

Matter bispectrum of large-scale structure: Three-dimensional comparison between theoretical models and numerical simulations

Andrei Lazanu,^{1,*} Tommaso Giannantonio,^{2,1,†} Marcel Schmittfull,^{3,‡} and E. P. S. Shellard^{1,§}

¹*Centre for Theoretical Cosmology, DAMTP, University of Cambridge, CB3 0WA, United Kingdom*

²*Kavli Institute for Cosmology Cambridge, Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*

³*Berkeley Center for Cosmological Physics, Department of Physics and Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720, USA*

(Received 18 November 2015; published 20 April 2016)

We study the matter bispectrum of the large-scale structure by comparing different perturbative and phenomenological models with measurements from N -body simulations obtained with a modal bispectrum estimator. Using shape and amplitude correlators, we directly compare simulated data with theoretical models over the full three-dimensional domain of the bispectrum, for different redshifts and scales. We review and investigate the main perturbative methods in the literature that predict the one-loop bispectrum: standard perturbation theory, effective field theory, resummed Lagrangian and renormalized perturbation theory, calculating the latter also at two loops for some triangle configurations. We find that effective field theory (EFT) succeeds in extending the range of validity furthest into the mildly nonlinear regime, albeit at the price of free extra parameters requiring calibration on simulations: EFT is found to be accurate to 5% up to a scale of $k_{\max}^* \approx 0.4 h/\text{Mpc}$ at $z = 1$, compared with $k_{\max}^* \approx 0.2 h/\text{Mpc}$ at $z = 1$ for most other one-loop perturbative methods. For the more phenomenological halo model, we confirm that despite its validity in the deeply nonlinear regime it has a deficit of power on intermediate scales, which worsens at higher redshifts (the maximum deficit in the amplitude correlator is $\sim 20\%$ at $z = 1$, and up to 40% at $z = 2$); this issue is ameliorated, but not solved, by combined halo-perturbative models. We show from simulations that in this transition region there is a strong squeezed bispectrum component that is significantly underestimated in the halo model at earlier redshifts. We thus propose a phenomenological method for alleviating this deficit, which we develop into a simple phenomenological “three-shape” benchmark model based on the three fundamental shapes we have obtained from studying the halo model. When calibrated on the simulations, this three-shape benchmark model accurately describes the bispectrum on all scales and redshifts considered, providing a prototype bispectrum HALOFIT-like methodology that could be used to describe and test parameter dependencies.

DOI: [10.1103/PhysRevD.93.083517](https://doi.org/10.1103/PhysRevD.93.083517)

I. INTRODUCTION

The Λ CDM model has so far been successful in describing the properties of the Universe, as recently confirmed by the latest *Planck* satellite results [1,2]. The initial conditions of this model are based on the assumption that all the structure in the Universe was generated by quantum fluctuations at primordial times, during an inflationary phase [3,4]. The physics of inflation has been extensively studied in recent years and many scenarios have been proposed [5–9]; distinguishing between the numerous existing models is one of the ultimate goals of cosmology. This problem can be tackled observationally by studying the properties of the perturbations at later times: the cosmic microwave

background (CMB) and the large-scale structure of the Universe (LSS).

CMB anisotropies have provided in the past two decades a wealth of cosmological information, which has been exploited with increasing efficiency by subsequent observational campaigns, up to the exquisite accuracy of the latest results from the *Planck* satellite [2]. The CMB has also provided some of the strongest constraints on inflation. On the one hand, the shape of the CMB two-point statistics (power spectrum) is directly related to the power spectrum of perturbations at the end of inflation, whose parameters and features can thus be accurately constrained [10]. On the other hand, many inflationary models predict a significant non-Gaussian component in the distribution of primordial perturbations [11]: higher-order statistics of the CMB anisotropies, such as the three-point correlation function (bispectrum), have provided strict constraints on such models [12].

Nevertheless, the CMB can primarily supply only two-dimensional data from the surface of last scattering, which

*A.Lazanu@damtp.cam.ac.uk

†T.Giannantonio@ast.cam.ac.uk

‡M.Schmittfull@berkeley.edu

§E.P.S.Shellard@damtp.cam.ac.uk

in temperature has been already almost fully exploited to the limit of cosmic variance by *Planck*. The LSS, traced by current and upcoming galaxy surveys, contains much more information than the CMB due to its three-dimensional nature, and it can thus provide further complementary insight on cosmology across cosmic time. In principle, there is roughly a 1000-fold increase in the number of modes available compared to the CMB [13], but this information is more challenging to extract due to the more limited theoretical understanding of the LSS physics in the low-redshift universe, where additional complexity is added by nonlinear structure formation, the relationship between dark and visible matter (galaxy bias), and redshift-space effects [14]. Indeed, the modelling of galaxy clustering is first based on a description of dark matter clustering; the clustering of collapsed dark matter halos is then defined by introducing halo bias, while the connection to observable galaxies can be made by using halo occupation distribution [15] or halo-abundance matching [16] methods, calibrated on N -body simulations. In this paper, we only describe the clustering of dark matter in real space, leaving the connection to galaxy observables, including bias and redshift-space effects, to subsequent work.

Galaxy surveys like SDSS [17] and BOSS [18,19] have dramatically increased our understanding of the Universe. Ongoing and future surveys, like DES [20,21], LSST [22], *Euclid* [23], DESI [24], *WFIRST* [25] and the proposed SPHEREX mission [26] are expected to increase the precision of the measurements even further.

To date, most cosmological implications from large-scale structure data have been drawn from the power spectrum of galaxies. At linear level, the matter power spectrum encodes all the information available if the primordial random fluctuations are Gaussian. The power spectrum is also sensitive to some classes of primordial non-Gaussianity (PNG) via the scale-dependent galaxy bias [27,28], which has been widely used to obtain competitive PNG constraints [29–34].

However, in order to fully exploit the LSS information and to test all types of PNG, it is important to also study higher-order statistics, such as the bispectrum [35–41]. Even for Gaussian initial conditions, where the primordial bispectrum is zero, nonlinear coupling between Fourier modes produces a nonzero bispectrum due to gravitational collapse [42]. This gravitational bispectrum must be well understood in order to be able to separate the primordial component and to constrain the physics of inflation. At the same time, it can provide additional cosmological information [43], for example on the growth of structure [44–46], and bias parameters [44,47–52].

Modeling the evolution of matter density perturbations beyond linear scales is a complex problem. On relatively large scales, in the quasilinear regime, significant progress has been made using perturbative methods. Arguably, the most common procedure is Eulerian standard perturbation

theory (SPT) [53–56], where the growth of structure is described by a set of differential equations in terms of the present-time density perturbations, expanded to the desired order. Alternatively, in Lagrangian perturbation theory (LPT) [57–63] the fluid equations are written in terms of the initial density perturbations via a displacement field, which reduces to the Zel’dovich approximation at linear order [64]. Both methods have advantages and shortcomings [65–68]; in particular, SPT has a narrow range of validity at low redshift, and its series expansion shows poor convergence properties. LPT has the additional drawback that its perturbative approach cannot predict clustering beyond shell crossing. For these reasons, recent years have seen a proliferation of further developments: SPT has been reformulated in the language of field theory by Refs. [69–74], reorganizing the series expansion in terms of vertices and propagators, and improving its convergence properties [renormalized perturbation theory (RPT)]; this has been later simplified to the MPTBREEZE scheme [75]. Related developments include the large- N expansion [76], the closure theory [77], and renormalization group approaches [78,79]. A resummation technique in Lagrangian space (RLPT) was developed by Ref. [80]; subsequent extensions were developed by Refs. [80–87]. Most recently, the effective field theory of LSS (EFTofLSS) has been developed by Refs. [13,88–95], based on the idea that the contribution of small-scale physics to the quasilinear perturbations can be encapsulated into a set of additional, unknown source terms in the equations of motion, whose value can be fixed by comparison with N -body simulations.

In the fully nonlinear regime, perturbation theories necessarily break down and numerical N -body simulations have to be used to calibrate phenomenological models of gravitational clustering, such as the halo model [96–98]. This formalism is based on the approximation that all matter in the Universe is in the form of spherical halos with a universal density profile and without substructure, and it can be used to describe the matter power spectrum and bispectrum relatively accurately (typically better than 10% at $k < 1 h/\text{Mpc}$ at $z = 0$) [99]. It is however difficult to significantly improve the halo model accuracy beyond the limits set by its underlying assumptions, especially on intermediate scales. For this reason, Refs. [85,100,101] combined a revised version of the halo model, valid on small scales, with perturbative recipes that are more accurate on quasilinear scales. Reference [102] also proposed a halo model extension that improves its accuracy at the cost of 12 extra parameters.

A more drastic approach was introduced by Refs. [103,104], where the physically motivated small-scale one-halo term was replaced with a series expansion in the even powers of k , with free parameters to be calibrated on N -body simulations. It is possible to extend these ideas even further into the direction of phenomenology at the cost

of a reduced physical understanding: the HALOFIT method [105,106] achieves a higher accuracy matter power spectrum by combining halo model-inspired templates with numerous heuristic parameters fit to N -body simulations while, in the ultimate numerical and agnostic approach, matter clustering is directly calculated by interpolating over a grid of N -body simulations spanning a range of different cosmologies [107]. No bispectrum counterpart exists to date for these numerical methods.

At the same time, there has been progress in N -body simulations studies and bispectrum estimators [40,43, 108–111]. In contrast with the standard brute-force method of measuring the bispectrum for all possible triangular configurations, Refs. [110,112] applied to the LSS the modal decomposition of the bispectrum introduced for CMB studies by Refs. [40,113], thus developing a significantly faster and more efficient estimator. A simplified version tailored to estimating the projection of the simulation bispectrum on the tree-level prediction was presented in Ref. [111].

Relatively few measurements of the bispectrum from galaxy surveys exist [114–121]. The state of the art results have recently been obtained by Refs. [122,123] from the BOSS luminous red galaxies. These data have been used to improve the power spectrum constraints on galaxy bias and structure growth; however, to date no primordial non-Gaussianity constraints exist from the LSS bispectrum.

In this paper, we make the first comprehensive comparison of models describing the matter bispectrum as a function of scale and redshift. We review a selection of different models from the literature and we analyze their accuracy on different scales by comparing their predictions with direct estimates of the bispectrum from N -body simulations. We compare two classes of models: methods based on perturbative approaches and phenomenological halo models. The perturbative models considered are tree-level, nonlinear tree-level, SPT, RPT, RLPT, and EFT (all at one loop). In the nonlinear regime, we investigate the standard halo model and a modified halo model combined with EFT, based on the method by Ref. [101]. We base our analysis on a full three-dimensional comparison of the shapes and amplitudes of the bispectra, which allows us to compare all the triangular configurations in the bispectra at once, rather than confronting individual slices in specific configuration limits, as usual with previous work. This approach is relevant to observational forecasts of predicted signal-to-noise where the full statistical significance requires summation over all triangle configurations.

We then develop a simple phenomenological model based on the three fundamental shapes of the halo model components, which provides a good global fit to the simulations. We quantitatively compare the simulations with each of the theoretical models considered and we discuss their advantages and limitations. We also show how to numerically calculate the two-loop bispectrum in

the MPTBREEZE formalism in an infrared-safe manner and we present the results for several scaled triangular configurations.

The plan of this paper is as follows. After a brief introduction to the bispectrum and its three-dimensional estimators in Sec. II, we review the theoretical models we consider in Sec. III (perturbation theory) and Sec. IV (nonlinear and phenomenological models). We then describe the N -body simulations in Sec. V. In Sec. VI we discuss the measured bispectrum shapes, and use this to introduce the phenomenological two-halo boost and three-shape benchmark model. We next present the results of the comparison between the different theoretical models and simulations in Sec. VII, before concluding in Sec. VIII. Several appendixes provide details of the considered models.

II. BISPECTRUM INTRODUCTION

The statistical analysis of random fields, such as the matter density perturbation $\delta \equiv (\rho - \bar{\rho})/\bar{\rho}$, where ρ is the matter density of mean $\bar{\rho}$, involves measuring its N -point correlation functions in real space, or its N -spectra in Fourier space. We consider here the power spectrum and bispectrum, which are defined as

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P(k) \quad (1)$$

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(k_1, k_2, k_3), \quad (2)$$

where δ_D is the Dirac delta function. For statistically homogeneous and isotropic cosmologies, to which we restrict our attention here, the bispectrum only depends on the wave numbers k_1, k_2, k_3 . While the power spectrum is a 1D quantity, as it is simply a real function of the wave number k , the bispectrum is a more complex 3D quantity, as it is a real function of k_1, k_2, k_3 . The bispectrum therefore contains more information, but it is also more cumbersome to study, and it thus requires relatively more advanced techniques to be measured and exploited. We introduce in this section our method for analyzing the full 3D matter bispectrum, and for comparing its observations with theoretical models.

A. Shape and amplitude correlators

In order to compare the observed or simulated bispectra with the corresponding theoretical predictions, we define the signal-to-noise weighted scalar product between two bispectrum shapes i and j [112,124]:

$$\langle B_i, B_j \rangle \equiv \frac{V}{\pi} \int_{\mathcal{V}_B} dV_k \frac{k_1 k_2 k_3 B_i(k_1, k_2, k_3) B_j(k_1, k_2, k_3)}{P(k_1) P(k_2) P(k_3)}, \quad (3)$$

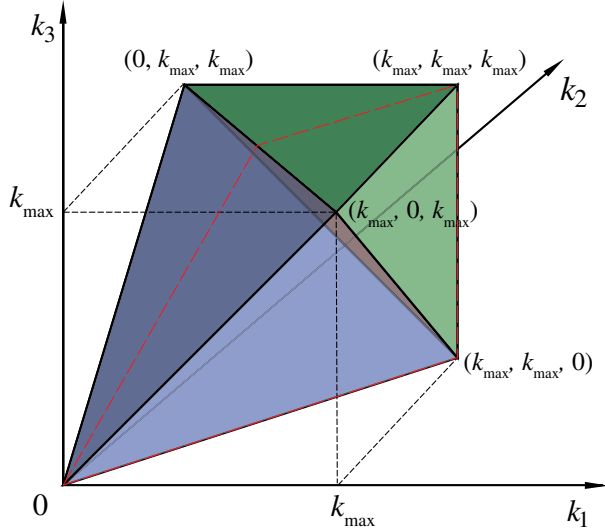


FIG. 1. The tetrapyd bispectrum domain consists of a tetrahedral region (blue) defined by the wave vector triangle condition in Eq. (2), together with a pyramidal region (green) bounded by the resolution limit k_{\max} . For the autocorrelator bispectrum this has a sixfold symmetry, so to illustrate the internal structure of the bispectrum (equilateral limit) we will split the tetrapyd across the vertical plane given by the red dashed lines, removing the front half as shown in Fig. 2.

where the integration domain \mathcal{V}_B is the tetrahedral region of volume V satisfying the triangle condition on the wave numbers k_1 , k_2 and k_3 (such that $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$), together with a chosen resolution limit $k_1, k_2, k_3 < k_{\max}$. The bispectrum domain is the union of a tetrahedron with a triangular pyramid on top (denoted the “tetrapyd”) and is illustrated in Fig. 1. The inner product Eq. (3) provides a natural definition for the signal-to-noise (SN) weighted bispectrum,

$$B_i^{\text{SN}}(k_1, k_2, k_3) \equiv \sqrt{\frac{k_1 k_2 k_3}{P(k_1)P(k_2)P(k_3)}} B_i(k_1, k_2, k_3), \quad (4)$$

where we use the measured (or HALOFIT) power spectrum $P_{\text{NL}}(k)$ for wave numbers in the quasilinear and nonlinear regimes (rather than the linear power spectrum P_{lin}). The SN-weighted bispectrum B_i^{SN} is the relevant quantity observationally if the matter bispectrum could be measured directly, providing optimal forecasts for an ideal survey (i.e. one without experimental noise or systematics). To develop an intuitive understanding of the distinct gravitational bispectrum contributions, we will plot the SN-weighted bispectrum in three dimensions on half the tetrapyd domain as shown in Fig. 2. Although the full tetrapyd has a sixfold symmetry for the isotropic bispectrum of Eq. (2), leaving this redundancy allows us to view B_i^{SN} from equilateral, flattened and squeezed limits simultaneously. (Future work will include bispectrum cross-correlators, e.g., the matter-matter-halo bispectrum where Fig. 2 shows the

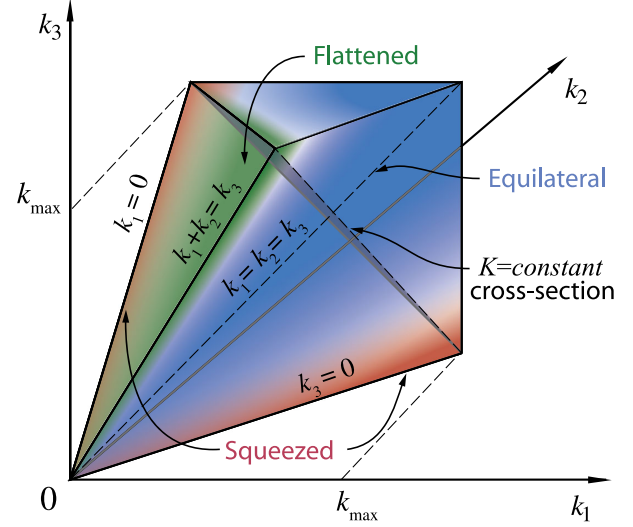


FIG. 2. The split 3D tetrapyd region used to illustrate the SN-weighted bispectrum showing only the back half with $k_1 < k_2$. Color-coded regions show the location of the squeezed (red), flattened (green) and equilateral or constant (blue) shape signals. In the bispectrum Ansatz Eq. (10) the shape $S(k_1, k_2, k_3)$ is defined on the $K \equiv k_1 + k_2 + k_3 = \text{const}$ cross-sectional planes, while the scale dependence $f(K)$ is given along the dashed diagonal $k_1 = k_2 = k_3$.

complete domain, as for recent CMB polarization results [12].)

Based on the scalar product of Eq. (3), we define the *shape correlator* (or cosine) by

$$S(B_i, B_j) \equiv \frac{\langle B_i, B_j \rangle}{\sqrt{\langle B_i, B_i \rangle \langle B_j, B_j \rangle}}, \quad (5)$$

which is restricted to $-1 \leq S \leq 1$. In the following, we will typically calculate the shape correlators between theoretical and simulated bispectra, to which the shapes i, j will correspond respectively.

In order to measure how well the magnitude of the theoretical bispectra i fit the (simulated) data j , we define the *amplitude correlator* as

$$\mathcal{A}(B_i, B_j) \equiv \sqrt{\frac{\langle B_i, B_i \rangle}{\langle B_j, B_j \rangle}}. \quad (6)$$

We can thus introduce a single quantity that combines the shape and amplitude information, the *total correlator*, defined as

$$\begin{aligned} \mathcal{T}(B_i, B_j) &\equiv 1 - \sqrt{\frac{\langle B_j - B_i, B_j - B_i \rangle}{\langle B_j, B_j \rangle}} \\ &= 1 - \sqrt{1 - 2S(B_i, B_j)\mathcal{A}(B_i, B_j) + \mathcal{A}^2(B_i, B_j)}. \end{aligned} \quad (7)$$

This total correlator offers an excellent means by which to determine the overall goodness of fit as we essentially measure the magnitude of the residual $B_i - B_j$ relative to the measured bispectrum $|B_j|$. If $B_i = B_j$, this is zero and the total correlator is $\mathcal{T} = 1$. If B_i and B_j are misaligned ($\mathcal{S} < 1$) or differ in amplitude ($\mathcal{A} \neq 1$), the residual $B_i - B_j$ is nonzero and the total correlator $\mathcal{T} < 1$. For increasing relative bispectrum residual, the total correlator always decreases. [Note that this is a more stringent test than the shape correlator of Eq. (5) alone because \mathcal{S} appears under a square root in Eq. (7)].

It is possible to relate the total correlator \mathcal{T} to the χ^2 goodness of fit determined between the theoretical bispectrum B_i and the estimated (or simulated) bispectrum B_j , as [125]

$$\begin{aligned} \chi^2 &= \sum_{k_1, k_2, k_3} \frac{[B_j(k_1, k_2, k_3) - B_i(k_1, k_2, k_3)]^2}{\text{var}(B_i)} \\ &= \langle B_j - B_i, B_j - B_i \rangle, \end{aligned} \quad (8)$$

so that χ^2 and the total correlator \mathcal{T} are simply linked by

$$\chi^2 = [1 - \mathcal{T}(B_i, B_j)]^2 \langle B_j, B_j \rangle. \quad (9)$$

As we are using a small number of simulations of limited resolution, in the following we will consider the total correlator \mathcal{T} together with its uncertainty as a measurement of the goodness of fit of each model. In principle, the use of $\chi^2/\text{d.o.f.}$ may be more suitable than \mathcal{T} to distinguish overfitting ($\chi^2/\text{d.o.f.} < 1$) from poor model performance ($\chi^2/\text{d.o.f.} > 1$). However, our focus here is to determine the k_{max} at which the model starts to become a poor description of our present simulations, which corresponds to the k_{max} where \mathcal{T} becomes significantly smaller than unity (given the estimated errors between simulations).

The three correlators here, \mathcal{S} , \mathcal{A} and \mathcal{T} , are all cumulative functions of k_{max} , which is the resolution cutoff used in the scalar product of Eq. (3). We therefore obtain an overall integrated measure of how well a particular theory matches simulations (or observations) up to k_{max} .

B. Three canonical shape functions

As we shall see in subsequent sections of this paper, we are able to obtain an accurate global description of the nonlinear gravitational bispectrum from a sum over a limited number of simple bispectrum shapes, provided that we have the flexibility to modify an overall scale-dependent amplitude. For this reason, we consider the following nontrivial bispectrum *Ansatz*:

$$B(k_1, k_2, k_3) = f(K)S(k_1, k_2, k_3), \quad (10)$$

where $K \equiv k_1 + k_2 + k_3$, and the ‘‘shape function’’ S is taken, in turn, to be a separable function of the form

$$S(k_1, k_2, k_3) = A(k_1)B(k_2)C(k_3) + \text{perms.} \quad (11)$$

This separation between transverse $K = \text{const}$ slices and the K -dependent diagonal is illustrated in Fig. 2.

The separable *Ansatz* [Eq. (10)] is motivated in part by comparison with primordial non-Gaussian models, for which we define the shape function S by taking out an overall scaling $(k_1 k_2 k_3)^{-2}$ after which S is (almost) scale invariant, that is, independent of the summed wave number K along the tetrapyd diagonal. For this reason, most primordial bispectra depend only on the two degrees of freedom transverse to the diagonal and can be completely defined by the shape S on the triangular surface $K = \text{const}$.

At late times, this simple separation of variables [Eq. (10)] may not apply accurately because of the scale-dependent transfer functions, which means that perturbations with different wave numbers k_i receive different amplifications. Nevertheless, this is encoded in the turnover of the late-time linear matter power spectrum $P_{\text{lin}}(k)$, which can still be used to create a separable (though scale-dependent) shape function, e.g. as we will see for the tree-level gravitational bispectrum. For this reason, the separable description [Eq. (10)] can still prove very useful if physically well-motivated shapes $S(k_1, k_2, k_3)$ are chosen and an overall scaling dependence $f(K)$ is allowed.

The three basic separable bispectrum shape functions $S(k_1, k_2, k_3)$ we shall employ are the constant shape [126], the squeezed (or local) shape [127–129] and the tree-level (or flattened) shape from standard perturbation theory discussed earlier. These three functions are essentially weighting functions for specific triangular configurations; that is, constant treats all triangles equally across the tetrapyd, squeezed favors those along the edges, and flattened those near the faces, as illustrated in Fig. 2 (qualitatively encompassing the commonly discussed equilateral, local and orthogonal shapes respectively). The constant shape is simply given by

$$S^{\text{const}}(k_1, k_2, k_3) = 1 \text{ (Mpc/h)}^6. \quad (12)$$

Physically, the constant bispectrum is produced by a random set of point sources, together with an appropriate scaling dependence $f(K)$. It is our first approximation to the bispectrum of the nonlinear virialized end products of gravitational collapse assumed in halo models, with a SN-weighted version illustrated in Fig. 3(a).

The second shape is squeezed and we shall define it as

$$\begin{aligned} S^{\text{squeez}}(k_1, k_2, k_3) &= \frac{1}{3} [P_{\text{lin}}(k_1)P_{\text{lin}}(k_2) + P_{\text{lin}}(k_2)P_{\text{lin}}(k_3) \\ &\quad + P_{\text{lin}}(k_3)P_{\text{lin}}(k_1)], \end{aligned} \quad (13)$$

which incorporates the scale dependence of the transfer functions within the linear power spectrum $P_{\text{lin}}(k)$. It is illustrated in Fig. 3(b). This squeezed shape is motivated by

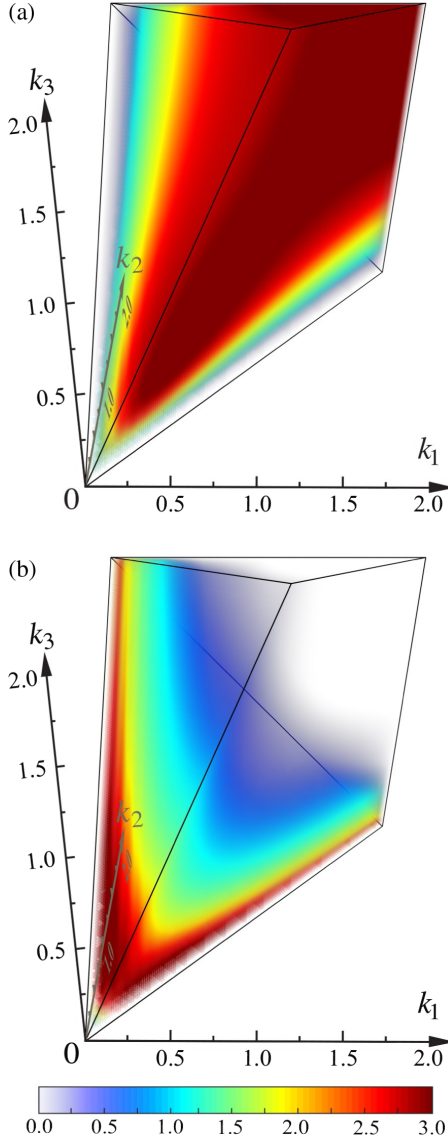


FIG. 3. (a) The SN-weighted constant bispectrum of Eq. (12) with a broadly equilateral signal shown together with (b) the squeezed or local model [Eq. (13)] with high signal at the edges near $k_i \approx 0$ (shown at redshift $z = 0$). Note that the plotted constant bispectrum does not have a constant cross-sectional shape because of the nonuniform signal-to-noise weighting [Eq. (4)] particularly near the edges; here S^{const} in Eq. (12) is multiplied by $f(K) = K^3$ (the color scale is normalized).

