Halo stochasticity from exclusion and nonlinear clustering

Baldauf, Tobias; Seljak, Uroš; Smith, Robert E.; Hamaus, Nico; Desjacques, Vincent

PHYSICAL REVIEW D **88,** 083507 (2013)

Halo stochasticity from exclusion and nonlinear clustering

Tobias Baldauf,[1,*] Uroš Seljak,[1,2,3] Robert E. Smith,[4] Nico Hamaus,[5,6] and Vincent Desjacques[7]

[1]*Institut für Theoretische Physik, Universität Zürich, 8057 Zürich, Switzerland*
[2]*Physics Department and Lawrence Berkeley National Laboratory, University of California, Berkeley 94720, California, USA*
[3]*Institute for the Early Universe, Ewha University, Seoul 120-750, South Korea*
[4]*Max-Planck-Institut für Astrophysik, 85748 Garching, Germany*
[5]*Institut d'Astrophysique de Paris, Université Pierre et Marie Curie, Paris 75014, France*
[6]*Department of Physics, University of Illinois at Urbana-Champaign, Urbana 61801, Illinois, USA*
[7]*Département de Physique Théorique & Center for Astroparticle Physics, Université de Genève, 1211 Genève, Switzerland*

(Received 21 May 2013; published 10 October 2013)

The clustering of galaxies in ongoing and upcoming galaxy surveys contains a wealth of cosmological information, but extracting this information is a nontrivial task since galaxies and their host haloes are stochastic tracers of the nonlinear matter density field. This stochasticity is usually modeled as the Poisson shot noise, which is constant as a function of wave number with amplitude given by $1/\bar{n}$, where $\bar{n}$ is the number density of galaxies. Here we use dark matter haloes in N-body simulations to show evidence for deviations from this simple behavior and develop models that explain the behavior of the stochasticity on large scales. First, haloes are extended, nonoverlapping objects, i.e., their correlation function needs to go to -1 on small scales. This leads to a negative correction to the stochasticity relative to the Poisson value at low wave number k, decreasing to zero for wave numbers large compared to the inverse exclusion scale. Second, haloes show a nonlinear enhancement of clustering outside the exclusion scale, leading to a positive stochasticity correction. Both of these effects go to zero for high k, making the stochasticity scale dependent even for $k < 0.1h$ Mpc^{-1}. We show that the corrections in the low-k regime are the same in Eulerian and Lagrangian space, but that the transition scale is pushed to smaller scales for haloes observed at present time (Eulerian space), relative to the initial conditions (Lagrangian space). These corrections vary with halo mass, and we present approximate scalings with halo mass and redshift. We also discuss simple applications of these effects to galaxy samples with nonvanishing satellite fraction, where the stochasticity can again deviate strongly from the fiducial Poisson expectation. Overall, these effects affect the clustering of galaxies at a level of a few percent even on very large scales and need to be modeled properly if we want to extract high precision cosmological information from the upcoming galaxy redshift surveys.

DOI: 10.1103/PhysRevD.88.083507 PACS numbers: 98.80.−k, 98.65.Dx, 98.80.Es

I. INTRODUCTION

The three-dimensional distribution of galaxies has the potential to tell us a lot about the physics governing our Universe. However, the imprint of the composition and history of our Universe on its structure is usually quantified in terms of the linear power spectrum. From there it is a fair way to go to connect to the distribution of luminous objects. Due to the stochastic nature of the initial conditions, the comparison between theory and observation has to be made at a statistical level. Thus, it has become common practice to reduce the data to n-spectra and to find a way to push the theory as far as possible in order to make predictions for the observed spectra. This means that the theoretical prediction needs to account for the fact that galaxies are only sampling the underlying matter distribution. While their distribution is clearly related to the matter distribution, there are a number of distinct features present in the galaxy distribution that are related to their discrete nature and the fact that galaxies form preferentially in high-density regions.

Due to the complicated nature of galaxy formation, cosmological constraints from galaxy surveys are usually obtained using a bias model [1]. The simplest local bias models [2] assume a proportionality between the galaxy and matter overdensities. As we will review in detail below in Sec. II A, the auto-power spectrum of a sample of N particles in a volume V is expected to have an additional *scale-independent* shot noise (SN) component V/N. We will refer to this Poisson prediction as fiducial stochasticity. On top of the fiducial Poisson shot noise, there are further contributions to the halo power spectrum that are white only over a limited range of wave numbers and lead to modifications in the $k \to 0$ limit. The latter will be referred to as stochasticity corrections. For instance, the studies of [3,4] found evidence for a sub-Poissonian noise in the halo distribution in N-body simulations (see also [5,6]) and used this concept to increase the information content extractable from surveys by weighting haloes accordingly. Subsequently, this approach was used to

*baldauf@physik.uzh.ch

improve constraints on primordial non-Gaussianity [7] and redshift space distortions [8].

Thus far, the origin of these stochasticity corrections has not been understood consistently. However, some authors noted that realistic bias models would at some point need to account for the finite size of haloes and the resulting exclusion effects [9]. The effect of halo exclusion on the power spectrum was previously discussed in [10] in a an Eulerian setting. Here, we will argue that the exclusion effect cannot be seen in isolation but has to be combined with the nonlinear clustering, which can lead to positive corrections on large scales. This approach partially alleviates the longstanding problem of nonvanishing contributions of the perturbative bias model on the largest scales, where perturbative corrections are considered unphysical. This paper aims at shedding light on the stochasticity properties of halo and galaxy samples and tries to quantify them where possible.

The paper breaks down as follows: We begin in Sec. II with a short review of the standard Poisson shot noise for a sample of discrete tracers. Then, in Sec. III we consider some simple toy models to understand the effects of exclusion on the power spectrum, before we go on to discuss more realistic models for the clustering of dark matter haloes in Sec. IV. In Sec. V we study the stochasticity and correlation function for a sample of dark matter haloes and a Halo Occupation Distribution (HOD) galaxy sample in N-body simulations. Finally, we summarize our findings in Sec. VI.

II. DISCRETE TRACERS

A. Correlation and power spectrum

The overdensity of discrete tracer particles (dark matter haloes, galaxies, etc.) can generically be written as

$$\delta^{(\mathrm{d})}(\boldsymbol{r}) = \frac{n(\boldsymbol{r})}{\bar{n}} - 1 = \frac{1}{\bar{n}}\sum_i \delta^{(\mathrm{D})}(\boldsymbol{r} - \boldsymbol{r}_i) - 1, \tag{1}$$

where $\bar{n}$ is the mean number density of the pointlike objects, whereas $n(\boldsymbol{r})$ is their local number density. The two-point correlation of this fluctuation field is the expectation value

$$\begin{aligned}\langle \delta^{(\mathrm{d})}(\boldsymbol{r})\delta^{(\mathrm{d})}(0)\rangle &= \frac{1}{\bar{n}^2}\left\langle \sum_{i,j}\delta^{(\mathrm{D})}(\boldsymbol{r}-\boldsymbol{r}_i)\delta^{(\mathrm{D})}(\boldsymbol{r}_j)\right\rangle - \frac{1}{\bar{n}}\left\langle \sum_i \delta^{(\mathrm{D})}(\boldsymbol{r}-\boldsymbol{r}_i)\right\rangle - \frac{1}{\bar{n}}\left\langle \sum_j \delta^{(\mathrm{D})}(\boldsymbol{r}_j)\right\rangle + 1 \\ &= \frac{1}{\bar{n}^2}\delta^{(\mathrm{D})}(\boldsymbol{r})\left\langle \sum_i \delta^{(\mathrm{D})}(\boldsymbol{r}-\boldsymbol{r}_i)\right\rangle + \frac{1}{\bar{n}^2}\left\langle \sum_{i\neq j}\delta^{(\mathrm{D})}(\boldsymbol{r}-\boldsymbol{r}_i)\delta^{(\mathrm{D})}(\boldsymbol{r}_j)\right\rangle - 1 \\ &= \frac{1}{\bar{n}}\delta^{(\mathrm{D})}(\boldsymbol{r}) + \frac{1}{\bar{n}^2}\left\langle \sum_{i\neq j}\delta^{(\mathrm{D})}(\boldsymbol{r}-\boldsymbol{r}_i)\delta^{(\mathrm{D})}(\boldsymbol{r}_j)\right\rangle - 1 \\ &= \frac{1}{\bar{n}}\delta^{(\mathrm{D})}(\boldsymbol{r}) + \xi^{(\mathrm{d})}(r).\end{aligned} \tag{2}$$

We split the sum into an $i = j$ and an $i \neq j$ part, corresponding to the correlation of the discrete particles with themselves and the correlation between different particles, respectively. The second term in the last equality is the reduced two-point correlation function of the tracers. The first term arises owing to "self-pairs," which are usually ignored in the calculation of real space correlations. Taking the Fourier transform of the last expression, the power spectrum of the discrete tracers is

$$P^{(\mathrm{d})}(k) = \frac{1}{\bar{n}} + \int \mathrm{d}^3 r \xi^{(\mathrm{d})}(r)\exp[\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}]. \tag{3}$$

Self-pairs contribute the usual Poisson white noise $1/\bar{n}$. The only requirement is that the power spectrum be positive definite. This implies that the Fourier transform of the two-point correlation $\xi^{(\mathrm{d})}(r)$ can be anything equal or greater than $-1/\bar{n}$. In the limit $k \to 0$, in particular, the power spectrum tends towards

$$P^{(\mathrm{d})}(k) \xrightarrow{k\to 0} \frac{1}{\bar{n}} + \int \mathrm{d}^3 r \xi^{(\mathrm{d})}(r), \tag{4}$$

where the integral of $\xi^{(\mathrm{d})}(r)$ over the whole space can be positive, zero, or negative (but greater than $-1/\bar{n}$) depending on the nature of the discrete tracers. This can lead to super-Poisson or sub-Poisson white noise in the low-k limit.

At $k = 0$, the power spectrum is $P^{(\mathrm{d})}(0) = 0$ because the fluctuation field $\delta^{(\mathrm{d})}(\boldsymbol{r})$ is defined relative to the mean number density; hence, $\langle\delta^{(\mathrm{d})}\rangle = 0$. This implies that $P^{(\mathrm{d})}(k)$ drops precipitously on very large scales (so it must be discontinuous at $k = 0$) regardless of the value of $\int \mathrm{d}^3 r \xi^{(\mathrm{d})}(r)$. To convince ourselves that this is indeed the case, we can write the Fourier modes of the tracer fluctuation field as

$$\delta^{(\mathrm{d})}(\boldsymbol{k}) = \frac{1}{\bar{n}}\sum_i \exp[\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i] - \int \mathrm{d}^3 r \exp[\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}]. \tag{5}$$

To calculate $\delta^{(\mathrm{d})}$ $(k = 0)$ (which is formally the difference between two infinite quantities), we first assume $N, V \gg 1$ at

fixed average number density $\bar{n} \equiv N/V$ and then take the limit $N, V \to \infty$. We thus have, for the Fourier transform of the density field,

$$\delta^{(\mathrm{d})}(\boldsymbol{k}) = \frac{1}{\bar{n}} \sum_i \exp[\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r}_i] - V\delta^{(\mathrm{K})}_{\boldsymbol{k},\boldsymbol{0}}, \tag{6}$$

which for $\boldsymbol{k} = 0$ yields

$$\delta^{(\mathrm{d})}(\boldsymbol{0}) = \frac{V}{N}N - V = 0. \tag{7}$$

This obviously holds also for a finite number N of tracers in a finite volume V. Therefore, the fact that $\int \mathrm{d}^3 r \xi^{(\mathrm{d})}(r)$ can be different from zero has nothing to do with the fact that $\langle \delta^{(\mathrm{d})} \rangle = 0$, nor with the so-called "integral constraint" that appears when measuring an excess of pairs relative to a random distribution in a finite volume [11–13].

B. The effect of exclusion with clustering

Let us now account for the fact that haloes are the center of an ensemble of particles, which by definition cannot overlap, and that these are clustered. Exclusion means it is forbidden to have two haloes closer than the sum of their radii R. This fact can be accounted for by writing the correlation function of the discrete tracers as

$$\xi^{(\mathrm{d})}_{\mathrm{hh}}(r) = \begin{cases} -1 & \text{for } r < R \\ \xi^{(\mathrm{c})}_{\mathrm{hh}}(r) & \text{for } r \geq R, \end{cases} \tag{8}$$

where the fictitious continuous correlation function $\xi^{(\mathrm{c})}_{\mathrm{hh}}(r)$ is defined for $r \in [0, \infty]$ and would, for instance, be related to the matter correlation function by the local bias model (see Sec. IV A below). Enforcing this step at the exclusion radius is certainly overly simplistic, since any triaxiality or variation of radius within the sample will smooth out this step. We will come back to this issue later.

For generic continuous clustering models, we can write the Fourier transform of the correlation function as

$$\begin{aligned} &\int_0^\infty \mathrm{d}^3 r \xi^{(\mathrm{d})}_{\mathrm{hh}}(r) j_0(kr) \\ &\quad = -\int_0^R \mathrm{d}^3 r j_0(kr) + \int_R^\infty \mathrm{d}^3 r \xi^{(\mathrm{c})}_{\mathrm{hh}}(r) j_0(kr) \\ &\quad = -V_{\mathrm{excl}} W_R(k) - \int_0^R \mathrm{d}^3 r \xi^{(\mathrm{c})}_{\mathrm{hh}}(r) j_0(kr) \\ &\qquad + \int_0^\infty \mathrm{d}^3 r \xi^{(\mathrm{c})}_{\mathrm{hh}}(r) j_0(kr) \\ &\quad = -V_{\mathrm{excl}} W_R(k) - V_{\mathrm{excl}}[W_R * P^{(\mathrm{c})}_{\mathrm{hh}}](k) + P^{(\mathrm{c})}_{\mathrm{hh}}(k), \end{aligned} \tag{9}$$

where the exclusion volume is $V_{\mathrm{excl}} = 4\pi R^3/3$, j_0 is the zeroth-order spherical Bessel function, the Fourier transform of the top-hat window is given by

$$W_R(k) = 3\frac{\sin(kR) - kR\cos(kR)}{(kR)^3}, \tag{10}$$

and the notation $[A * B](k)$ describes a convolution integral

$$[A * B](k) = \int \frac{\mathrm{d}^3 q}{(2\pi)^3} A(q) B(|\boldsymbol{k} - \boldsymbol{q}|). \tag{11}$$

We also defined the continuous power spectrum as the full Fourier transform of the continuous correlation function

$$P^{(\mathrm{c})}_{\mathrm{hh}}(k) = \int_0^\infty \mathrm{d}^3 r \xi^{(\mathrm{c})}_{\mathrm{hh}}(r) j_0(kr). \tag{12}$$

Combining the above results with the fiducial stochasticity contribution, we finally have, for the power spectrum of the discrete tracers,

$$P^{(\mathrm{d})}_{\mathrm{hh}}(k) = \frac{1}{\bar{n}} + P^{(\mathrm{c})}_{\mathrm{hh}}(k) - V_{\mathrm{excl}} W_R(k) - V_{\mathrm{excl}}[W_R * P^{(\mathrm{c})}_{\mathrm{hh}}](k). \tag{13}$$

This equation is the basis of our paper, and we will thus explore it in detail.

It is common practice to ignore the exclusion window and to approximate the continuous power spectrum by the linear local bias model, which yields, for the power spectrum of the discrete tracers in the Poisson model,

$$P^{(\mathrm{d})}_{\mathrm{hh}}(k) = \frac{1}{\bar{n}} + b_1^2 P_{\mathrm{lin}}(k). \tag{14}$$

This needs to be modified because of exclusion and nonlinear effects. In practice, for $k > 0$, it is difficult to separate the effects. Here we will formally define the stochasticity effects discussed in this paper as a stochasticity power spectrum [4],

$$\begin{aligned} &(2\pi)^3 \delta^{(\mathrm{D})}(\boldsymbol{k} + \boldsymbol{k}') C(k) \\ &\quad = \langle [\delta_{\mathrm{h}}(\boldsymbol{k}) - b_1 \delta_{\mathrm{m}}(\boldsymbol{k})][\delta_{\mathrm{h}}(\boldsymbol{k}') - b_1 \delta_{\mathrm{m}}(\boldsymbol{k}')] \rangle \\ &\quad = (2\pi)^3 \delta^{(\mathrm{D})}(\boldsymbol{k} + \boldsymbol{k}')[P_{\mathrm{hh}}(k) - 2b_1 P_{\mathrm{hm}}(k) + b_1^2 P_{\mathrm{mm}}(k)], \end{aligned} \tag{15}$$

where $b_1 = P_{\mathrm{hm}}(k)/P_{\mathrm{mm}}(k)$ is the first-order bias from the cross correlation in the low-k limit. We then have in the low-k limit,

$$P^{(\mathrm{d})}_{\mathrm{hh}}(k) = C(k) + b_1^2 P_{\mathrm{lin}}(k). \tag{16}$$

One could make this generally valid at all k by defining $b(k) = P_{\mathrm{hm}}(k)/P_{\mathrm{mm}}(k)$, but we will not do this here and instead will explore physically motivated models of nonlinear bias. In this paper we are interested in the stochasticity power spectrum $C(k)$ and, in particular, its limit as $k \to 0$.

