is shown on the bottom left, and the absolute difference between the two responses is shown on the bottom right.

is similar to the PDF discrepancy. Since the underdense cells, by definition, have the lowest particle-per-volume ratio, they represent regions of the simulation box where the worst resolution is achieved. This could explain why both the PDF and its separate universe response is less converged for underdense cells. This explanation is consistent with the fact that the underdense halo cells exhibit agreement between the separate universe and clustering methods, since isolated halos are less affected by changes in simulation resolution than isolated dark matter particles are.

VII. CONCLUSION

In this paper, we have measured the one-point statistics for the matter density field in N-body simulations, smoothed on an approximately uniform mass scale using Voronoi tessellation (Fig. 3). We also measured the one-point statistics for the Voronoi densities of simulation halos. Using separate universe simulations, we measured the linear response of the Voronoi density PDF to the presence of large-scale matter fluctuations (Fig. 4). We verify that this linear response accurately predicts the cross-correlation of local PDFs in simulation subboxes with the matter density perturbations in those subboxes (Fig. 6). For the matter, this verification represents a test of the separate universe method over an unprecedented range of densities. We also tested the sensitivity of these simulations results to changes in resolution (Fig. 8).

We interpreted the linear, separate universe responses of the matter PDF as the cosmic bias of regions of the density field, smoothed over a uniform mass scale, and selected based on their density. For the halos, we interpreted the separate universe PDF response as the bias of halos selected based on their Voronoi cell size. This can be thought of as a new kind of assembly bias.

By tracking the Voronoi volumes of simulation dark matter particles, we were able to determine the evolution of each individual Voronoi cell (Fig. 7). The mapping from densities of Voronoi cells at late times back to their early-time densities is highly stochastic. However, we found the mapping for a median Voronoi cell with density below the virial density is qualitatively similar to the spherical collapse mapping. Further progress on modeling the global and position-dependent Voronoi density PDFs may require considering aspherical dynamics, and studying the bulk flows of matter into halos and out of cosmic voids.

Observationally, it may be possible to extract an analogous observable to the matter Voronoi density PDF from weak lensing maps. For example, this could be achieved by partitioning a lensing map into regions with a fixed mean amount of lensing convergence. The one-point distribution of the sizes, or densities, of these regions could then be computed. For galaxy surveys, a catalog of observed galaxies can easily be Voronoi tessellated directly. However, extracting cosmological information from galaxy tessellations is limited by redshift errors, and possibly redshift space distortions [15]. The former issue may be dealt with through advanced statistical