“local” non-Gaussianity in which perturbation fields are simply squared, and where the leading contribution has a large wavelength mode affecting nonlinearity on small scales (i.e., for squeezed triangles with $k_1 \ll k_2, k_3$). However, Eq. (13) regularizes the related scale-invariant primordial local shape,

$$S^{\text{local}}(k_1, k_2, k_3) = \frac{1}{3} \left(\frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_3 k_1} + \frac{k_3^2}{k_1 k_2} \right), \quad (14)$$

which behaves poorly because it diverges for very squeezed triangles.

Finally, the third flattened shape is the tree-level gravitational bispectrum given by [53]

$$S^{\text{tree}}(k_1, k_2, k_3) = 2P_{\text{lin}}(k_1)P_{\text{lin}}(k_2)F_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) + 2 \text{ perms.}, \quad (15)$$

where the kernel $F_2^{(s)}$ can be expressed as

$$F_2^{(s)}(\mathbf{q}_1, \mathbf{q}_2) = \frac{5}{7} + \frac{1}{2} \frac{\mathbf{q}_1 \cdot \mathbf{q}_2}{q_1 q_2} \left(\frac{q_1}{q_2} + \frac{q_2}{q_1} \right) + \frac{2}{7} \frac{(\mathbf{q}_1 \cdot \mathbf{q}_2)^2}{q_1^2 q_2^2}, \quad (16)$$

which, although not immediately apparent, is also a separable shape of the form of Eq. (11). Equation (16) represents the leading-order gravitational non-Gaussianity generated by nonlinear terms in the equations of motion. As we shall see, the scaling dependence $f(K)$ in Eq. (10) allows us to approximately incorporate higher-order perturbative corrections. However, the actual gravitational bispectrum is more closely approximated if the tree-level shape [Eq. (15)] is modified by employing the nonlinear power spectrum [130], given by the HALOFIT method [106,131]:

$$S^{\text{treeNL}}(k_1, k_2, k_3) = 2P_{\text{NL}}(k_1)P_{\text{NL}}(k_2)F_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) + 2 \text{ perms.} \quad (17)$$

For this reason, we will generally employ this improved flattened bispectrum as the third shape in our subsequent modeling. Both the tree-level [Eq. (15)] and the nonlinear tree-level [Eq. (17)] shapes are illustrated in Fig. 4.

C. Scale-dependent or “sliced” correlators

Having given the key shapes S^i that we will use to describe gravitational non-Gaussianity using the separable *Ansatz* Eq. (10), we must also define a scale-dependent correlator that can be used to test the accuracy of this approximation. To determine this we need a more “localized,” binned (or “sliced”) correlator, which only integrates over the transverse degrees of freedom on the $K = \text{constant}$ surfaces, modifying Eq. (3) to have the restricted domain of integration,

$$\langle B_i, B_j \rangle_K^S \equiv \frac{V}{\pi} \int_{\Delta V_B} dV_k \frac{k_1 k_2 k_3 B_i(k_1, k_2, k_3) B_j(k_1, k_2, k_3)}{P_\delta(k_1) P_\delta(k_2) P_\delta(k_3)}, \quad (18)$$

such that the integral is now evaluated in a specific thin slice of the tetrahedron with

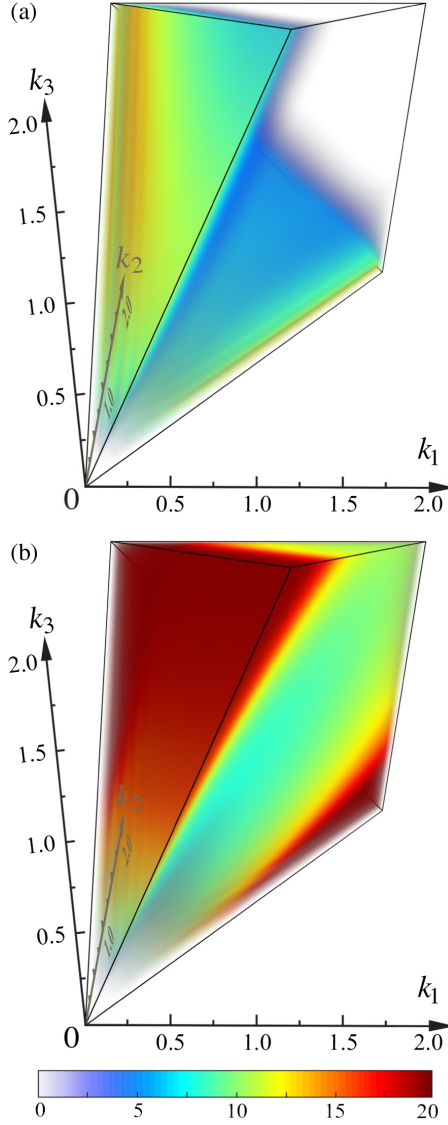


FIG. 4. Flattened shapes: (a) the SN-weighted tree-level bispectrum of Eq. (15) compared with (b) the nonlinear tree-level model [Eq. (17)], both shown at redshift $z = 2$. Note that this flattened shape is dominated by signal on the outer tetrapyd face (front left) where $k_1 + k_2 \approx k_3$ (see Fig. 2 for the geometry). The nonlinear tree-level amplitude is substantially higher than the tree level, but they share an excellent binned shape correlation [Eq. (19)], which always remains above 99%.

$$K < k_1 + k_2 + k_3 < K + \Delta K,$$

and where the index S denotes slice. Substituting the localized inner product definitions in the correlators [Eqs. (5)–(7)], this allows us to define the *sliced* correlators \mathcal{S}^S , \mathcal{A}^S and \mathcal{T}^S ; for example, the binned shape correlator becomes

$$\mathcal{S}^S(K) \equiv \frac{\langle B_i, B_j \rangle_K^S}{\sqrt{\langle B_i, B_i \rangle_K^S \langle B_j, B_j \rangle_K^S}}. \quad (19)$$

Importantly, if we find a good binned shape correlation $\mathcal{S}^S(K) \approx 1$ between our target model (or simulation) and the canonical shapes above [Eqs. (12), (13), (15)], then we can use the binned amplitude correlation \mathcal{T}^S to determine the overall scale dependence $f(K)$ in our separable *Ansatz* of Eq. (10). Later in Sec. VI we will combine these in a “three-shape benchmark” model and establish that it can achieve an excellent fit to simulations, thus dramatically reducing the number of degrees of freedom required to accurately describe the matter bispectrum.

III. PERTURBATION THEORY FOR LARGE-SCALE STRUCTURE

The amplitude of the matter density fluctuations in the Universe δ is small at early times and on large scales, so that $\delta \ll 1$ and linear dynamics suffices for an accurate modeling in this regime. At later times and on smaller scales, perturbations grow under gravity, making linear theory increasingly inaccurate. Various methods exist that can extend the range of validity of the model, accurately describing the large-scale structure to smaller scales and later times than linear theory.

A first possible approach, which we consider in this section, is to extend linear theory perturbatively, by expanding the evolution equations to higher order. This leads directly to standard (Eulerian) perturbation theory (SPT), which we review in Sec. III A; we next summarize more recent developments, which extend the range of validity by improving the SPT expansion convergence, removing divergences, and adding counterterms. The methods we consider are effective field theory (EFT) in Sec. III B, renormalized (Eulerian) perturbation theory (RPT) in Sec. III C, and resummed Lagrangian perturbation theory (RLPT) in Sec. III D. For each method, we provide a more complete review in the Appendixes A, B, C, and D respectively. We discuss possible extensions to two loops in Sec. III E and Appendix E. We finally discuss the shapes of the perturbation theory bispectra in Sec. III F.

A. Standard perturbation theory

Eulerian standard perturbation theory is derived by expanding the evolution equations for the dark matter density and velocity fields as a series of the linearly evolved density field δ_1 . In analogy with field theory, the resulting expansion for the power spectrum and bispectrum can be grouped to loop orders according to the number of δ_1 ’s involved. We present here in the following the expressions for the SPT matter power spectrum and bispectrum, whose derivation is summarized in Appendix A; see also Ref. [56] for a comprehensive review.

The tree-level (zero-loop) power spectrum is simply given by the linear power spectrum:

$$P_{\text{tree}}^{\text{SPT}}(k, z) = P_{11}(k, z) = D^2(z)P_{\text{lin}}(k), \quad (20)$$

where $D(z)$ is the linear growth function normalized to 1 today. This can be evaluated numerically by evolving the primordial fluctuations through the Boltzmann equations through codes such as CAMB [132]. The one-loop contribution can be obtained from two diagrams and has the following form [55]:

$$P_{1\text{-loop}}^{\text{SPT}}(k, z) = P_{13}(k, z) + P_{22}(k, z), \quad (21)$$

where the two contributions have the following expressions:

$$P_{13}(k, z) = D^4(z) \int \frac{d^3q}{(2\pi)^3} 6P_{\text{lin}}(k)P_{\text{lin}}(q)F_3^{(s)}(\mathbf{k}, \mathbf{q}, -\mathbf{q}) \quad (22)$$

$$P_{22}(k, z) = D^4(z) \int \frac{d^3q}{(2\pi)^3} 2P_{\text{lin}}(q)P_{\text{lin}}(|\mathbf{k} - \mathbf{q}|) \times [F_2^{(s)}(\mathbf{q}, \mathbf{k} - \mathbf{q})]^2, \quad (23)$$

and where the kernels $F_n^{(s)}$ are defined in Appendix A.

The tree-level bispectrum has the following expression [53]:

$$B_{\text{tree}}^{\text{SPT}}(k_1, k_2, k_3, z) = 2D^4(z)P_{\text{lin}}(k_1)P_{\text{lin}}(k_2)F_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) + 2 \text{ perms.} \quad (24)$$

In order to improve the accuracy of the tree-level bispectrum, Ref. [130] proposed simply replacing the linear power spectrum in the tree-level formula with the nonlinear power spectrum estimated e.g. with the HALOFIT method [106,131]. This heuristically extends the range of validity of the model, and is what we call ‘‘nonlinear tree level’’ bispectrum [see Eq. (17)]. Reference [130] showed that a further improvement can be achieved if, in addition to using the nonlinear power spectrum, the F_2 kernel is replaced with a modified version F_2^{eff} , which includes six free parameters that are fit to N -body simulations. Later, this method was extended by Ref. [133] by adding three extra parameters, as described in Appendix F, and recalibrated on more precise N -body data over an extended range. This is what we indicate as the ‘‘nine-parameter fit’’ model in the discussion below.

For the one-loop bispectrum, there are four diagrams that can be drawn [134]:

$$B_{1\text{-loop}}^{\text{SPT}} = B_{222} + B_{321}^{(I)} + B_{321}^{(II)} + B_{411}. \quad (25)$$

These have the following expressions:

$$B_{222}(k_1, k_2, k_3, z) = 8D^6(z) \int_{\mathbf{q}} P_{\text{lin}}(q)P_{\text{lin}}(|\mathbf{k}_2 - \mathbf{q}|) \times P_{\text{lin}}(|\mathbf{k}_3 + \mathbf{q}|)F_2^{(s)}(-\mathbf{q}, \mathbf{k}_3 + \mathbf{q}) \times F_2^{(s)}(\mathbf{k}_3 + \mathbf{q}, \mathbf{k}_2 - \mathbf{q})F_2^{(s)}(\mathbf{k}_2 - \mathbf{q}, \mathbf{q}) \quad (26)$$

$$B_{321}^{(I)}(k_1, k_2, k_3, z) = 6D^6(z)P_{\text{lin}}(k_3) \int_{\mathbf{q}} P_{\text{lin}}(|\mathbf{k}_2 - \mathbf{q}|) \times P_{\text{lin}}(q)F_3^{(s)}(-\mathbf{q}, -\mathbf{k}_2 + \mathbf{q}, -\mathbf{k}_3) \times F_2^{(s)}(\mathbf{k}_2 - \mathbf{q}, \mathbf{q}) + 5 \text{ perms.} \quad (27)$$

$$B_{321}^{(II)}(k_1, k_2, k_3, z) = 6D^6(z)P_{\text{lin}}(k_2)P_{\text{lin}}(k_3)F_2^{(s)}(\mathbf{k}_2, \mathbf{k}_3) \times \int_{\mathbf{q}} P_{\text{lin}}(q)F_3^{(s)}(\mathbf{k}_3, \mathbf{q}, -\mathbf{q}) + 5 \text{ perms.} \quad (28)$$

$$B_{411}(k_1, k_2, k_3, z) = 12D^6(z)P_{\text{lin}}(k_2)P_{\text{lin}}(k_3) \times \int_{\mathbf{q}} P_{\text{lin}}(q)F_4^{(s)}(\mathbf{q}, -\mathbf{q}, -\mathbf{k}_2, -\mathbf{k}_3) + 2 \text{ perms.,} \quad (29)$$

where $\int_{\mathbf{q}} \equiv \int \frac{d^3q}{(2\pi)^3}$. The numerical integration of the expressions above is nontrivial, and we discuss the necessary procedures in Appendix A.

It is known [56] that SPT only succeeds in extending the range of validity of linear theory by a small amount at low redshift, while it overpredicts the power seen in N -body simulations on smaller scales. This is because the SPT loop corrections are integrated over all k modes, including scales that are not in the linear regime, which are actually suppressed in reality compared to SPT [104]. Furthermore, the convergence of the SPT expansion is problematic, as it relies on the near cancellation of large positive and negative terms, so that increasing the loop order does not necessarily improve the accuracy of the expansion, especially at low redshift.

B. Effective field theory

Some of the problems of SPT mentioned in the previous section are addressed by the effective field theory of LSS (EFTofLSS, or simply EFT). At nonlinear level, the Fourier modes do not evolve independently any more, and hence small-scale fluctuations can influence much larger scales. The basic assumption of EFT is to introduce additional free parameters that describe the effect of nonperturbative small-scale physics onto the larger observable scales. The SPT expansion can only be expected to work when the density contrast is small, $\delta \ll 1$, so that its range of validity at low redshift becomes increasingly limited. Nevertheless, even when this condition is not satisfied,

the gravitational potential is still small and can be used to produce a valid perturbative expansion. Based on this fact, EFTofLSS has been developed in Refs. [13,88].

This method consists of adding to the equations of motion an effective stress-energy tensor $\tau_{\mu\nu}$, induced by short wavelength modes. This has the effect of adding corrections to the fluid equations, with terms corresponding to the speed of sound, viscosity and stochastic pressure. As we describe in more detail in Appendix B, the EFT method leads to additional contributions to the SPT matter power spectrum and bispectrum, with free parameters to be calibrated with N -body simulations.

At one loop, one term is added to the SPT matter power spectrum [135]:

$$P^{\text{EFT}}(k, z) = P^{\text{SPT}}(k, z) + P_{c_s}(k, z), \quad (30)$$

where

$$P_{c_s}(k, z) = -2(2\pi)c_{s(1)}^2 \frac{k^2}{k_{\text{NL}}^2} D^{2+\zeta}(z) P_{\text{lin}}(k). \quad (31)$$

Here the parameters $c_{s(1)}$ and ζ are fit to N -body simulations, and k_{NL} is defined as the scale where the perturbative *Ansatz* ($\delta \ll 1$) breaks down.

Likewise, one term is added to the one-loop SPT matter bispectrum:

$$B^{\text{EFT}}(k_1, k_2, k_3, z) = B^{\text{SPT}}(k_1, k_2, k_3, z) + B_{c_s}(k_1, k_2, k_3, z), \quad (32)$$

where

$$\begin{aligned} B_{c_s}(k_1, k_2, k_3, z) &= [2P_{\text{lin}}(k_1)P_{\text{lin}}(k_2)\tilde{F}_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) + 2 \text{ perms.}]D(z)^{4+\zeta} \\ &\quad - [2\bar{c}_1 k_1^2 P_{\text{lin}}(k_1)P_{\text{lin}}(k_2)F_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) \\ &\quad + 5 \text{ perms}]D(z)^{4+\zeta}. \end{aligned} \quad (33)$$

Here, $\bar{c}_1 = 2\pi \frac{c_{s(1)}^2}{k_{\text{NL}}}$, and the sound speed parameter $c_{s(1)}$ is fixed at the power spectrum level only, so that the bispectrum includes no extra free parameters. \tilde{F} is given by Eq. (B3).

The additional EFT terms effectively subtract the excess power that is present in the SPT results, so that an accurate modeling can be achieved over an extended range of scales.

C. Renormalized perturbation theory

The renormalized perturbation theory (RPT) model has been developed in Refs. [70–75]. This method uses the formalism of the SPT and reorganizes the infinite expansion differently using an idea from Ref. [69]. As described in more detail in Appendix C, this approach is based on the study of the nonlinear propagator connecting the initial

with the evolved fields describing density and velocity perturbations. In this way, the perturbative expansion can be written as a series of the nonlinear propagator. This infinite series can be resummed, yielding the RPT expressions for power spectrum and bispectrum at any number of loops. Compared to SPT, this method has the advantage that all the contributions involved are positive and the resummation of the propagator terms gives a well-defined perturbative expansion in the nonlinear regime. However, the expressions involved are complicated and the solutions are computationally demanding, the solution involving solving numerically a set of integro-differential equations. Moreover, more than one loop is required to obtain an accurate result, even on mildly nonlinear scales.

In order to solve these problems, Refs. [73,75] proposed a method that simplifies the calculation dramatically. The scheme is called $\text{MPT}_{\text{BREEZE}}$ and in this formalism only the late-time propagator is calculated and hence no time integrations are required. As described in Appendix C, the $\text{MPT}_{\text{BREEZE}}$ power spectrum contributions can be expressed in terms of their SPT counterparts as follows:

$$P_{\text{tree}}^{\text{MPT}_{\text{BREEZE}}}(k, z) = P_{11}(k, z) \exp[2f(k)D^2(z)] \quad (34)$$

$$P_{1\text{-loop}}^{\text{MPT}_{\text{BREEZE}}}(k, z) = P_{22}(k, z) \exp[2f(k)D^2(z)], \quad (35)$$

where the function $f(k)$ is given in Eq. (C16).

The bispectrum contributions can be treated in a similar manner [74], and the result up to one loop is given in terms of the SPT one-loop contributions [Eqs. (26), (27)]:

$$\begin{aligned} B^{\text{MPT}_{\text{BREEZE}}}(k_1, k_2, k_3, z) &= (B_{\text{tree}}^{\text{SPT}} + B_{222} + B_{321}')(k_1, k_2, k_3, z) \\ &\quad \times \exp[(f(k_1) + f(k_2) + f(k_3))D^2(z)]. \end{aligned} \quad (36)$$

The main advantages of RPT and its $\text{MPT}_{\text{BREEZE}}$ variant are that the expansion series becomes positive definite, so that no cancellation occurs and each successive term improves the range of validity of the theory; and the exponential prefactor term, which effectively suppresses the theory outside its range of validity, thus avoiding some of the SPT problems.

D. Resummed Lagrangian perturbation theory

Alternatively, perturbation theory can be derived as a function of the Lagrangian coordinates of the initial conditions. As the observable statistical quantities (power spectra and bispectra) are always defined in the evolved (Eulerian) coordinates, Lagrangian perturbation theory (LPT) has to deal with the evolution of the displacement field Ψ , which relates the two coordinate systems.

By expanding the evolved density and velocity perturbations as a series of Ψ , it is possible to calculate perturbative predictions for power spectrum and

bispectrum at any chosen order, although the calculations are complex [57–63]. A general drawback of LPT is that this method cannot describe accurately the physics of shell crossing, as particles continue to stream according to their initial velocity; thus dark matter halos never collapse, and LPT presents a power deficit on small scales.

More recently Ref. [80] used the cumulant expansion theorem to obtain a simpler resummed expression for the polyspectra, called resummed Lagrangian perturbation theory (RLPT). This method yields a resummed series expansion similar to, but simpler than, RPT.

We summarize the LPT and RLPT methods in Appendix D; the final results are the RLPT power spectrum [80]

$$P^{\text{RLPT}}(k) = \exp \left[-\frac{k^2}{6\pi^2} \int dp P_{\text{lin}}(p) \right] \times \left[P_{\text{lin}}(k) + P_{1\text{-loop}}^{\text{SPT}}(k) + \frac{k^2}{6\pi^2} P_{\text{lin}}(k) \int dp P_{\text{lin}}(p) \right], \quad (37)$$

where $P_{1\text{-loop}}^{\text{SPT}}$ is the one-loop SPT term (without the tree-level term); and the bispectrum [84]

$$B^{\text{RLPT}}(k_1, k_2, k_3) = \exp \left[-\frac{k_1^2 + k_2^2 + k_3^2}{12\pi^2} \int dp P_{\text{lin}}(p) \right] \times \left[B_{\text{tree}}^{\text{SPT}} + B_{1\text{-loop}}^{\text{SPT}} + \frac{k_1^2 + k_2^2 + k_3^2}{12\pi^2} B_{\text{tree}}^{\text{SPT}} \right] \times \int dp P_{\text{lin}}(p). \quad (38)$$

From Eqs. (37), (38) it is evident that the RLPT power spectrum and bispectrum reduce back to SPT if the exponential prefactor is expanded to first order. Furthermore, this prefactor is similar to the RPT results: in both cases, the theory decays rapidly to zero outside its range of validity. Thus this method is not expected to yield realistic predictions in the fully nonlinear regime where the exponential cutoff dominates, but only on quasilinear scales.

E. Going to two loops and estimating perturbation residuals

So far, most of the LSS perturbation theory work has considered up to two loops in the power spectrum [65,135] and up to one loop in the bispectrum. The reason has been mainly computational, but there is also a theoretical constraint: perturbation theories are expected only to work close to the linear regime, as they rely on perturbing around small density fluctuations. Even in the EFTofLSS approach, which significantly extends the range of validity of the nonlinear power spectrum over SPT [135], it is not

possible to push the model further to scales associated with dark matter halos: in the fully nonlinear regime only phenomenological halo models and fits to N -body simulations can be used. Nevertheless, it is interesting to investigate how far into intermediate nonlinear scales perturbation theory can be extrapolated.

Since the bispectrum is a three-dimensional quantity, its expansion at two loops requires the computation of challenging six-dimensional integrals. Moreover, the integrals involved have divergences that cancel between different terms, so these must be identified and eliminated before numerical computation to ensure convergence. Fortunately, in the MPTBREEZE formalism, the number of terms that appear is reduced; as described in Appendix E, we have therefore calculated analytically the terms involved, and we have eliminated the divergences based on the ideas developed in Refs. [136–138]. Unfortunately, due to the complexity of the integrals, we have not been able to perform the full three-dimensional bispectrum calculation as in the other one-loop cases. The analytic divergence-free expressions obtained are presented in Appendix E, while here we show in Fig. 5 three triangle configurations: equilateral, squeezed and flattened, also with a comparison between all the tree-level and one-loop perturbative methods at $z = 0$. The EFT bispectrum is expected to be accurate up to higher k than one-loop RPT, as discussed by Ref. [138] and as shown in Sec. VII below; therefore, knowing that the RPT approach is a convergent expansion with the precision increasing as the number of loops is increased, we can estimate the range of validity of the one-loop and two-loop MPTBREEZE results by comparison with EFT. In Fig. 5, we observe that the two-loop MPTBREEZE bispectrum closely follows the EFT prediction for an extra 0.04 h/Mpc more than the one-loop bispectrum in the equilateral and flattened cases, while the squeezed limit shows a more modest improvement. It is therefore clear that extending MPTBREEZE to two loops in the quantitative comparisons of Sec. VII would significantly improve its range of validity, but we decide not to pursue this for consistency with the other PT methods, and because of the huge analytic and numerical challenges which seem to be entailed.