What are the corrections arising from the exclusion and deviations from the local bias model? In the low-k limit, the window function scales as $W_R(k) \xrightarrow{k \to 0} 1 - k^2R^2/10$. Hence, the convolution integral leads to a constant term plus corrections scaling as k^2R^2 times moments of the continuous power spectrum,

$$[W_R * P_{\rm hh}^{\rm (c)}](k) \xrightarrow{k\to 0} \int \frac{{\rm d}^3q}{(2\pi)^3} P_{\rm hh}^{\rm (c)}(q) W_R(q) + k^2R^2 \int \frac{{\rm d}^3q}{(2\pi)^3} P_{\rm hh}^{\rm (c)}(q) \times \left[W_R(q) \left(\frac{1}{(qR)^2} - \frac{1}{6} \right) - \frac{\sin(qR)}{(qR)^3} \right]. \quad (17)$$

Thus, irrespective of the shape of the continuous power spectrum on large scales, exclusion always introduces a white (k^0) correction on large scales.

Figure 1 illustrates the behavior of the correlation function of discrete tracers. In the very popular local bias model, the clustering of dark matter haloes is modeled at leading order as $P_{\rm hh}^{\rm (c)}(k) = b_1^2 P_{\rm lin}(k)$. In configuration space this leads to $\xi_{\rm hh}^{\rm (c)}(r) = b_1^2 \xi_{\rm lin}(r)$, shown by the black dashed line. We will consider this linear bias model as the fiducial model on top of which we define corrections. Nonlinear halo clustering suggests an enhancement proportional to higher powers of the linear correlation function as exemplified by the red dashed line. Our above arguments suggest that this clustering model, if at all, can only be true outside the exclusion radius. Inside this radius, the probability to find another halo is zero, leading to $\xi_{\rm hh}^{\rm (d)}(r < R) = -1$.

An intuitive understanding of the corrections can be obtained in the $k \to 0$ limit, where the halo power spectrum

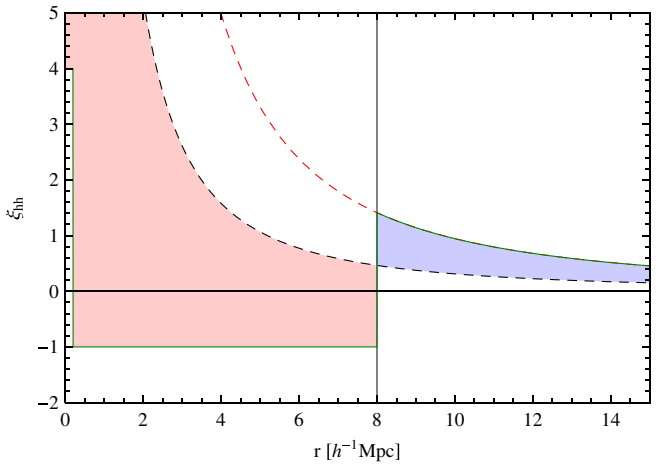

FIG. 1 (color online). Cartoon version of the correlation function of discrete tracers. Continuous linear correlation function (black dashed) and nonlinear correlation function (red dashed). The true correlation function of discrete tracers (green solid line) agrees with the nonlinear continuous correlation function outside the exclusion scale and is -1 below, except for the delta function at the origin arising from discreteness. Thus, there are two corrections compared to the continuous linear bias model—a negative correction inside the exclusion radius (red shaded) and a positive one outside the exclusion radius due to nonlinear clustering (blue shaded).

is given by an integral over the correlation function and can thus be written as

$$P_{\rm hh}^{\rm (d)}(k) \xrightarrow{k\to 0} \frac{1}{\bar{n}} - V_{\rm excl} - b_1^2 \int_0^R {\rm d}^3r \xi_{\rm lin}(r) + \int_R^\infty {\rm d}^3r [\xi_{\rm hh,NL}^{\rm (c)}(r) - b_1^2 \xi_{\rm lin}(r)], \quad (18)$$

where we introduced $\xi_{\rm hh,NL}^{\rm (c)}(r)$ to account for generic nonlinear continuous models of the halo clustering. The red and blue shaded regions in Fig. 1 show the negative and positive corrections with respect to the linear bias model for which we would have, in absence of exclusion, $P_{\rm hh}^{\rm (d)}(k) \xrightarrow{k\to 0} 1/\bar{n}$. Note that the nonlinear halo-halo correlation function could in principle be smaller than the linear bias prediction. Our above notion of a positive correction arising from the nonlinear correction outside the exclusion radius is solely based on local bias arguments. In general, this statement should be relaxed (for an example, see Appendix A) and the blue region could have either sign.

III. TOY MODELS

To show that the exclusion can indeed lower the stochasticity we perform a simple numerical experiment. We consider a set of hard sphere haloes of radius $R/2$. For this purpose, we distribute N particles randomly in a cubic box ensuring that $|\boldsymbol{x}_i - \boldsymbol{x}_j| > R$ for all pairs of particles (i, j). The corresponding correlation function is expected to be zero except for scales $r < R$, where $\xi = -1$ due to exclusion. Thus we expect the fiducial stochasticity to be lowered by $4\pi R^3/3$ in the $k \to 0$ limit. For an intuitive derivation of the corrections to the power spectrum we will consider a fixed number of particles N in a finite volume V. Using Eq. (6) the auto-power spectrum of the tracer particles can be written as [14]

$$\begin{aligned} P^{\rm (d)}(k) &= \frac{1}{V} \langle \delta^{\rm (d)}(\boldsymbol{k}) \delta^{\rm (d)}(-\boldsymbol{k}) \rangle \\ &= \frac{V}{N^2} \sum_{i=j} \langle \exp[{\rm i}\boldsymbol{k} \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j)] \rangle \\ &\quad + \frac{V}{N^2} \sum_{i\neq j} \langle \exp[{\rm i}\boldsymbol{k} \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j)] \rangle - V\delta_{\boldsymbol{k},\boldsymbol{0}}^{\rm (K)} \\ &= \frac{1}{\bar{n}} + \frac{V}{N^2} \sum_{i\neq j} \langle \exp[{\rm i}\boldsymbol{k} \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j)] \rangle - V\delta_{\boldsymbol{k},\boldsymbol{0}}^{\rm (K)}. \end{aligned} \quad (19)$$

This yields for the hard sphere sample, which we consider as a proxy for excluded haloes

$$P_{\rm hh}^{\rm (d)}(k) = \frac{1}{\bar{n}_{\rm h}} - \frac{4\pi R^3}{3} W_R(k). \quad (20)$$

In Fig. 2 we show the power spectrum of this toy halo sample for $R = 8h^{-1}$ Mpc, $N = 800$ and

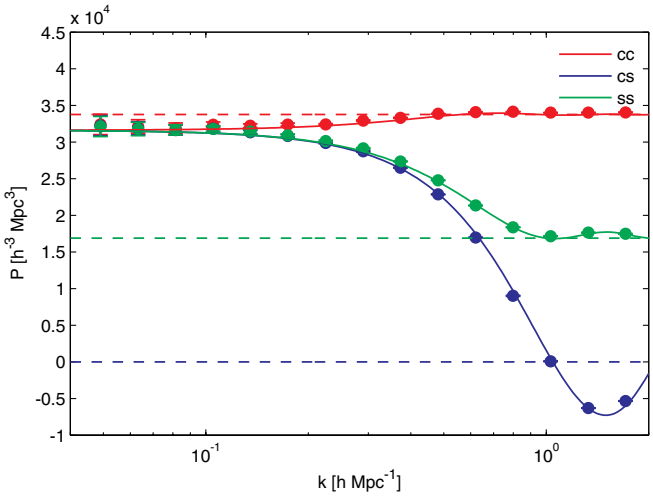

FIG. 2 (color online). Power spectrum of a randomly distributed halo sample obeying exclusion (top, red points) and corresponding model with (red solid line) and without (top, red dashed line) exclusion. In a second step we populate these haloes with $N_{\rm gal} = 2$ satellite galaxies, and calculate the auto power spectrum of the satellite galaxies (central, green points) and their cross-power spectrum with the halo centers (bottom, blue points). The corresponding solid lines show our model predictions, whereas the dashed lines show the naive expectation of Poisson shot noise.

$V = 300^3 h^{-3}$ Mpc^3. We clearly see that the measured power follows the exclusion corrected stochasticity. The window is close to unity on large scales and decays at $k \approx 1/R$; i.e., the fiducial shot noise is recovered for high k. This is a first indication for stochasticity not being scale independent. Note that the above derivations are only true in the limit, where the total exclusion volume is small compared to the total volume and thus allows for a quasi random distribution (about 0.8% volume coverage in our case).

A. Satellite galaxies

Galaxies are believed to populate dark matter haloes. Let us consider the simple case that each of the dark matter haloes under consideration hosts a central galaxy that, as the name suggests, coincides with the halo center plus a fixed number $N_{\rm s,h}$ of satellite galaxies, such that the total number of satellite galaxies is given by $N_{\rm s} = N_{\rm s,h} N_{\rm h}$. For simplicity, we will assume that the galaxies are distributed according to a profile $\rho_{\rm s}(r)$ with typical scale $R_{\rm s}$ around the centers of the host halo centers. For long wavelength modes $k < 1/R_{\rm s}$, the $N_{\rm s,h}$ galaxies within one halo are effectively one particle, which is why on large scales we expect the stochasticity of the satellite galaxy sample to be equal to that of the host haloes, and only for scales $k > 1/R_{\rm s}$ can the modes probe the distinct nature of the particles and the stochasticity goes to $1/\bar{n}_{\rm s}$.

We can evaluate our model Eq. (19) to obtain the satellite-satellite power spectrum

$$\begin{aligned} P_{\rm ss}^{\rm (d)}(k) &= \frac{V}{N_{\rm s}} + \frac{V}{N_{\rm s}^2} \sum_{\rm h_i} \sum_{{\rm s}_j \in {\rm h}_i} \sum_{{\rm s}_l \neq {\rm s}_j \in {\rm h}_i} \langle \exp[{\rm i}\boldsymbol{k} \cdot (\boldsymbol{r}_j - \boldsymbol{r}_l)] \rangle \\ &\quad + \frac{V}{N_{\rm s}^2} \sum_{\rm h_i} \sum_{{\rm s}_j \in {\rm h}_i} \sum_{{\rm h}_m \neq {\rm h}_i} \sum_{{\rm s}_l \in {\rm h}_m} \langle \exp[{\rm i}\boldsymbol{k} \cdot (\boldsymbol{r}_j - \boldsymbol{r}_l)] \rangle \\ &= \frac{V}{N_{\rm s}} + \frac{V}{N_{\rm s}} (N_{\rm s,h} - 1) \langle \exp[{\rm i}kR_{\rm s}\mu] \rangle^2 \\ &\quad - \frac{4\pi R^3}{3} u_{\rm s}^2(k) W_R(k) \\ &= \frac{1}{\bar{n}_{\rm s}} [1 + (N_{\rm s,h} - 1) u_{\rm s}^2(k)] - \frac{4\pi R^3}{3} u_{\rm s}^2(k) W_R(k). \end{aligned} \tag{21}$$

Here μ is the cosine of the angle between $\boldsymbol{k}$ and $\Delta \boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$ that is averaged over, $u_{\rm s}(k)$ is the normalized Fourier transform of the galaxy profile $\rho_{\rm s}(r)$. For definiteness we will assume a delta function profile $\rho(r) = \delta^{\rm (D)}(r - R_{\rm s})/r^2$ corresponding to $u_{\rm s}(k) = j_0(kR_{\rm s})$, where j_0 is the zeroth order spherical Bessel function. The two terms in the above equation correspond to the one and two halo terms in the halo model [15,16], the profile and the fiducial shot noise arise from correlations between particles in the same halo, whereas the exclusion term is dominated by the correlation between distinct haloes. The results of the numerical experiment are shown in Fig. 2 as the green points. The model prediction is shown as the green solid line and describes the simulation measurement very well. On small scales the power is dominated by the fiducial galaxy shot noise and on large scales the host halo stochasticity dominates

$$\begin{aligned} P_{\rm ss}^{\rm (d)}(k \ll 1/R_{\rm s}, 1/R) &= \frac{1}{\bar{n}_{\rm h}} - \frac{4\pi R^3}{3}, \\ P_{\rm ss}^{\rm (d)}(k \gg 1/R_{\rm s}, 1/R) &= \frac{1}{\bar{n}_{\rm s}}. \end{aligned} \tag{22}$$

While the distribution of satellite galaxies on a sphere of fixed radius around the halo center is very peculiar and unrealistic, the qualitative behavior is the same for all profiles with finite support. In the case studied above, the corrections to the fiducial galaxy shot noise $1/\bar{n}_{\rm h}$ are always positive. This is due to the high satellite fraction. As we will discuss below, this behavior might be completely different for galaxy samples with small satellite fraction, where the exclusion effect can be more important than the enhancement due to the satellites.

B. Central and satellite galaxies

We can also consider the cross-power spectrum between halo centers (central galaxies) and the satellite galaxies. In this case there is no Poisson shot noise, since the samples are nonoverlapping and the power is dominated by a one-halo term describing the radial distribution of the satellites around the halo center

$$P_{\rm cs}^{\rm (d)}(k) = \frac{V}{N_{\rm h}N_{\rm s}}\sum_{\rm h_i}\sum_{\rm s_j\in h_i}\langle\exp[{\rm i}\boldsymbol{k}\cdot(\boldsymbol{r}_j-\boldsymbol{r}_i)]\rangle + \frac{V}{N_{\rm h}N_{\rm s}}\sum_{\rm h_i}\sum_{\rm h_j\neq h_i}\sum_{\rm s_l\in h_j}\langle\exp[{\rm i}\boldsymbol{k}\cdot(\boldsymbol{r}_l-\boldsymbol{r}_i)]\rangle = \frac{1}{\bar{n}_{\rm h}}u_{\rm s}(k) - \frac{4\pi R^3}{3}u_{\rm s}(k)W_R(k). \quad (23)$$

Here we have again a one halo term arising from correlations of the halo center with satellites in the same halo and a two halo term arising from the correlation of satellite galaxies in one halo with the center of another halo. The comparison with the result of the numerical experiment in Fig. 2 shows very good agreement.

The above discussion is overly simplified as we assume all haloes to be of the same mass and to host the same number of galaxies. Any realistic galaxy sample will be hosted by a range of halo masses (i.e. a range of exclusion radii) and the number of galaxies per halo will also be a function of mass.

The total galaxy power spectrum of the combined central and satellite samples can be obtained as a combination of the central-central, central-satellite and satellite-satellite contributions

$$P_{\rm gg}(k) = (1-f_{\rm s})^2P_{\rm cc}(k) + 2f_{\rm s}(1-f_{\rm s})P_{\rm cs}(k) + f_{\rm s}^2P_{\rm ss}(k), \quad (24)$$

where $f_{\rm s} = N_{\rm s}/(N_{\rm c}+N_{\rm s})$ is the satellite fraction. For realistic satellite fractions for SDSS LRGs [17] $f_{\rm s}\approx 0.1$, the weighting of the central-central power spectrum dominates over the contributions from the central-satellite and satellite-satellite power spectra by factors of 9 and 81, respectively. A more realistic galaxy sample based on a HOD population of dark matter haloes in a N-body simulation will be discussed in Sec. V F.

C. Toy model with clustering

Haloes are clustered, i.e., there is an enhanced probability to find two collapsed objects in the vicinity of each other to finding them widely separated. Let us discuss the influence of this phenomenological result on our toy model. For the sake of simplicity let us assume that haloes always come in pairs, i.e., that there is a second halo outside the exclusion scale at typical separation $R_{\rm clust}$. This will similarly to satellite galaxies residing in one halo, lead to a *positive* k^0 term on large scales, that decays for $k > 1/R_{\rm clust}$. In a more realistic setting, not all haloes will come in pairs, some of them will be single objects, others will come in clusters of n-haloes. Furthermore not all of them will be separated by exactly the clustering scale.

Some authors argued that any large scale k^0 behavior in the perturbation theory description of biasing is unphysical and should be suppressed by constant but aggressive smoothing [18] or by a k-dependent smoothing [19]. Based on the above considerations, we argue that such terms are just a result of the clustering of haloes and thus *not* unphysical. Whether the magnitude of these effects can be covered by a perturbative treatment such as second-order bias combined with perturbation theory, is a different question, which we will pick up later in Sec. IV B. For now, let us note that such a k^0 term is also predicted by biasing models that go beyond the local bias model as for instance the correlation of thresholded regions as discussed briefly in the next subsection and for instance in [20]. Generally the clustering scale exceeds the exclusion scale and thus one should expect that the enhancement due to clustering decays at lower k than the suppression due to exclusion. This is actually what happens in the simulations as we will show in Sec. V.