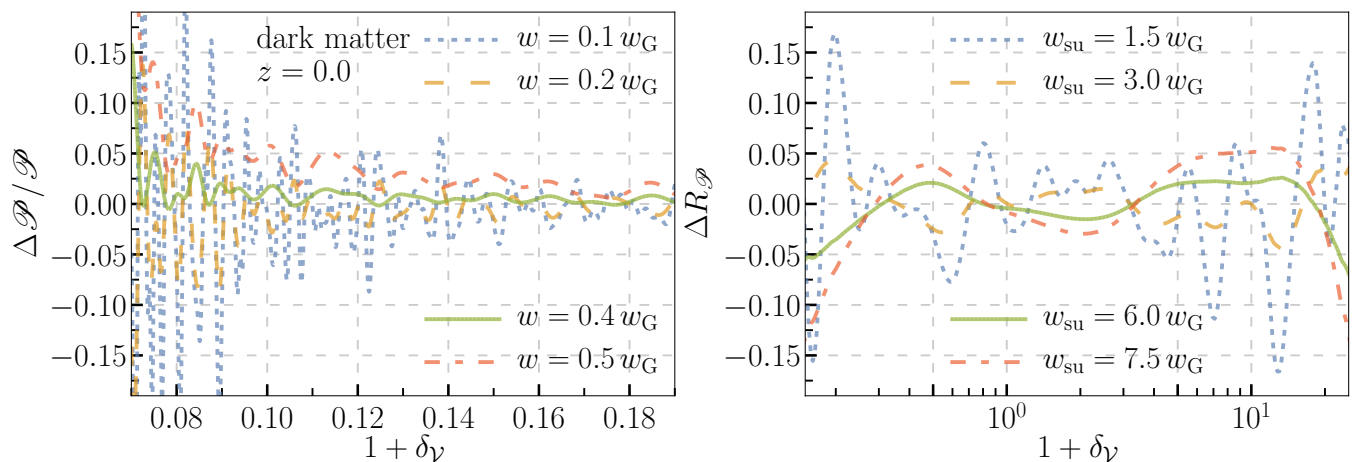


FIG. 9. Comparison of the halo Voronoi density PDF (left) and its separate universe response (right) when computed with different KDE smoothing widths. The values chosen for our reported results are $w = 0.3 w_G$ and $w_{su} = 4.5 w_G$ and the differences above are computed with respect to these reference widths. From the plots, choosing a narrower width yields results that oscillate around those obtained with our chosen widths, demonstrating they have higher variance but that our results are not statistically biased with respect to them. Choosing a wider width yields statistically biased results, which the narrower width results do not oscillate around.

methods for improving observational redshifts, such as hierarchical modeling [61, 62]. The effects of redshift space distortions need to be determined in simulations or mock observational data sets.

Our investigation was limited to adiabatic perturbations in the standard Λ CDM cosmology. In this context, long-wavelength density perturbations undergo a universal growth, which is independent of scale in the linear regime. It has been previously demonstrated that extensions to Λ CDM, including dynamical dark energy [47, 63], massive neutrinos [64], radiation [65], as well as primordial non-Gaussianity [66, 67], lead to scale-dependent effects on large scales. Such effects would also appear in the shapes of PDF responses considered here, and cause them to become scale dependent on large scales. The Voronoi density PDF and its linear response may be a useful observable for targeting specific regions or features of the density field that are sensitive to these scale-dependent effects.

ACKNOWLEDGMENTS

We would like to thank Matias Zaldarriaga and Ben Wandelt for useful discussions concerning this work. Results in this paper were obtained using the high-performance computing system at the Institute for Advanced Computational Science at Stony Brook University. Our figures were made using the PYTHON package MATPLOTLIB [68], and many of our calculation were carried out using SCIPY [69] and NUMPY [70]. D.J. is supported by Grants No. NSF PHY-1620628 and DOE DE-SC0017848. M.L. is supported by Grant No. DOE DE-SC0017848.

Appendix A: KDE width selection for the PDF and its separate universe response

The choice of KDE width is important because the level of smoothing controls the variance and statistical bias of the estimator. A width that is too narrow results in excess variance, while a width that is too wide results in a highly biased estimator. According to Eq. (29), the bias of $\mathcal{P}_{\log}$ is proportional to the second derivative of the distribution. Converting to $\mathcal{P}$ using Eq. (27) contributes bias proportional to the first derivative. Since the slope of the PDF is very steep in the low-density tail, this part of the PDF is not as reliably estimated compared with higher densities.

In Fig. 9, we demonstrate how the optimal kernel width can be chosen simply by inspecting results obtained using different widths. We first estimated the PDF in the global cosmology with a KDE width that is too narrow, $w = 0.1 w_G$, which results in excess variance but negligible bias. We then recomputed the KDE using a wider smoothing width. If the KDE with a broader kernel width is not too biased, the fractional difference between the two estimators should oscillate around zero. In Fig. 9, we can clearly see that this is the case for the halo PDF using $w = 0.3 w_G$. Repeating this process with wider kernel widths, eventually the bias of the estimator becomes noticeable. For example, with $w = 0.4 w_G$, the bias is around 3% at $\delta_V = -0.93$, and with $w = 0.5 w_G$ the bias is 15% at the same Voronoi density. If we computed the fractional difference between the PDFs estimated using these wider kernels and the PDFs estimated using the narrower kernel widths, they would not oscillate around zero. The goal is to choose the widest kernel without any significant estimator bias.

We only show the results at low density in Fig. 9, since that is where the estimator bias is most severe.