Controlled perturbative expansions become increasingly accurate as the number of loops is increased, so a criterion for determining where perturbation theory at a given order breaks down is to calculate the next-order contribution and find where they become significant. In Table I we show the value of the wave number where the higher order expansion deviates by more than 10% (20%) from the lower order. Hence, we compare SPT with tree level, the MPTBREEZE at one loop to the tree level and the two-loop MPTBREEZE bispectrum to its one-loop counterpart. For completeness, we also determine the effect of the counterterm in EFT which corrects SPT. At one loop we evaluate deviations with the total correlator \mathcal{T} , but at two-loop order we

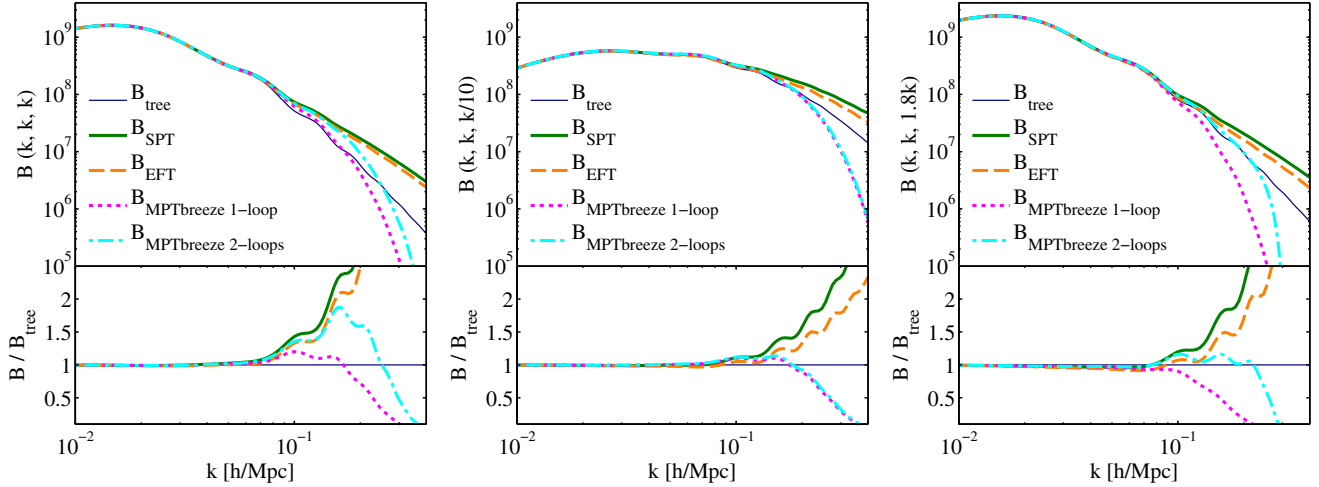


FIG. 5. Equilateral (left), squeezed (middle) and flattened bispectra (right) from perturbation theories at $z = 0$. We show the theoretical predictions of the tree-level bispectrum, SPT, EFT and the one- and two-loop MPTBREEZE bispectra. For this last model, we observe that the wave number at which the theory starts decaying increases significantly when adding the two-loop terms in the case of the equilateral and flattened configurations, closely following the EFT model down to smaller scales, while for the squeezed configuration the improvement is negligible. Bispectra are plotted in units of $(\text{Mpc}/h)^6$ throughout the paper.

determine the worst case amongst the three limiting configurations evaluated. Table I indicates that the tree-level bispectrum is in fact valid only for small wave numbers $k \lesssim 0.1 h/\text{Mpc}$ at $z = 0$ and $k \lesssim 0.2 h/\text{Mpc}$ at $z = 2$, with one-loop contributions apparently offering only a small incremental improvement. However, the comparison of SPT results with the EFT controlled expansion indicates that it may be possible to extrapolate perturbative expansions considerably further. As we shall see in Sec. VII, there is an unexpectedly good correspondence between some perturbative bispectra and the results of numerical simulations, going well beyond the thresholds estimated in Table I.

F. Shapes of perturbative bispectrum models

We conclude this section by describing the shapes of the various terms appearing in the different perturbative approaches using the binned shape correlator \mathcal{S}^S , defined on $K = \text{const}$ slices in Eq. (18). We determine \mathcal{S}^S for each perturbative model against the tree-level, squeezed and

constant shapes in Sec. II B. The results of this comparison are illustrated in Fig. 6. In the SPT and EFT bispectra, the tree-level term is always present, and so inevitably the flat tree-level shape dominates the large-scale results. For this reason, we restrict our attention to an analysis of the one-loop SPT terms and EFT counterterms separately, in order to achieve a better understanding of the underlying shape corrections. This also simplifies the figures, because in this way there is no mixture of different powers of the growth factor and it is sufficient to test the shapes of these terms at $z = 0$. The left panel of Fig. 6 shows the shape correlators in scale-invariant slices of $K \equiv k_1 + k_2 + k_3 = \text{constant}$ for the sum of the positive one-loop terms of SPT (thick lines) and the negative contributions (thin lines). The central panel represents the EFT counterterm for the tree level, $-B_{c_s}$, and the right panel shows the shapes of the MPTBREEZE bispectrum. Figure 6 shows strong correlations with the tree-level shape in the range $0.1 h/\text{Mpc} < k < 0.5 h/\text{Mpc}$ and beyond, with only the exponential cutoff in the MPTbreeze affecting the correlation.

Since the tree-level shape correlator is so dominant with respect to the others, we conclude that the perturbative approaches are indistinguishable in shape from the tree-level shape Eq. (15) in each scale-invariant K -bin. This is for the relevant range of scales probed by this analysis, with the possible exception of some small deviations appearing in the one-loop SPT terms at small k . Overall, Fig. 6 implies these one-loop correction terms are not adding any qualitatively new shape degrees of freedom; thus perturbative methods can be well approximated in terms of the tree-level shape using the separable *Ansatz*:

$$B_{\text{PT}}(k_1, k_2, k_3) = f(K) \mathcal{S}^{\text{tree}}(k_1, k_2, k_3), \quad (39)$$

TABLE I. Domain of validity for perturbation theory: wave number k_{max}^* where the two perturbative expansions are compared show relative deviations greater than 10% (20%).

| Threshold 10% (20%) | Perturbation theories | | |
|---------------------|----------------------------|-------------|-------------|
| | k_{max}^* (h/Mpc) | | |
| Theory | $z = 0$ | $z = 1$ | $z = 2$ |
| SPT/tree-level | 0.07 (0.08) | 0.08 (0.12) | 0.12 (0.14) |
| EFT/SPT | 0.12 (0.41) | 0.41 (0.93) | 0.77 (1.52) |
| RPT 1-loop/tree | 0.08 (0.10) | 0.09 (0.14) | 0.13 (0.20) |
| RPT 2-loops/1-loop | 0.09 (0.11) | 0.13 (0.16) | 0.19 (0.23) |

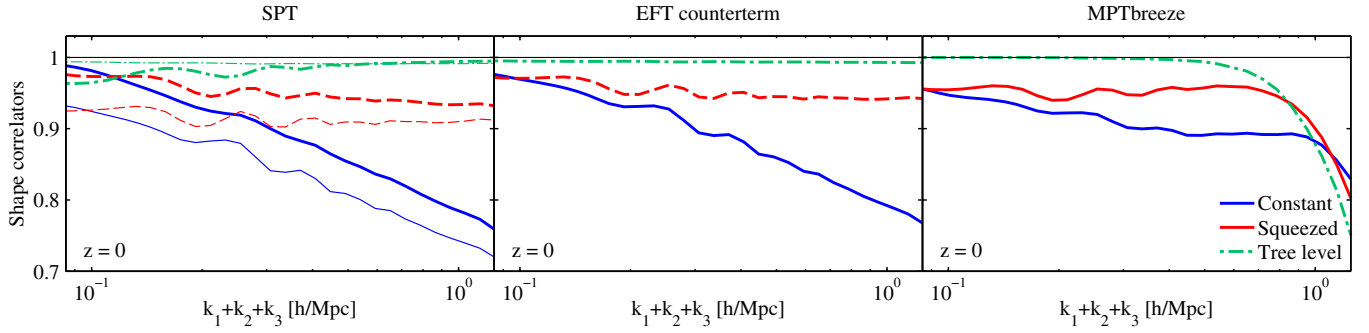


FIG. 6. Shapes of the perturbation theory bispectra. For each of the theoretical bispectra considered, we show the shape correlators in k slices \mathcal{S}^S [Eq. (18)] with respect to the constant, squeezed and tree-level shapes [Eqs. (12)–(15)]. The left panel shows the one-loop SPT shape correlators, the central panel shows the EFT counterterm ($-B_{c_s}$) shapes, and the right panel refers to the MPTBREEZE one-loop shapes. All panels refer to $z = 0$. In the case of SPT (left), the thick lines represent the sum of the positive terms of the one-loop expansion [B_{222} and $B_{321}^{(I)}$], while the thin lines refer to the sum of the negative terms [$B_{321}^{(II)}$ and B_{411}].

where $K = k_1 + k_2 + k_3$ and $f(K)$ is an appropriate scale-dependent function defined in Eq. (10). We will use this result in the construction of the phenomenological benchmark model in Sec. V below.

IV. NONPERTURBATIVE MODELS OF LARGE-SCALE STRUCTURE

A. Halo model basics

We next extend the clustering modeling deeper into the nonlinear regime using the halo model of the large-scale structure [96–98]. This framework is based on the assumption that all the matter in the Universe is concentrated into discrete regions called halos. As summarized in the review in Ref. [99], the matter power spectrum in this model is described by two contributions:

$$P(k, z) = P_{1h}(k, z) + P_{2h}(k, z), \quad (40)$$

where the one- and two-halo terms describe contributions from dark matter particle pairs that reside in the same or in different halos respectively, given by

$$P_{1h}(k, z) = \int_0^\infty dm n(m, z) \left(\frac{m}{\bar{\rho}}\right)^2 u^2(k|m, z), \quad (41)$$

$$\begin{aligned} P_{2h}(k, z) &= \int_0^\infty dm_1 n(m_1, z) \left(\frac{m_1}{\bar{\rho}}\right) u(k|m_1, z) \\ &\times \int_0^\infty dm_2 n(m_2, z) \left(\frac{m_2}{\bar{\rho}}\right) u(k|m_2, z) \\ &\times P_h(k|m_1, m_2, z). \end{aligned} \quad (42)$$

Here $\bar{\rho}$ is the mean density of the Universe today, and the one- and two-halo terms can be calculated once the following ingredients are specified: the halo mass function $n(m, z)$; the Fourier transform of the halo profile $u(k|m, z)$;

and the halo power spectrum $P_h(k|m_1, m_2, z)$, which we describe in Appendix G below.

Likewise, the matter bispectrum can be expressed as a sum of three terms:

$$\begin{aligned} B(k_1, k_2, k_3, z) &= B_{1h}(k_1, k_2, k_3, z) + B_{2h}(k_1, k_2, k_3, z) \\ &+ B_{3h}(k_1, k_2, k_3, z), \end{aligned} \quad (43)$$

where the one-, two-, and three-halo contributions refer to dark matter particle triplets residing in one, two, or three halos, given by

$$B_{1h}(k_1, k_2, k_3, z) = \int_0^\infty dm n(m, z) \left(\frac{m}{\bar{\rho}}\right)^3 \prod_{i=1}^3 u(k_i|m, z), \quad (44)$$

$$\begin{aligned} B_{2h}(k_1, k_2, k_3, z) &= \left[\int_0^\infty dm_1 n(m_1, z) \left(\frac{m_1}{\bar{\rho}}\right) u(k_1|m_1, z) \right. \\ &\times \int_0^\infty dm_2 n(m_2, z) \left(\frac{m_2}{\bar{\rho}}\right)^2 u(k_2|m_2, z) u(k_3|m_2, z) \\ &\times P_h(k_1|m_1, m_2, z) \left. \right] + 2 \text{ cyc.}, \end{aligned} \quad (45)$$

$$\begin{aligned} B_{3h}(k_1, k_2, k_3, z) &= \prod_{i=1}^3 \left[\int_0^\infty dm_i n(m_i, z) \left(\frac{m_i}{\bar{\rho}}\right) u(k_i|m_i, z) \right] \\ &\times B_h(k_1, k_2, k_3|m_1, m_2, m_3, z). \end{aligned} \quad (46)$$

Here B_h is the halo bispectrum, which we describe in Appendix G below.

B. Combined halo-PT model

As we show below in Sec. VII, the halo model provides a good description of N -body simulations in the fully nonlinear regime; however, some well-known shortcomings of this formalism are that [99,139,140]: (i) in the transition between linear and nonlinear scales, the halo model description is less accurate, and in the mildly nonlinear regime, perturbative methods are often more successful; (ii) in the linear limit, the nonlinear contributions P_{1h} , B_{1h} , B_{2h} do not vanish, leading to excess power with respect to linear theory for $k \rightarrow 0$; (iii) at higher redshift, as the fraction of matter in virialized structures decreases, the accuracy of the halo model degrades rapidly.

The issues (i) and (ii) are addressed by a combined formalism developed by Valageas and Nishimichi (VN) [100,101], which we briefly summarize here; we will call this model the “halo-PT model” in later sections.

1. Power spectrum

The one- and two-halo power spectrum terms can be combined with perturbation theory as follows [100]:

$$P_{1h}^{\text{VN}}(k) = \int_0^\infty dm n(m) \left(\frac{m}{\bar{\rho}}\right)^2 [u^2(k|m) - W_f^2(kq_m)], \quad (47)$$

$$P_{2h}^{\text{VN}}(k) = F_{2h}(1/k) P_{\text{PT}}(k), \quad (48)$$

where $q_m = R_f = [3m/(4\pi\bar{\rho})]^{1/3}$ is the Lagrangian radius of a halo of mass m , F_{2h} describes the probability that two particles at this Fourier space separation are in distinct halos, and $P_{\text{PT}}(k)$ is the nonlinear matter power spectrum in perturbation theory, e.g. SPT or EFT. With respect to the standard halo model presented in Sec. IV above, the one-halo term is modified by subtracting the filter function $W_f^2(kq_m)$, which ensures that the one-halo term vanishes in the limit $k \rightarrow 0$; the two-halo term is based on a perturbation theory of choice, corrected by the probabilistic prefactor F_{2h} given in Eq. (H3). The derivation of this model is summarized in Appendix H.

2. Bispectrum

Using a similar approach, Ref. [101] derived a combined model for the bispectrum. In analogy with the power spectrum case, the only term that should contribute to the bispectrum on very large scales is the three-halo term. Hence, that is the only perturbative contribution, while the one- and two-halo terms are nonperturbative.

The one-halo bispectrum term is

$$B_{1h}^{\text{VN}}(k_1, k_2, k_3) = \int_0^\infty dm n(m) \left(\frac{m}{\bar{\rho}}\right)^3 \times \prod_{i=1}^3 [u(k_i|m) - W_f(k_i q_m)]. \quad (49)$$

This function has the correct behavior on large scales, as its slope is at least $B_{1h}^{\text{VN}}(k_1, k_2, k_3) \propto k_j^2$ for any $k_j \rightarrow 0$.

The full result for the two-halo bispectrum is

$$B_{2h}^{\text{VN}}(k_1, k_2, k_3) = \int_0^\infty dm_1 n(m_1) \left(\frac{m_1}{\bar{\rho}}\right) [u(k_1|m_1) - W_f(k_1 q_{m_1})] \times \int_0^\infty dm_2 n(m_2) \left(\frac{m_2}{\bar{\rho}}\right)^2 [u(k_2|m_2) - W_f(k_2 q_{m_2})] \times [u(k_3|m_2) - W_f(k_3 q_{m_3})] P_{hh}(k_1|m_1, m_2) + 2 \text{ cyc.} \quad (50)$$

Here $P_{hh}(k_1|m_1, m_2) = b(m_1)b(m_2)P_{\text{lin}}(k_1)$. This result was however found to be unsatisfactory [101], because it scales as $B_{2h}^{\text{VN}} \sim k_1^2 P(k_1)$ for $k_1 \rightarrow 0$, while a scaling $\propto P(k_1)$ is expected; this implies that the approximations made in the derivation of this term are not accurate enough. Reference [101] therefore changes tack and argues for an alternative result that scales more appropriately in the large-scale limit, by replacing the halo with the matter power spectrum, and removing one prefactor:

$$B_{2h'}^{\text{VN}}(k_1, k_2, k_3) = \int_0^\infty dm_1 n(m_1) \left(\frac{m_1}{\bar{\rho}}\right) [u(k_1|m_1) - W_f(k_1 q_{m_1})] \times \int_0^\infty dm_2 n(m_2) \left(\frac{m_2}{\bar{\rho}}\right)^2 [u(k_2|m_2) - W_f(k_2 q_{m_2})] \times P_{\text{lin}}(k_1) + 2 \text{ cyc.} \quad (51)$$

The large-scale limit of this result is $B_{2h'}^{\text{VN}} \sim P(k_j)$ for $k_j \rightarrow 0$, as desired. Notice however that the rougher approximations assumed while deriving Eq. (51) make the accurate prediction of this term more uncertain.

Finally, the three-halo bispectrum is obtained with a perturbative approach. Similarly to the two-halo power spectrum, this contribution should match the tree-level bispectrum on very large scales. The probability that the three wave vectors belong to different halos can be approximated by

$$F_{3h}(k_1, k_2, k_3) = \int_0^{\nu_{k_1}} d\nu_1 \int_0^{\nu_{k_2}} d\nu_2 \times \int_0^{\nu_{k_3}} d\nu_3 f(\nu_1) f(\nu_2) f(\nu_3). \quad (52)$$

The three-halo bispectrum can then be written as

$$B_{3h}^{\text{VN}}(k_1, k_2, k_3) = F_{3h}(1/k_1, 1/k_2, 1/k_3) B_{\text{PT}}(k_1, k_2, k_3), \quad (53)$$

where $B_{\text{PT}}(k_1, k_2, k_3)$ is the matter bispectrum in the perturbative method of choice. In practice, the probabilistic

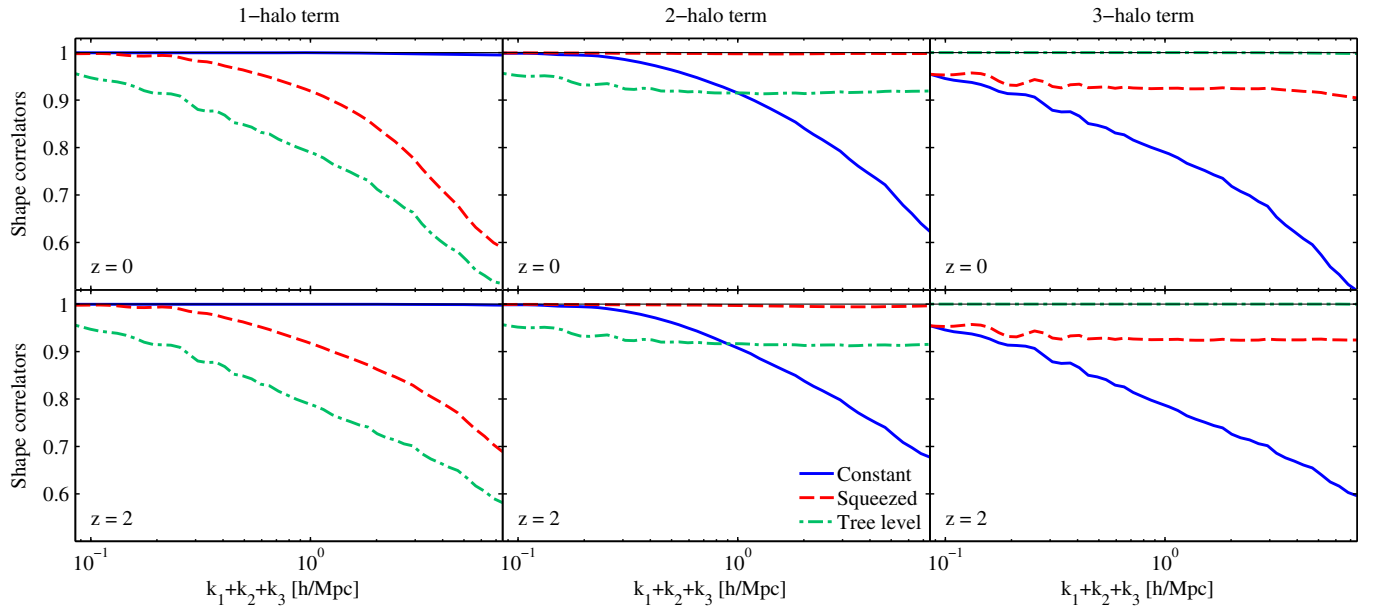


FIG. 7. Shapes of the halo model bispectrum. We show the correlation of the three components of the halo model with the constant, squeezed and tree-level shapes at redshifts $z = 0$ (upper panels) and $z = 2$ (lower panels). The left panels show that the one-halo term has a constant shape [Eq. (12)], the central panels demonstrate that the two-halo term is nearly fully correlated with the squeezed shape [Eq. (13)], and the right panels indicate that the three-halo term has the same shape as the tree-level bispectrum [Eq. (15)]. These results hold independent of scale and redshift.

prefactor appears to be neglected and set to unity for the bispectrum case [101].

C. Halo model shapes

By analogy with the shape investigation of perturbation theory bispectra we described in Sec. III F, we characterize here the shapes of the distinct halo model contributions, each of which has been evaluated numerically for a specific set of cosmological parameters (see Sec. V). In Fig. 7 we show the binned shape correlator results \mathcal{S}^S [Eq. (18)], by projecting the three halo model bispectrum components onto the canonical constant, squeezed and tree-level shapes [Eqs. (12)–(15)], defined on slices of $K = k_1 + k_2 + k_3 = \text{constant}$, for redshifts $z = \{0, 2\}$. The respective panels of Fig. 7 showing the one-, two-, and three-halo terms demonstrate that they are maximally correlated with the constant, squeezed, and tree-level shapes respectively, on all scales, and independently of redshift. This clear observation confirms the accuracy of the separable *Ansatz* [Eq. (10)] and the completeness of our canonical three shapes [Eqs. (12)–(15)] when characterizing the degrees of freedom needed to describe the standard halo bispectrum. This motivates us to find simple fitting functions $f_i(K)$ for each of the three halo model components.

1. One-halo term

Given the excellent shape correlation between the one-halo bispectrum [Eq. (44)] and the constant shape [Eq. (12)] that we observe in Fig. 7, we note that this term can be approximated by

$$B_{1h}(k_1, k_2, k_3) = f_{1h}(K) S^{\text{const}}(k_1, k_2, k_3), \quad (54)$$

where $K \equiv k_1 + k_2 + k_3$. Because of the constant cross-sectional form of Eq. (54), without loss of generality we can focus exclusively on the equilateral case to find a good fit. In Fig. 8 (top panel) we illustrate the equilateral one-halo bispectrum obtained from Eq. (44) at $z = \{0, 1, 2, 3\}$, compared with the following square-Lorentzian fitting function we introduce:

$$f_{1h}(K) = \frac{A}{[1 + bK^2]^2}, \quad (55)$$

where A and b are functions of redshift z through the perturbation growth factor $D(z)$. We first fit A , b for each redshift separately, and then we obtain two overall redshift-dependent fitting functions, taking account of the growth factors in the following form:

$$A = \frac{2.45 \times 10^6 D(z)^8}{0.8 + 0.2D(z)^{-3}} \quad (56)$$

$$b = 0.054 D(z)^{2.2} h^{-2} \text{ Mpc}^2. \quad (57)$$

We can see in Fig. 8 (top panel) for the equilateral case, and in Fig. 9 over the full 3D domain, that this is a good approximation of the full one-halo term.

While this phenomenological fit may not be particularly well-motivated physically, it does illustrate that once the one-halo shape has been identified, then a relatively simple

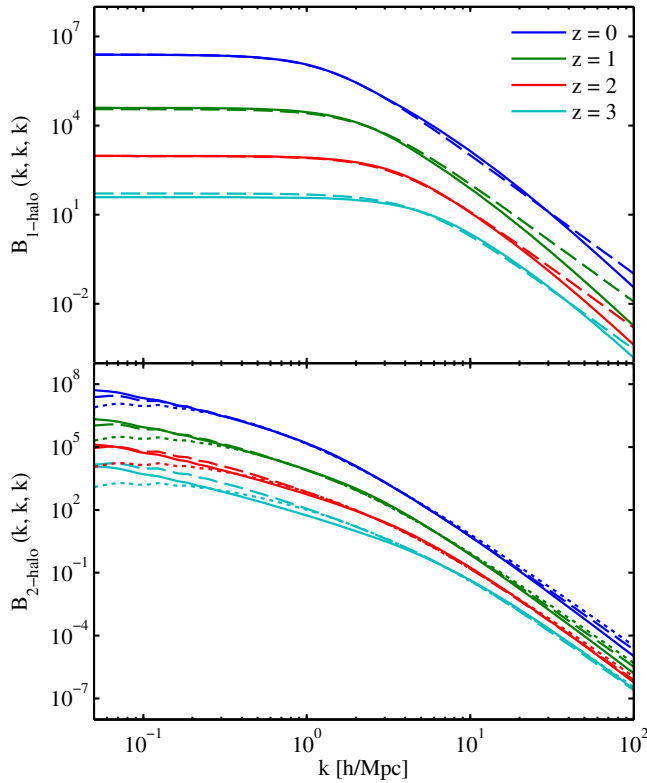


FIG. 8. Equilateral one-halo (top panel) and two-halo (bottom panel) bispectra at $z = \{0, 1, 2, 3\}$ (solid lines, from top to bottom), compared with the corresponding fitting function from Eqs. (55), (59) (dashed lines). The dotted lines refer to the corrected two-halo fitting function of Eq. (60).

combination of growth factors can be used to describe the scale-dependent amplitude for the relevant wave number range around $K \sim 1 h/\text{Mpc}$. Alternatively, it is sufficient to model the one-halo bispectrum directly by evaluating Eq. (44) for equilateral values only: $k_1 = k_2 = k_3$. More significantly, knowing empirically that *Ansätze* like Eq. (55) are accurate may offer insight which leads to a much simpler mathematical derivation of the individual halo contributions.