D. Density threshold bias

The spherical collapse model suggests that spherical Lagrangian regions exceeding the critical collapse density $\delta_{\rm c}\approx 1.686$ segregate from the background expansion and form gravitationally bound objects. Hence, the clustering statistics of regions above threshold can tell us something about the clustering of dark matter haloes and galaxies. The study of [1] considered the correlation function of regions whose density exceeds a certain value in a Gaussian random field, smoothed with a top-hat window of scale R. At the same time this paper pioneered bias models, which are nothing but a large-scale expansion of the full correlation function of thresholded regions. Let us see how nonlinear or nonperturbative clustering can affect the power spectrum on the largest scales in the full model.

The root mean square overdensity within the smoothed regions is given by

$$\sigma_R^2 = \frac{1}{2\pi^2}\int {\rm d}k k^2 W_R^2(k)P_{\rm lin}(k). \quad (25)$$

The correlation of thresholded regions can be calculated exactly employing the two-point probability density function of Gaussian random fields. For simplicity, we will consider regions of a fixed overdensity rather than regions above threshold. The peak height $\nu = \delta_{\rm c}/\sigma_R$ can be chosen based on the spherical collapse argument. For the correlation function of regions of fixed overdensity one obtains [1,20]

$$1+\xi_{\rm tr}(r) = \frac{1}{\bar{n}_{\rm tr}^2}\frac{1}{(2\pi)\sqrt{1-\xi_R^2(r)/\sigma_R^4}} \times \exp\left[-\nu^2\frac{1-\xi_R(r)/\sigma_R^2}{1-\xi_R^2(r)/\sigma_R^4}\right], \quad (26)$$

where $\bar{n}_{\rm tr} = 1/\sqrt{(2\pi)\sigma_R^2}\exp[-\nu^2/2]$. Here $\xi_R(r)$ is the linear correlation function smoothed on scale R. In the large distance, small correlation limit, the correlation

function of the thresholded regions can be approximated by a linearly biased version of the linear correlation function $\xi_{\rm tr}(r) = b_{1,\rm tr}^2 \xi_R(r)$ with $b_{1,\rm tr} \approx \nu/\sigma_R$. If one is interested in an accurate description of the nonperturbative correlation function on smaller scales, higher orders in the expansion need to be considered. Comparing the expansion of the correlation of thresholded regions in powers of the smoothed correlation to the full nonperturbative result in Eq. (26), we can investigate the convergence properties of the linear bias model. The left panel of Fig. 3 shows the correlation function of thresholded regions for a Gaussian random field smoothed on $R = 4h^{-1}$ Mpc and the linearly biased versions of the smoothed and unsmoothed linear correlation functions.

We can now Fourier transform the correlation function of thresholded regions and subtract out the linearly biased power spectrum to obtain the correction introduced by the nonlinear clustering

$$\Delta P_{\rm tr}(k) = {\rm FT}[\xi_{\rm tr}](k) - b_{1,\rm tr}^2 P_{\rm lin}(k). \tag{27}$$

As we show in Fig. 3, there is a nonvanishing correction in the $k \to 0$ limit that is approximately constant on large scales and goes to zero on small scales. The presence of such a correction was discussed in a slightly different context in [20].

In its original form, the thresholded regions are a continuous field and thus do not include any exclusion. One could however imagine that the patches defining the smoothing scale do not overlap. In this case the correlation function of thresholded regions should go to -1 for $r < 2R$. To show how the exclusion scale affects the correction in the power spectrum we consider a few exclusion radii smaller than two smoothing radii. As is obvious in Fig. 3, increasing the smoothing scale first reduces the scale-independent correction on large scales, compensates for it completely, and then eventually leads to a negative scale-independent correction for $r = 2R$.

E. Peak bias model

While the thresholded regions provide a continuous bias model, the peak model [21,22] goes beyond in identifying a discrete set of points and providing the correlation function of these points. Most studies of the peak model to date have focused on the large separation limit [23–25], where closed form expressions for the peak correlation in terms of the underlying linear correlation function and its derivatives are possible. However, [26] calculated the one-dimensional peak correlation function for a set of power law power spectra and [27] computed the two-dimensional peak correlation function for peaks in the CMB. The reason for the restriction to one or two dimensions owes to the dimensionality of the covariance matrix that needs to be inverted for the calculation of the peak correlation. In a one-dimensional field the covariance matrix is a six by six matrix (field amplitude, first and second derivative at two points).

Here we consider realistic ΛCDM power spectra in three dimensions, smooth them on a realistic Lagrangian scale $R = 2h^{-1}$ Mpc and evaluate the exact nonperturbative one-dimensional correlation of peaks following the approach of [26] (see Appendix B for a brief review). Note that the correlation of field derivatives diverges for top-hat smoothing, which is why we follow common praxis and employ a Gaussian smoothing. The Gaussian smoothing makes the correlators of field derivatives well behaved but beyond that there is no physical motivation to employ

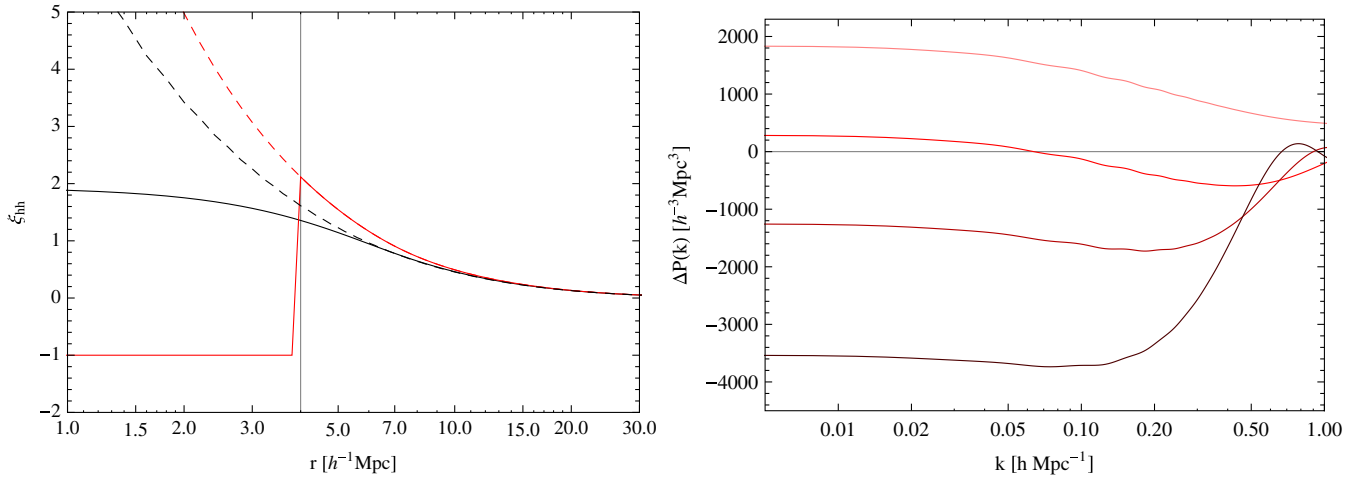

FIG. 3 (color online). Kaiser bias [1] in configuration and Fourier space. *Left panel*: Unsmoothed (black dashed) and $R = 4h^{-1}$ Mpc smoothed (black solid) linearly biased matter correlation functions $b_{1,\rm tr}^2 \xi(r)$ and continuous correlation function of the thresholded regions $\xi_{\rm tr}(r)$ (red dashed). The red solid line shows a simple implementation of exclusion imposed on the correlation function of the thresholded regions. *Right panel*: Power spectrum correction arising from the nonlinear biasing (top line) and effect of increasing exclusion for $R = 0$, 4, 6, $8h^{-1}$ Mpc from top to bottom.

this filter. We study a range of peak heights ν and also a range of bin widths in ν. The peak correlation function for four different bin widths is shown in the left panel of Fig. 4.

The first remarkable observation is that peaks of a fixed height don't seem to obey exclusion, only after considering a finite width in peak height, we can observe that the correlation function goes to -1 on small scales. The transition scale to the fixed peak height case increases with bin with, i.e., wider bins have a larger exclusion region. As above for the thresholded regions we can expand the peak correlation function in the large distance limit and obtain a bias expansion that has contributions from the underlying matter correlation function *and* correlation functions of the derivatives. Doing that, it becomes obvious, that the linear matter bias is only assumed outside of the BAO scale and that there is a distinct scale-dependent bias that is partially described by the derivative terms in the bias expansion. Fourier transforming the full peak correlation function and subtracting out the linear biased power spectrum we obtain the correction shown in the right panel of Fig. 4. The qualitative behavior agrees with the result obtained above for the thresholded sample with an *ad hoc* exclusion scale. On large scales there is a combination of clustering and exclusion effects, the clustering decays first and then also the exclusion correction goes to zero. Note that the projected one-dimensional matter power spectrum scales as k^0 one large scales and thus doesn't vanish in the $k \to 0$ limit. This fact makes the distinction between clustering and stochasticity terms in the one-dimensional peak model very difficult. We hope to report on results for the full three-dimensional peak model in the near future.

IV. QUANTIFYING THE CORRECTIONS

Let us now try to quantify the stochasticity corrections for a realistic halo sample. We expect the effect to be time independent in the $k \to 0$ limit if the same sample of particles is evolved under gravity. Thus, to minimize the influence of nonlinearities, we will consider the protohaloes in Lagrangian space. In numerical studies of the effect, we will later define the protohalo as the initial ensemble of particles that form the Friends-of-Friends (FoF) haloes in our final output at redshift $z_f = 0$.

A. Continuous halo power spectrum from local bias

The local Lagrangian bias model assumes that the initial halo density field can be written as a Taylor series in the matter fluctuations at the same Lagrangian position $\boldsymbol{q}$

$$\delta_h(\boldsymbol{q}, \eta_i) = b_1^{(L)}(\eta_i)\delta(\boldsymbol{q}, \eta_i) + \frac{b_2^{(L)}(\eta_i)}{2!}\delta^2(\boldsymbol{q}, \eta_i) + \frac{b_3^{(L)}(\eta_i)}{3!}\delta^3(\boldsymbol{q}, \eta_i) + \cdots, \tag{28}$$

where η_i is the conformal time of the initial conditions and $\boldsymbol{q}$ is the Lagrangian coordinate. We will follow the approach of [28] where the smoothing scale is an unobservable scale, which should not affect n-point clustering statistics on scales exceeding the smoothing scale. The above model can be used as the starting point for a coevolution of haloes and dark matter, which finally leads to a Eulerian bias prescription. Recently, such a calculation was shown to correctly predict nonlocal Eulerian bias terms [29,30]. The peak background split (PBS) [31] makes predictions for the Lagrangian bias parameters in the above

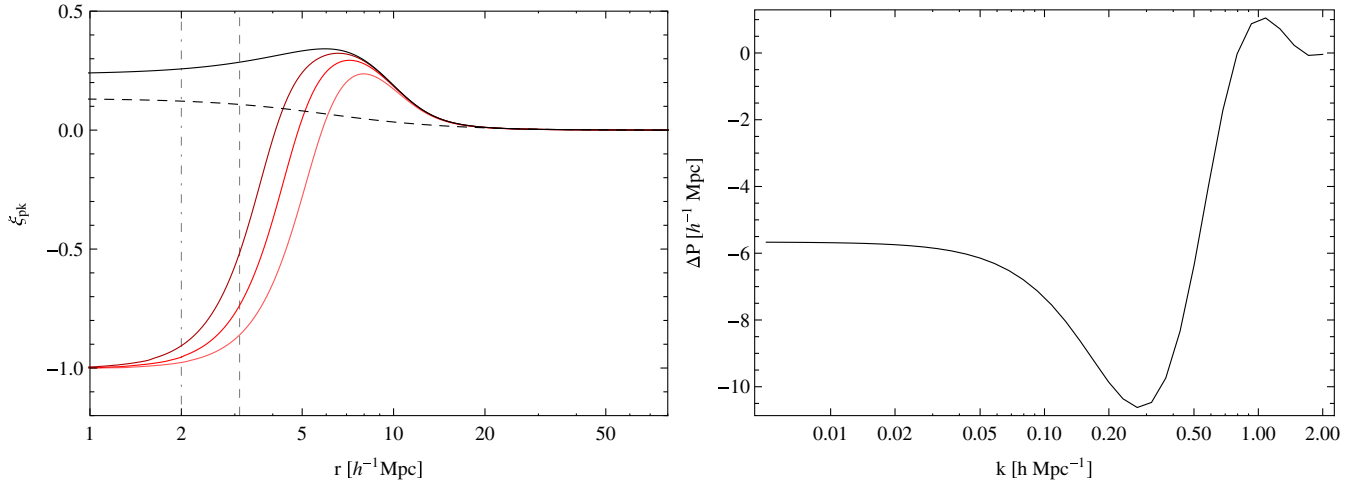

FIG. 4 (color online). Clustering of peaks in a one-dimensional skewer through a density field smoothed with a Gaussian filter of scale $R = 2h^{-1}$ Mpc ($M \approx 8.6 \times 10^{12} h^{-1} M_\odot$). *Left panel*: For fixed peak height the correlation function flattens out on small scales (black), but with increasing bin width the exclusion becomes stronger. The width of the bin in peak height increases from dark to light red. The linear local bias expansion is the same for all of these models and is shown by the dashed line. For reference we overplot the Gaussian smoothing (dash-dotted) and the top-hat smoothing scale containing the same mass (dashed). *Right panel*: Corresponding stochasticity correction $\Delta P_{pk}(k) = \mathrm{FT}[\xi_{pk}](k) - b_{1,pk}^2 P_{lin}(k)$ for the fiducial bin width.

equation, and the corresponding late time Eulerian local bias parameters can then be obtained based on the spherical collapse model. There is some evidence that the peak model yields a better description of some aspects of the initial halo clustering than the local Lagrangian bias model. While we will briefly discuss these effects in Appendix A, we refrain from using this model for the modeling of the stochasticity corrections, since the peak bias expansion beyond leading order has not been studied in great detail and its implementation goes beyond the scope of this study.

For the continuous power spectrum in the initial conditions the local Lagrangian bias model predicts

$$P_{\rm hh}^{(\rm c)}(k,\eta_{\rm i}) = (b_1^{(\rm L)})^2 D^2(\eta_{\rm i}) P_{\rm lin,0}(k) + \frac{1}{2}(b_2^{(\rm L)})^2 D^4(\eta_{\rm i}) I_{22}(k), \tag{29}$$

where $D(\eta)$ is the linear growth factor and the scale-dependent bias correction is described by

$$I_{22}(k) = \int \frac{{\rm d}^3 q}{(2\pi)^3} P_{\rm lin,0}(q) P_{\rm lin,0}(|\boldsymbol{k}-\boldsymbol{q}|). \tag{30}$$

This term leads to a positive k^0 contribution in the low-k regime. In this sense it deviates from typical perturbative contributions to the power spectrum, which start to dominate on small scales. For this reason, this term was partially absorbed into the shot noise by [28]. We will explicitly consider the term, since it describes the effect of nonlinear clustering and is responsible for super-Poissonian stochasticity. The cross-power spectrum between haloes and matter is given by

$$P_{\rm hm}^{(\rm c)}(k,\eta_{\rm i}) = b_1^{(\rm L)}(\eta_{\rm i}) D^2(\eta_{\rm i}) P_{\rm lin,0}(k) \tag{31}$$

and does obviously not contain any second-order bias corrections. This statement remains true if higher-order biasing schemes are considered, since higher-order biases only renormalize the bare bias parameters [28].

Truncating the bias expansion is only valid if $\langle\delta^2\rangle \ll 1$, which is certainly satisfied on large scales in the initial conditions, but not necessarily on the scales relevant for halo clustering outside the exclusion radius. On these scales one might have to consider all the higher-order local bias parameters. It is beneficial to calculate this effect in configuration space where the local bias model leads to a power series in the linear correlation function

$$\xi^{(\rm c)}(r) = \sum \frac{(b_i^{(\rm L)})^2}{i!} D^{2i}(\eta_{\rm i}) \xi_{\rm lin}^i(r) \overset{\nu\to\infty}{\longrightarrow} \exp\left[\frac{\nu}{\delta_{\rm c}}\xi_{\rm lin}(r)\right]. \tag{32}$$

The limit applies only in the high-peak limit and Press-Schechter bias parameters [32]. We will restrict ourselves to the quadratic bias model since it can account for the main effects and since using higher-order biasing schemes also requires more parameters to be determined. The Press-Schechter and Sheth-Tormen prescriptions provide a rough guideline for the scaling of bias with mass and redshift, but fail to provide correct predictions for the bias amplitude. Thus we obtain the bias parameters from fits to observables not affected by stochasticity (such as the halo-matter cross-power spectrum) and obtaining higher-order biases would require higher-order spectra such as the bispectrum. Since our discussion is mostly in Lagrangian space we will drop the superscripts E and L from now on and absorb the growth factors into the linear power spectra and correlation functions.