We have shown the same method for kernel width selection for the separate universe PDF response in the right panel of Fig. 9. Here, we used $w = 4.5 w_G$ as our reference. The narrower KDE widths oscillate around the reference response, while the broader widths yield a noticeable estimator bias. In this case, we have computed the absolute difference instead of the fractional difference, since the response has zero crossings. The effects of statistical bias are much milder for the response, allowing for a much wider smoothing width. This is due to the partial cancellation of the estimator bias when subtracting the overdense and underdense separate universe PDFs.

For the PDF clustering biases presented in Sec. VC, the cancellation of oversmoothing effects is not as good. To optimize the kernel widths for the subboxes, we relied on Eq. (25). The global universe PDF's slope, which appears in the background shift term of Eq. (25), benefits from a similar level of estimator bias cancellation as the separate universe response. The optimal smoothing of the subbox data can then be chosen by computing the clustering for both $\mathcal{P}_{\text{loc}}$ and $\mathcal{P}_{\text{su}}$, and selecting the widest width for which Eq. (25) is satisfied within statistical error. We found a narrower width of $w_{\text{su}} = 2 w_G$ is required to satisfy this condition, which causes excess variance in the clustering method results.

Appendix B: Bias and variance cancellation for separate universe response observables

Consider an observable $\mathcal{O}$ with separate universe counterpart $\mathcal{O}_{\text{su}}(\delta_L)$ in the presence of a large-scale density perturbation δ_L . Now suppose we have a biased estimator $\hat{\mathcal{O}}_{\text{su}}$ for $\mathcal{O}_{\text{su}}$ such that,

$$\hat{\mathcal{O}}_{\text{su}}(\delta_L) = \mathcal{O} + \Delta\mathcal{O} + \mathcal{O}(R_{\mathcal{O}} + \Delta R_{\mathcal{O}})\delta_L, \quad (\text{B1})$$

where $\mathcal{O}$ represents the true mean value of the observable in the global cosmology and $R_{\mathcal{O}}$ is the true value of its linear response. The term $\Delta\mathcal{O}$ represents errors that are uncorrelated with δ_L while $\Delta R_{\mathcal{O}}$ represents errors that are linearly correlated with δ_L . Since the estimator is biased, the mean values of $\Delta\mathcal{O}$ and $\Delta R_{\mathcal{O}}$ over many samples do not vanish. Computing the estimator for the separate universe response,

$$\begin{aligned} \hat{R}_{\mathcal{O}} &= \frac{\hat{\mathcal{O}}_{\text{su}}(+\delta_L) - \hat{\mathcal{O}}_{\text{su}}(-\delta_L)}{2\delta_L\hat{\mathcal{O}}} \\ &\simeq R_{\mathcal{O}} + \Delta R_{\mathcal{O}} - \frac{R_{\mathcal{O}}}{\mathcal{O}}\Delta\mathcal{O}. \end{aligned} \quad (\text{B2})$$

The bias and variance of this estimator are,

$$\langle \hat{R}_{\mathcal{O}} - R_{\mathcal{O}} \rangle \simeq \langle \Delta R_{\mathcal{O}} \rangle - \frac{R_{\mathcal{O}}}{\mathcal{O}}\langle \Delta\mathcal{O} \rangle, \quad (\text{B3})$$

$$\begin{aligned} \langle \hat{R}_{\mathcal{O}}^2 \rangle - \langle \hat{R}_{\mathcal{O}} \rangle^2 &= \sigma^2(\Delta R_{\mathcal{O}}) + \left(\frac{R_{\mathcal{O}}}{\mathcal{O}} \right)^2 \sigma^2(\Delta\mathcal{O}) \\ &\quad - 2 \frac{R_{\mathcal{O}}}{\mathcal{O}} \sigma(\Delta R_{\mathcal{O}}) \sigma(\Delta\mathcal{O}) \rho(\Delta\mathcal{O}, \Delta R_{\mathcal{O}}), \end{aligned} \quad (\text{B4})$$

where $\sigma^2(\Delta\mathcal{O})$ and $\sigma^2(\Delta R_{\mathcal{O}})$ are the variances of $\Delta\mathcal{O}$ and $\Delta R_{\mathcal{O}}$ respectively, and $\rho(\Delta\mathcal{O}, \Delta R_{\mathcal{O}})$ is their correlation coefficient. In the situations where,