2. Two-halo term

As seen in Fig. 7, the two-halo bispectrum [Eq. (45)] is strongly correlated on all $K = \text{const}$ slices with the squeezed shape $S^{\text{squeez}}(k_1, k_2, k_3)$ constructed from products of the power spectrum defined in Eq. (13). This means that we can write

$$B_{2h}(k_1, k_2, k_3) = f_{2h}(K) S^{\text{squeez}}(k_1, k_2, k_3). \quad (58)$$

In order to obtain a phenomenological fit, we consider again the equilateral configuration, which we show in Fig. 8 (bottom panel). From this simple analysis, we find that a useful fitting function valid for the redshift range considered is

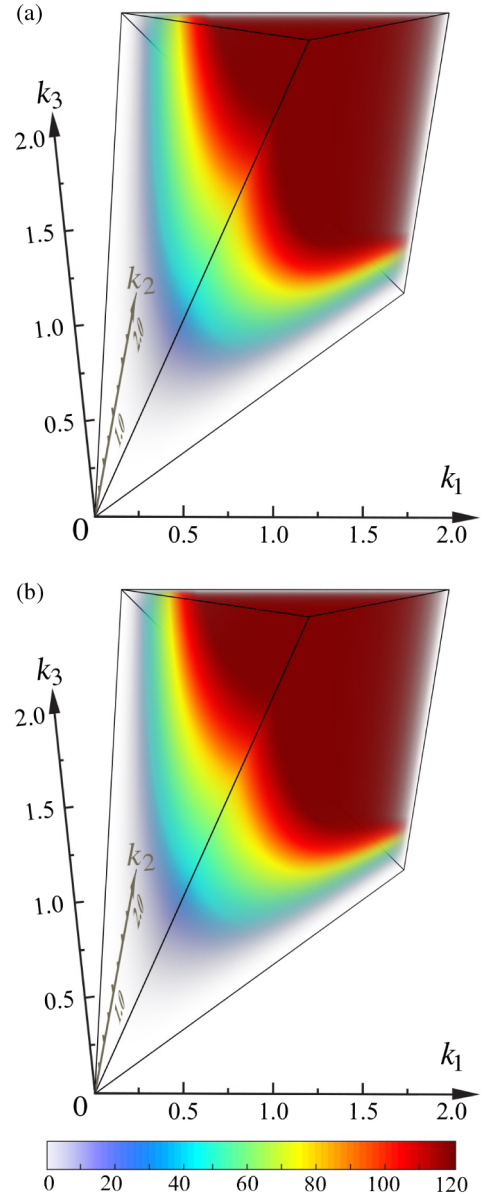


FIG. 9. The SN-weighted one-halo bispectrum of Eq. (44) (upper panel) compared at $z = 0$ with the one-halo constant shape *Ansatz* of Eq. (54) with scale dependence $f_{1h}(K)$ given by Eqs. (55), (57) (lower panel). This fit is visually hard to distinguish reflecting the high total correlation achieved over all length scales (and redshifts). The cross-sectional shape does not appear constant because of the SN weighting [Eq. (3)].

$$f_{2h}(K) = \frac{155}{1 + 26.2h^2 \text{Mpc}^{-2} D(z)^{-8/3} K^{-2}}, \quad (59)$$

where it should be noted that the squeezed shape form already includes a $D^4(z)$ redshift dependence from the linear power spectrum in Eq. (13).

However, as discussed above in Sec. IV B 2, the standard two-halo term causes some large-scale power excess in the full bispectrum, because it does not decay appropriately as

$k \rightarrow 0$; thus the full bispectrum does not recover the tree-level form on large scales. We can modify our fitting function in order to solve this issue, by considering the functional form

$$f_{2h}(K) = \frac{C}{(1 + DK^{-1})^3}. \quad (60)$$

This function is chosen to decay more rapidly on very large scales, as in that regime there should be no contribution from the two-halo term. By fitting the full two-halo term at different redshifts and considering the halo-PT VN model, we obtain

$$C = 240 \quad (61)$$

$$D = 2.5h \text{ Mpc}^{-1} D(z)^{-4/3}. \quad (62)$$

We can see in Fig. 8 (bottom panel) for the equilateral case, and in Fig. 10 over the full 3D domain, that this is a good approximation of the two-halo bispectrum term.

Nevertheless, despite this improvement at $z=0$ as $k \rightarrow 0$ we will show later that the two-halo model does not predict the appropriate growth rates at redshifts $z > 0$ when compared to simulations.

3. Three-halo term

The three-halo term [Eq. (46)] has a good shape correlation with the tree-level bispectrum [Eq. (15)], because it is essentially constructed out of this solution

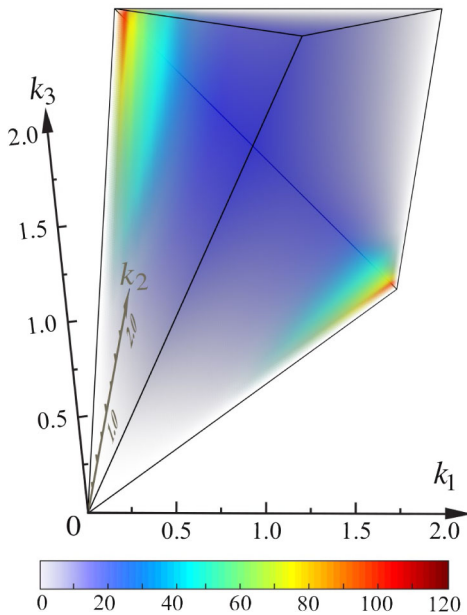


FIG. 10. The SN-weighted two-halo bispectrum of Eq. (45) at $z = 0$, showing the strongly squeezed signal. Like the one-halo bispectrum shown in Fig. 9, an excellent fit to this model can be obtained with the separable *Ansatz* of Eq. (58) using the standard squeezed shape [Eq. (13)].

or its one-loop extensions, all of which share the same highly correlated flattened shape (as discussed in Sec. III F). Hence, as we have discussed previously, the three-halo term can be expressed again with a simple fitting function [Eq. (39)] using the tree-level shape S^{tree} . The standard halo model effectively identifies the three-halo term with the tree-level bispectrum so we can take the fitting function to be unity. Extensions taking a perturbative result with one-loop corrections can also be described at high accuracy with Eq. (39) but with nontrivial scaling $f(K)$ (e.g. to simplify the halo-PT VN model which uses B^{EFT}). Since the closely related nonlinear tree-level bispectrum S^{treeNL} given in Eq. (17) provides a better approximation to the perturbative models, we can more conveniently use this as our base tree-level *Ansatz*:

$$B_{3h}(k_1, k_2, k_3) = f_{3h}(K) S^{\text{treeNL}}(k_1, k_2, k_3). \quad (63)$$

Both tree-level and nonlinear tree-level shapes are plotted in Fig. 4. We will employ Eq. (63) when developing the phenomenological three-shape model in Sec. V.

V. POLYSPECTRA FROM SIMULATIONS

A. N -body simulations

We use the N -body simulations with Gaussian initial conditions described in detail in Ref. [110]. The simulations contain 512^3 particles that are evolved from an initial redshift of $z = 49$ until today using the N -body GADGET-3 code [141,142] with second-order Lagrangian Perturbation Theory (2LPT) initial conditions [143,144]. These yield a less-than-2% accuracy in the bispectrum, as shown in Ref. [145]. The simulations are run using a flat Λ CDM universe with the following WMAP7 [146] parameters: baryon energy density $\Omega_b h^2 = 0.0226$; dark matter energy density $\Omega_c h^2 = 0.11$; cosmological constant energy density $\Omega_\Lambda = 0.734$; dimensionless Hubble constant $h = 0.71$; optical depth $\tau = 0.088$; amplitude of primordial perturbations $\Delta_{\mathcal{R}}^2(k_0) = 2.43 \times 10^{-9}$; and scalar spectral index $n_s(k_0) = 0.963$, where $k_0 = 0.002h \text{ Mpc}^{-1}$. We use simulations of three different box sizes of 1600, 400 and 100 Mpc/h respectively; the first one has glass Gaussian initial conditions and the other two have regular grid initial conditions. We denote the simulations using their names from Ref. [110]: $G512g$, G_{400}^{512} , G_{100}^{512} . Given the fixed number of particles, the three box sizes lead to the following wave number ranges: $[0.0039, 0.5] h/\text{Mpc}$, $[0.016, 2.0] h/\text{Mpc}$ and $[0.062, 8.0] h/\text{Mpc}$ respectively. For each box size, three independent realizations are available.

We combine the power spectra and bispectra from the different simulation boxes as follows. As it can be seen in Fig. 6 of Ref. [110], where the matter power spectra from the three simulations considered are compared to the HALOFIT model, at any redshift $z > 0$ the power spectrum of simulation $G512g$ only follows the HALOFIT model up to

$k_{\max} \approx 0.2 h/\text{Mpc}$; however, for $k \gtrsim 0.1 h/\text{Mpc}$, the simulation G_{400}^{512} matches HALOFIT more closely. The same behavior is seen at larger k for the G_{400}^{512} and G_{100}^{512} boxes. Therefore, we combine the power spectra and bispectra from the simulations in order to use each simulation in the range of scales where its results are the closest to HALOFIT, and we apply a smooth transition between the different boxes. We define a smoothing function $H(k)$ in the range $k \in [k_s, k_e]$ of the form

$$H(k) = \frac{1 - \sin\left(\pi \frac{k - k_e/2 - k_s/2}{k_e - k_s}\right)}{2}. \quad (64)$$

As we have three realizations for each of the simulations, we match each realization $i = 1, 2, 3$ from each simulation with the same i realization in the other simulations, thus obtaining three combined realizations of the power spectra and bispectra over the full k range we consider. We have checked that modifying the smoothing function has only a small impact on the overall results. This procedure allows us to use an overall large simulation data set covering the entire region of interest in wave vector space with three realizations. However, larger errors appear in the interior of the domain where the transition between the simulations occurs.

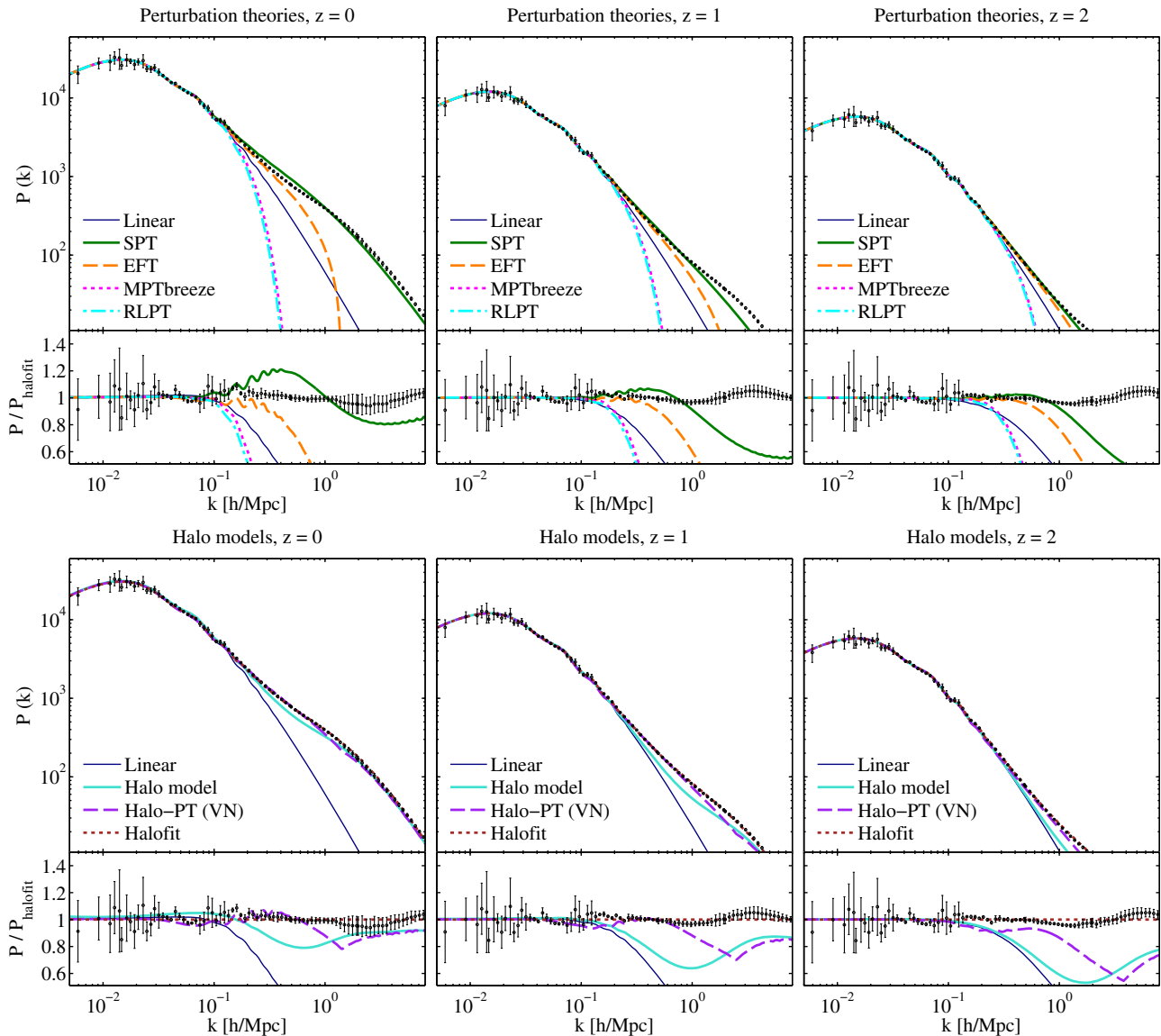


FIG. 11. Overview of the matter power spectra predicted by the range of theoretical models we consider, compared with data measured from N -body simulations. The data points are combined from N -body simulations with three different box sizes. The upper and lower rows refer to perturbation theories and halo models respectively. The columns refer to $z = 0, 1$ and 2 from left to right; in each plot, the main upper panel shows the power spectra comparison, while the smaller lower panel shows the residuals with respect to the HALOFIT prediction.

B. Power spectrum

We estimate the power spectrum of the simulations in each k -bin by averaging the squared absolute value of the matter overdensity $|\delta_{\mathbf{q}}|^2$ over all modes that fall into the shell with distance k from the origin (i.e. over \mathbf{q} with $|\mathbf{q}| - \Delta k/2 \leq k < |\mathbf{q}| + \Delta k/2$, where Δk is the bin width). We compare in Fig. 11 the power spectrum measured from the simulations with the models that we consider: linear theory; the nonlinear power spectrum from HALOFIT, EFT, MPTBREEZE and RLPT at one loop; the halo model; and the combined halo-PT model (based on EFT). The lower panels of Fig. 11 show the power spectrum residuals with respect to the HALOFIT model.

Focusing first on the perturbative methods, we note that they increase their range of validity to higher k modes as the redshift is increased, as expected. We confirm that SPT presents excess power in the quasilinear regime, departing from the simulations by more than 10% at $k \approx 0.15$ h/Mpc at $z = 0$. The SPT excess power is however reduced at higher redshifts, as expected given that the one-loop corrections have a higher growth rate compared to the tree level. The EFT method can extend the range of validity by subtracting the SPT excess power. However, the scale range over which EFT is accurate strongly depends on which simulations were used to calibrate the counterterm, and over which range of scales and redshifts. In the present case, the c_s^2 counterterm we are using was calibrated by Ref. [138] with the $G512g$ simulation box we are presenting at $z = 0$; therefore, there is no guarantee that this same counterterm will be accurate at higher k over the smaller-box simulations G_{400}^{512} , G_{100}^{512} , and at $z > 0$. Indeed, it is likely that a refitting of c_s^2 over the combined range of simulations we are using would improve the EFT model accuracy over an extended range of k and z . The MPTBREEZE and RLPT approaches include an exponential cutoff: this reduces the range in which the model is accurate to 10% to $k < 0.10$ h/Mpc at $z = 0$; nonetheless, these models feature an improved accuracy in the mildly nonlinear regime before the cutoff sets in, although the precision of our N -body simulations does not allow detailed quantitative statements at the percent level.

We then consider the halo models: we see that at $z = 0$ this formalism provides a good description of the matter power spectrum on small scales and in the range $k \in [0.01, 0.2]$ h/Mpc , after which we find the well-known power deficit in the transition region between the one- and two-halo terms. The model performs again better at smaller scales ($k \gtrsim 2$ h/Mpc at $z = 0$), reaching an accuracy of $\sim 10\%$. On very large scales, the halo model amplitude exceeds the simulations, as the one-halo term does not decay to zero as it physically should. By moving to higher redshifts, we see that at $z = 2$ the halo model provides a worse description of the simulations at intermediate and small scales, as the power deficit in the transition region is exacerbated. This is because the total fraction of dark

matter particles that belong to collapsed structures is drastically reduced at this redshift, which undermines the assumptions underlying the halo model approach. On large scales on the other hand, the excess power nearly disappears at high redshift, due to the quick decay of the one-halo term as a function of z .

The combined halo-PT model based on EFT succeeds in removing the excess power seen on large scales at $z = 0$; as we discuss below, this excess will appear even more evidently in the bispectrum. This model is also partly successful in reducing the power deficit on intermediate scales, thanks to the extra power that is added there from the perturbative term. However, due to the negative counterterm, the EFT power spectrum prediction becomes negative on small scales ($k \gtrsim 1$ h/Mpc at $z = 0$). After this point, we base the halo-PT model on the SPT prediction: this is the reason for the cusp we see in the halo-PT model residuals in the nonlinear regime.

We finally note that the results from the simulations are in good agreement with the nonlinear HALOFIT power spectrum, as they are within 10% accuracy over the entire k -range considered at all z .

C. Modal bispectrum methodology

We next follow the modal decomposition method to reconstruct the dark matter bispectrum, using the method developed by Refs. [112, 113, 147]. In this approach, the full 3D bispectrum $B(k_1, k_2, k_3)$ is expanded on an orthonormal basis defined on the same tetrapyd domain $Q_n(k_1, k_2, k_3)$, with $n = 0, \dots, n_{\max}$. In this way, the full bispectrum information is encoded in the expansion coefficients β_n^Q , and the bispectrum estimator \hat{B} can be written as

$$\frac{\hat{B}(k_1, k_2, k_3) \sqrt{k_1 k_2 k_3}}{\sqrt{P(k_1)P(k_2)P(k_3)}} = \sum_{n=0}^{n_{\max}-1} \beta_n^Q Q_n(k_1, k_2, k_3). \quad (65)$$

We note that the left-hand side is the signal-to-noise weighted bispectrum $B_i^{\text{SN}}(k_1, k_2, k_3)$ defined in Eq. (4). The accuracy of this estimator is regulated by the dimension of the expansion basis, n_{\max} ; for the smooth bispectra that are typical of the LSS, Ref. [110] demonstrated that the choice $n_{\max} \sim 100$ suffices to achieve a convergence of the total bispectrum signal-to-noise; i.e. considering higher n_{\max} has negligible effect on the matter bispectrum. This highlights the benefits of the modal method: once the basis Q_n is chosen, the entire three-dimensional bispectrum information can be simply compressed in a set of ~ 100 numbers.

Reference [113] tested several different choices of the basis Q_n , demonstrating that the modal method successfully reconstructs the bispectrum in all cases. The most suitable choice for Q_n is however built from a set of tetrahedral polynomials $q_p(x)$, which are analogues of the

Legendre polynomials on the unit interval. In more detail, the basis Q_n can be written as

$$Q_n(x, y, z) = q_{\{r\}}(x)q_{\{s\}}(y)q_{\{t\}}(z), \quad (66)$$

where $n = r + s + t$, $\{rst\}$ means symmetrization over the three indices, and the order of the permutations is taken as in Ref. [113]. In turn, the tetrahedral polynomials of order n , $q_n(x)$, can be generated by taking the determinant

$$q_n(x) = \frac{1}{\mathcal{N}} \begin{vmatrix} 1/2 & 7/24 & \cdots & w_n \\ 7/24 & 1/5 & \cdots & w_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n-1} & w_n & \cdots & w_{2n-1} \\ 1 & x & \cdots & x^n \end{vmatrix}, \quad (67)$$

where

$$w_n = \frac{n+6}{2(n+3)(n+2)}, \quad (68)$$

and the normalization \mathcal{N} is chosen so that the polynomials $q_n(x)$ are orthonormal with respect to the product:

$$\langle q_n, q_m \rangle = \int_0^1 q_n(x)q_m(x) \frac{1}{2}x(4-3x)dx = \delta_{nm}. \quad (69)$$

D. Bispectrum reconstruction from simulations

A modal reconstruction for the matter bispectrum $B_i^{\text{SN}}(k_1, k_2, k_3)$ [Eq. (65)] was obtained using the mode functions [Eq. (67)] for the full array of simulations described in Sec. VA. This decomposition and its validation were described in detail in Ref. [110]: a relatively small number of modes were sufficient to recover the full bispectrum at the required resolutions, that is, using 120 modes for the *G512g* simulations and 50 modes for the other two simulations. We focus attention here on the low-redshift regime $z < 3$ where the bispectrum is accessible to current and future galaxy surveys and where nonlinearities become important. To obtain the full bispectrum across the widest range of scales we combined and averaged all the simulation bispectra, interpolating in overlapping regions using the same prescription as that described for the power spectrum. Error bars for bispectrum correlators were estimated by determining variances from the different simulations.

In Fig. 12 we plot the full three-dimensional matter bispectrum we have obtained across the tetrapyd domain for $0.02 \text{ h/Mpc} \leq k \leq 2 \text{ h/Mpc}$ and at four different redshifts $z = \{0, 1, 2, 3\}$. The color scheme is scaled using the growth factor $D(z)$ such that the tree-level bispectrum would appear constant in the perturbative regime. These

plots range from quasilinear to highly nonlinear regions and several qualitative observations about the nature and evolution of the matter bispectrum are immediately apparent.

At the higher redshifts $z = 2, 3$ shown in Figs. 12(a)–(b), a flattened signal is dominant up to $K \equiv \sum_i k_i \lesssim 4, 3.5 \text{ h/Mpc}$ respectively (i.e. the tetrahedron region). This is consistent with the flattened tree-level shape [Eq. (15)] which is shown in Fig. 4(a) at $z = 2$, but at much lower amplitude on a more sensitive scale. This means the flattened signal extrapolates with growing amplitude well beyond the perturbative regime at these redshifts (e.g. from Table I $K \lesssim 0.6 \text{ h/Mpc}$ at $z = 2$). We focus further on the perturbative regime with $K \lesssim 1 \text{ h/Mpc}$ in Sec. VII A. For larger K , the bispectrum is dominated by a nearly uniform signal associated with halo formation (i.e. the top pyramidal region with $K \gtrsim 4 \text{ h/Mpc}$). Also in Figs. 12(a)–(b), we note that a significant squeezed signal is visible for $1 \text{ h/Mpc} \lesssim K \lesssim 4 \text{ h/Mpc}$ (on the left and bottom tetrapyd edges), which can be compared with Fig. 10.

At the lower redshifts $z = 0, 1$ in Figs. 12(c)–(d), the strong halo signal grows to become completely dominant for $K \gtrsim 1 \text{ h/Mpc}$ (saturating the color scheme with $B_{\text{max}}^{\text{SN}} \approx 350$). At $z = 0$, this “constant” halo signal is so large the other contributions seem to be absent (compare with Fig. 9). However, this apparent suppression of flattened and squeezed signals at $z = 0$ is only relative, due to the signal-to-noise weighting [Eq. (4)] with the nonlinear power spectrum $P_{\text{NL}}(k)$. This deeply nonlinear nature of perturbations today is reflected in the greater difficulty of matching phenomenological models to simulations at low redshift.

VI. TOWARDS A THREE-SHAPE BISPECTRUM BENCHMARK MODEL

In this section we analyze the measured bispectrum to identify the shape degrees of freedom required for its accurate construction. We study the growth rates of each of these contributions, highlighting differences with the standard halo model, particularly for the squeezed shape. We use these results to guide the development of simple phenomenological bispectrum models: the two-halo boost model and the three-shape benchmark model.