B. Theory including clustering and exclusion

We can now use the bias model introduced above to evaluate the discrete power spectrum Eq. (13). We have

$$P^{(\rm d)}(k) = \frac{1}{\bar{n}} + b_1^2 P_{\rm lin}(k) + \frac{1}{2} b_2^2 I_{22}(k) - b_1^2 V_{\rm excl}[W_R * P_{\rm lin}](k) - \frac{1}{2} b_2^2 V_{\rm excl}[W_R * I_{22}](k) - V_{\rm excl} W_R(k). \tag{33}$$

The splitting of I_{22}, the nonlinear clustering term arising from b_2, is somewhat counterintuitive, since we expect this term to be active only outside the exclusion scale. Furthermore, there is no corresponding term in the halo-matter or matter-matter power spectra that would cancel the continuous I_{22}. Thus we combine the continuous part and the exclusion correction for the nonlinear clustering term into a positive correction whose small-scale contributions have been removed,

$$P^{(\rm d)}(k) = \frac{1}{\bar{n}} + b_1^2 P_{\rm lin}(k) + \frac{1}{2} b_2^2 I_{22}(k,R) - b_1^2 V_{\rm excl}[W_R * P_{\rm lin}](k) - V_{\rm excl} W_R(k). \tag{34}$$

Here we defined the correction term,

$$I_{22}(k,R) = \int_R^\infty {\rm d}^3 r \xi^2(r) j_0(kr). \tag{35}$$

A simpler version of Eq. (34) has been presented in [10], where the nonlinear clustering is neglected and the results are presented in Eulerian rather than Lagrangian space.

In the $k \to 0$ limit the Fourier transform simplifies to a spatial average over the correlation function

$$P^{(\rm d)}(k) \overset{k\to 0}{\longrightarrow} \frac{1}{\bar{n}} + \frac{1}{2} b_2^2 \int_R^\infty {\rm d}^3 r \xi^2(r) - b_1^2 \int_0^R {\rm d}^3 r \xi(r) - V_{\rm excl}, \tag{36}$$

where the linear bias term vanishes due to $P_{\rm lin}(k) \longrightarrow k \to 00$. The fact that the integral over ξ^2 runs only from the exclusion scale to infinity mitigates the smoothing dependence of the correction, since smoothing on the scale of the halo affects the correlation function only on the halo scale, which is by definition smaller than the exclusion scale.

TABLE I. Mean masses, exclusion radii, first and second-order Lagrangian and Eulerian bias parameters for our $z = 0$ halo sample. Masses are in units of $h^{-1}M_\odot$ and radii in units of h^{-1} Mpc. The mass dependence of these parameters is also plotted in Fig. 6. Note that the second-order bias parameter is just a phenomenological fitting parameter used to get a reasonable representation of the correlation function. We do not claim that our fitting procedure yields an accurate second-order bias parameter for the sample. For this purpose on has to employ the bispectrum, which yields second-order bias parameters that are in much better agreement with the peak-background split expectation but fail to describe the correlation function.

Bin	M	$R_{\rm excl}^{\rm (L)}$	$b_1^{\rm (L)}$	b_2	$b_1^{\rm (E)}$
I	1.14×10^{13}	2.6	7.80	1174.5	1.19
II	1.27×10^{13}	2.8	8.49	1206.0	1.22
III	1.42×10^{13}	2.8	9.24	1287.3	1.25
IV	1.62×10^{13}	3.0	10.70	1342.9	1.27
V	1.89×10^{13}	3.6	13.17	1357.7	1.31
VI	2.25×10^{13}	3.9	14.94	1404.2	1.38
VII	2.80×10^{13}	4.2	16.65	1635.7	1.45
VIII	3.72×10^{13}	4.9	21.92	1587.5	1.58
IX	5.65×10^{13}	5.8	29.42	1439.8	1.78
X	1.66×10^{14}	8.7	56.94	1346.5i	2.54

C. Stochasticity matrix

We will now consider the power spectrum for a set of nonoverlapping halo mass bins. We will consider their auto-power spectra and cross-power spectra between different halo mass bins i and j and denote this quantity P_{ij}, whereas the cross-power between a certain mass bin and the matter is denoted $P_{i\delta}$. The sum over equal pairs in Equation (19) is only present for the auto-power spectra and thus the $1/\bar{n}$ shot noise affects only the diagonal entries of the power spectrum matrix $P_{ij}(k)$. On the other hand, exclusion affects also the off-diagonal matrix entries, since by definition also haloes of different mass are distinct objects and can thus not overlap. Furthermore, different mass haloes are affected by nonlinear clustering, since the probability to find any sort of massive object ($M > M_*$) in the vicinity of a massive object is enhanced. For simplicity we will employ equal number density mass bins, which all have the same fiducial shot noise $1/\bar{n}$.

When trying to extract the amplitude and scale dependence of the noise, we need to remove all the contributions due to linear bias from the halo power spectra. For this purpose, we will employ the stochasticity matrix as defined in [4] (see also Eq. (15) for the definition of the diagonal),

$$(2\pi)^3\delta^{(\rm D)}(\boldsymbol{k}+\boldsymbol{k}')C_{ij}(k) = \langle[\delta_i(\boldsymbol{k}) - b_{1,i}\delta(\boldsymbol{k})][\delta_j(\boldsymbol{k}') - b_{1,j}\delta(\boldsymbol{k}')]\rangle, \quad (37)$$

such that in terms of the power spectra, we have

$$C_{ij}(k) = P_{ij}(k) - b_{1,i}P_{\delta j}(k) - b_{1,j}P_{\delta i}(k) + b_{1,i}b_{1,j}P_{\delta\delta}(k). \quad (38)$$

The model introduced above can be straightforwardly generalized to multiple mass bins and their respective cross-power spectra by the following replacements $b_1^2 \to b_{1,i}b_{1,j}$, $b_2^2 \to b_{2,i}b_{2,j}$ and $R \to R_{ij} = (R_i + R_j)/2$. The exact form of the combined exclusion radius is somewhat debatable, but for now we will employ the arithmetic mean. The resulting correction to the linear local bias model is given by

$$C_{ij}(k) = \frac{1}{\bar{n}}\delta_{ij}^{\rm (K)} + \frac{1}{2}b_{2,i}b_{2,j}I_{22}(k,R_{ij}) - V_{{\rm excl},ij}W_{R_{ij}}(k) - b_{1,i}b_{1,j}V_{{\rm excl},ij}[W_{R_{ij}} * P_{\rm lin}](k), \quad (39)$$

where $V_{ij} = 4\pi/3R_{ij}^3$. We see that the definition of the stochasticity matrix removes all occurrences of the linearly biased power spectrum. In Eulerian space both $P_{\rm hm}$ and $P_{\rm hh}$ would have an additional contribution from b_2I_{12}, where I_{12} describes the cross correlation between nonlinear bias and nonlinear matter clustering. The term is defined as

$$I_{12}(k) = \int\frac{{\rm d}^3q}{(2\pi)^3}P_{{\rm lin},0}(q)P_{{\rm lin},0}(|\boldsymbol{k}-\boldsymbol{q}|)F_2(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q}), \quad (40)$$

where $F_2(\boldsymbol{q}_1,\boldsymbol{q}_2)$ is the standard perturbation mode coupling kernel [33]. The definition of the stochasticity matrix also removes all occurrences of I_{12}.

V. EVALUATIONS AND COMPARISON TO SIMULATIONS

A. The simulations and halo sample

Our numerical results are based on the Zürich horizon zHORIZON simulations, a suite of 30 pure dissipationless dark matter simulations of the ΛCDM cosmology in which the matter density field is sampled by $N_{\rm p} = 750^3$ dark matter particles. The box length of $1500h^{-1}$ Mpc, together with the WMAP3 [34] inspired cosmological parameters ($\Omega_{\rm m} = 0.25$, $\Omega_\Lambda = 0.75$, $n_{\rm s} = 1$, $\sigma_8 = 0.8$), then implies a particle mass of $M_{\rm p} = 5.55 \times 10^{11}h^{-1}M_\odot$. The total simulation volume is $V \approx 100h^{-3}$ Gpc^3 and enables precision studies of the clustering statistics on scales up to a few hundred comoving megaparsecs.

The simulations were carried out on the ZBOX2 and ZBOX3 computer clusters of the Institute for Theoretical Physics at the University of Zurich using the publicly available GADGET-II code [35]. The force softening length of the simulations used for this work was set to $60h^{-1}$ kpc, consequently limiting our considerations to larger scales. The transfer function at redshift $z_{\rm f} = 0$ was calculated using the CMBFAST code of [36] and then rescaled to the initial redshift $z_{\rm i} = 49$ using the linear growth factor. For each simulation, a realization of the power spectrum and the corresponding gravitational potential were calculated. Particles were then placed on a Cartesian grid of spacing $\Delta x = 2h^{-1}$ Mpc and displaced according to a second-order Lagrangian perturbation theory. The displacements and initial conditions were

computed with the 2LPT code of [37], which leads to slightly non-Gaussian initial conditions.

Gravitationally bound structures are identified at redshift $z_f = 0$ using the B-FoF algorithm kindly provided by Volker Springel with a linking length of 0.2 mean interparticle spacings. Haloes with less than 20 particles were rejected such that we resolve haloes with $M > 1.2 \times 10^{13} h^{-1} M_\odot$. The halo particles are then traced back to the initial conditions at $z_i = 49$ and the corresponding center of mass is identified. We split the halo sample into ten equal number density bins with a number density of $\bar{n} = 3.72 \times 10^{-5} h^3$ Mpc^{-3}, leading to a shot noise contribution to the power at the level of $P_{SN} \approx 2.7 \times 10^4 h^{-3}$ Mpc^3.

B. The correlation function in Lagrangian space

The corrections to the halo power spectrum in our model are motivated by certain features in the halo-halo correlation function. While the fiducial stochasticity affects the correlation function only at the origin, the two other effects, exclusion and nonlinear clustering, should be clearly visible in the correlation function at finite distances. For this purpose we measure the correlation function of the traced-back haloes for our ten halo mass bins using direct pair counting.

In Fig. 5 we show the correlation function for mass bin V. The log-linear plot clearly shows that the correlation function is -1 on small scales and shows a smooth transition to positive values around the exclusion scale visualized by the vertical black line. The exclusion scale is fitted both in the initial and final conditions as 0.8 times the maximum in the correlation function and is shown in Fig. 6. The ratio between the initial and final exclusion radii is roughly 3 for all mass bins. The spherical collapse model suggests that haloes collapse by a factor 5, but there is no reason to believe that protohaloes that are in direct contact in Lagrangian space are still touching each other in Eulerian space. Thus it is reasonable to expect a somewhat smaller reduction in the exclusion scale. On large scales $30h^{-1}$ Mpc $< r < 90h^{-1}$ Mpc the correlation function is reasonably well described by linear bias shown in the Figure as a dot-dashed line. We infer the linear bias parameter from the ratio of halo-matter cross-power spectrum and matter power spectrum on scales $k < 1.5 \times 10^{-2} h$ Mpc^{-1}

$$\hat{b}_{1,\mathrm{hm}} = \frac{\hat{P}_{\mathrm{hm}}}{\hat{P}_{\mathrm{mm}}}. \tag{41}$$

See Fig. 14 and Appendix A for why we have to restrict the linear bias fitting to large scales even in the initial conditions. The advantage of the cross-power spectrum is that it should be free of stochasticity contributions and fully described by linear bias [38] on large scales.

There is a clear enhancement of the data in Fig. 5 compared to the linear bias model on small scales. Thus we consider the quadratic bias model,

$$\xi^{(c)}_{\mathrm{hh}}(r) = b_1^2 \xi_{\mathrm{mm,lin}}(r) + \frac{1}{2} b_2^2 \xi^2_{\mathrm{mm,lin}}(r), \tag{42}$$

and fit for the quadratic bias parameter on scales exceeding the maximum. The resulting continuous correlation function is shown as the dashed line in Fig. 5. It does not fully account for the enhanced clustering outside the exclusion scale. The inferred bias parameters are given in Table I and are shown in Fig. 6 and will be discussed in more detail below. Note that the above fit is performed using an unsmoothed version of the linear correlation function, whereas the local bias model relies on an explicit

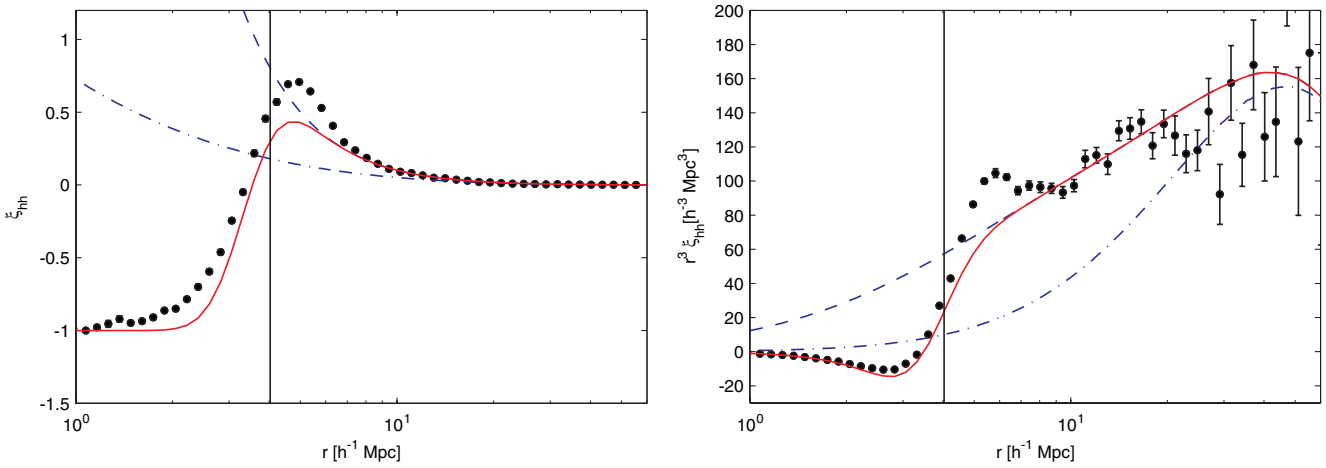

FIG. 5 (color online). Example of the halo-halo correlation function of the traced back haloes for mass bin V. The vertical solid line is the fitted exclusion radius. The dot-dashed line shows the linear bias contribution, whereas the dashed line shows linear plus quadratic bias. Note that the second order bias parameter was fitted to the correlation function and does deviate quite strongly from the PBS prediction. The red solid line shows a simple model for halo exclusion Eq. (42). In the right panel we show the integrand of the Fourier transform $r^3 \xi_{\mathrm{hh}}(r)$, which is of essential importance for the stochasticity modeling.

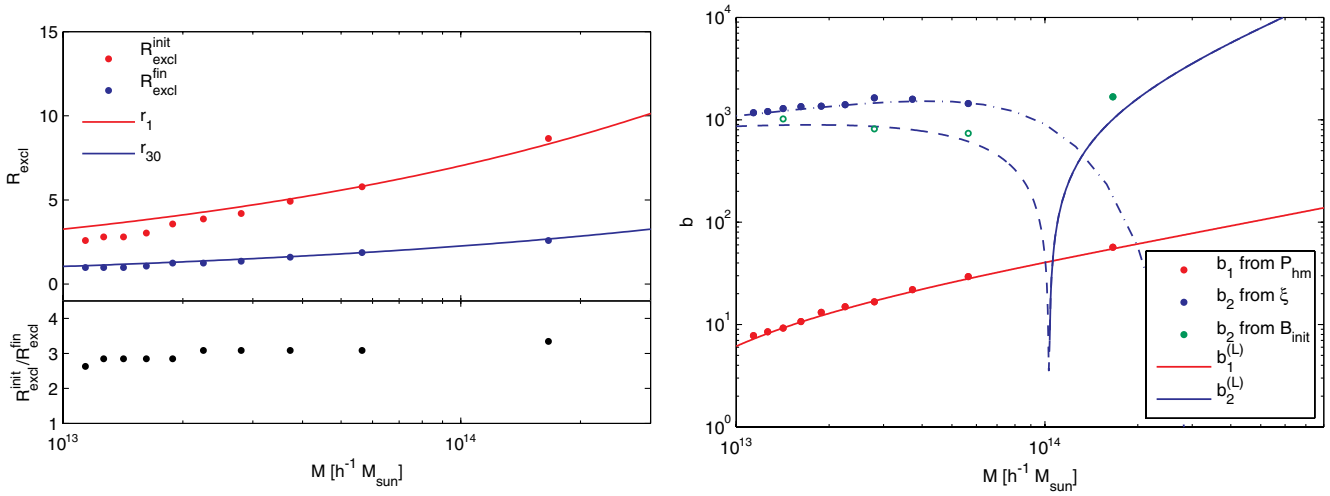

FIG. 6 (color online). *Upper left panel*: Exclusion radii measured in the initial and final halo-halo correlation functions. The exclusion radius is defined as a fixed fraction of the maximum in the halo-halo correlation function (see Fig. 5). The red line shows the naïve estimate of the exclusion radius $R = (3M/4\pi\bar{\rho})^{1/3}$, whereas the blue line shows $R = (3M/4\pi\bar{\rho}(1+\delta))^{1/3}$ with $\delta \approx 30$. This fitted overdensity is clearly distinct from the spherical collapse prediction of $\delta = 180$. *Lower left panel*: Ratio of the initial and final exclusion radii. *Right panel*: Bias parameters fitted from the halo-matter cross-power spectrum (b_1), the halo-matter bispectrum and the correlation function (b_2). Note that b_2 fitted from the correlation function deviates strongly from the value predicted by the peak-background split (blue solid and dashed for positive and negative) and the value inferred from the bispectrum. The dash-dotted lines show a fit used for extrapolation purposes.

smoothing scale. Here, we argue that the smoothing scale for the local bias model should be related to the Lagrangian scale of the haloes and thus be smaller than the exclusion scale. In this case, the smoothing scale does typically not affect the scale dependence of the correlation function (except for around the BAO scale).