$$\langle \Delta R_{\mathcal{O}} \rangle \mathcal{O} \simeq \langle \Delta\mathcal{O} \rangle R_{\mathcal{O}}, \quad (\text{B5})$$

there is significant cancellation of the estimator bias of $\hat{R}_{\mathcal{O}}$. In situations where,

$$\rho(\Delta\mathcal{O}, \Delta R_{\mathcal{O}}) \simeq 1, \quad (\text{B6})$$

there is significant cancellation of the variance of $\hat{R}_{\mathcal{O}}$.

Whether or not these cancellations occur in practice depends on the details of the observable and the estimator under consideration. For example, in Fig. 8, over the range of densities $1 < 1 + \delta_V < 10$, we see that the finite resolution effects give negligible systematic errors to the one-point statistics in the global cosmology. Over the same range of densities, the separate universe response does have a systemically lower response in the higher resolution simulation. Over the range of densities $10 < 1 + \delta_V < 100$, there is a systematic error in the shape of the global PDF, while for the separate universe response, the systematic errors have canceled. At low densities, $\delta_V < 1$, the systematic error for the response is well approximate by the fractional systematic error for the global universe PDF, which suggests $\langle \Delta R_{\mathcal{P}} \rangle \mathcal{P} \simeq \langle \Delta\mathcal{P} \rangle (R_{\mathcal{P}} + 1)$ in this regime. Predicting whether or not there will be any bias or variance cancellation requires an accurate model for both the global universe observable and its separate universe response.