A. Simulation bispectrum shapes

We first analyze the shapes of the bispectra measured from N -body simulations, in analogy with the investigation of the perturbative and halo model shapes we presented in Figs. 6, 7 above. We calculate the sliced or binned shape correlators $\mathcal{S}^S(K)$ between the N -body matter bispectrum and the tree-level [Eq. (15)], squeezed [Eq. (13)], and constant [Eq. (12)] shapes to determine whether, in combination, these three canonical shapes are sufficient to

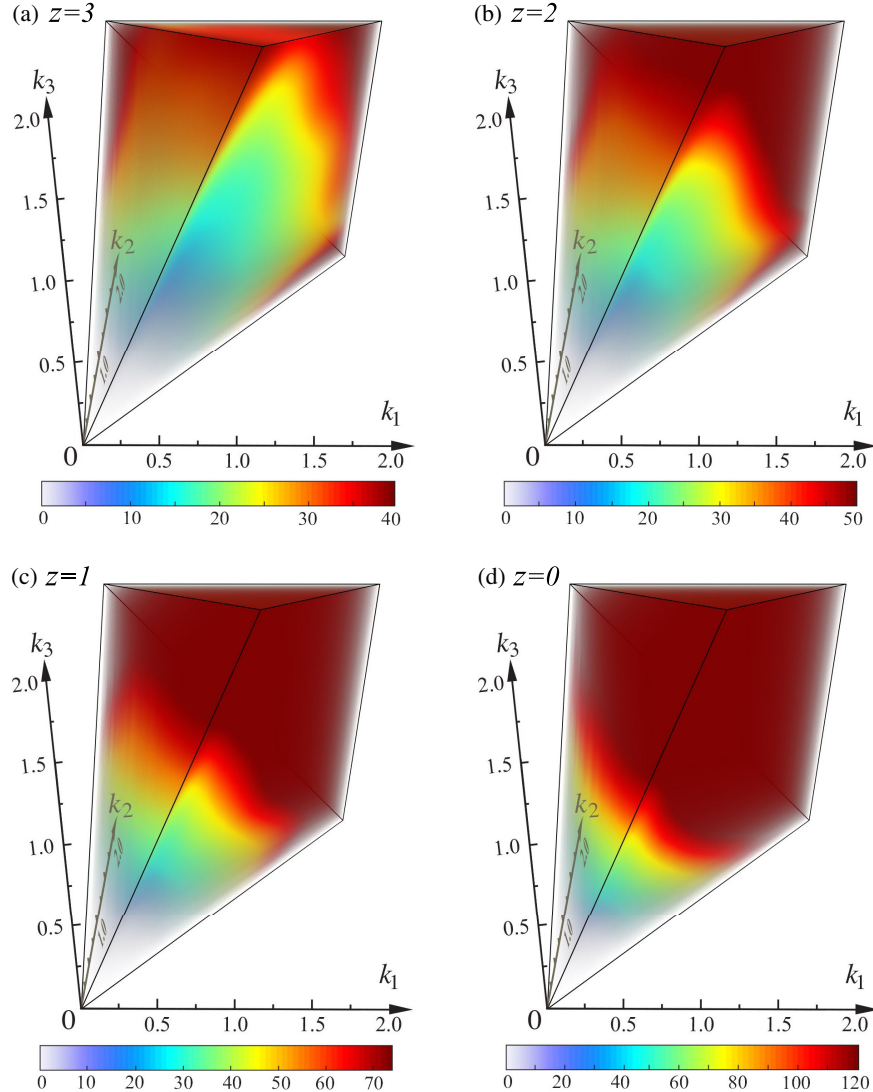


FIG. 12. Evolution of the SN-weighted 3D bispectrum from N -body simulations into the nonlinear regime with $k_i \leq 2 h/\text{Mpc}$ at redshifts (a) $z = 3$, (b) $z = 2$, (c) $z = 1$, and (d) $z = 0$. The bispectrum color scheme is scaled with the growth factor $D(z)$ and the tetrahedral geometry of the bispectrum domain is illustrated in Fig. 2. Note the presence of both a strong flattened and squeezed signal shape at redshifts $z = 2, 3$ [front left face of tetrapyd and lower edge respectively in panels (a), (b)]. At lower redshift this is overtaken by a strong uniform or one-halo signal throughout the interior region for $k \gtrsim 1 h/\text{Mpc}$ [front right face in panel (d)]. The color scale is fixed at $z = 3$ in (a) to encompass all values up to the maximum. It is then scaled with the growth rate expected for the tree-level signal to aid physical interpretation and reveal nonlinear growth rates. This means at small scales in (d) at $z = 0$ the color scale is saturated, which is useful to highlight features at intermediate scales.

describe the actual bispectrum. The panels of Fig. 13 show a consistent behavior across the range of redshifts considered. We know that, on large scales, perturbations approach linearity and therefore the tree-level bispectrum is expected to be a good approximation to the N -body data. The plots show that this is indeed the case, as on these scales ($K \lesssim 0.5 h/\text{Mpc}$ at $z = 0$) there is a high correlation between the simulated bispectrum and the tree-level shape. The scales up to which the bispectrum is completely dominated by the tree-level shape move significantly to larger values of K as the redshift increases, as expected. On small scales, Fig. 13 shows that deep into the nonlinear

regime ($K \gtrsim 3 h/\text{Mpc}$ at $z = 0$) the constant shape dominates, which closely corresponds to the one-halo model discussed in Sec. VII (and as shown previously in Ref. [110]). On intermediate scales, there are several competing contributions of comparable magnitude in the transition between constant and flattened regimes. Nevertheless, Fig. 13 reveals that at all redshifts there is a range of wave numbers where the squeezed shape exhibits the highest correlation, which is a new result. These quantitative shape correlation results confirm the qualitative picture developed from the evolution of the 3D bispectrum reconstructions shown in Fig. 12.

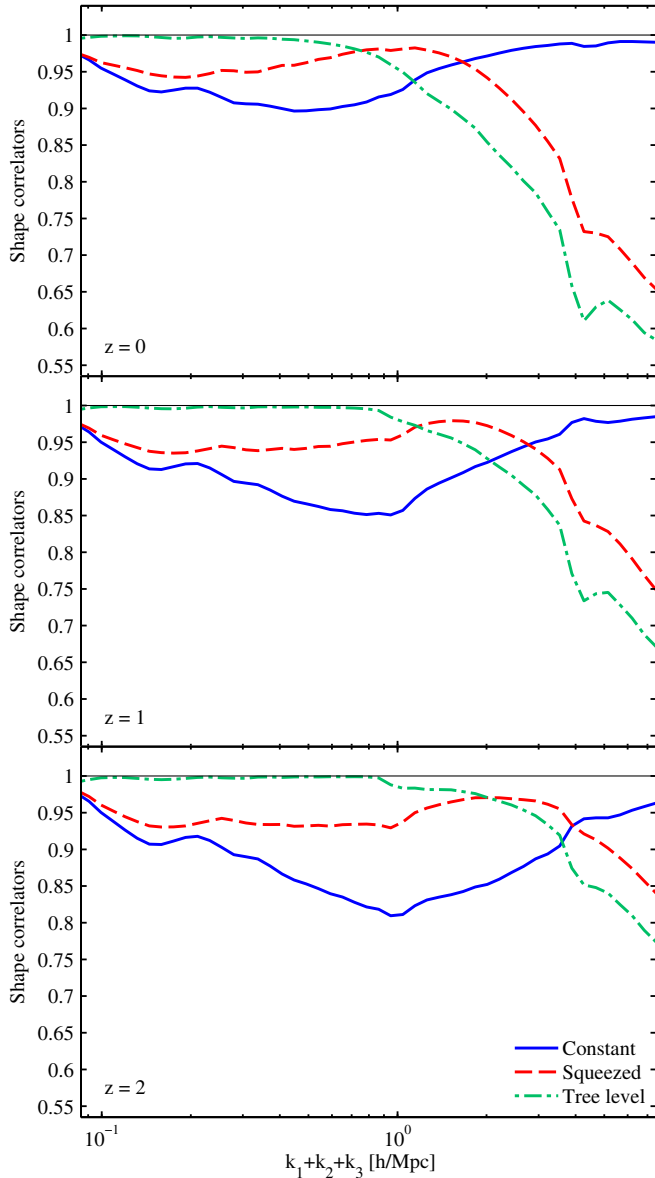


FIG. 13. Sliced shape correlations of the measured N -body bispectrum with the three canonical shapes: constant [Eq. (12)], squeezed [Eq. (13)] and tree-level [Eq. (15)] shown at redshifts $z = \{0, 1, 2\}$ (upper to lower panels). The sliced or binned shape correlator on a given $K = k_1 + k_2 + k_3$ slice is defined in Eq. (19).

These observations can be interpreted using the halo model formalism for which the basic underlying physical assumptions appear to be corroborated qualitatively. On large scales, the three-halo term is dominant because in this regime the particle triplets over which the bispectrum is estimated should typically be in different halos, thus reflecting the large-scale quasilinear bispectrum predicted by perturbation theory. As shown in Sec. IV C, at small K the tree-level shape is the most important contribution to the observed bispectrum. On small scales, the three particles are typically in the same nonlinear virialized halo,

and hence the one-halo component dominates; this has a constant shape, which we confirm to be the leading observed bispectrum shape in the high- K limit. The two-halo term contributes over intermediate length scales, where two particles are in one halo and the third particle is elsewhere; this corresponds to the squeezed shape, which indeed we find to be dominating the bispectrum on intermediate scales (though with a larger contribution for $z > 0$ than expected in the standard halo model).

As a further illustration, we show in Fig. 14 the equilateral bispectrum ($k_1 = k_2 = k_3$) of the halo model at $z = 0$ and $z = 2$ compared with the measured equilateral N -body bispectrum. Here we can see more clearly the three terms contributing to the halo model and how the two-halo term provides the most significant contribution at intermediate scales at $z = 0$. However, a deficit emerges relative to N -body simulations at $z = 2$ where the predicted two-halo term no longer dominates over the one- and three-halo terms.

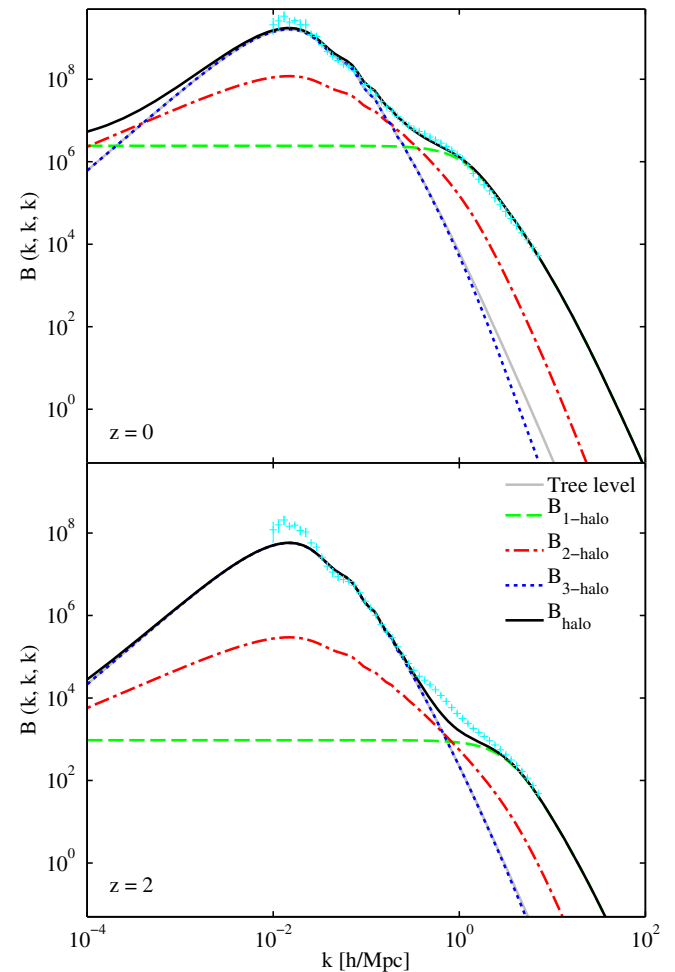


FIG. 14. Equilateral configuration of the halo model bispectrum at $z = 0$ (top panel) and $z = 2$ (bottom panel), showing the contributions of the three components of the halo model, contrasted with the measurements from N -body simulations (cyan points). Note the emerging deficit on intermediate scales at $z = 2$.

B. Two-halo boost model

Based on the observation that the halo model has a deficit at intermediate scales, which is found for bispectrum slices in different configurations and becomes more severe as the redshift increases, we have explored simple phenomenological ways of improving the model. The two-halo term of the halo model has its highest and most important contribution where the deficit is worst.

As a first simple method to improve the agreement between the model and the simulations, we increase the contribution of the two-halo term at higher redshifts in order to compensate for the deficit. We find that a “boosted” two-halo term can provide a much better fit to numerical simulations for redshifts $z > 0$ by multiplying the existing two-halo term by the heuristic factor $D(z)^{-1.7}$. We determine this “best-fit” factor by computing the total correlator \mathcal{T} of Eq. (7) between the model and the simulations separately at each redshift, and then obtaining the scaling law by maximizing the correlator \mathcal{T} . We show in Fig. 15 that the function $D(z)^{-1.7}$ describes well the numerical values found over the relevant redshift range. This simple method solves the power deficit in the intermediate regime but we discussed previously how the halo model already has an excess of power as $k \rightarrow 0$, driven by the combination of one- and two-halo terms (for $z > 0$). Therefore, there is a quantitative problem with simply boosting the two-halo term because it increases the excess on very large scales. In Sec. VII B, we will make direct comparisons with the standard halo and other models.

C. Two-shape time-shift model

In Ref. [110] using tree-level and constant bispectrum shapes it was already recognized that simple phenomenological models of the bispectrum could be constructed; this was motivated by explaining the different growth rates of primordial non-Gaussian shapes in terms of an initial time offset. This time-shift model relies on the fact that in the nonlinear regime the matter bispectrum can be approximated by the constant bispectrum using the following *Ansatz* [consistent with Eq. (10)]:

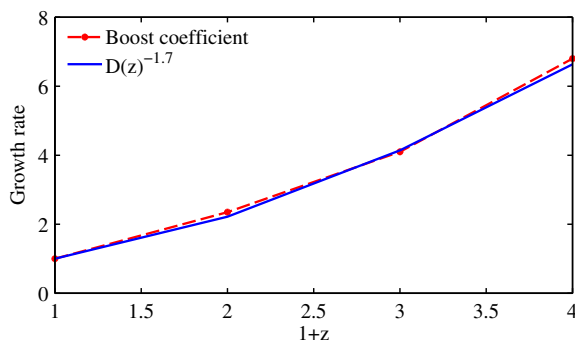


FIG. 15. Best-fit boost coefficient to simulations for the two-halo term compared to $D(z)^{-1.7}$.

$$B_{\text{const}}(k_1, k_2, k_3) = c_1 D(z)^{n_h} K^\nu, \quad (70)$$

with two free parameters, an amplitude c_1 and a growth rate n_h determined from simulations, plus a scale dependence $\nu \approx -1.7$ for equilateral configurations in the one-halo model [101,139]. This two-shape model was further improved by replacing the tree-level bispectrum [Eq. (15)] with the nonlinear tree-level bispectrum [Eq. (17)], i.e. the tree-level bispectrum calculated with the nonlinear power spectrum from simulations:

$$B_{T\text{-shift}}(k_1, k_2, k_3) = c_1 D(z)^{n_h} K^\nu + S^{\text{treeNL}}(k_1, k_2, k_3) \quad (71)$$

While this model produced a reasonable description of the matter bispectrum in terms of the shape correlation \mathcal{S} (see Ref. [110]), our more detailed analysis here with the binned shape correlator \mathcal{S}^S has revealed the possibility of further improvement on intermediate scales by extending the model with the additional squeezed shape of Eq. (13).

D. Three-shape bispectrum model

Based on the three shapes we identified in the halo model in Sec. IV C, we propose a more general benchmark model that incorporates the physical behavior of all these components, but with rescaled growth factors to provide an improved quantitative fit to simulations. As shown in Fig. 7, the one-, two- and three-halo terms have a high shape correlation with the constant, squeezed and tree-level shapes respectively on slices of constant $K \equiv k_1 + k_2 + k_3$. Since these shapes also describe the measured matter bispectrum (see Fig. 13), we can construct it as a scale-dependent sum of three templates:

$$\begin{aligned} B_{3\text{-shape}}(k_1, k_2, k_3) &= \sum_{i=1}^3 f_i(K) S^i(k_1, k_2, k_3) \\ &= f_{1h}(K) S^{\text{const}}(k_1, k_2, k_3) \\ &\quad + f_{2h}(K) S^{\text{squeez}}(k_1, k_2, k_3) \\ &\quad + f_{3h}(K) S^{\text{treeNL}}(k_1, k_2, k_3), \end{aligned} \quad (72)$$

where the nonlinear tree-level, squeezed, and constant shapes S^{treeNL} , S^{squeez} , S^{const} are defined in Eqs. (12), (13) and (17) respectively, and the amplitudes f_{1h} , f_{2h} , f_{3h} were discussed in Sec. IV C in the context of the halo model.¹

¹An even simpler three-shape model can be obtained by substituting the linear tree level [Eq. (15)] for the flattened three-halo shape; it provides a satisfactory fit to the simulations. In this simple scenario, the fitting functions f_{1h} [Eq. (55)] and $f_{3h} = 1$ are given by the standard halo model, while for the two-halo term we allow an improved fit and growth scaling [Eq. (60)] with coefficients $C = 240D(z)^{-1}$ and $D = 2.35 h\text{Mpc}^{-1} D(z)^{-1}$. However, the three-shape model of Eq. (72) above provides an improved fit in the flattened limit in the quasilinear regime.

We know that the one-halo term provides an adequate description of the matter bispectrum on small scales, so we fix the amplitude f_{1h} to the simple functional fit of Eq. (55) for the one-halo model presented in Sec. IV C. On the largest scales, where the three-halo term is dominant, we know that the tree-level shape [Eq. (63)] provides an excellent fit to simulations. However, on intermediates scales, while the shape correlation remains good beyond the strictly perturbative regime (see Fig. 13), its amplitude is insufficient, as can be seen by comparing Fig. 4 with Fig. 12. For this reason, we have chosen the nonlinear tree-level form [Eq. (17)] instead because of its higher amplitude and the fact that it is a better approximation to one-loop perturbative expansions. Nevertheless, it is well known that introducing the nonlinear power spectra into halo models generically causes excess power at low redshifts $z \approx 0$, so we need a prescription for cutting off the flattened shape in nonlinear regions (see, for example, the discussion about the combined halo-PT model in Sec. IV B or the discussion of halo exclusion in Ref. [148]). In order to keep this three-halo suppression as simple as possible we take an exponential form:

$$f_{3h} = \exp(-K/E), \quad (73)$$

where we fit E to simulations at several redshifts to obtain an appropriate amplitude and growth rate; in principle, it should be linked to the nonlinear scale k_{NL} satisfying $k^3 P_{\text{lin}}(k, z) = 2\pi^2$. Finally, for the squeezed shape scaling f_{2h} we do not use the two-halo model amplitude, but instead the prescription of Eq. (60) with the two free parameters C and D obtained from simulations (see discussion in Sec. IV C). By matching f_{2h} to the excess in the measured bispectrum at redshifts $z = \{0, 1, 2, 3\}$, together with the cutoff scale in f_{3h} , we obtain the following approximate fit for the coefficients C , D and E :

$$\begin{aligned} C &= 140D(z)^{-5/4} \\ D &= 1.9 \text{ hMpc}^{-1} D(z)^{-3/2} \\ E &= 7.5k_{\text{NL}}(z). \end{aligned} \quad (74)$$

We emphasize that this is different from the previous two-halo fits of Eqs. (61)–(62), because these were obtained by fitting to the two-halo model predictions, which underestimate power for $z > 0$. This is illustrated starkly at $z = 2$ in Fig. 17, where we compare the standard two-halo model prediction with the squeezed shape of Eq. (60) with best-fit simulation parameters of Eq. (74). We also note that for redshifts $z > 1$, the length scale E moves rapidly to large $K \gg 1 \text{ h/Mpc}$, so the exponential suppression f_{3h} term [Eq. (73)] acts primarily to reduce power in the $z = 0$ bispectrum and is less relevant elsewhere.

In Fig. 16 we plot the value of the binned amplitude $|B|^S(K)$ for the three-shape model of Eq. (72), which we

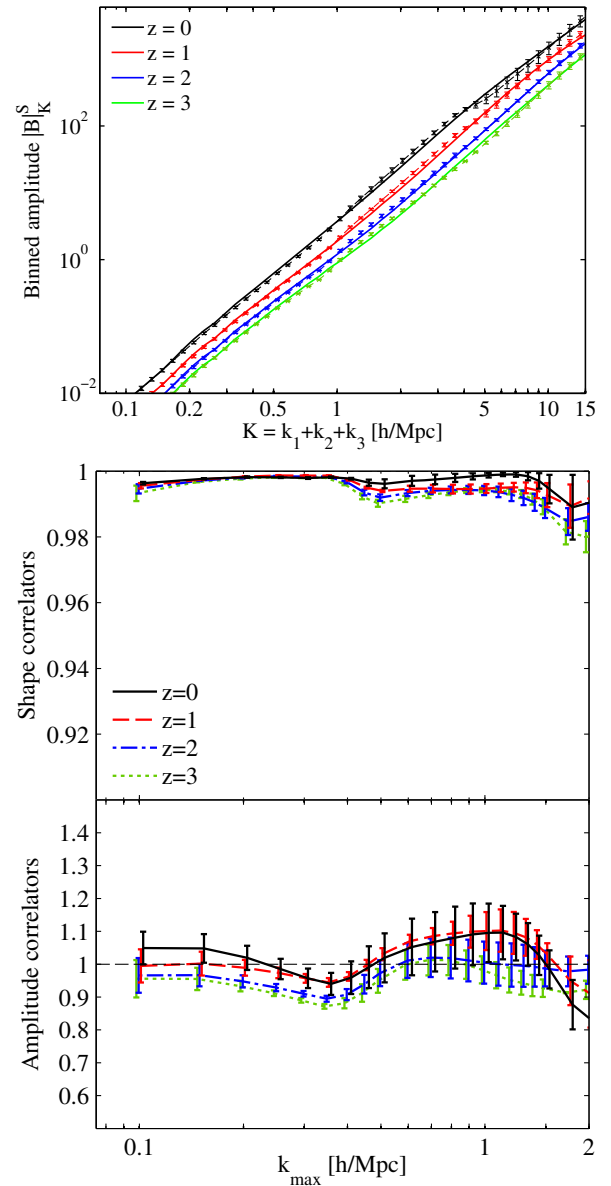


FIG. 16. Comparison between the measured N -body matter bispectrum and the three-shape model. The top panel shows the binned amplitude $|B|^S(K)$ from the simulations (points and dashed lines) and from the fitted three-shape model (solid lines) at redshifts $z = \{0, 1, 2, 3\}$. The middle and bottom panels show a relative comparison between the simulations and the benchmark model, using the binned shape and amplitude correlators, \mathcal{S}^S and \mathcal{A}^S . These results demonstrate that the three-shape model exhibits a high shape correlation on all scales and describes the simulated data well.

compare directly to the measured bispectrum from simulations; we also show the binned shape \mathcal{S}^S and amplitude \mathcal{A}^S correlators between the model and N -body bispectrum. The plots show a good fit using the three-scale model across all scales $k > 0.1 \text{ h/Mpc}$ and all redshifts. The shape correlations in this range are approximately 99% or higher and the amplitude correlator is within 10% of the

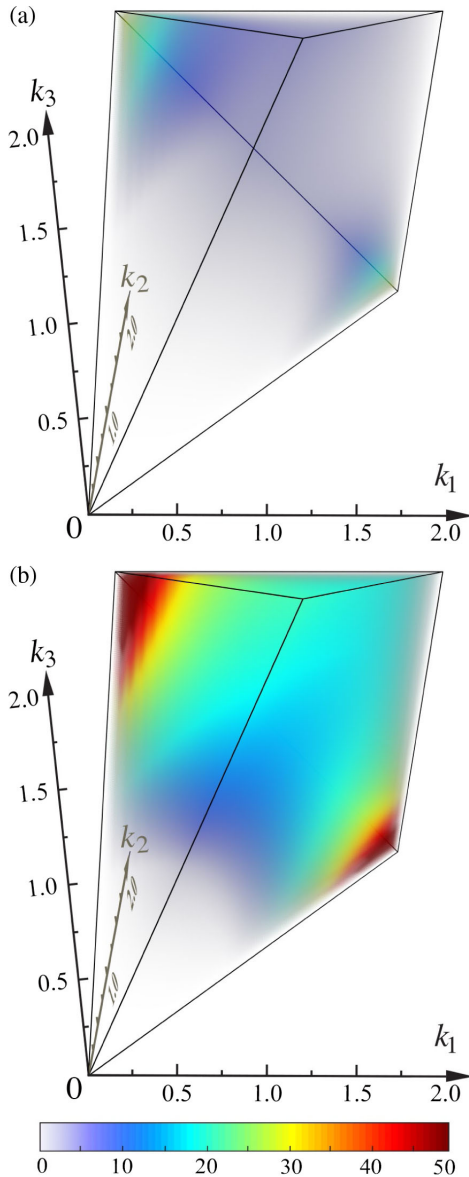


FIG. 17. The SN-weighted two-halo bispectrum [Eq. (45)] (upper panel) at $z = 2$ compared to the best-fit two-halo squeezed shape *Ansatz* [Eq. (58)]; this allows the three-shape benchmark model to accurately match the simulation data shown in Fig. 12. The two-halo model clearly exhibits a large deficit and does not describe squeezed contributions adequately at higher redshift.

measured bispectrum (consistent given present simulation uncertainties). These correlation results are in line with expectations for a good fit for an $n_{\max} = 50$ eigenfunction decomposition [Eq. (65)] (see validation discussions in Ref. [110]). We note that given the high shape correlations, we could introduce additional degrees of freedom in f_{1h} , f_{3h} to improve this quantitative fit further, but our purpose first is to demonstrate the efficacy of this simple approach.

Employing this new three-shape model as a benchmark has several advantages over using the simulated bispectra directly, though we will use both in subsequent discussions.

First, it smooths out any systematic discontinuities appearing where the simulations are joined together. Secondly, it allows direct comparisons with theoretical models without performing eigenfunction decompositions on the latter, so residual offsets do not have to be subtracted. And finally the model is simple, capturing the most important features of the halo model without requiring computationally costly reevaluations at all wave number combinations (k_1, k_2, k_3) , and thus it can be seen as an initial step towards a full HALOFIT-style phenomenological model of the matter bispectrum.

E. Directions for further improvement

The three-shape benchmark model achieves a high degree of correlation with the full bispectrum from N -body simulations; however undoubtedly further improvements of this model can be achieved in the future, not least by deriving some key results from first principles, such as the modified two-halo growth rates. In principle, showing that the matter bispectrum is well approximated by the separable form of Eq. (72) should considerably simplify mathematical modeling.