In our above discussion we have assumed a sharp transition between the exclusion regime and the clustering regime. This is certainly unphysical as is obvious in Fig. 5. The smoothness is probably caused by a number of phenomena, for instance mass variation within the mass bins (should be quite small $<1.3h^{-1}$ Mpc for the highest-mass bin which has $R \approx 8h^{-1}$ Mpc and even smaller for the lower-mass bins) or alignment of triaxial haloes. The lack of a physically motivated, working model for the transition forces us to employ a somewhat *ad hoc* functional form for the step, which is based on a lognormal distribution of halo distances (see Appendix C),

$$\xi_{\rm hh}^{\rm (d)}(r) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\log_{10}(r/R)}{\sqrt{2}\sigma}\right)\right][\xi_{\rm hh}^{\rm (c)}(r) + 1] - 1. \qquad (43)$$

For alternative implementations of exclusion windows in the context of the halo model see [39,40]. The resulting shape of the correlation function is shown as the red solid line, where the smoothing was chosen to be $\sigma \approx 0.09$ and seems to be quite independent of halo mass. The model clearly underestimates the peak in the data in the log-linear plot. Our final goal is to construct an accurate model for the effect of exclusion and nonlinear clustering on the power spectrum. Thus, we should not only check the validity of our model on plots of the correlation function itself but also on the integrand in the Fourier integrals $r^3\xi_{\rm hh}(r)\mathrm{d}\ln r$. We do so in the right panel of Fig. 5, where it is obvious that the model does not reproduce the exact shape of the correlation function outside of the exclusion scale. However, we certainly improved over the naive linear biasing on all scales and obtained a reasonable parametrization of exclusion and nonlinear clustering effects.

Let us now come back to the mass dependence of the inferred bias and exclusion parameters. As we show in Fig. 6, the linear bias b_1 is in very good agreement with the bias parameters inferred from a Sheth-Tormen mass function [41] rescaled to the initial conditions at $z_{\rm i} = 49$. For the second-order bias we compare the measurement from the correlation function to measurements from the bispectrum of the protohaloes and the second-order bias inferred from the Sheth-Tormen mass function. The latter agrees reasonably well with the bispectrum measurement reproducing the zero crossing in the theoretical bias function. Note that the bispectrum measurement (for details of the approach see [29]) uses only large-scale information and is thus a clean probe for second-order bias. Isolating second-order bias effects in the correlation function is less straightforward. With decreasing scale, higher and higher bias parameters become important, and to our knowledge there is no established scale down to which a certain order of bias can be trusted to a given precision. Our fitting procedure was led by the goal of obtaining a good parametrization of the correlation function, which could subsequently be used to calculate the corresponding power spectra and the corrections to the linear bias model. Note

that due to the functional form of Eq. (42) this fitting approach allows inference of the magnitude of b_2, but not its sign. The second-order bias parameters obtained in this way deviate significantly from the theoretical bias function and the bispectrum measurement. The most severe failure of the model is the imaginary b_2 for the highest-mass bin. In this case the deviation is connected to corrections arising from the peak constraint, as we explain in Appendix A. We find that the initial Lagrangian second-order bias parameter in Fig. 6 can be roughly fitted as follows,

$$b_2 = 1100\left(\frac{M}{10^{13}h^{-1}M_\odot}\right)^{0.35}\exp\left[-\left(\frac{M}{10^{14}h^{-1}M_\odot}\right)^2\right]. \quad (44)$$

We will use this fitting function for extrapolation in mass and redshift in Sec. V D. The initial second-order bias parameters for halo samples identified at higher redshifts ($z = 0.5$ and $z = 1$) are roughly the same as for the $z = 0$ halo sample.

Although we will use this b_2 fit to predict the resulting stochasticity corrections, we do not believe that the observed scale dependence of the correlation function is solely a second-order bias effect. We checked an expansion of Eq. (42) to higher orders in the correlation function using peak-background split bias parameters. Even up to tenth order there is no considerable improvement in the fit. Thus we argue that the enhancement is a nonperturbative effect (e.g. peak bias) and consider the ξ^2 scale dependence as a reasonably well working phenomenological parametrization rather than a physical truth. We hope to shed more light on this issue in a forthcoming paper.

C. Stochasticity matrix

In Fig. 7 we show the diagonals of the stochasticity matrix measured in our simulations in Lagrangian space ($z_i = 49$) and Eulerian space ($z_f = 0$). The most remarkable observation in this plot is the agreement between the results, given the different amplitude of the growth factors and the linear bias parameters at these two times. This is a result of the fact, that gravity cannot introduce or alter k^0 dependencies [42] due to mass and momentum conservation. This can for instance be seen in the low-k limit of standard perturbation theory [33]: The mode coupling term P_{22} is a gravity-gravity correlator and thus scales as k^4, whereas the propagator term P_{13} is a gravity-initial condition correlator and scales as $k^2 P_{\rm lin}$.

For the highest mass bin there is a clear suppression of the noise level on the largest scales which then asymptotes to the fiducial value $1/\bar{n}_h$ at a scale $k \approx 1/R \approx 0.3h$ Mpc^{-1}. Since the radius of the halo shrinks during collapse, this scale is found at a higher wave number in Eulerian space. For the less massive haloes the behavior is not completely monotonic. On large scales we find a noise level slightly exceeding the fiducial value. Going to higher wave numbers the fiducial value is crossed, the residual reaches a minimum and finally asymptotes to $1/\bar{n}$. This behavior can be explained as follows: the clustering scale exceeds the exclusion scale and as we have argued in Sec. III C, the enhanced correlation on the clustering scale leads to a positive contribution on the largest scales that decays for lower wave numbers than the negative exclusion correction.

The wave number at which the stochasticity asymptotes to its fiducial value in the final field is at fairly high wave

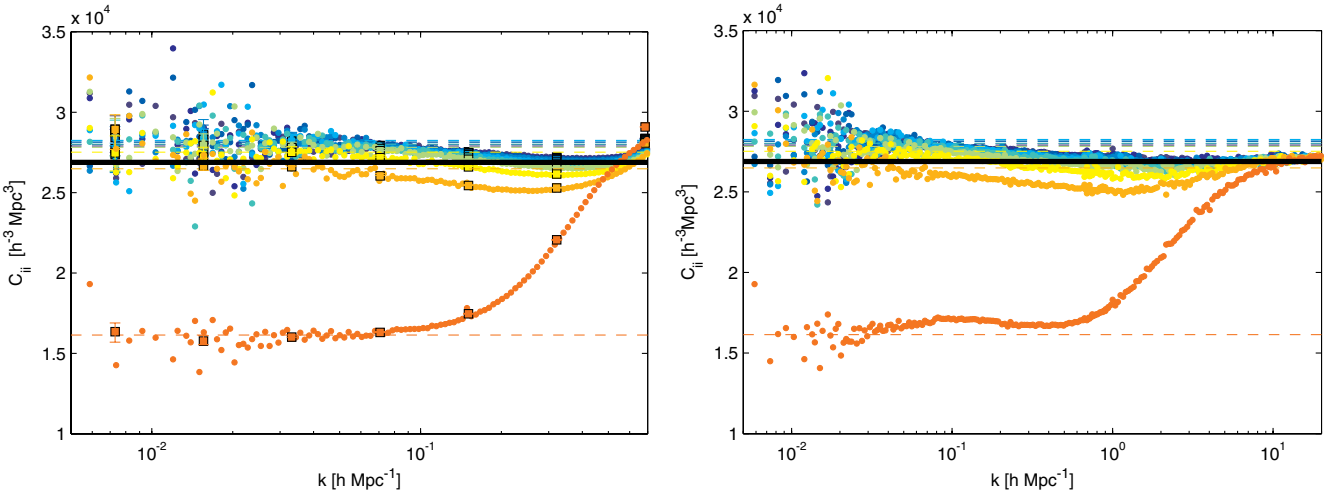

FIG. 7 (color online). Diagonals of the stochasticity matrix σ_{ij}. *Left panel*: Initial conditions $z_i = 49$. *Right panel*: Final field $z_f = 0$. There is remarkable agreement in the large-scale amplitude between initial conditions and final field besides the strong difference in the bias parameters and growth factors. We highlight this fact by the horizontal dashed lines that have the same amplitude in both panels and are matched to the large-scale stochasticity matrix in the initial conditions. For both panels, there is clear evidence for stochasticity going to fiducial $1/\bar{n}$ for high wave numbers, and a modification due to exclusion and clustering for $k \leq 1/R$, where R is the scale of the halo at the corresponding redshift. Note the different scaling of the k scale in the two plots. The mass increases from blue to orange, i.e., top to bottom.

numbers, exceeding the Nyquist frequencies of both $N_c = 512$ and $N_c = 1024$ grids. To probe smaller scales we employ a mapping technique [43,44] that allows us to resolve small scales without having to increase the grid size beyond $N_c = 512$. The technique consists of splitting the box into n parts per dimension and adding these parts to the same grid. This allows inference of each lth mode but also increases the Nyquist frequency by a factor l. We use several different mapping factors $l = 4, 6, 12, 20, 50$ to probe all the scales up to $k \approx 20h$ Mpc^{-1}.

In Fig. 8 we show the diagonals of the stochasticity matrix for one and two mass bins respectively. For the one-bin case we select all the haloes in our simulation, effectively combining all the ten mass bins resulting in $M_{1\mathrm{bin}} = 3.84 \times 10^{13} h^{-1} M_\odot$. For the two-bin case we combine the five lightest and the five heaviest mass bins, resulting in masses $M_{2\mathrm{bin,I}} = 1.47 \times 10^{13} h^{-1} M_\odot$ and $M_{2\mathrm{bin,II}} = 6.21 \times 10^{13} h^{-1} M_\odot$. The plot shows both the initial condition and the final stochasticity and the two agree very well in the low-k limit. The stochasticity correction does not depend on the fiducial stochasticity and thus the scale dependence of the total stochasticity is more pronounced in the wider mass bins due to their lower fiducial shot noise. Interestingly the stochasticity correction for the 1-bin case vanishes in the low-k limit, but is negative in the intermediate regime. This is a sign of a perfect cancellation between exclusion and nonlinear clustering. The final stochasticity seems to be a k-rescaled version of the initial stochasticity. Finally, let us stress that the power spectrum of the wide bins cannot be obtained by a summation of the contributing bin-power spectra from the ten bin case since one has to account for the off-diagonal components of the stochasticity matrix.

In the left panel of Fig. 10 we compare our model Eq. (39) to the measured stochasticity matrix of the ten bins in the initial conditions. The only modification to the model is that we replace the hard cutoff by the smoothed transition Eq. (43). We employ the parameters obtained in the fit to the corresponding halo correlation functions in Sec. V B. The data points in the plot are copied from Fig. 7. Given the differences between the model and the correlation functions in Fig. 5, there is a reasonably good agreement both in large-scale amplitude and scale dependence of the stochasticity correction. We can conclude that while being a relatively crude fit to the correlation function, our model can account for the major effects, exclusion and nonlinear clustering. The drawback is that this model lives in Lagrangian space and cannot be straightforwardly applied to the halo power spectrum in Eulerian space.

In Fig. 15 we show the off-diagonal terms of the stochasticity matrix in the initial conditions and our corresponding model predictions. As for the diagonals discussed above, the corrections can be either positive or negative, depending on whether exclusion or nonlinear clustering dominates The model predictions are in reasonable agreement with the measurements except for the highest mass bin. This failure is connected to the fact that the b_2 parameter for the highest mass bin is imaginary, i.e., we have to set the second-order bias term in the cross correlations to zero. This is a severe problem of our overly simplistic model, which is related to the importance of the peak effects for the correlation function of the highest mass bin (see Appendix A).

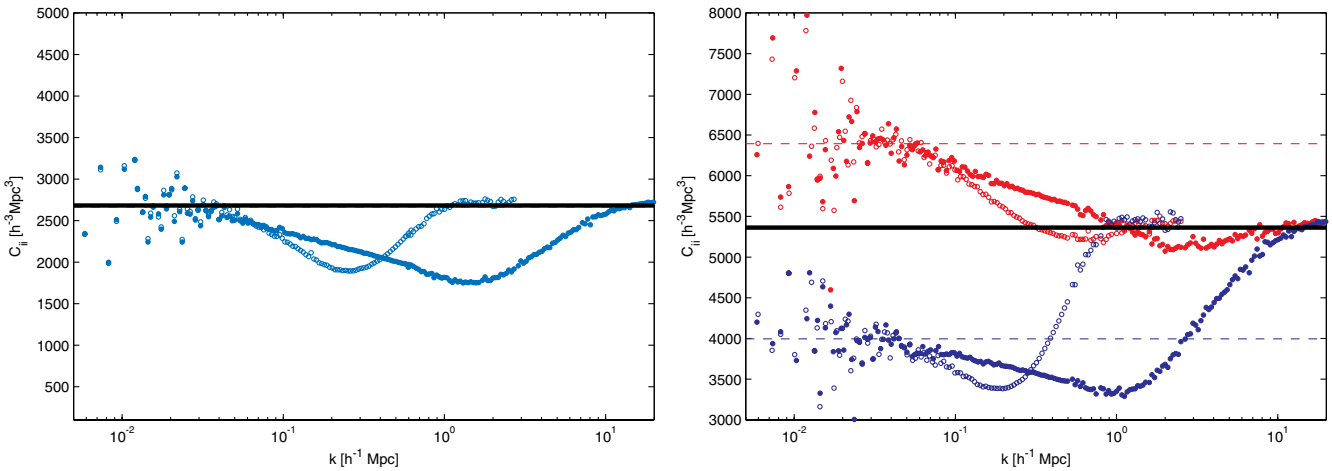

FIG. 8 (color online). *Left panel*: Diagonals of the stochasticity matrix for one mass bin containing all the haloes in our simulation. Open points show the initial condition measurement, whereas the filled points show the final value. The horizontal thick solid line shows the fiducial shot noise $1/\bar{n} = 2680h^{-1}$ Mpc. The Eulerian bias is $b_1^{(\mathrm{E})} = 1.49$. On the largest scales there seems to be a cancellation between the exclusion and nonlinear clustering contributions, resulting in no net correction to the fiducial shot noise. *Right panel*: Same as left panel but for splitting our haloes into two mass bins with equal number density. The upper red points are the measurement for the lighter, lower-bias bin $b_1^{(\mathrm{E})} = 1.25$ and the lower blue points are for the more massive, higher-bias bin $b_1^{(\mathrm{E})} = 1.74$. The fiducial shot noise is $1/\bar{n} = 5362h^{-1}$ Mpc. Note that both mass bins show a significant scale dependence of the stochasticity.