-
- [1] Ž. Ivezić *et al.* (LSST), LSST: from Science Drivers to Reference Design and Anticipated Data Products, *Astrophys. J.* **873**, 111 (2019), arXiv:0805.2366 [astro-ph].
 - [2] R. Laureijs *et al.* (EUCLID), Euclid Definition Study Report, (2011), arXiv:1110.3193 [astro-ph.CO].
 - [3] D. Spergel *et al.*, Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report, (2015), arXiv:1503.03757 [astro-ph.IM].
 - [4] A. Aghamousa *et al.* (DESI), The DESI Experiment Part I: Science, Targeting, and Survey Design, (2016), arXiv:1611.00036 [astro-ph.IM].
 - [5] N. Aghanim *et al.* (Planck), Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, A6 (2020), arXiv:1807.06209 [astro-ph.CO].
 - [6] G. D’Amico, J. Gleyzes, N. Kokron, K. Markovic, L. Senatore, P. Zhang, F. Beutler, and H. Gil-Marín, The Cosmological Analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure, *JCAP* **05**, 005, arXiv:1909.05271 [astro-ph.CO].
 - [7] T. Colas, G. D’Amico, L. Senatore, P. Zhang, and F. Beutler, Efficient Cosmological Analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure, *JCAP* **06**, 001, arXiv:1909.07951 [astro-ph.CO].
 - [8] M. M. Ivanov, M. Simonović, and M. Zaldarriaga, Cosmological Parameters and Neutrino Masses from the Final Planck and Full-Shape BOSS Data, *Phys. Rev. D* **101**, 083504 (2020), arXiv:1912.08208 [astro-ph.CO].
 - [9] O. H. Philcox, M. M. Ivanov, M. Simonović, and M. Zaldarriaga, Combining Full-Shape and BAO Analyses of Galaxy Power Spectra: A 1.6% CMB-independent constraint on H_0 , *JCAP* **05**, 032, arXiv:2002.04035 [astro-ph.CO].
 - [10] R. van de Weygaert, Fragmenting the universe. 3. The Construction and statistics of 3-D Voronoi tessellations, (1993).
 - [11] R. van de Weygaert and W. Schaap, The Cosmic Web: Geometric Analysis, *Lect. Notes Phys.* **665**, 291 (2009), arXiv:0708.1441 [astro-ph].
 - [12] M. A. Aragon-Calvo, B. J. Jones, R. van de Weygaert, and M. van der Hulst, The Multiscale Morphology Filter: Identifying and Extracting Spatial Patterns in the Galaxy Distribution, *Astron. Astrophys.* **474**, 315 (2004), arXiv:0705.2072 [astro-ph].
 - [13] P. Busch and S. D. M. White, The Tessellation-Level-Tree: characterizing the nested hierarchy of density peaks and their spatial distribution in cosmological N-body simulations, *Mon. Not. Roy. Astron. Soc.* **493**, 5693 (2020), arXiv:1910.08829 [astro-ph.CO].
 - [14] S. Shandarin, S. Habib, and K. Heitmann, The Cosmic Web, Multi-Stream Flows, and Tessellations, *Phys. Rev. D* **85**, 083005 (2012), arXiv:1111.2366 [astro-ph.CO].
 - [15] M. C. Cooper, J. A. Newman, D. S. Madgwick, B. F. Gerke, R.-B. Yan, and M. Davis, Measuring galaxy environments with deep redshift surveys, *Astrophys. J.* **634**, 833 (2005), arXiv:astro-ph/0506518.
 - [16] D. Jamieson and M. Loverde, The position-dependent matter density PDF, (2020), arXiv:2010.07235 [astro-ph.CO].
 - [17] P. J. E. Peebles, *The large-scale structure of the universe* (1980).
 - [18] H. Gil-Marín, J. Noreña, L. Verde, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider, The power spectrum and bispectrum of SDSS DR11 BOSS galaxies – I. Bias and gravity, *Mon. Not. Roy. Astron. Soc.* **451**, 539 (2015), arXiv:1407.5668 [astro-ph.CO].
 - [19] H. Gil-Marín, L. Verde, J. Noreña, A. J. Cuesta, L. Samushia, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider, The power spectrum and bispectrum of SDSS DR11 BOSS galaxies – II. Cosmological interpretation, *Mon. Not. Roy. Astron. Soc.* **452**, 1914 (2015), arXiv:1408.0027 [astro-ph.CO].
 - [20] H. Gil-Marín, W. J. Percival, L. Verde, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, S. A. Rodríguez-Torres, and M. D. Olmstead, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies, *Mon. Not. Roy. Astron. Soc.* **465**, 1757 (2017), arXiv:1606.00439 [astro-ph.CO].
 - [21] Z. Slepian *et al.*, Detection of baryon acoustic oscillation features in the large-scale three-point correlation function of SDSS BOSS DR12 CMASS galaxies, *Mon. Not. Roy. Astron. Soc.* **469**, 1738 (2017), arXiv:1607.06097 [astro-ph.CO].
 - [22] T. Abbott *et al.* (DES), Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing, *Phys. Rev. D* **98**, 043526 (2018), arXiv:1708.01530 [astro-ph.CO].
 - [23] D. W. Pearson and L. Samushia, A Detection of the Baryon Acoustic Oscillation features in the SDSS BOSS DR12 Galaxy Bispectrum, *Mon. Not. Roy. Astron. Soc.* **478**, 4500 (2018), arXiv:1712.04970 [astro-ph.CO].
 - [24] H. Gil-Marín *et al.*, The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0, *Mon. Not. Roy. Astron. Soc.* **498**, 2492 (2020), arXiv:2007.08994 [astro-ph.CO].