One improvement that can be incorporated into the model is to replace the nonlinear tree-level shape [Eq. (17)] with specific one- and two-loop perturbative expansions. However, while this approach could extend the tree-level shape further into the nonlinear regime, it requires prescriptions for suppressing the two- and one-halo terms more strongly to avoid overprediction. This is similar in spirit to the suppression of the perturbative bispectrum contribution in the halo-PT model by Ref. [101]; but it is clear that an exponential cutoff where the perturbative expansion breaks down is likely too aggressive, since Fig. 13 shows that the tree-level shape is present up to relatively high wave vector, $k \sim 1 h/\text{Mpc}$.

Clearly further improvement of the three-shape model can be achieved through more extensive comparisons with higher-resolution N -body simulations, over a finer grid of scales and redshifts. The quality of fits obtained in the squeezed and flattened limits are constrained in accuracy by the restricted *Ansätze* chosen, allowing only three redshift-dependent parameters. The likely outcome is a finer tuning of a larger number of phenomenological free parameters, again in the spirit of the HALOFIT method, with extensive surveys required to uncover dependencies on cosmological parameters.

A final point of interest is the question whether the three-shape model we introduced satisfies well-known constraints in the squeezed limit. For example, Ref. [149] derived a consistency relation between the integrated squeezed-limit bispectrum and a response function derived from the power spectrum. In the case of our three-shape model [Eq. (72)], the tree-level shape term satisfies the consistency relation automatically, as was demonstrated by Ref. [149] for tree-level SPT. We know that the

squeezed- and constant-shape terms of the benchmark model are similar to two- and one-halo terms of the standard halo model; furthermore, as we show in Sec. VII below, our model performs well compared with the N -body simulations in the squeezed limit over the configurations we have tested, so that it is unlikely that there is any large inconsistency. However, a more quantitative test of the consistency relation would require a full numerical evaluation of the integrated bispectrum, which we leave for future investigation.

VII. BISPECTRUM MODEL COMPARISON WITH SIMULATIONS

We next use the N -body simulations to compare the accuracy of the different theoretical bispectrum models described in the previous sections, both perturbative and nonperturbative. We present this model comparison in two ways: we first directly compare the simulated and theoretical bispectra over a range of representative triangular configurations (equilateral, squeezed, and flattened), and we then use the full three-dimensional amplitude and shape correlators presented in Sec. II.

At high redshift, all models are expected to perform well over an extended range of scales, as the fluctuations are nearly linear, the power spectrum is linear and the bispectrum can be described by the tree-level expression. At lower redshifts, nonlinearities become more important and significant differences appear between the models. In the comparisons, we concentrate on redshifts $z = \{0, 1, 2\}$, as these span the observable redshift range of most current and future observations from galaxy surveys. We investigate the perturbative methods and the halo models separately, because the perturbative methods decay quickly in the nonlinear regime and therefore their predictions for high k are of no interest; we present the comparison of PT models on scales $k \leq 0.4 h/\text{Mpc}$ only. On the other hand, the phenomenological models, which are either based on or at least inspired by halo models, are expected to perform well even in the fully nonlinear regime; in this case we extend the model comparison up to the smallest scales accessible to the present simulations, i.e. $k \leq 7.8 h/\text{Mpc}$.

A. Testing alternative perturbative approaches

We first qualitatively compare perturbative bispectrum predictions with the matter bispectrum measured from simulations. In Fig. 18 we plot most of these predictions at redshift $z = 2$ in three dimensions for wave numbers $0.02 h/\text{Mpc} < k < 0.6 h/\text{Mpc}$, together with the actual N -body bispectrum (upper left). We choose $z = 2$ so that the perturbative models decay at higher k , and more of the signal is visible, but the overall behavior is comparable at lower z . The N -body bispectrum shows a dominant flattened signal over the whole domain, which grows in amplitude as k increases. Qualitatively, this measured

signal matches well the one-loop SPT and EFT models, as well as the nonlinear tree-level bispectrum, in regions well beyond the strictly perturbative regime. However, the tree-level and MPTBREEZE predictions are appreciably lower for large $k \approx 1 h/\text{Mpc}$ (with the latter exponentially suppressed for large k by prescription).

We confirm these observations for three specific limiting cases in Fig. 19 with a comparison of PT bispectra amplitudes with measured values: from top to bottom we show the equilateral, squeezed, and flattened triangle configurations. In addition to the bispectrum amplitudes, in each case we also plot residuals with respect to the tree-level model. Figure 19 demonstrates that all models converge to the tree level for $k \lesssim 0.1 h/\text{Mpc}$ at $z = 0$ in agreement with simulations; the range of validity of the tree-level theory increases for higher redshift and for flatter shapes. For the phenomenological models, we note that both the simple nonlinear tree-level model and the nine-parameter tree-level fit both increase the range over which there is agreement with simulations. For $z \geq 1$ these two cases are nearly indistinguishable and both show a similar deficit in power for larger k . In principle the nine-parameter model does provide a better match to the $z = 0$ bispectrum; however, it also exhibits large oscillations which originate through the slope parameter n for a power spectrum with baryon acoustic oscillations (BAO) features, as noted and circumvented in Ref. [133]. While it is possible to remove these unwanted oscillations of n with a spline smoothing, we do not apply this extra processing step here for simplicity.

For the one-loop perturbative models plotted in Fig. 19, all approaches agree in the strictly perturbative regime at $z = 0$. However, beyond this regime for larger k , SPT generally overestimates the bispectrum, while the RLPT and RPT MPTBREEZE models underestimate it. The EFT approach lies in between the SPT and RLPT curves, and typically extends the range of agreement with simulations. This trend is also apparent at higher redshift with the exception of the squeezed limit where even SPT falls below the measured bispectrum in the quasilinear regime.

Given these interesting observations, we have undertaken a comprehensive quantitative comparative analysis using the integrated amplitude correlator \mathcal{A} [Eq. (6)] and the shape correlator \mathcal{S} [Eq. (5)], the results of which are plotted in Fig. 20. This corresponds to a signal-to-noise weighted integration over all triangular configurations up to a given resolution k_{max} , rather than the specific limiting configurations Fig. 19. Here, we directly compare the theoretical predictions B_{theory}^j to the three-shape benchmark model $B_{3\text{-shape}}$ given in Eq. (72) with parameters given in Eq. (74), which provides an excellent fit and a smoother representation of the actual bispectrum from simulations (see Sec. VID).

We estimate the uncertainties on the correlators as follows. From each simulation realization i , we obtain the amplitude and shape correlators $\mathcal{A}(B_{\text{sim}}^i, B_{3\text{-shape}})$,

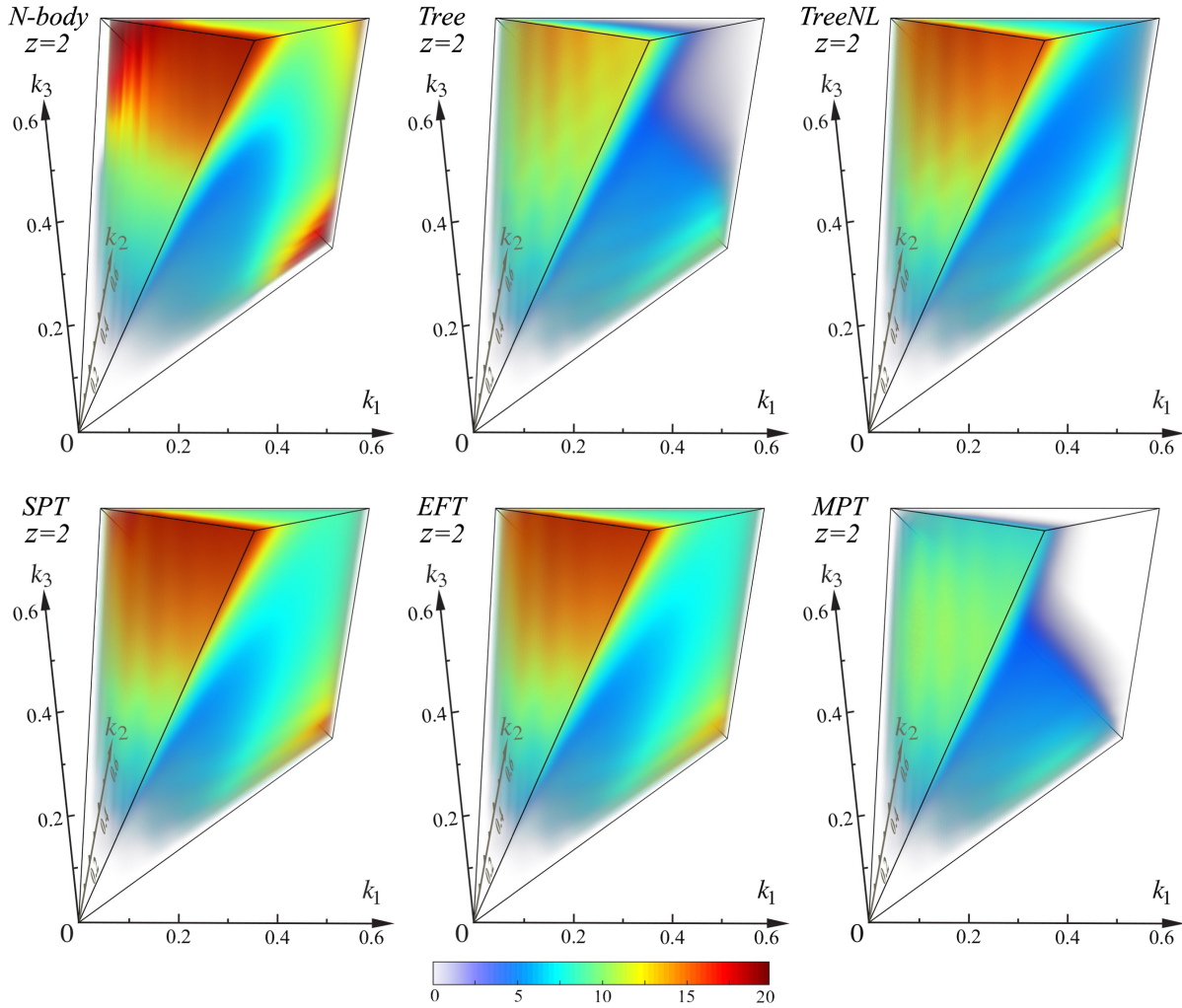


FIG. 18. Comparison at redshift $z = 2$ of the SN-weighted bispectrum for perturbative models with the simulation data (top left): the perturbative models are respectively tree-level bispectrum (top center), nonlinear tree level (top right), standard one-loop perturbation theory SPT (bottom left), one-loop effective field theory EFT (bottom center) and renormalized perturbation theory MPT (bottom right); RLPT is not plotted as it appears very similar to MPT. Note that all perturbation theories have signal concentrated at flattened triangles (from front face), and so are highly correlated with the tree-level bispectrum shape of Eq. (15), when using the binned shape correlator [Eq. (19)]. The N -body bispectrum also exhibits a squeezed signal for $k \gtrsim 0.4 h/\text{Mpc}$. We have chosen $z = 2$ so that the PT models decay at higher k and there is more signal to display, but the general behavior is similar at lower z .

$\mathcal{S}(B_{\text{sim}}^i, B_{3\text{-shape}})$ as a function of k_{max} . For each value of k_{max} , we can thus derive the mean and standard deviation of the correlators: $\mu_{\mathcal{A}}$, $\sigma_{\mathcal{A}}$, and similarly for the shape. As we are comparing all theoretical models with the benchmark three-shape model, the total uncertainty σ_{tot} on the correlators between each model j and the benchmark, $\mathcal{A}(B_{\text{theory}}^j, B_{3\text{-shape}})$, will be larger than the variance $\sigma_{\mathcal{A}}$ obtained from the scatter of $\mathcal{A}(B_{\text{sim}}^i, B_{3\text{-shape}})$; this is because of the small k -dependent discrepancy that exists between the simulations and the smooth benchmark model. The grey shaded areas in Fig. 20 represent two different estimates of σ_{tot} , as follows. The light grey area represents a conservative error estimate obtained by adding the error bars of the simulation to the deviation from one of the mean of the correlator, i.e. assuming $\sigma_{\text{tot}} = |\mu_{\mathcal{A}} - 1| + \sigma_{\mathcal{A}}$, while

the darker grey area represents the part of the benchmark model outside the $1\sigma_{\mathcal{A}}$ error bars, i.e. assuming

$$\sigma_{\text{tot}} = \begin{cases} 0 & \text{if } 1 \in [\mu_{\mathcal{A}} - \sigma_{\mathcal{A}}, \mu_{\mathcal{A}} + \sigma_{\mathcal{A}}] \\ \mu_{\mathcal{A}} - \sigma_{\mathcal{A}} - 1 & \text{if } \mu_{\mathcal{A}} - \sigma_{\mathcal{A}} > 1 \\ 1 - \mu_{\mathcal{A}} - \sigma_{\mathcal{A}} & \text{if } \mu_{\mathcal{A}} + \sigma_{\mathcal{A}} < 1. \end{cases} \quad (75)$$

The same reasoning applies to the shape correlators \mathcal{S} , with the difference that $\mu_{\mathcal{S}} \leq 1$.

From Fig. 20, we note that there are always high shape correlations well beyond the perturbative regime. For example, at $z = 0$ all theories have a shape correlation greater than 99% up to $k < 0.2 h/\text{Mpc}$, even when there are variations of $\mathcal{O}(20\%)$ in the amplitude correlator. These remarkably high shape correlations imply that bispectrum

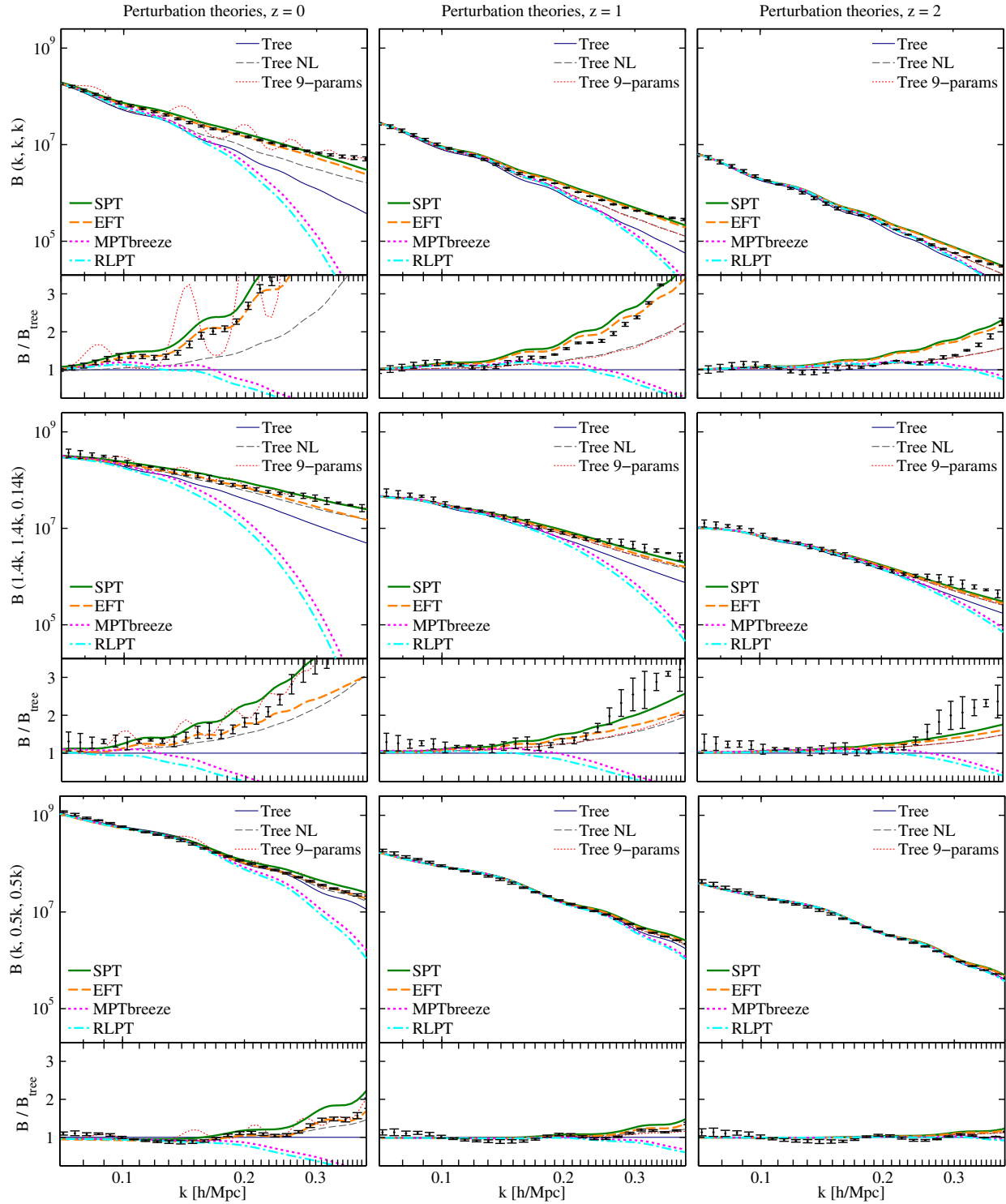


FIG. 19. Comparison of perturbation theory models of the matter bispectrum with N -body simulations, at redshifts 0, 1, 2 (left to right), for the equilateral, squeezed, and flattened configurations (top to bottom). The lower panels show the residuals with respect to the tree-level model.

estimators that measure the projection of the full bispectrum on these theoretical shapes (like in Ref. [111]) should yield a high proportion of the total bispectrum signal-to-noise. Since the shape correlator is not as discerning a tool

for distinguishing between different perturbative models, we focus most attention on amplitude deviations.

We also employ the total correlator \mathcal{T} , which combines the information of amplitude and shape, in order to directly

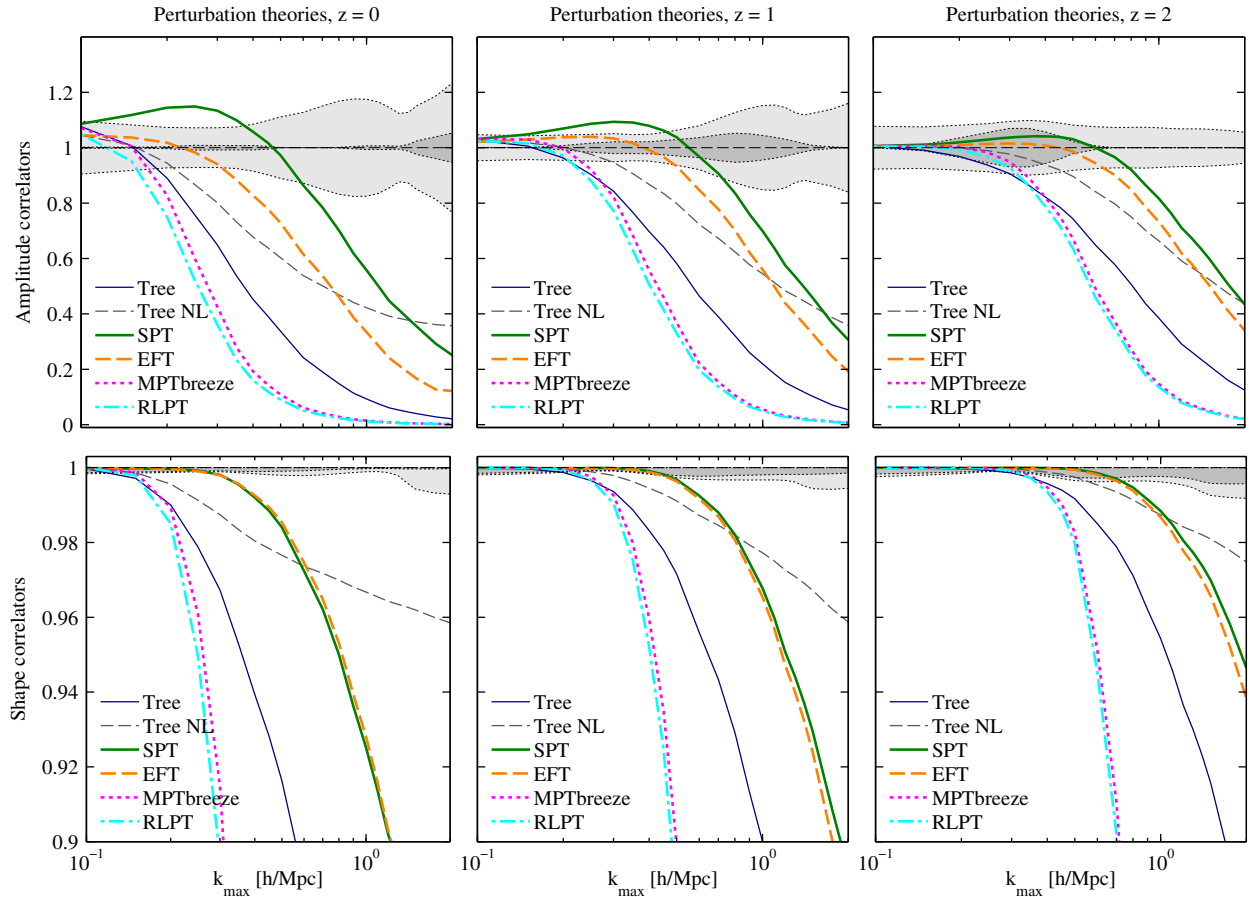


FIG. 20. The amplitude \mathcal{A} (top row) and shape \mathcal{S} (bottom row) correlators at redshifts 0, 1, 2 for the perturbative methods, obtained by comparing with the benchmark model. The shaded areas represent error estimates between the benchmark model and the simulations and are explained in the main body of the paper.

estimate the range of validity of each model as a function of redshift. In Table II we give the maximum wave number k_{\max}^* at which the total correlator between each model and the benchmark model deviates from unity by more than a fixed threshold of 10% (and 5%). While we show results at the three redshifts considered, $z = \{0, 1, 2\}$, an important caveat is that the comparison at $z = 0$ is more approximate, due to the less-than-perfect match between the simulations and the benchmark model; we therefore do not report the 5% results at $z = 0$, and choose to focus primarily on the results at $z = \{1, 2\}$ in the following discussion. A striking feature of Table II is the wide range of wave numbers for which there is good correspondence between theoretical predictions and the measured bispectrum, well beyond expectations for the limits of the perturbative regime estimated in Table I. This shows that even where these theories are no longer expected to be accurate, they can nevertheless be successfully extrapolated into the nonlinear regime for phenomenological modeling.

The tree-level [Eq. (15)] and the nonlinear tree-level [Eq. (17)] models are the simplest approximations to the matter bispectrum, and their range of validity can be

verified from Fig. 20: at $z = 1$ we find $k_{\max}^* = 0.22 h/\text{Mpc}$ for the tree level and $k_{\max}^* = 0.30 h/\text{Mpc}$ for the nonlinear tree level (at 10%). The nonlinear bispectrum improves faster than the linear one at higher redshifts: the tree level increases by roughly $0.05 h/\text{Mpc}$ at each redshift, while the nonlinear tree level increases by $> 0.1 h/\text{Mpc}$.

The one-loop SPT bispectrum adds four extra terms to the tree-level shape. Two of them give positive contributions and the other two negative contributions. As seen in Fig. 20, at low redshift the additional SPT contributions tend to overshoot the measured bispectrum, apparently lowering the value of k_{\max}^* up to which predictions are accurate (see Table II). However, at $z = 2$ the overshoot remains within bounds, extending the fit up as far as $k < 0.66 h/\text{Mpc}$ in the case of the 10% threshold (almost accidentally at this specific redshift, possibly because of additional squeezed contributions in the measured bispectrum). In general, SPT predicts an excess of power on quasilinear scales, before finally decaying in the fully nonlinear regime. This overshoot phenomenon appears because the loop integrals involved require integrating

TABLE II. Wave number k_{\max}^* where the total correlator \mathcal{T} [Eq. (7)] between the perturbative theory and the benchmark model deviates by more than 10% (5%) from unity. In the case of $z = 0$, we only report the 10% results, as the accuracy of the benchmark model is lower.

| Perturbation theories | | | |
|-----------------------|----------------------|-------------|-------------|
| Threshold 10% (5%) | k_{\max}^* (h/Mpc) | | |
| Theory | $z = 0$ | $z = 1$ | $z = 2$ |
| Tree level | 0.13 | 0.22 (0.17) | 0.27 (0.20) |
| NL tree level | 0.17 | 0.30 (0.22) | 0.42 (0.31) |
| SPT | 0.11 | 0.37 (0.14) | 0.66 (0.49) |
| EFT | 0.29 | 0.45 (0.36) | 0.60 (0.50) |
| MPTBREEZE | 0.16 | 0.24 (0.21) | 0.32 (0.28) |
| RLPT | 0.15 | 0.22 (0.19) | 0.30 (0.26) |

momenta over an infinite range, a regime in which the basic assumption $\delta \ll 1$ is no longer valid. Despite this problem, the shape correlation is excellent up to $k \sim 0.3$ h/Mpc, improving significantly over the tree-level result. We also note that evidence for the amplitude overshoot is not very strong from our simulations because they have rather large uncertainty on \mathcal{A} , especially at $z = 0$.