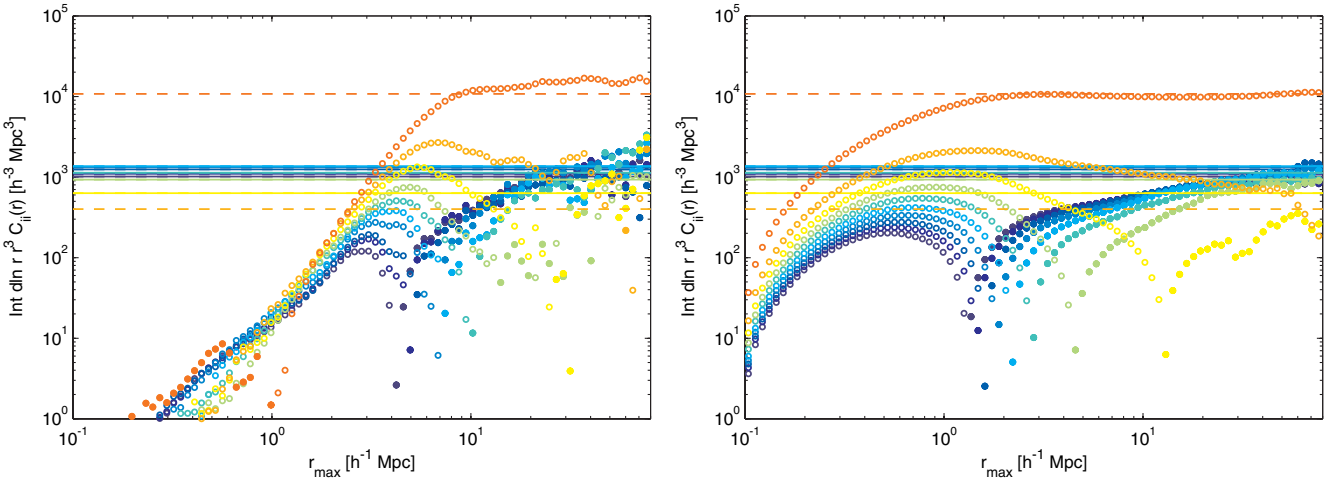

FIG. 9 (color online). Cumulative contributions to the stochasticity correction up to scale r [see Eq. (45)] in configuration space for our ten mass bin sample. Open symbols show negative contributions and filled symbols positive contributions. *Left panel*: Initial conditions. *Right panel*: Final configuration at $z = 0$. We clearly see that the corrections in the initial conditions are dominated on larger scales than in the final configuration, where the negative corrections are clearly in the one-halo regime, where both the halo-matter and matter-matter correlation functions are highly nonlinear. This gives further motivation for the modeling of the effect in Lagrangian space. The horizontal solid (dashed) lines show the positive (negative) stochasticity corrections inferred from Fig. 7.

Let us try to gain some more insight on where the stochasticity corrections arise in Lagrangian and Eulerian space. For this purpose we consider the configuration space version of the diagonal of the stochasticity matrix defined in Eq. (38),

$$C_{ii}(r) = \xi_{ii}(r) - 2b_{1,i}\xi_{i\delta}(r) + b_{1,i}^2\xi_{\delta\delta}(r). \tag{45}$$

The stochasticity level in the $k \to 0$ limit is then given by

$$C_{ij}(k) \xrightarrow{k\to 0} \int_0^\infty \mathrm{d}\ln r r^3 \sigma_{ij}(r). \tag{46}$$

In Fig. 9, we show the above integral as a function of the upper integration boundary where the full large scale stochasticity correction would be obtained by taking this boundary to infinity. Comparing the contributions in the initial conditions and the final configuration, we clearly see that the large-scale stochasticity arises on different scales at the two times. We clearly see that the negative stochasticity corrections are dominated at much smaller scales in the final configuration as compared to the initial conditions. At these scales the halo-matter and matter-matter correlation functions are dominated by the one halo term, i.e., the halo profile, which complicates quantitative predictions of the exclusion effect in the final configuration and motivates our Lagrangian approach.

D. Redshift and mass dependence of the correction

The subtraction of the fiducial $1/\bar{n}$ shot noise from the power spectrum will lead to a biased estimate of the continuous halo power spectrum. Thus, estimating the bias as

$$\hat{b}_{1,\mathrm{hh}} = \sqrt{\frac{\hat{P}_{\mathrm{hh}} - 1/\bar{n}}{\hat{P}_{\mathrm{mm}}}} \tag{47}$$

will lead to a flawed estimate of the bias. Indeed, it has been found in simulations that the biases estimated from the auto- and cross-power spectra are generally not in agreement [45,46]. Studying for instance Table I in [46] we see that $\hat{b}_{1,\mathrm{hh}}$ exceeds $\hat{b}_{1,\mathrm{hm}}$ for low-mass haloes at redshift 0 indicating that the fiducial shot noise subtraction underestimates the true noise level. For high-mass objects the opposite happens; the bias from the cross-power exceeds the bias from the auto-power indicating that the employed $1/\bar{n}$ shot noise subtraction overestimates the true noise level.

Let us try to understand this effect in more detail. Based on our model, subtraction of the fiducial shot noise on large scales leaves us with the linear bias term plus the stochasticity correction,

$$\hat{P}_{\mathrm{hh}}(k) - \frac{1}{\bar{n}} = \Delta P_{\mathrm{hh}}(k) + b_1^2 P_{\mathrm{lin}}(k) = \hat{b}_{1,\mathrm{hh}}^2 \hat{P}_{\mathrm{mm}}(k), \tag{48}$$

where in absence of shot noise in the cross-power spectrum $b_1 = \hat{b}_{1,\mathrm{hm}}$. Thus we have for systematic error on the linear bias parameter,

$$\frac{\Delta b_1}{b_1} = \frac{\hat{b}_{1,\mathrm{hh}}}{b_1} - 1 \approx \frac{1}{2}\frac{\Delta P_{\mathrm{hh}}}{b_1^2 P_{\mathrm{mm}}}. \tag{49}$$

Consequently, the ratio $\hat{b}_{1,\mathrm{hh}}/\hat{b}_{1,\mathrm{hm}}$ is a function of mass and redshift due to the mass and redshift dependence of the parameters of the model. We show the k dependence of the bias correction in Fig. 10. Since we don't have a reliable model to relate the scale-dependent stochasticity matrix

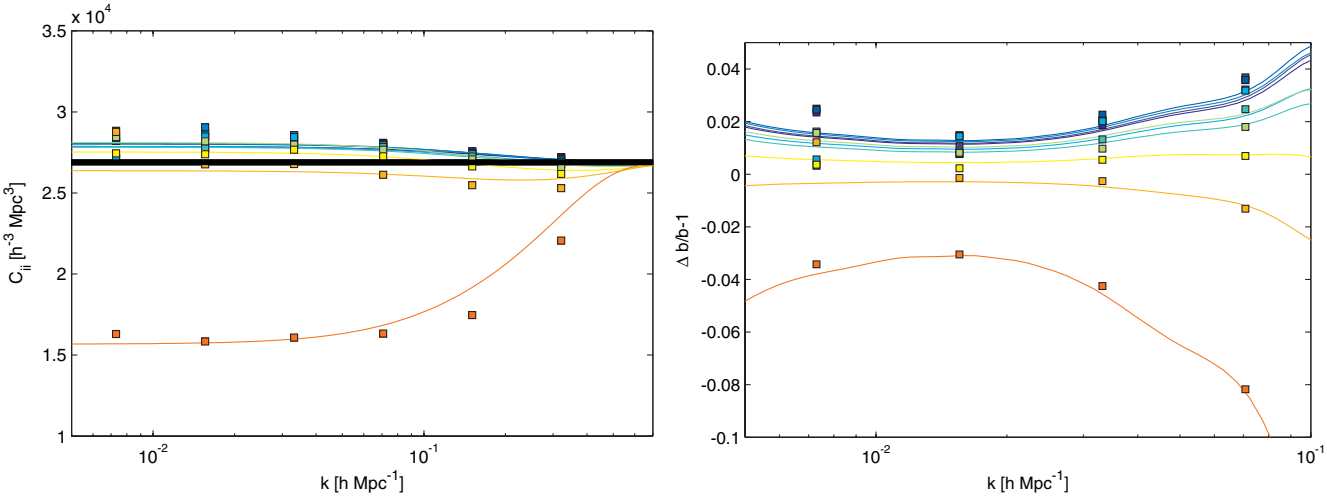

FIG. 10 (color online). *Left panel*: Theoretical prediction for the diagonals of the stochasticity matrix for the ten mass bins. The parameters b_2 and R are measured from the correlation function. The points are the coarsely binned simulation measurement taken from the left panel of Fig. 7. *Right panel*: Bias correction as defined in Eq. (48). While there is clear scale dependence for weakly nonlinear wave numbers $k > 0.1h$ Mpc^{-1}, the correction is fairly flat on large scales, where the cosmic variance errors are largest. It could thus be easily interpreted as a higher or lower bias value. The points show the relative deviation of the bias inferred from the halo-halo power spectrum in the final configuration from the bias inferred from the halo-matter cross-power spectrum.

from Lagrangian space to the one in Eulerian space we employ the scale-dependent stochasticity model from Lagrangian space but divide by the present day linear power spectrum. This procedure should provide a reasonable estimate for the bias corrections on large scales. The linear bias is usually estimated close to the peak of the linear power spectrum, where it is approximately flat and where nonlinear corrections are believed to be negligible. As a result, the bias correction is also fairly flat and would lead to a 1% overestimation of bias for low mass objects and a 3–4% underestimation for clusters. This behavior can qualitatively explain the deviations found in [46].

In Fig. 11 we show the amplitude of the low-k limit of the stochasticity correction for ten equal halo mass bins at redshifts $z = 0$, 0.5, 1. We overplot the theoretical expectation based on our model, linear bias parameters from the peak background split and second-order bias parameters obtained from our phenomenological b_2 relation in Eq. (44). In particular, we calculate the Lagrangian bias parameters and exclusion radii corresponding to the halo samples at $z = 0$, 0.5, 1 and use them to predict the stochasticity correction. As a general result we can see that there is a negative correction for high masses and a positive correction for low masses with a zero-crossing scale that decreases with increasing redshift. The model captures the trends in the measurements relatively well. We are also overplotting the low-k amplitude of the SN for the one- and two-bin samples at $z = 0$ as the squares and diamonds. Besides the fact that these bins are much wider and thus have lower fiducial shot noise, the low-k amplitude is in accordance with the model and also narrower mass bins of the same mass. This fact supports our conjecture, that the stochasticity correction does not depend on the fiducial shot noise, but rather on mass (via the exclusion scale and the linear and nonlinear bias parameters). The negative correction for high masses is dominated by the exclusion term, whose amplitude depends on the linear bias parameter and the exclusion scale. The latter is a function of mass but not a function of redshift, whereas the bias increases with redshift and thus the negative

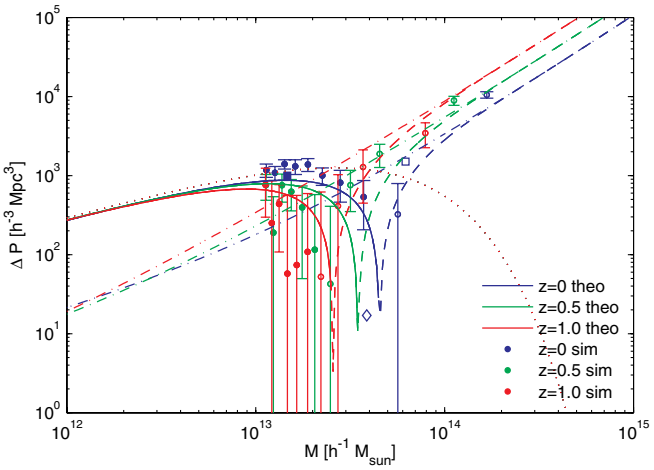

FIG. 11 (color online). Dependence of the stochasticity correction on halo mass and redshift. The lines are based on linear bias parameters from the peak-background split and a fitting function modeling b_2. The dashed lines and open points describe negative values. We also include the low-k limits of the one- and two-bin splitting as the diamond and squares, respectively. The dotted line shows the positive low mass correction arising from nonlinear biasing whereas the dash-dotted blue, green and red lines show the negative high-mass exclusion corrections for the three different redshifts.

correction at high masses also increases with redshift. The positive correction at the low-mass end depends on the second-order bias parameter. In our fits to the correlation function we found that this parameter is roughly constant for the three different redshifts under consideration.

E. Eigensystem and combination of mass bins

The stochasticity matrix can be diagonalized as

$$\sum_j C_{ij} V_j^{(l)} = \lambda^{(l)} V_i^{(l)}, \tag{50}$$

where $V_i^{(l)}$ are the eigenvectors and $\lambda^{(l)}$ the corresponding eigenvalues. The eigenvector corresponding to a low eigenvalue can be used as a weighting function in order to construct a halo sample that has the lowest possible stochasticity contamination [4]. Furthermore, the eigenvalues allow for a clean separation of the exclusion and clustering contributions to the total noise correction. We show the measurement and comparison to our model in Fig. 12. The data show pronounced low and high eigenvalues with most of the eigenvalues identical to the fiducial shot noise. The high eigenvalue is probably related to the nonlinear clustering and the low eigenvalue to exclusion. The model also predicts eight of the ten eigenvalues to agree with the fiducial shot noise as well as one high and one low eigenvalue. The exact agreement is not perfect, which is probably due to an imperfect representation of the off-diagonal stochasticity terms. The main problem with the off-diagonals is to estimate the exclusion radii. The right panel of Fig. 12 shows the eigenvectors corresponding to the eigenvalues. The eigenvector corresponding to the low eigenvalue is clearly connected to the mass, i.e., the exclusion volume, whereas the eigenvector corresponding to the high eigenvalue is clearly connected to the second-order bias.

So far we have concentrated on the stochasticity correction for narrow mass bins and quantified them in terms of the corresponding stochasticity matrix. If all the corrections were linear in the parameters of the model, all we needed to do is to calculate the corresponding mean parameters of the sample and use them to calculate the stochasticity correction for the combined sample. However, the corrections are in general a nonlinear function of the parameters. The wider the mass bins the less exact is a bulk description by a set of mean parameters. It should be more exact to consider subbins and combine them. Thus we need to calculate the stochasticity correction for narrow mass bins $M \in [\underline{M}_i, \bar{M}_i] i = 1, \ldots, h$. Then, when considering samples that span a wide range of halo masses or realistic galaxy samples, we need to weight the prediction for the stochasticity correction accordingly. The weighted density field is then

$$\tilde{\delta} = \frac{\sum_i w_i \delta_i}{\sum_i w_i} \tag{51}$$

and the corresponding noise level of the combined sample is given by

$$\tilde{C} = \frac{\sum_{ij} w_i C_{ij} w_j}{(\sum_i w_i)^2}. \tag{52}$$

For halo samples the weighting is given by the mean number density in a mass bin,

$$w_i = \int_{\underline{M}}^{\bar{M}} \mathrm{d}M n(M). \tag{53}$$

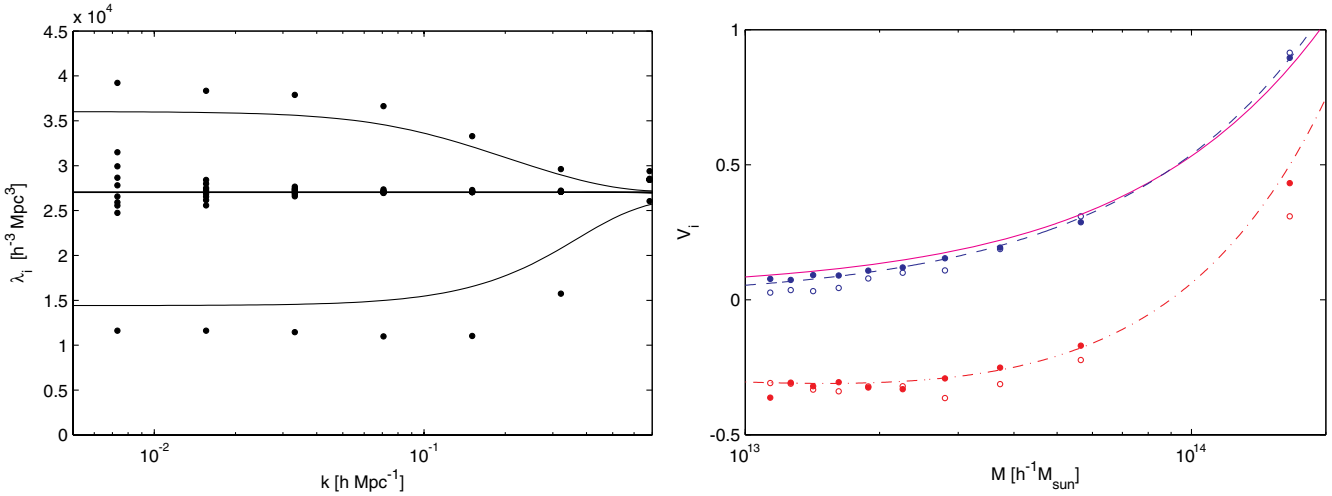

FIG. 12 (color online). *Left panel*: Eigenvalues of the stochasticity matrix for the traced back $z = 0$ haloes. The lines show our prediction based on the modelling of the initial correlation function whereas the points are based on the diagonalization of the measured stochasticity matrix of the traced back haloes. *Right panel*: Eigenvectors in the $k \to 0$ limit. The blue (upper) and red (lower) filled points show the measured eigenvectors for the lowest and highest eigenvalue in the initial conditions, whereas the open points show the respective prediction of our model. The blue (dashed) line shows mass weighting, the red (dash-dotted) line, b_2-weighting, and the magenta (solid) line shows the modified mass weighting proposed by [4].

A similar weighting scheme can be derived for galaxies for the two-halo term in the context of the halo model.