- [25] A. Yang and W. C. Saslaw, The galaxy counts-in-cells distribution from the SDSS, *Astrophys. J.* **729**, 123 (2011), arXiv:1009.0013 [astro-ph.CO].
- [26] J. Bel and C. Marinoni, Determination of the abundance of cosmic matter via the cell count moments of the galaxy distribution, *Astron. Astrophys.* **563**, A36 (2014), arXiv:1310.3196 [astro-ph.CO].
- [27] L. Hurtado-Gil *et al.*, The best fit for the observed galaxy counts-in-cell distribution function, *Astronomy & Astrophysics* **601**, A40 (2017).
- [28] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles* (1988).
- [29] Y. Jing, Correcting for the alias effect when measuring the power spectrum using FFT, *Astrophys. J.* **620**, 559 (2005), arXiv:astro-ph/0409240.
- [30] M. C. Neyrinck, ZOBOV: a parameter-free void-finding algorithm, *Mon. Not. Roy. Astron. Soc.* **386**, 2101 (2008), arXiv:0712.3049 [astro-ph].
- [31] P. Sutter, G. Lavaux, N. Hamaus, A. Pisani, B. D. Elt, M. S. Warren, F. Villaescusa-Navarro, P. Zivick, Q. Mao, and B. B. Thompson, VIDE: The Void IDentification and Examination toolkit, *Astron. Comput.* **9**, 1 (2015), arXiv:1406.1191 [astro-ph.CO].
- [32] F. Bernardeau and R. van de Weygaert, A New method for accurate velocity statistics estimation, *Mon. Not. Roy. Astron. Soc.* **279**, 693 (1996), arXiv:astro-ph/9508125.
- [33] M. C. Neyrinck, Quantifying distortions of the Lagrangian dark-matter mesh in cosmology, *Mon. Not. Roy. Astron. Soc.* **428**, 141 (2013), arXiv:1204.1326 [astro-ph.CO].
- [34] A. Paranjape and S. Alam, Voronoi volume function: A new probe of cosmology and galaxy evolution, *Mon. Not. Roy. Astron. Soc.* **495**, 3233 (2020), arXiv:2001.08760 [astro-ph.CO].
- [35] K. C. Chan and N. Hamaus, Volume Statistics as a Probe of Large-Scale Structure, (2020), arXiv:2010.13955 [astro-ph.CO].
- [36] G. Lejeune Dirichlet, Über die reduction der positiven quadratischen formen mit drei unbestimmten ganzen zahlen., *Journal für die reine und angewandte Mathematik* **40**, 209 (1850).
- [37] G. Voronoi, Nouvelles applications des paramètres continus à la théorie des formes quadratiques. premier mémoire. sur quelques propriétés des formes quadratiques positives parfaites., *J. Reine Angew. Math* **133**, 97 (1908).
- [38] E. Sirko, Initial conditions to cosmologicaln-body simulations, or, how to run an ensemble of simulations, *The Astrophysical Journal* **634**, 728–743 (2005).
- [39] W. Hu, C.-T. Chiang, Y. Li, and M. LoVerde, Separating the Universe into real and fake energy densities, *Phys. Rev. D* **94**, 023002 (2016), arXiv:1605.01412 [astro-ph.CO].
- [40] F. R. Bouchet, S. Colombi, E. Hivon, and R. Juszkiewicz, Perturbative Lagrangian approach to gravitational instability, *Astron. Astrophys.* **296**, 575 (1995), arXiv:astro-ph/9406013.
- [41] M. Crocce, S. Pueblas, and R. Scoccimarro, Transients from Initial Conditions in Cosmological Simulations, *Mon. Not. Roy. Astron. Soc.* **373**, 369 (2006), arXiv:astro-ph/0606505 [astro-ph].
- [42] D. Blas, J. Lesgourgues, and T. Tram, The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes, *JCAP* **1107**, 034, arXiv:1104.2933 [astro-ph.CO].
- [43] V. Springel, The Cosmological simulation code GADGET-2, *Mon. Not. Roy. Astron. Soc.* **364**, 1105 (2005), arXiv:astro-ph/0505010 [astro-ph].
- [44] D. Jamieson and M. Loverde, Separate Universe Void Bias, *Phys. Rev. D* **100**, 123528 (2019), arXiv:1909.05313 [astro-ph.CO].
- [45] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, The Rockstar Phase-Space Temporal Halo Finder and the Velocity Offsets of Cluster Cores, *Astrophys. J.* **762**, 109 (2013), arXiv:1110.4372 [astro-ph.CO].
- [46] Y. Li, W. Hu, and M. Takada, Separate universe consistency relation and calibration of halo bias, *Physical Review D* **93**, 10.1103/physrevd.93.063507 (2016).
- [47] D. Jamieson and M. Loverde, Quintessential isocurvature in separate universe simulations, *Physical Review D* **100**, 10.1103/physrevd.100.023516 (2019).
- [48] C. H. Rycroft, Voro++: A three-dimensional voronoi cell library in c++, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **19**, 041111 (2009), <https://doi.org/10.1063/1.3215722>.
- [49] F. Bernardeau and L. Kofman, Properties of the cosmological density distribution function, *Astrophys. J.* **443**, 479 (1995), arXiv:astro-ph/9403028.
- [50] P. Coles and B. Jones, A Lognormal model for the cosmological mass distribution, *Mon. Not. Roy. Astron. Soc.* **248**, 1 (1991).
- [51] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, London, 1986).
- [52] T. Baldauf *et al.*, Linear response to long wavelength fluctuations using curvature simulations, *Journal of Cosmology and Astroparticle Physics* **2016** (09), 007–007.
- [53] K. C. Chan, Y. Li, M. Biagetti, and N. Hamaus, Measurement of Void Bias using Separate Universe Simulations, (2019), arXiv:1909.03736 [astro-ph.CO].
- [54] M. Frigo and S. G. Johnson, The design and implementation of fftw3, *Proceedings of the IEEE* **93**, 216 (2005).
- [55] Y. Li, W. Hu, and M. Takada, Super-sample covariance in simulations, *Physical Review D* **89**, 10.1103/physrevd.89.083519 (2014).