The one-loop EFT bispectrum includes one counterterm, which increases the accuracy of the model due to the one free parameter that is introduced and fitted at the level of the power spectrum. In Fig. 20 we observe that this method provides substantially improved agreement with the simulations, albeit at the cost of an extra parameter, which was calibrated on the power spectrum of N -body simulations, assuming a specific cosmological model. This counterterm effectively removes excess power provided by SPT in the quasilinear regime and the results that we obtain from the three-dimensional comparison are consistent with the improved agreement found in Ref. [138]. The EFT method appears to work well up to $k_{\max}^* = 0.45$ h/Mpc at $z = 1$ and $k_{\max}^* = 0.60$ h/Mpc at $z = 2$. However, we must proceed cautiously before using such projections because the detailed correspondence in the equilateral and squeezed limits shown in Fig. 19 is not as encouraging. (We also observe additional correlated squeezed signals emerging on these scales in the measured bispectrum which require more sophisticated joint fitting.) At higher redshift, the contribution of the counterterm becomes less significant, because the growth rate of the term is $\propto D^{7.1}(z)$ compared to $\propto D^6(z)$ from the one-loop SPT terms. Although one can in principle add another three additional counterterms for the one-loop EFT bispectrum, we have found that the improvement in the accuracy is modest relative to the cost of introducing these further free parameters.

The RPT approach (MPTBREEZE formalism) at one loop solves the SPT excess by cutting off terms appropriately with an exponential function, as can be seen in Fig. 20. Compared to SPT, all terms are positive to any number of loops, and so this is a convergent expansion. With accuracy

increasing with number of loops, the amplitude on all scales should always approach the measured bispectrum from below. We see in Table II that the RPT method appears to be accurate to 10% at $k < 0.24$ h/Mpc at $z = 1$, improving to $k < 0.32$ h/Mpc at $z = 2$. The main improvement of MPTBREEZE compared with the other methods arises on large scales, before the exponential damping begins. The disadvantage of this suppression is that it precludes any extrapolations into the nonlinear regime.

The RLPT results we have obtained are similar to RPT, although the validity range is marginally smaller due to the increased power suppression; in this case we find $k_{\max}^* = 0.22$ h/Mpc at $z = 1$, and $k_{\max}^* = 0.30$ h/Mpc at $z = 2$.

We conclude that all one-loop perturbative methods match simulations at present precision within the expected perturbative regime. In terms of phenomenological extrapolation into the nonlinear regime, the EFT method goes furthest (once the counterterm coefficient has been appropriately fitted). Both RLPT and RPT undershoot the measured bispectrum in this regime by construction, while SPT generically overshoots. On the other hand, the nonlinear tree-level bispectrum Eq. (17) provides a useful projection to larger k which has the advantage of being much simpler to calculate.

B. Testing phenomenological halo models

By analogy with the discussion of the PT methods above, we first make qualitative comparisons of the phenomenological halo models with the measured bispectrum. In Fig. 21, we plot these bispectra in three dimensions at two redshifts $z = 0, 2$. While the standard halo model provides a reasonable fit at $z = 0$, it reveals a large deficit on intermediate scales $k \sim 1$ h/Mpc. This is corrected in the three-shape model by using the nonlinear tree-level bispectrum and adopting a different growth rate for the squeezed signal at higher redshift. In Fig. 22 we offer a more detailed picture in the limiting equilateral, squeezed and flattened configurations, also showing residuals relative to the standard halo model. From Fig. 21, we can see that for all configurations the standard halo model provides a good match to the N -body data on both linear and fully nonlinear scales, while a more significant mismatch appears in the transition regime at redshifts $z > 0$. The problem may be due in part to the approximate nature of the assumption in the halo model about all the matter in the Universe being in collapsed halos, while other sources of inaccuracy are the spherical shapes of the halos as well as neglecting their internal substructure; it is an issue acknowledged in the literature both for the power spectrum and the bispectrum [100–102,148]. We confirm that this mismatch becomes more severe at higher redshift: for example, at $z = 2$ there is up to a factor of three mismatch on these intermediate scales.

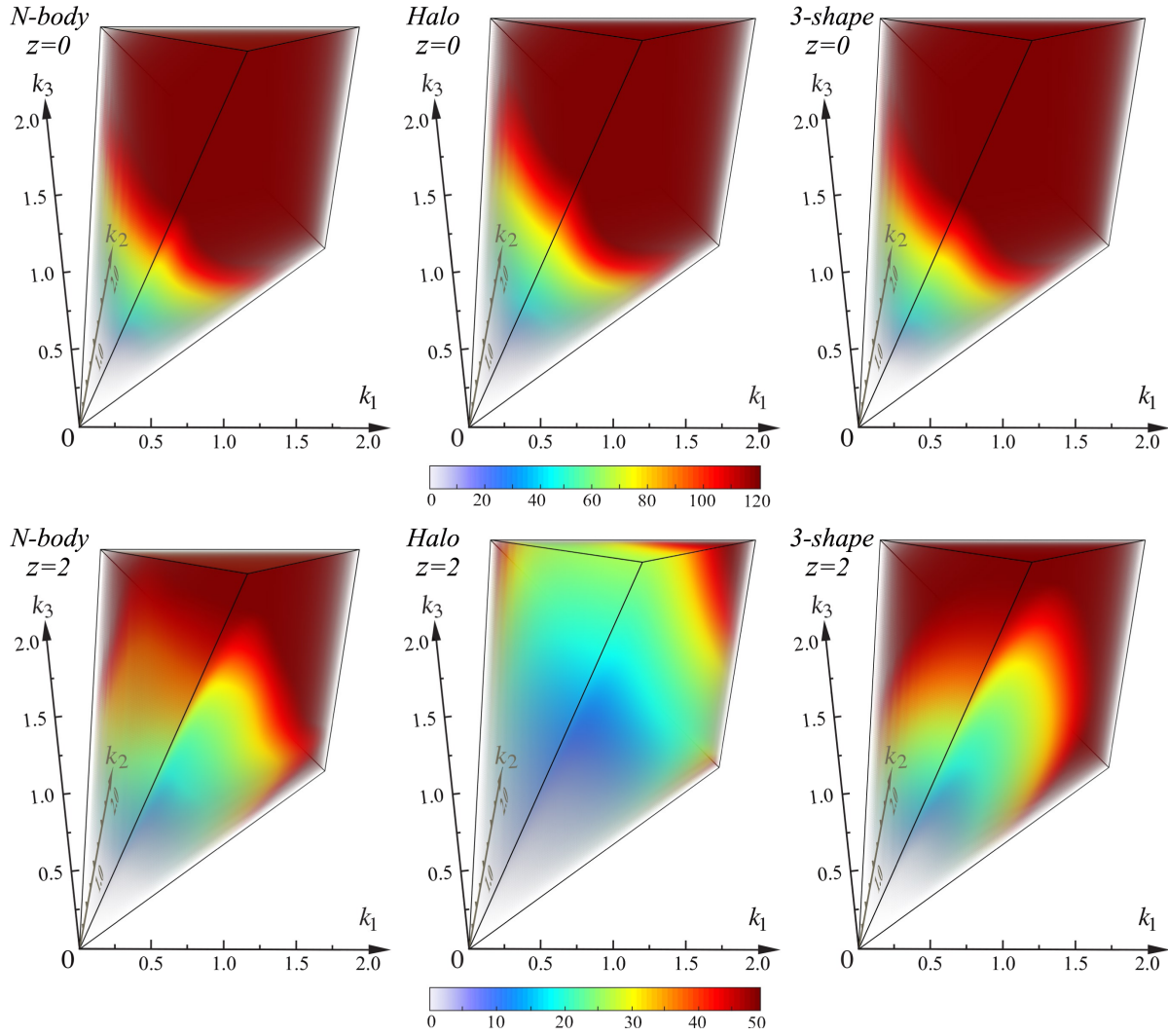


FIG. 21. Comparison between N -body simulation bispectrum (left panels) with the standard halo bispectrum model Eqs. (44)–(46) (centre panels) and the three-shape benchmark model Eq. (72) (right panels) shown at two redshifts $z = 1, 2$. The standard halo model is effectively normalized to fit the measured bispectrum at $z = 0$, which is also achieved well by the phenomenological three-shape model (upper panels). However, at higher redshift $z = 2$ the halo model exhibits the wrong growth rates for the flattened three-halo and squeezed two-halo configurations, yielding a substantial deficit (lower panel center); the measured bispectrum behavior can be accommodated in the three-shape benchmark model (lower panel right).

The other phenomenological models we consider attempt to improve the behavior in the transition region in different ways, and with varying degrees of success; they are also plotted in Fig. 22. The combined halo-PT model provides some improvement at $z = 0$ for flattened configurations, but it fails to significantly improve the situation at higher redshifts and especially in the squeezed limit. The phenomenological two-halo boost and three-shape benchmark models improve the N -body results over a broader range of redshifts and configurations, largely by increasing the relative amplitude of the two-halo term at $z > 0$. The three-shape benchmark, in particular, achieves a satisfactory fit in all limits and at all redshifts using only the restricted *Ansatz* [Eq. (72)] by also increasing power in the flattened limit with the nonlinear tree-level bispectrum.

We now turn to a full three-dimensional analysis with the amplitude (\mathcal{A}) and shape (\mathcal{S}) correlators plotted in Fig. 23 for redshifts $z = 0, 1, 2$; as in the previous subsection, we again compare to the three-shape benchmark model with best-fit parameters of Eq. (74). We also determine where the accuracy of different phenomenological models and fits break down in Table III.

It is apparent from Fig. 23 that the standard halo model offers an insightful description of the matter bispectrum in the nonlinear regime at redshift $z = 0$; the shape correlation is above 99% everywhere investigated and the amplitude deviates by less than 15% from the measured simulation bispectrum over the range $0.4 h/\text{Mpc} < k < 8 h/\text{Mpc}$. Nevertheless, we observe some excess power on large scales, e.g. at $k_{\text{max}} \sim 0.1 h/\text{Mpc}$, which is a well-known

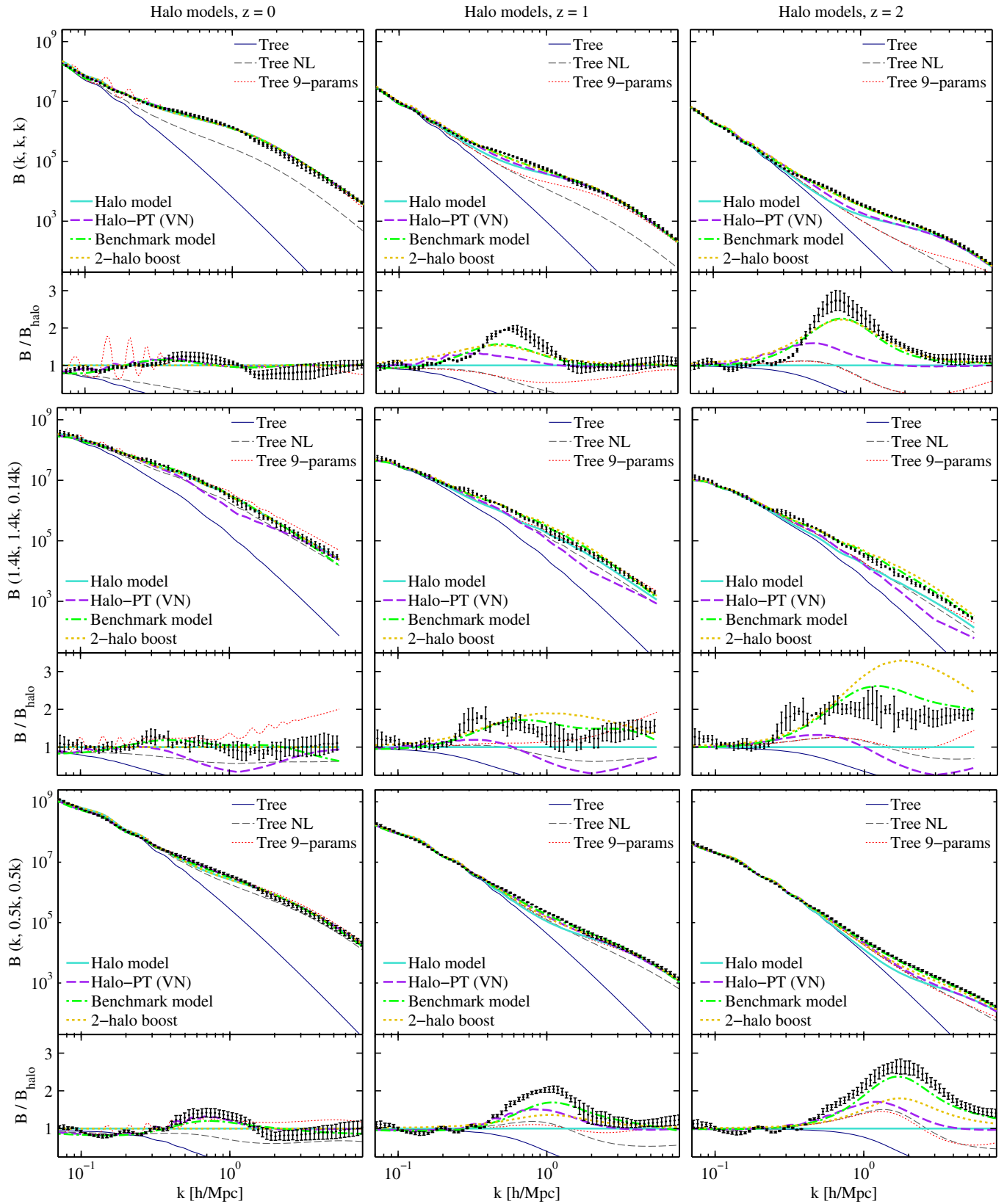


FIG. 22. Comparison of phenomenological nonlinear models of the matter bispectrum with N -body simulations, at redshifts 0, 1, 2 (left to right), for the equilateral, squeezed, and flattened configurations (top to bottom). The lower panels show the substantial residuals with respect to the standard halo model for $z > 0$, demonstrating that the simple three-shape benchmark model provides a good fit to the N -body matter bispectrum for all three limits and redshifts.

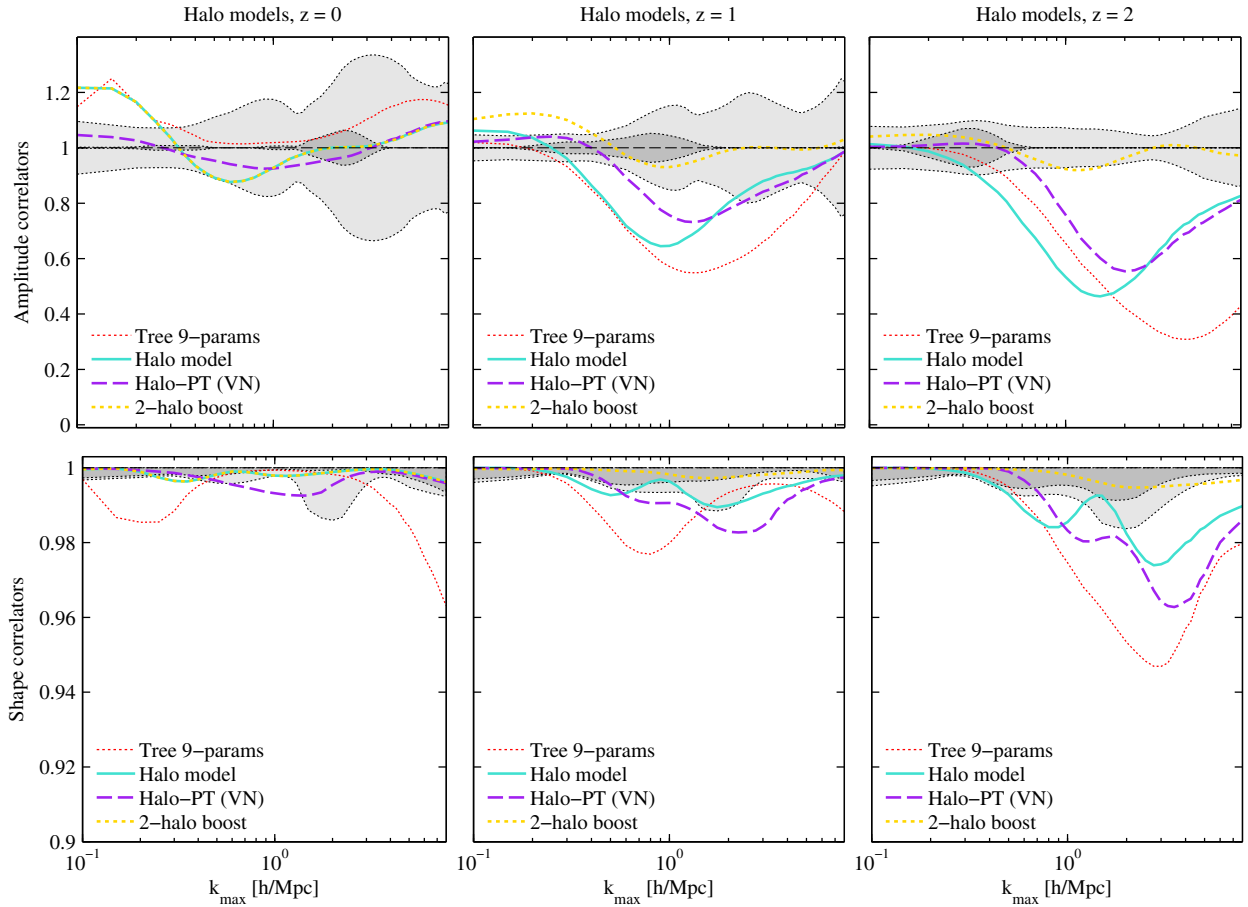


FIG. 23. The amplitude \mathcal{A} (top row) and shape \mathcal{S} (bottom row) correlators at redshifts 0, 1, 2 for the phenomenological halo models, obtained by comparing with the three-shape benchmark model. The shaded areas represent error estimates between the three-shape benchmark model and the simulations and are explained in Sec. VII A.

problem of the standard halo model, due to the one-halo term approaching a constant and the two-halo term not vanishing as $k \rightarrow 0$. The large-scale excess is less important as the redshift is increased. However, we see in Fig. 23 that there is a new problem on intermediate scales where an amplitude deficit emerges, which increases significantly as a function of redshift: in the transition regime, the amplitude correlator decreases from 0.9 at $z = 0$ to 0.65 at $z = 1$, and 0.45 at $z = 2$. As discussed in previous sections, this is primarily due to an underprediction of the two-halo component (squeezed shape) in this k -range. Moreover, the lowest point in the transition regime shifts to higher k at higher redshift, from $k \approx 0.5 h/\text{Mpc}$ at $z = 0$ to $k \approx 1.5 h/\text{Mpc}$ at $z = 2$ (see Fig. 8 for an illustration of this in the equilateral configuration). In the strongly nonlinear regime, after the two-halo component has decayed and the one-halo term becomes dominant, the halo model again approaches the simulations. As we discussed above in Sec. V, a possible way of solving this problem is by boosting the two-halo component, which peaks exactly in the regime of interest; this leads to the two-halo boost model also shown

in Fig. 23, which entails a minimal cost of introducing additional power on large scales.

The power excess produced on linear scales by the standard halo model is corrected in the combined halo-PT model of Sec. IV B. As this model can use any perturbative theory on linear scales, we choose to use EFT, because we found it in the previous section to offer the most extended range of validity. In this prescription, the two- and three-halo terms of the halo model are switched on as the perturbation theory is decaying. Hence, at $z = 0$, this model provides the best fit across all scales considered; in the strongly nonlinear regime, the model converges towards the standard halo model result, because on small scales the improved two-halo and the improved one-halo terms are the same as their standard counterparts. Nonetheless, for $z > 0$, the combined halo-PT model has the same problem as the standard halo model, as there is a deficit in the transition regime, though marginally weaker. In this model, the improved one- and especially the two-halo terms are heavily suppressed on large and intermediate scales. This is not visible in Fig. 23 because most of the signal comes from EFT on these scales, making it more

TABLE III. Wave number k_{\max}^* where the amplitude deviation for phenomenological halo models is greater than 20% when compared to the three-shape benchmark model matched to simulations. (The small k excess problem of the standard halo model is ignored.) At $z = 0$ all models agree within 20% over the entire range of scales.

| Phenomenological halo models | | | |
|------------------------------|----------------------|---------|---------|
| Threshold 20% | k_{\max}^* [h/Mpc] | | |
| Theory | $z = 0$ | $z = 1$ | $z = 2$ |
| Standard halo model | > 8 | 0.47 | 0.51 |
| Combined halo-PT model | > 8 | 0.48 | 0.68 |
| 9-parameter fit | > 8 | 0.82 | 0.90 |

challenging to solve the deficit by a simple boost of the improved two-halo term.

The nine-parameter fit, which is based on the simple tree-level model, fitted to $k \leq 0.4$ h/Mpc and for $z \leq 1.5$ is fairly accurate when extrapolated across the full domain at $z = 0$. (In principle, improvements could be obtained by refitting the parameters to higher redshifts and further into the nonlinear regime, though the model does not naturally include the squeezed and constant shapes required.) Spurious peaks appearing at $z = 0$ are produced by the BAO features of the power spectrum, as discussed previously. However, at $z = 1, 2$ this model becomes increasingly inaccurate at large k with its amplitude decreasing in a similar fashion to the nonlinear tree-level bispectrum. Nevertheless, the nine-parameter model produces an accurate result up to $k_{\max} \sim 0.8$ h/Mpc for all the redshifts considered.

As for perturbation theories, in Table III we present the maximum value of the wave number k_{\max}^* for which the phenomenological halo models show good agreement, that is, by considering the point where the amplitude correlator deviates by more than 20% from unity. The numerical results of the table confirm the general trends discussed above. In contrast to the PT case, here the agreement between models and simulated data becomes worse at higher redshift, as the basic assumptions underlying the halo model become less valid. At higher redshifts, a secondary range of validity exists at high k after the transition region, which is visible from Fig. 23 but not reported in the table.

Among the alternative phenomenological models we tested, we conclude that the combined halo-PT model based on EFT is the most accurate, offering a physically well-motivated attempt to solve problems of the standard halo model. Nevertheless, like the standard halo model, it also does not exhibit appropriate growth rates for the two-halo contribution at high redshift and, further, the prescription for transitioning between EFT and the other halo contributions deserves closer scrutiny. From a phenomenological point of view there is a straightforward means to improve the theory by boosting the two-halo

term at higher redshifts, as in the three-shape benchmark model.

VIII. CONCLUSIONS

The bispectrum of large-scale structure has so far been a relatively neglected observable, due to the high cost of measuring it with most current suboptimal estimators, and the relative complexity of its modeling and interpretation. This is however bound to change in the current age of precision cosmology and ever-larger galaxy surveys, as the combination of two- and three-point statistics can improve the constraining power of the upcoming data, by breaking the existing degeneracies between cosmological and astrophysical parameters. The ultimate goal of large-scale structure bispectrum measurements is its potential to constrain models of the early Universe via their non-Gaussian contribution to the primordial density perturbations, thus complementing and improving existing CMB constraints [38,49].

Achieving these ambitious objectives will require efforts on multiple fronts. A first issue shared with power spectrum analysis is the endeavor to improve the theoretical modeling as far as possible into the nonlinear regime; other outstanding points include making the bispectrum estimation faster and more efficient, and developing a comprehensive method for comparing bispectrum predictions with observations.

In this paper we have made progress on all these fronts. Firstly, we studied how accurately different theoretical models for the matter bispectrum work on different scales, by comparing them with N -body simulations and introducing a new simplified phenomenological model based on three canonical bispectrum shapes. Secondly, we have used for our study the efficient modal bispectrum estimator by Ref. [110], which allowed us to reconstruct the full three-dimensional bispectrum information based on ~ 100 modes only. Thirdly, we have introduced the amplitude, shape, and total correlators as instruments to estimate the overall goodness of match between a bispectrum model and measurements across its full three-dimensional domain, thus greatly simplifying the process of model comparison and parameter estimation.

The different bispectrum models we considered can be divided into two categories: methods based on perturbation techniques, and phenomenological models based on or inspired by the halo model. The perturbative methods assume a small departure from linear scales, when the density fluctuations are small, and therefore have limited range of validity. Multiple approaches exist for increasing the scales of validity of perturbative theories, such as effective field theories, and resummed perturbation theories. We have confirmed that such one-loop recipes manage to accurately model nonlinearities up to $k_{\max} \approx 0.15$ h/Mpc at $z = 0$ for the matter bispectrum and further at higher redshift ($k_{\max} \approx 0.4$ h/Mpc at $z = 2$). This is

already beyond the expectations for the strictly perturbative regime, but some methods appear to be amenable for even more ambitious extrapolations into the nonlinear regime, with effective field theory predictions apparently showing good agreement to $k \approx 0.3 h/\text{Mpc}$ at $z = 0$, though at the cost of introducing free extra parameters calibrated to simulations. The much simpler nonlinear tree-level bispectrum also offered useful nonlinear projections out to $k_{\text{max}} \approx 0.17 h/\text{Mpc}$ at $z = 0$.

In addition, we have derived for the first time the expressions of the two-loop MPTBREEZE bispectrum in an infrared-safe manner, demonstrating that it is analytically and numerically tractable, even if computationally challenging. We have shown the improvement in the wave number range over the one-loop calculation for three triangle shape configurations.

From a different perspective, the halo models rely on models of matter collapse in order to describe nonlinearities from a phenomenological point of view. In that sense, they are valid much further beyond the scales that can be modeled by perturbation theories, and can match simulations reasonably well in the strongly nonlinear regime at $z = 0$. The combined halo-PT model [101] represents a compromise between the two approaches. It relies on a perturbative method on large scales, chosen here to be the EFT, where the halo model is not accurate, while relying on the halo model on nonlinear scales. For these reasons, we found that the halo-PT model gives the most accurate predictions on all scales at $z = 0$. Nevertheless, at higher redshifts, a significant deficit appears at intermediate scales for all halo models.