F. A realistic galaxy sample

Let us now see how the stochasticity matrix behaves for a realistic galaxy sample. In Halo Occupation Distribution (HOD) models [47,48] the occupation number $N_g(M)$ is usually split into a central and a satellite component $N_g = N_c + N_s$. In Fig. 13 we show the stochasticity of the Luminous Red Galaxy (LRG) sample described in [49,50]. The total number density of the LRGs is $\bar{n}_g = 7.97 \times 10^{-5} h^{-3}$ Mpc^3 corresponding to a fiducial shot noise of $1/\bar{n}_g \approx 1.25 \times 10^4 h^{-3}$ Mpc^3. The effective stochasticity level for the full sample is $\mathrm{SN}_{\mathrm{eff}} = 1.09 \times 10^4 h^{-3}$ Mpc^3, corresponding to a correction of $\Delta P_{gg} = -1.8 \times 10^3 h^{-3}$ Mpc^3. The satellite fraction of the galaxy sample is 4.9%. Let us try to understand the total correction based on the constituent central, satellite and central-satellite cross-power spectra. The sum of these three components weighted according to Eq. (24) agrees with the measured stochasticity of the full sample. The central-central power spectrum dominates the negative stochasticity correction on large scales with a weighted correction of $(1 - f_s)^2 \Delta P_{cc} = -2100 h^{-3}$ Mpc^3 The satellite-satellite power spectrum has a positive one halo contribution on large scales that contributes a weighted correction of $f_s^2 \Delta P_{ss} = +570 h^{-3}$ Mpc^3 The central-satellite cross-power spectrum changes sign but contributes about $\Delta P_{cs} = -370 h^{-3}$ Mpc^3 at $k = 0.03h$ Mpc^{-1}. The amplitude of these corrections could in principle be understood based on a accurate model for the stochasticity correction of the host haloes and the halo model. In this context the corrections are given as [15,16]

$$P_{cc}^{(1h)}(k) = \frac{1}{\bar{n}_c} \tag{54}$$

$$P_{ss}^{(1h)}(k) = \frac{1}{\bar{n}_s} + \frac{1}{\bar{n}_s^2} \int \mathrm{d}M n(M) N_{s,h}(M) [N_{s,h}(M) - 1] u^2(k|M) \Theta(N_{s,h} - 1) \tag{55}$$

$$P_{cs}^{(1h)}(k) = \frac{1}{\bar{n}_c \bar{n}_s} \int \mathrm{d}M n(M) N_{c,h}(M) N_{s,h}(M) u(k|M) \Theta(N_{s,h} - 1) \tag{56}$$

$$P_{cc}^{(2h)}(k) = \frac{1}{\bar{n}_c^2} \int \mathrm{d}M n(M) N_{c,h}(M) \int \mathrm{d}M' n(M') N_{c,h}(M') P_{hh}(k|M, M') \tag{57}$$

$$P_{ss}^{(2h)}(k) = \frac{1}{\bar{n}_s^2} \int \mathrm{d}M n(M) N_{s,h}(M) u(k|M) \int \mathrm{d}M' n(M') N_{s,h}(M') u(k|M') P_{hh}(k|M, M') \tag{58}$$

$$P_{cs}^{(2h)}(k) = \frac{1}{\bar{n}_s \bar{n}_c} \int \mathrm{d}M n(M) N_{c,h}(M) \int \mathrm{d}M' n(M') N_{s,h}(M') u(k|M') P_{hh}(k|M, M'). \tag{59}$$

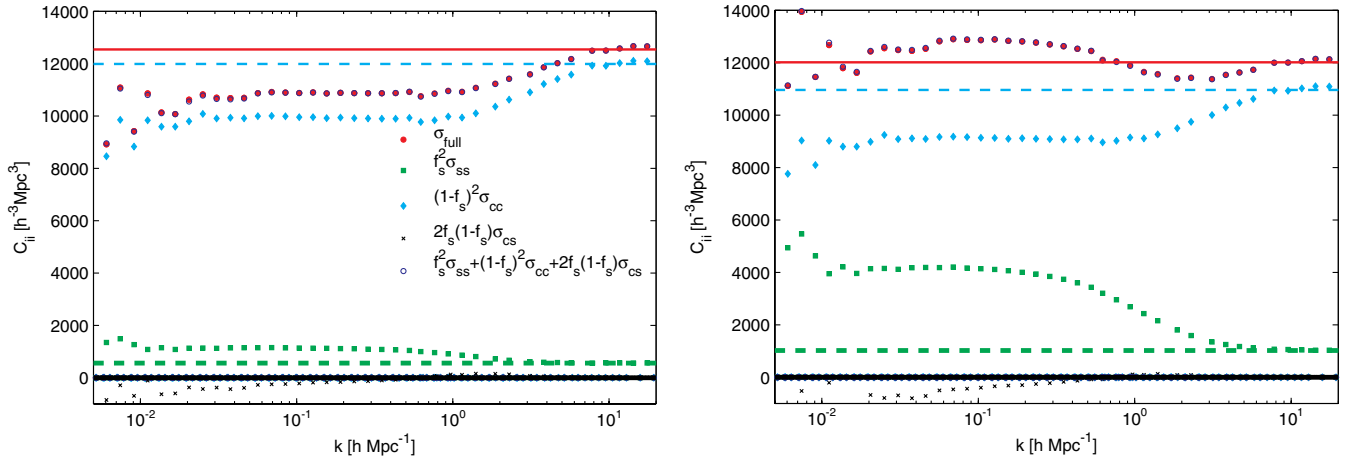

FIG. 13 (color online). Stochasticity of an HOD implementation of a Luminous Red Galaxy sample. The full sample is shown by the red, filled circles and the bare shot noise is shown by the top thin solid red line. We split the sample into the central-central (cyan diamonds and cyan dashed thin line), satellite-satellite (green squares and green thick dashed line) and central-satellite (black crosses and black thick solid line) contributions. The constituent stochasticity levels are weighted according to their contribution to the full galaxy power spectrum [see Eq. (23)] and their sum is given by the open circles, which coincide with the full sample. *Left panel*: Halo occupation distribution model with a satellite fraction of $f_s = 4.9\%$ ([49,50]). *Right panel*: Same as left panel, but for a satellite fraction of $f_s = 8.5\%$.

On large scales we have $u(k|M) \to 1$. Furthermore the halo-halo power spectra can be again split into a linear bias part $b(M)b(M')P(k)$ and a correction term accounting for the discreteness of the host haloes. For the central galaxy sample our model yields a correction of $\Delta P_{\rm cc} \approx -1000h^{-3}$ Mpc^3. More accurate predictions would require a better model of the stochasticity corrections, which in turn requires a better model of exclusion and nonlinear biasing.

We also consider a slightly modified galaxy sample with a larger satellite fraction $f_{\rm s} = 8.47\%$. For this purpose we create a copy of each satellite galaxy at twice its separation from the host halo center. The resulting stochasticity properties are shown in the right panel of Fig. 13. In contrast to the previous case the actual stochasticity now exceeds the fiducial shot noise due to the strong positive contribution of the satellite-satellite one halo term.

VI. CONCLUSIONS

In this paper we discuss effects of the discreteness and nonlinear clustering of haloes on their stochasticity in the power spectrum. The standard model for stochasticity is the Poisson shot noise model with stochasticity given as the inverse of the number density of galaxies, $1/\bar{n}$. Motivated by the results in [4], we study the distribution of haloes in Lagrangian space and estimate the effect of exclusion and nonlinear clustering of protohaloes on the stochasticity. These induce corrections relative to $1/\bar{n}$ in the low-k limit. Exclusion lowers and nonlinear clustering enhances the large-scale stochasticity. The total value of the large-scale stochasticity depends on which of the two effects is stronger, but the amplitude of the correction does not directly depend on the abundance of the sample. These stochasticity corrections must decay to zero for high k, implying they are scale dependent in the intermediate regime. The transition scale is related to either the exclusion scale of the halo sample under consideration or to the nonlinear clustering scale. At the final time (Eulerian space), these transition scales shrink due to the nonlinear collapse, but the low-k amplitude of the stochasticity agrees with Lagrangian space, as expected from mass and momentum conservation.

While the presented model can explain the observed trend of modified stochasticity at a qualitative level, the quantitative agreement is not perfect. This is related to our imperfect modeling of the Lagrangian halo correlation function with a local bias ansatz. A more realistic modeling might be possible in the full framework of peak biasing in three dimensions, as the one-dimensional results in Sec. III E indicate.

We also discuss the effects of satellite galaxies, when a galaxy sample with a nonvanishing satellite fraction is considered. In this case the stochasticity can dramatically deviate from the auxiliary $1/\bar{n}$ value on large scales. In this case one has to identify the number density of host haloes to infer the stochasticity on large scales and account for the fact that on scales below the typical scale of the satellite profile there is a transition to the fiducial Poisson shot noise of the galaxy sample.

Finally, we consider the stochasticity matrix of haloes of different mass. We show that diagonalization of this matrix gives rise to one eigenvalue with a low amplitude, with the eigenvector that approximately scales with the halo mass. This provides an explanation to the stochasticity suppression with mass weighting explored in [3,4]. It would be interesting to explore how the stochasticity corrections imprint themselves in the halo bispectrum, which is a promising probe of inflationary physics [51] and whose measurement becomes realistic in present and upcoming surveys [52].

ACKNOWLEDGMENTS

The authors would like to thank Niayesh Afshordi, Kwan Chuen Chan, Donghui Jeong, Patrick McDonald, Teppei Okumura, Fabian Schmidt, Ravi Sheth, Zvonimir Vlah, and Jaiyul Yoo for useful discussions. The simulations were carried out on the ZBOX3 supercomputer at the Institute for Theoretical Physics of the University of Zurich. This work is supported in part by NASA ATP Grant No. NNX12AG71G, Swiss National Foundation (SNF) under Contract No. 200021-116696/1, WCU Grant No. R32-10130, National Science Foundation Grant No. 1066293. T. B. acknowledges the hospitality of the Aspen Center for Physics. R. E. S. acknowledges support from the ERC under advanced Grant No. 246797, "GALFORMOD," from the European Research Council. V. D. acknowledges support from the Swiss National Sciences Foundation.

APPENDIX A: PERTURBATIVE PEAK EFFECTS IN THE INITIAL CORRELATION FUNCTION

As pointed out by [45], there is strong numerical evidence for the presence of k^2-terms in the linear bias of protohalo. Such terms in fact appear in all the Lagrangian bias factors predicted by the peak model, as can be seen from a large scale expansion of the peak correlation function [23,24]. In the particular case of the peak-matter cross correlation, the linear bias expansion is exact on all scales [25] and agrees with the average density profile derived in [21]. Under the assumption that haloes are represented by peaks, we thus have for the halo-matter power spectrum in the initial conditions,

$$P_{\rm hm}(k) = b_1(1 + \kappa_\delta k^2) W_{{\rm G},R_{\rm pk}}(k) P(k), \qquad \text{(A1)}$$

where $W_{{\rm G},R_{\rm pk}}$ is the Gaussian filter of scale $R_{\rm pk}$. Note that the peak smoothing scale is not necessarily related to the exclusion scale. Similarly one can expand the peak-peak correlation function on large scales and obtain a

scale-dependent linear bias relation that would relate to the halo-halo power spectrum of the form

$$P_{\rm hh}(k) \approx b_1^2(1+\kappa_\delta k^2)^2 W^2_{{\rm G},R_{\rm pk}}(k)P(k). \qquad \text{(A2)}$$

One has to be careful when considering the correspondence between large scales in the correlation function and small wave numbers in the power spectrum. As we stressed above, the amplitude of the low-k power spectrum is tightly coupled to the small scale correlation function. Thus, there is no reason to believe that the large-scale expansion of the correlation function will yield a correct description of the low-k power spectrum. As we will discuss in more detail below, a correct description of the low-k power spectrum in the peak model requires higher order bias expansions or a nonperturbative evaluation of the full peak correlation function.

Given the functional form of the halo-matter power spectrum in the initial conditions we can fit for the linear bias b_1, the relative peak bias κ_δ and the smoothing scale $R_{\rm pk}$. We show the corresponding scale-dependent bias in the left panel of Fig. 14. This plot also shows the linear local (i.e. scale independent) bias parameter used in the main text as the horizontal dashed line.

We can use the inferred bias parameters and peak smoothing scale to calculate the corresponding imprint on the halo-halo correlation function. As we show in the right panel of Fig. 14, the correlation function deviates significantly from the naïve linear scale-independent bias prediction shown as the horizontal dash-dotted line. On scales exceeding $r \approx 25h^{-1}$ Mpc, the measured correlation function is in much better correspondence with the linear peak bias shown as the solid red line. Below $25h^{-1}$ Mpc the peak correlation fails to predict the scale dependence and does actually worse than the linear scale-independent bias. This had to be expected, since the linear peak bias is only accurate on large scales, and on smaller scales higher-order bias corrections need to be taken into consideration. In fact, comparison between the linear peak bias and the full numerical evaluation of the one-dimensional peak model, suggests that 20–$30h^{-1}$ Mpc is a typical breakdown scale for the linear peak bias. We will explore the convergence properties of the perturbative peak model in more detail in a forthcoming paper.

While accurately predicting the functional form of the scale-dependent bias, the mass dependence of the coefficients deviates from the predictions of the peak model. The main reason for discussing the peak corrections here, is that the presence of the peak corrections invalidates the simple b_2 fitting procedure following Eq. (42). In this simple approach we consider the linear bias from the low-k limit of the cross-power spectrum and consider the positive correction $b_2^2\xi^2$ on top of it. The right panel of Fig. 14 rather suggests that the quadratic bias corrections need to be considered on top of the linear peak bias. Furthermore the peak model predicts several second-order bias contributions in contrast to the one bias contribution arising from the local model. Fitting the second-order peak bias parameters would require higher-order spectra, such as the protohalo bispectrum and even the second-order biasing might not be sufficient to explain the upturn in the halo-halo correlation function. We thus restrict ourselves to the simple local model and stress that the quadratic bias parameters are a phenomenological fit rather than a true quadratic bias and we thus don't expect them to be in accordance with the peak-background split prediction.

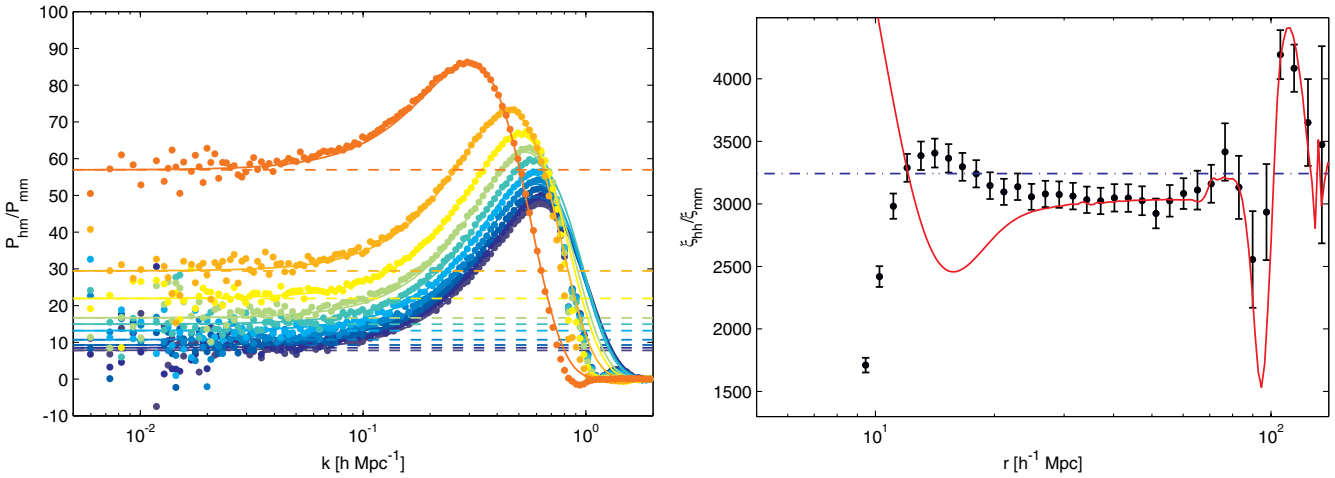

FIG. 14 (color online). *Left panel*: Protohalo-matter cross-power spectrum. Horizontal dashed lines show the linear bias and solid lines a fit using the functional form suggested by the peak model. *Right panel*: Initial condition halo-halo correlation function divided by linear initial condition correlation function for mass bin X with linear scale-independent bias (horizontal dash-dotted) and linear peak bias (red solid). Around the BAO scale and down to $r = 30h^{-1}$ Mpc, the peak model describes the ratio of halo and matter correlation functions much better, but then it fails miserably.