- [56] C.-T. Chiang *et al.*, Position-dependent power spectrum of the large-scale structure: a novel method to measure the squeezed-limit bispectrum, *Journal of Cosmology and Astroparticle Physics* **2014** (05), 048–048.
- [57] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, Large scale structure of the universe and cosmological perturbation theory, *Phys. Rept.* **367**, 1 (2002), arXiv:astro-ph/0112551.
- [58] T. Lam and R. Sheth, Perturbation theory and excursion set estimates of the probability distribution function of dark matter, and a method for reconstructing the initial distribution function, *Mon. Not. Roy. Astron. Soc.* **386**, 407 (2008), arXiv:0711.5029 [astro-ph].
- [59] A. G. Doroshkevich, The space structure of perturbations and the origin of rotation of galaxies in the theory of fluctuation., *Astrofizika* **6**, 581 (1970).
- [60] E. Jennings and D. Jennings, Non-linear stochastic growth rates and redshift space distortions, *Mon. Not. Roy. Astron. Soc.* **449**, 3407 (2015), arXiv:1502.02052 [astro-ph.CO].
- [61] J. Jasche and B. D. Wandelt, Bayesian inference from photometric redshift surveys, *Mon. Not. Roy. Astron. Soc.* **425**, 1042 (2012), arXiv:1106.2757 [astro-ph.CO].
- [62] A. I. Malz and D. W. Hogg, How to obtain the redshift distribution from probabilistic redshift estimates, (2020), arXiv:2007.12178 [astro-ph.CO].
- [63] C.-T. Chiang *et al.*, Quintessential scale dependence from separate universe simulations, *Physical Review D* **94**, 10.1103/physrevd.94.123502 (2016).
- [64] C.-T. Chiang *et al.*, Scale-dependent bias and bispectrum in neutrino separate universe simulations, *Physical Review D* **97**, 10.1103/physrevd.97.123526 (2018).
- [65] C. Shiveshwarkar, D. Jamieson, and M. Loverde, .
- [66] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects, *Phys. Rev. D* **77**, 123514 (2008), arXiv:0710.4560 [astro-ph].
- [67] A. Barreira, G. Cabass, F. Schmidt, A. Pillepich, and D. Nelson, Galaxy bias and primordial non-Gaussianity: insights from galaxy formation simulations with IllustrisTNG, *JCAP* **12**, 013, arXiv:2006.09368 [astro-ph.CO].
- [68] J. D. Hunter, Matplotlib: A 2d graphics environment, *Computing in Science & Engineering* **9**, 90 (2007).
- [69] P. Virtanen *et al.*, Scipy 1.0: fundamental algorithms for scientific computing in python, *Nature Methods* **17**, 261–272 (2020).
- [70] C. R. Harris *et al.*, Array programming with numpy, *Nature* **585**, 357–362 (2020).