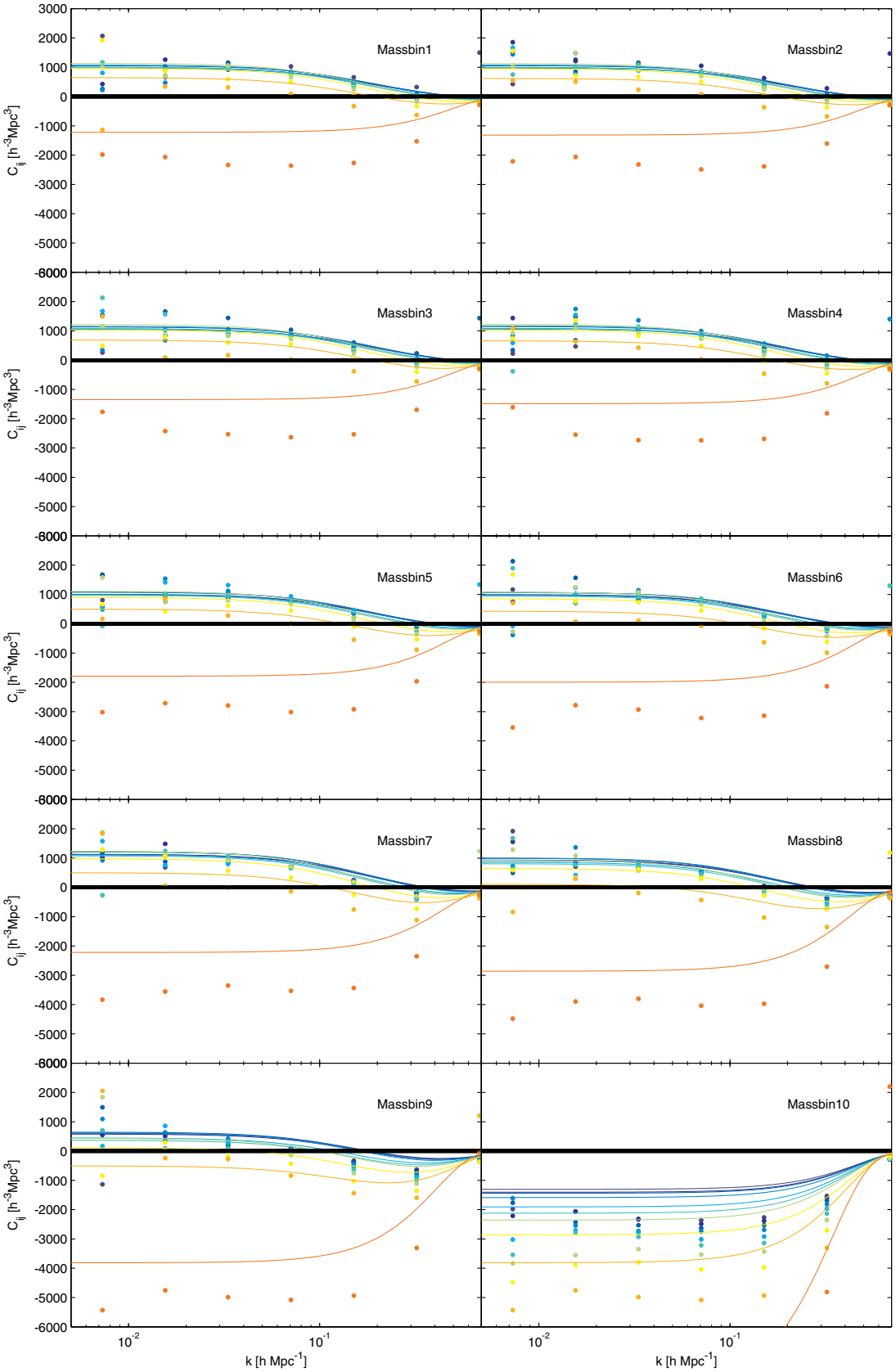

FIG. 15 (color online). Cross terms of the stochasticity matrix of the $z = 0$ haloes traced back to the initial conditions. The text in the panel refers to the mass bin with which we cross correlate all the other bins. Note that for reference we also include the diagonals of the matrix with the fiducial shot noise subtracted out. Mass increases from top to bottom, i.e., blue to orange.

APPENDIX B: THE ONE-DIMENSIONAL PEAK MODEL

Here we review the major steps in the derivation of the one-dimensional peak model following [26]. As mentioned before, the peak model associates maxima of the density fields with the formation sites of dark matter haloes. Thus, we are interested in the clustering statistics of these points. For simplicity we will only consider a one-dimensional field, which would be for a example a skewer through the full three-dimensional cosmological density field. The number density of peaks can be written as a sum over delta functions at the peak positions $x_{\rm pk}$,

$$n_{\rm pk}(x) = \sum_{\rm pk} \delta^{\rm (D)}(x - x_{\rm pk}), \tag{B1}$$

and $\delta_{\rm pk}(x) = n_{\rm pk}(x)/\bar{n}_{\rm pk} - 1$. Using that the first derivative of the field vanishes at the peak position, we can expand the density field around the peak position

$$\delta(x) \approx \delta(x_{\rm pk}) + \frac{1}{2}\delta''(x_{\rm pk})(x - x_{\rm pk})^2. \tag{B2}$$

Taking the derivative, we obtain

$$\delta'(x) \approx \delta''(x_{\rm pk})(x - x_{\rm pk}). \tag{B3}$$

Using the transformation properties of the Dirac delta, we have

$$\delta^{\rm (D)}(x - x_{\rm pk}) = \delta''(x_{\rm pk})\delta^{\rm (D)}(\delta'). \tag{B4}$$

This expression is known as the Kac-Rice formula [24,53]. The mean number density is readily obtained as an integral over the one-point probability density function of the field amplitude, slope, and curvature,

$$\bar{n}_{\rm pk} = \langle n_{\rm pk}(x)\rangle = \int {\rm d}\boldsymbol{y}\mathbb{P}_{\rm 1pt}(\boldsymbol{y})\delta''\delta^{\rm (D)}(\delta'), \tag{B5}$$

where $\boldsymbol{y} = (\delta, \delta', \delta'')$. The two-point correlation function is then given by

$$\begin{aligned}\langle\delta_{\rm pk}(x_1)\delta_{\rm pk}(x_2)\rangle &= \int {\rm d}\boldsymbol{Y}\mathbb{P}_{\rm 2pt}(|x_1 - x_2|;\boldsymbol{Y})\delta_{\rm pk}(x_1)\delta_{\rm pk}(x_2) \qquad \text{(B6)}\\ &= \frac{1}{\bar{n}_{\rm pk}^2}\int {\rm d}\boldsymbol{Y}\mathbb{P}_{\rm 2pt}(|x_1 - x_2|;\boldsymbol{Y})\delta_1''\delta_2''\delta^{\rm (D)}(\delta_1')\delta^{\rm (D)}(\delta_2') - 1, \qquad \text{(B7)}\end{aligned}$$

where $\boldsymbol{Y} = (\delta_1, \delta_1', \delta_1'', \delta_2, \delta_2', \delta_2'')$. The one- and two-point probability density function (PDF) are given by

$$\begin{aligned}\mathbb{P}_{\rm 1pt}(\boldsymbol{y}) &= \frac{1}{\sqrt{(2\pi)^3\det m}}\exp\left[-\frac{1}{2}ym^{-1}y^T\right],\\ \mathbb{P}_{\rm 2pt}(\boldsymbol{Y}) &= \frac{1}{\sqrt{(2\pi)^6\det M}}\exp\left[-\frac{1}{2}YM^{-1}Y^T\right].\end{aligned} \tag{B8}$$

The symmetric 6×6 covariance matrix of the field amplitude and derivatives $M_{ij} = \langle Y_i Y_j\rangle$ can then be written as

$$M = \begin{pmatrix} m & B(r)\\ B^{\rm T}(r) & m\end{pmatrix}, \tag{B9}$$

where the constituent block matrices are given by

$$\begin{aligned} m &= \begin{pmatrix} \sigma_0^2 & 0 & -\sigma_1^2\\ 0 & \sigma_1^2 & 0\\ -\sigma_1^2 & 0 & \sigma_2^2\end{pmatrix},\\ B(r) &= \begin{pmatrix} \xi_0(r) & -\xi_{1/2}(r) & -\xi_1(r)\\ \xi_{1/2}(r) & \xi_1(r) & -\xi_{3/2}(r)\\ -\xi_1(r) & \xi_{3/2}(r) & \xi_2(r)\end{pmatrix}.\end{aligned} \tag{B10}$$

The only remaining ingredient for the evaluation of Eq. (B7) are the correlators of field amplitudes and derivatives, which we obtain by smoothing the three-dimensional density field with a Gaussian filter and considering its values and derivatives along one coordinate axis, which we choose to be the z direction without loss of generality,

$$\begin{aligned}\xi_{(n+m)/2}(r) = \int \frac{{\rm d}^3k}{(2\pi)^3}(-1)^n({\rm i}\mu k)^{n+m}\\ \times\exp[{\rm i}\mu kr]P(k)W_{{\rm G},R_{\rm pk}}(k),\end{aligned} \tag{B11}$$

where $\mu = \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{z}}$. The moments of field amplitudes and derivatives are then given as $\sigma^2_{(n+m)/2} = \xi_{(n+m)/2}(0)$.

APPENDIX C: THE EXCLUSION KERNEL

Let us assume, that there is some scatter around the mean exclusion radius R of some sample due to triaxiality and a finite width of the mass bin. Let us furthermore assume that this scatter is lognormally distributed such that the PDF is given by

$$f(r) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left[-\frac{\ln^2(r/R)}{2\sigma^2}\right]. \tag{C1}$$

Let us now calculate the probability of finding a pair with an actual separation that is smaller than the scale under consideration,

$$F(r) = \int_0^r {\rm d}r' f(r') = \frac{1}{2}\left(1 + {\rm erf}\left[\frac{\ln(r/R)}{\sqrt{2}\sigma}\right]\right). \tag{C2}$$

To find a pair at separation r, its exclusion scale needs to be smaller than the actual separation and the total probability is given as a product of the fiducial probability of finding a pair in a nonexcluded sample and the probability that the actual exclusion scale is smaller than r,

$$\begin{aligned}\delta\tilde{\mathbb{P}}(r) &= F(r)\delta\mathbb{P}(r)\\ &= \bar{n}[1+\xi^{(\mathrm{d})}(r)]\delta V\\ &= F(r)\bar{n}[1+\xi^{(\mathrm{c})}(r)]\delta V. \end{aligned} \tag{C3}$$

We can now infer the correlation of the discrete tracers,

$$\xi^{(\mathrm{d})}(r) = F(r)[1+\xi^{(\mathrm{c})}(r)] - 1. \tag{C4}$$

In the main text we use the logarithm to base 10 for convenience rather than the natural logarithm. The corresponding scatters are related by a simple rescaling by a factor of $\log_{10} e \approx 0.43$.

[1] N. Kaiser, Astrophys. J. **284**, L9 (1984).
[2] J. N. Fry and E. Gaztanaga, Astrophys. J. **413**, 447 (1993).
[3] U. Seljak, N. Hamaus, and V. Desjacques, Phys. Rev. Lett. **103**, 091303 (2009).
[4] N. Hamaus, U. Seljak, V. Desjacques, R. E. Smith, and T. Baldauf, Phys. Rev. D **82**, 043515 (2010).
[5] R. Casas-Miranda, H. J. Mo, R. K. Sheth, and G. Boerner, Mon. Not. R. Astron. Soc. **333**, 730 (2002).
[6] M. Manera and E. Gaztañaga, Mon. Not. R. Astron. Soc. **415**, 383 (2011).
[7] N. Hamaus, U. Seljak, and V. Desjacques, Phys. Rev. D **84**, 083509 (2011).
[8] N. Hamaus, U. Seljak, and V. Desjacques, Phys. Rev. D **86**, 103513 (2012).
[9] R. K. Sheth and G. Lemson, Mon. Not. R. Astron. Soc. **304**, 767 (1999).
[10] R. E. Smith, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **75**, 063512 (2007).
[11] P. J. E. Peebles and E. J. Groth, Astron. Astrophys. **53**, 131 (1976).
[12] J. A. Peacock and D. Nicholson, Mon. Not. R. Astron. Soc. **253**, 307 (1991).
[13] A. Labatie, J.-L. Starck, M. Lachièze-Rey, and P. Arnalte-Mur, arXiv:1009.1232.
[14] The finite grid leads to $\delta(\mathrm{D})(\boldsymbol{k}-\boldsymbol{k}') = \frac{V}{(2\pi)^3}\delta(\mathrm{K})_{\boldsymbol{k},\boldsymbol{k}'}$ and consequently $V\delta^{(\mathrm{K})}_{\boldsymbol{k},\boldsymbol{k}'}P(k) = \langle\delta(\boldsymbol{k})\delta(-\boldsymbol{k}')\rangle$.
[15] U. Seljak, Mon. Not. R. Astron. Soc. **318**, 203 (2000).
[16] A. Cooray and R. Sheth, Phys. Rep. **372**, 1 (2002).
[17] D. J. Eisenstein *et al.*, Astron. J. **122**, 2267 (2001).
[18] N. Roth and C. Porciani, Mon. Not. R. Astron. Soc. **415**, 829 (2011).
[19] K. Chuen Chan and R. Scoccimarro, Phys. Rev. D **86**, 103519 (2012).
[20] J. Beltrán Jiménez and R. Durrer, Phys. Rev. D **83**, 103509 (2011).
[21] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).
[22] J. A. Peacock and A. F. Heavens, Mon. Not. R. Astron. Soc. **217**, 805 (1985).
[23] V. Desjacques, M. Crocce, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **82**, 103529 (2010).
[24] V. Desjacques, Phys. Rev. D **78**, 103503 (2008).
[25] V. Desjacques and R. K. Sheth, Phys. Rev. D **81**, 023526 (2010).
[26] S. L. Lumsden, A. F. Heavens, and J. A. Peacock, Mon. Not. R. Astron. Soc. **238**, 293 (1989).
[27] A. F. Heavens and R. K. Sheth, Mon. Not. R. Astron. Soc. **310**, 1062 (1999).
[28] P. McDonald, Phys. Rev. D **74**, 103512 (2006).
[29] T. Baldauf, U. Seljak, V. Desjacques, and P. McDonald, Phys. Rev. D **86**, 083540 (2012).
[30] K. C. Chan, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **85**, 083509 (2012).
[31] H. J. Mo and S. D. M. White, Mon. Not. R. Astron. Soc. **282**, 347 (1996).
[32] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).
[33] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, Phys. Rep. **367**, 1 (2002).
[34] D. N. Spergel *et al.*, Astrophys. J., Suppl. Ser. **170**, 377 (2007).
[35] V. Springel, Mon. Not. R. Astron. Soc. **364**, 1105 (2005).
[36] U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).
[37] M. Crocce, S. Pueblas, and R. Scoccimarro, Mon. Not. R. Astron. Soc. **373**, 369 (2006).
[38] N. Frusciante and R. K. Sheth, J. Cosmol. Astropart. Phys. 11 (2012) 016.
[39] J. L. Tinker, D. H. Weinberg, Z. Zheng, and I. Zehavi, Astrophys. J. **631**, 41 (2005).
[40] P. Valageas and T. Nishimichi, Astron. Astrophys. **527**, A87 (2011).
[41] R. Scoccimarro, R. K. Sheth, L. Hui, and B. Jain, Astrophys. J. **546**, 20 (2001).
[42] P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University, Princeton, NJ, 1980).
[43] A. Jenkins, C. S. Frenk, F. R. Pearce, P. A. Thomas, J. M. Colberg, S. D. M. White, H. M. P. Couchman, J. A. Peacock, G. Efstathiou, and A. H. Nelson, Astrophys. J. **499**, 20 (1998).
[44] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman, Mon. Not. R. Astron. Soc. **341**, 1311 (2003).
[45] A. Elia, A. D. Ludlow, and C. Porciani, Mon. Not. R. Astron. Soc. **421**, 3472 (2012).
[46] T. Okumura, U. Seljak, and V. Desjacques, J. Cosmol. Astropart. Phys. 11 (2012) 014.
[47] A. A. Berlind and D. H. Weinberg, Astrophys. J. **575**, 587 (2002).
[48] Z. Zheng, A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Davé, C. S. Frenk, N. Katz, and C. G. Lacey, Astrophys. J. **633**, 791 (2005).

[49] T. Baldauf, R. E. Smith, U. Seljak, and R. Mandelbaum, Phys. Rev. D **81**, 063531 (2010).
[50] R. Reyes, R. Mandelbaum, U. Seljak, T. Baldauf, J. E. Gunn, L. Lombriser, and R. E. Smith, Nature (London) **464**, 256 (2010).
[51] T. Baldauf, U. Seljak, and L. Senatore, J. Cosmol. Astropart. Phys. 4 (2011) 006.
[52] F. Marin *et al.*, arXiv:1303.6644.
[53] J. Adler and R. J. Taylor, *Random Fields and Geometry* (Springer, New York, 2007).