We have found that a simple way to solve this halo deficit problem is to increase the contribution of the squeezed or two-halo shape at $z > 0$, which we have found dominates in the transition regime. Driven by the observations from N -body simulations, we have generalized this idea, thus developing a simple phenomenological three-shape model that fits the simulations well over the full range of scales and redshifts considered. This benchmark model is based on the fundamental shapes of the halo model—tree-level, squeezed and constant shapes, corresponding to the three-, two- and one-halo terms respectively. This model can be seen as a first step towards the development of an accurate phenomenological model calibrated on N -body simulations, translating the idea behind the HALOFIT method to the bispectrum domain. This will be observationally relevant for weak gravitational lensing which is sensitive to the matter bispectrum.

Solving this two-halo deficit problem motivates our new benchmark model but it uncovers a more serious misconception in the standard halo approach built as it is on a hierarchical picture of structure formation. The basic premise that nonlinear halos form first and then using these to classify and calculate non-Gaussian structures may need to be carefully reconsidered. This is clear already from

the tree-level bispectrum, which is present at high redshifts $z > 30$ long before any halos form; fundamentally it is associated with the initial stage of gravitational collapse in the first dimension which causes “pancakelike” structures to form. The three-halo term accommodates this *a posteriori* by noting that the large-scale tree-level signal will be imprinted on the halo distribution. In the same manner, there will be a squeezed signal from the formation of filamentary structures (due to the onset of collapse in the second dimension), which again precedes halos on any given length scale. At present the two-halo model is flawed by assuming a hierarchical origin for this squeezed bispectrum contribution, and so it does not capture the appropriate growth rate at higher redshift. Our investigations here present quantitative bispectrum data in the relevant intermediate regime, which shows clear pathways ahead for improving the halo model (see also Ref. [85]), as well as mathematical simplifications due to the approximate separability of the underlying bispectrum.

Future developments of this work will on the one hand lead to a more comprehensive and accurate phenomenological model of the matter bispectrum, fitted on higher-resolution simulations, which will provide a bispectrum counterpart to the HALOFIT method. On the other hand, we will extend the modeling and the comparison to the case of biased tracers, i.e. dark matter halos and galaxies, to bridge the gap between modeling and observations by galaxy surveys. Finally, we plan to include the effects of primordial non-Gaussianity of different types to determine how it is amplified through gravitational collapse and how it can be optimally identified.

ACKNOWLEDGMENTS

We wish to thank Martín Crocce, James Fergusson, Donough Regan, Daniel Baumann, Cristiano Porciani, and Christian Wagner for useful discussions, and Eiichiro Komatsu and Emiliano Sefusatti for comments on an earlier version of this work. A. L., T. G., and E. P. S. S. are supported by STFC Grant No. ST/L000636/1. T. G. also acknowledges support from the Kavli Foundation. This work was undertaken on the COSMOS Shared Memory system at DAMTP, University of Cambridge, operated on behalf of the STFC DiRAC HPC Facility. This equipment is funded by BIS National E-infrastructure Capital Grant No. ST/J005673/1 and STFC Grants No. ST/H008586/1, No. ST/K00333X/1 and No. STM007065/1. The 3D bispectrum visualizations used the OSPRay ray-tracing rendering engine and we are grateful for support from Johannes Guenther and Gregory Johnson, Intel Corporation.

APPENDIX A: STANDARD EULERIAN PERTURBATION THEORY

We summarize here the derivation of the SPT power spectrum and bispectrum, following Ref. [56]. The density

contrast δ and the peculiar velocity \mathbf{u} are defined in terms of the average density and velocity \mathbf{v} as

$$\rho(\mathbf{x}, \tau) = \bar{\rho}(\tau)[1 + \delta(\mathbf{x}, \tau)] \quad (\text{A1})$$

$$\mathbf{v}(\mathbf{x}, \tau) = \mathcal{H}\mathbf{x} + \mathbf{u}(\tau). \quad (\text{A2})$$

Then the cosmological gravitational potential Φ satisfies the Poisson equation:

$$\nabla^2 \Phi(\mathbf{x}, \tau) = \frac{3}{2} \Omega_m(\tau) \mathcal{H}^2(\tau) \delta(\mathbf{x}, \tau). \quad (\text{A3})$$

If we define the momentum as $\mathbf{p} = am\mathbf{u}$, the particle number density in phase space $f(\mathbf{x}, \mathbf{p}, \tau)$ satisfies the Vlasov equation:

$$\frac{df}{d\tau} = \frac{\partial f}{\partial \tau} + \frac{\mathbf{p}}{ma} \cdot \nabla f - am \nabla \Phi \cdot \frac{\partial f}{\partial \mathbf{p}} = 0. \quad (\text{A4})$$

In order to obtain the spatial distribution of the particles, the moments of Eq. (A4) can be taken by appropriate integration in momentum space. The first three moments are of interest here:

$$\int d^3 \mathbf{p} f(\mathbf{x}, \mathbf{p}, \tau) = \rho(\mathbf{x}, \tau) \quad (\text{A5})$$

$$\int d^3 \mathbf{p} \frac{\mathbf{p}}{am} f(\mathbf{x}, \mathbf{p}, \tau) = \rho(\mathbf{x}, \tau) \mathbf{u}(\mathbf{x}, \tau) \quad (\text{A6})$$

$$\int d^3 \mathbf{p} \frac{p_i p_j}{am} f(\mathbf{x}, \mathbf{p}, \tau) = \rho(\mathbf{x}, \tau) u_i(\mathbf{x}, \tau) u_j(\mathbf{x}, \tau) \sigma_{ij}(\mathbf{x}, \tau). \quad (\text{A7})$$

Equation (A5) gives the continuity equation and Eqs. (A6), (A7) give the Euler equation, in analogy to fluid mechanics [150]:

$$\frac{\partial \delta(\mathbf{x}, \tau)}{\partial \tau} + \nabla \cdot [(1 + \delta(\mathbf{x}, \tau)) \mathbf{u}(\mathbf{x}, \tau)] = 0 \quad (\text{A8})$$

$$\begin{aligned} \frac{\partial \mathbf{u}(\mathbf{x}, \tau)}{\partial \tau} + \mathcal{H} \mathbf{u}(\mathbf{x}, \tau) + \mathbf{u}(\mathbf{x}, \tau) \cdot \nabla \mathbf{u}(\mathbf{x}, \tau) \\ = -\nabla \Phi(\mathbf{x}, \tau) - \frac{1}{\rho} \nabla_j (\rho \sigma_{ij}). \end{aligned} \quad (\text{A9})$$

On large scales, the Universe is expected to be smooth and hence Eqs. (A8), (A9) can be linearized. By defining the divergence and vorticity of the velocity field as

$$\theta(\mathbf{x}, \tau) = \nabla \cdot \mathbf{u}(\mathbf{x}, \tau) \quad (\text{A10})$$

$$\mathbf{w}(\mathbf{x}, \tau) = \nabla \times \mathbf{u}(\mathbf{x}, \tau), \quad (\text{A11})$$

it can be shown that the vorticity decays quickly due to the expansion of the Universe, and hence it can be ignored. δ and θ satisfy the following equations in Fourier space:

$$\begin{aligned} \frac{\partial \delta(\mathbf{k}, \tau)}{\partial \tau} + \theta(\mathbf{k}, \tau) = - \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 \delta_D(\mathbf{k} - \mathbf{k}_{12}) \alpha(\mathbf{k}_1, \mathbf{k}_2) \\ \times \theta(\mathbf{k}_1, \tau) \delta(\mathbf{k}_2, \tau) \end{aligned} \quad (\text{A12})$$

$$\begin{aligned} \frac{\partial \theta(\mathbf{k}, \tau)}{\partial \tau} + \mathcal{H} \theta(\mathbf{k}, \tau) + \frac{3}{2} \Omega_m \mathcal{H}^2 \delta(\mathbf{k}, \tau) \\ = - \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 \delta_D(\mathbf{k} - \mathbf{k}_{12}) \beta(\mathbf{k}_1, \mathbf{k}_2) \theta(\mathbf{k}_1, \tau) \theta(\mathbf{k}_2, \tau), \end{aligned} \quad (\text{A13})$$

where $\mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2$ and

$$\alpha(\mathbf{k}_1, \mathbf{k}_2) = \frac{\mathbf{k}_{12} \cdot \mathbf{k}_1}{k_1^2} \quad (\text{A14})$$

$$\beta(\mathbf{k}_1, \mathbf{k}_2) = \frac{k_{12}(\mathbf{k}_1 \cdot \mathbf{k}_2)}{2k_1^2 k_2^2}. \quad (\text{A15})$$

In a Λ CDM universe, Eqs. (A12), (A13) can be solved with the expansions

$$\delta(\mathbf{k}, \tau) = \sum_{n=1}^{\infty} D^n(a) \delta_n(\mathbf{k}) \quad (\text{A16})$$

$$\theta(\mathbf{k}, \tau) = -\mathcal{H} \sum_{n=1}^{\infty} D^n(a) \theta_n(\mathbf{k}), \quad (\text{A17})$$

with $D(a)$ the linear growth factor and δ_n and θ_n given in terms of the expansions

$$\begin{aligned} \delta_n(\mathbf{k}) = \int d^3 \mathbf{q}_1 \cdots \int d^3 \mathbf{q}_n F_n(\mathbf{q}_1 \cdots \mathbf{q}_n) \\ \times \delta_1(\mathbf{q}_1) \cdots \delta_1(\mathbf{q}_n) \delta_D(\mathbf{k} - \mathbf{q}_1 - \cdots - \mathbf{q}_n) \end{aligned} \quad (\text{A18})$$

$$\begin{aligned} \theta_n(\mathbf{k}) = \int d^3 \mathbf{q}_1 \cdots \int d^3 \mathbf{q}_n G_n(\mathbf{q}_1 \cdots \mathbf{q}_n) \\ \times \delta_1(\mathbf{q}_1) \cdots \delta_1(\mathbf{q}_n) \delta_D(\mathbf{k} - \mathbf{q}_1 - \cdots - \mathbf{q}_n). \end{aligned} \quad (\text{A19})$$

In what follows, $D(a)$ and $D(z)$ will be used interchangeably, using the relation between the scale factor and redshift $1 + z = \frac{1}{a}$. The kernels F_n and G_n are homogeneous functions of the wave vectors and are given in terms of α , β by the following recurrence relations:

$$F_n(\mathbf{q}_1, \dots, \mathbf{q}_n) = \sum_{m=1}^{n-1} \frac{G_m(\mathbf{q}_1, \dots, \mathbf{q}_m)}{(2n+3)(n-1)} [(2n+1)\alpha(\mathbf{k}_1, \mathbf{k}_2)F_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) + 2\beta(\mathbf{k}_1, \mathbf{k}_2)G_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n)] \quad (\text{A20})$$

$$G_n(\mathbf{q}_1, \dots, \mathbf{q}_n) = \sum_{m=1}^{n-1} \frac{G_m(\mathbf{q}_1, \dots, \mathbf{q}_m)}{(2n+3)(n-1)} [3\alpha(\mathbf{k}_1, \mathbf{k}_2)F_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) + 2n\beta(\mathbf{k}_1, \mathbf{k}_2)G_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n)], \quad (\text{A21})$$

where $F_1 = G_1 = 1$, $\mathbf{k}_1 = \mathbf{q}_1 + \dots + \mathbf{q}_m$, and $\mathbf{k}_2 = \mathbf{q}_{m+1} + \dots + \mathbf{q}_n$. For the correlation functions, the symmetrized versions of these functions are required, denoted $F_n^{(s)}$ and $G_n^{(s)}$. In up to one-loop calculations for the power and bispectrum, only the expressions up to $n = 4$ for F_n are required. The explicit expressions for F_3 and F_4 are given explicitly in Ref. [54]. For F_2 , the expression is given in Eq. (16) and here we show how it can be derived. Considering a matter-only universe, with $\Omega_m = 1$, $a = \tau^2$ and $\mathcal{H} = \frac{2}{\tau}$, Eqs. (A12), (A13) become

$$\delta' + \theta = -I_1[\delta, \theta] \quad (\text{A22})$$

$$\theta' + \frac{2}{\tau}\theta + \frac{6}{\tau^2}\delta = -I_2[\delta, \theta], \quad (\text{A23})$$

where I_1 and I_2 are the expressions on the rhs of Eqs. (A12), (A13). Expanding δ and θ to second order, one obtains the following equations:

$$\delta = \tau^2\delta_1 + \tau^4\delta_2 \quad (\text{A24})$$

$$\delta' = 2\tau\delta_1 + 4\tau^3\delta_2 \quad (\text{A25})$$

$$\theta = -2\tau\theta_1 - 2\tau^3\theta_2 \quad (\text{A26})$$

$$\delta = -2\theta_1 - 6\tau^2\theta_2. \quad (\text{A27})$$

For $n = 1$, I_1 and I_2 are second-order quantities and hence Eqs. (A22), (A23) are solved by $\theta_1 = \delta_1$. For $n = 2$, one has to use the first-order solutions for the integrals on the rhs of the expressions, and the following equations are obtained:

$$4\tau^3\delta_2 - 2\tau^3\theta_2 = -I_1[\tau^2\delta_1, -2\tau\delta_1] \quad (\text{A28})$$

$$-10\tau^2\theta_2 + 6\tau^2\delta_2 = -I_2[\tau^2\delta_1, -2\tau\delta_1]. \quad (\text{A29})$$

The above equations can be solved for δ_2 by substituting the integral expressions I_1 and I_2 and then α [Eq. (A14)] and β [Eq. (A15)]. One finds the integral expression

$$\begin{aligned} \delta_2(\mathbf{k}) &= \int d^3\mathbf{q}_1 \int d^3\mathbf{q}_2 \delta_D(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2) \\ &\times \left[\frac{5}{7} + \frac{2(\mathbf{q}_1 \cdot \mathbf{q}_2)^2}{7q_1^2 q_2^2} + \frac{\mathbf{q}_1 \cdot \mathbf{q}_2}{7} \left(\frac{6}{q_1^2} + \frac{1}{q_2^2} \right) \right]. \quad (\text{A30}) \end{aligned}$$

The expression of $F_2^{(s)}$ from Eq. (16) is finally obtained by symmetrization over the arguments \mathbf{q}_1 and \mathbf{q}_2 .

We have defined in Sec. II the matter power spectrum and bispectrum. Their expressions in SPT can be obtained by inserting the expansions of Eqs. (A16), (A18) into Eqs. (1), (2) respectively. The full expansion is then grouped according to the number of δ_1 's involved. This loop expansion can be interpreted in analogy with the loop diagrams from quantum field theory and this represents an intuitive manner of determining all the contributions at each order in the expansion. In this diagrammatic expansion, the exterior lines represent the arguments of the correlation function, vertices where n lines meet are the kernels F_n , and the interior lines represent wave vectors that are integrated over. As usual, the sum of wave vectors into any vertex should be 0 and numerical factors in front of each diagram represent its symmetry. The loop order represents the number of interior lines in each of its vertices. Following this procedure yields the power spectrum and bispectrum expressions presented in Sec. III A [Eqs. (20)–(29)].

The actual numerical evaluation of the integrals [Eqs. (20)–(29)] is nontrivial because the kernels may diverge. It has been shown [151,152] that the divergences exactly cancel each other when summing the whole contributions at each loop order together, both in the power spectrum and bispectrum, provided that the linear power spectrum grows slowly enough on very large scales. However, for the numerical evaluation, a method to remove the divergences should be used. For the power spectrum at one loop only, a convenient split of the integration regions has been used in Ref. [55] which solves the divergence problems. More recently, both the power spectrum and bispectrum divergences have been eliminated in Refs. [135–138]. We briefly explain this last method in the next paragraphs.

By considering Eqs. (22)–(29), it can be easily seen that divergences appear at $q = 0$ and $\mathbf{q} = \pm\mathbf{k}_i$. The basic idea of the method is to first perform a convenient change of variable in order to move the divergences to 0 and then, as the variable of integration spans all space, to do a symmetrization in $\mathbf{q} \leftrightarrow -\mathbf{q}$. For the power spectrum, this method yields

$$P_{1\text{-loop}}^{\text{SPT}}(k, z) = D^4(z) \int \frac{d^3q}{(2\pi)^3} [6P_{\text{lin}}(k)P_{\text{lin}}(q)F_3^{(s)}(\mathbf{k}, \mathbf{q}, -\mathbf{q}) + 2P_{\text{lin}}(q)P_{\text{lin}}(|\mathbf{k} - \mathbf{q}|)[F_2^{(s)}(\mathbf{q}, \mathbf{k} - \mathbf{q})]^2\Theta(|\mathbf{k} - \mathbf{q}| - q) + 2P_{\text{lin}}(q)P_{\text{lin}}(|\mathbf{k} + \mathbf{q}|)[F_2^{(s)}(-\mathbf{q}, \mathbf{k} + \mathbf{q})]^2\Theta(|\mathbf{k} + \mathbf{q}| - q)], \quad (\text{A31})$$

where Θ is the Heaviside step function. For the bispectrum, $B_{321}^{(II)}$ and B_{411} only have divergences at 0 and hence do not need any change of variable. The integrand b_{222} of B_{222} needs to be reexpressed in the following manner [137,138] (where we correct a typo in the original paper):

$$\begin{aligned} b_{222}^{(k_3 > k_1)} &= 8P_{\text{lin}}(q)P_{\text{lin}}(|\mathbf{k}_2 - \mathbf{q}|)P_{\text{lin}}(|\mathbf{k}_3 + \mathbf{q}|)F_2^{(s)}(-\mathbf{q}, \mathbf{k}_3 + \mathbf{q})F_2^{(s)}(\mathbf{k}_3 + \mathbf{q}, \mathbf{k}_2 - \mathbf{q}) \\ &\quad \times F_2^{(s)}(\mathbf{k}_2 - \mathbf{q}, \mathbf{q})\Theta(|\mathbf{k}_2 - \mathbf{q}| - q)\Theta(|\mathbf{k}_3 + \mathbf{q}| - q) \\ &\quad + 8P_{\text{lin}}(|\mathbf{k}_3 + \mathbf{q}|)P_{\text{lin}}(|-\mathbf{k}_1 + \mathbf{q}|)P_{\text{lin}}(q)F_2^{(s)}(\mathbf{k}_3 + \mathbf{q}, -\mathbf{q})F_2^{(s)}(-\mathbf{q}, -\mathbf{k}_1 + \mathbf{q}) \\ &\quad \times F_2^{(s)}(-\mathbf{k}_1 + \mathbf{q}, -\mathbf{q} - \mathbf{k}_3)\Theta(|-\mathbf{k}_1 + \mathbf{q}| - |\mathbf{k}_3 + \mathbf{q}|)\Theta(|\mathbf{k}_3 + \mathbf{q}| - \mathbf{q}) \\ &\quad + 8P_{\text{lin}}(|\mathbf{k}_2 - \mathbf{q}|)P_{\text{lin}}(q)P_{\text{lin}}(|\mathbf{k}_1 + \mathbf{q}|) \times F_2^{(s)}(-\mathbf{k}_2 + \mathbf{q}, -\mathbf{k}_1 - \mathbf{q}) \\ &\quad \times F_2^{(s)}(-\mathbf{k}_1 - \mathbf{q}, \mathbf{q})F_2^{(s)}(\mathbf{q}, \mathbf{k}_2 - \mathbf{q})\Theta(|\mathbf{k}_2 - \mathbf{q}| - q)\Theta(|\mathbf{k}_1 + \mathbf{q}| - |\mathbf{k}_2 - \mathbf{q}|) \\ &\quad + 8P_{\text{lin}}(|\mathbf{k}_2 - \mathbf{q}|)P_{\text{lin}}(\mathbf{q})P_{\text{lin}}(|\mathbf{k}_1 + \mathbf{q}|)F_2^{(s)}(-\mathbf{k}_2 + \mathbf{q}, -\mathbf{k}_1 - \mathbf{q})F_2^{(s)}(-\mathbf{k}_1 - \mathbf{q}, \mathbf{q}) \\ &\quad \times F_2^{(s)}(\mathbf{q}, \mathbf{k}_2 - \mathbf{q})\Theta(|\mathbf{k}_2 - \mathbf{q}| - q)\Theta(|\mathbf{k}_2 - \mathbf{q}| - |\mathbf{k}_1 + \mathbf{q}|) \end{aligned} \quad (\text{A32})$$

$$b_{222}^{(k_3 < k_1)} = b_{222}^{(k_3 > k_1)}|_{\mathbf{k}_1 \leftrightarrow \mathbf{k}_3}, \quad (\text{A33})$$

with the note that this expression is only valid under the integral sign due to the various remappings. Similarly, b_{321}^I becomes

$$b_{321}^I \rightarrow 2b_{321}^I\Theta(|\mathbf{k}_2 - \mathbf{q}| - q). \quad (\text{A34})$$

The sum of the four contributions is then calculated by performing the integrals directly. They can be calculated numerically fast using the multidimensional integrator CUBA [153].

APPENDIX B: EFFECTIVE FIELD THEORY

The equations governing this effective field theory are obtained by considering the collisionless Boltzmann equation in an expanding universe and smoothing it on a length scale Λ^{-1} . Hence, the theory is determined by the equations of motion of the long-wavelength modes, sourced by a stress-energy tensor. In the absence of the stress-energy tensor, the SPT Eqs. (A12)–(A13) are recovered. This stress-energy tensor only modifies the Euler equation for the velocity, by adding a term $-\frac{1}{\rho_i}\nabla\tau$ on the rhs of Eq. (A9); the continuity and Poisson equation remain unchanged, but valid for the long-wavelength modes only.

Therefore, the equations of motion are modified and the expansions for the density [Eq. (A16)] and velocity [Eq. (A17)] perturbations are modified to

$$\delta(k, a) = \sum_{n=1}^{\infty} D^n(a)\delta_n(\mathbf{k}) + \epsilon \sum_{n=1}^{\infty} D^{n+\zeta}(a)\tilde{\delta}_n(\mathbf{k}) \quad (\text{B1})$$

$$\theta(k, a) = \sum_{n=1}^{\infty} D^n(a)\theta_n(\mathbf{k}) + \epsilon \sum_{n=1}^{\infty} D^{n+\zeta}(a)\theta\delta_n(\mathbf{k}), \quad (\text{B2})$$

where δ_n and θ_n can be expressed in terms of the kernels F_n and G_n [Eqs. (A18)–(A21)], while the expressions with tildes can be expressed similarly in terms of \tilde{F}_n and \tilde{G}_n . The kernels with tildes satisfy however slightly more complicated recurrence relations, shown in Ref. [138]. ζ is a constant fixed from the scaling of the power spectrum.

In the case of the power spectrum, the lowest level counterterm that appears is the two-point correlation function between δ_1 and $\tilde{\delta}_1$. This can be expressed in terms of a single free parameter, the sound speed $c_{s(1)}^2$. Hence, the term of Eq. (31) is added to the linear and SPT one-loop terms [135]. The free parameter is fixed by fitting the one-loop EFT power spectrum with the non-linear power spectrum at a low value of k , where the SPT result is still valid, while ζ is fixed by looking at the redshift evolution of the power spectrum, and a value of $\zeta = 3.1$ is found to best fit simulations as well as scaling properties of the Universe.

For the bispectrum, we use the counterterm corresponding to the tree-level bispectrum. Only the $\tilde{F}_2^{(s)}$ kernel is required, which has the following expression:

$$\begin{aligned} \tilde{F}_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) = & -\frac{\bar{c}_1}{(1+\zeta)(7+2\zeta)} \left\{ \left(5 + \frac{113\zeta}{14} + \frac{17\zeta^2}{7} \right) (k_1^2 + k_2^2) + \left(7 + \frac{148\zeta}{7} + \frac{48\zeta^2}{7} \right) \mathbf{k}_1 \cdot \mathbf{k}_2 \right. \\ & + \left(2 + \frac{59\zeta}{7} + \frac{18\zeta^2}{7} \right) \left(\frac{1}{k_1^2} + \frac{1}{k_2^2} \right) (\mathbf{k}_1 \cdot \mathbf{k}_2)^2 + \left(\frac{7}{2} + \frac{9\zeta}{2} + \zeta^2 \right) \left(\frac{k_1^2}{k_2^2} + \frac{k_2^2}{k_1^2} \right) \mathbf{k}_1 \cdot \mathbf{k}_2 \\ & \left. + \left(\frac{20\zeta}{7} + \frac{8\zeta^2}{7} \right) \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^3}{k_1^2 k_2^2} \right\}. \end{aligned} \quad (\text{B3})$$

The bispectrum counterterm that follows has no extra free parameters in addition to those needed for the power spectrum, and can be expressed as given in Eq. (33).

Four counterterms corresponding to the one-loop bispectrum can be added to the one above. The counterterms have three free parameters. Their expressions are shown in Ref. [138], but the improvement in the accuracy of the bispectrum is modest, and we will thus disregard them.

APPENDIX C: RENORMALIZED PERTURBATION THEORY

Using the notation from SPT and defining $\eta \equiv \log a(\tau)$, we have the following two-component vector:

$$\Psi(\mathbf{k}, \eta) = (\delta(\mathbf{k}, \eta), -\theta(\mathbf{k}, \eta)/\mathcal{H}). \quad (\text{C1})$$

Equations (A12), (A13) may be recast in a matrix notation,

$$\begin{aligned} \partial_\eta \Psi_a(\mathbf{k}, \eta) + \Omega_{ab}(\mathbf{k}, \eta) \\ = \gamma_{abc}^{(s)}(\mathbf{k}, \mathbf{k}_1, \mathbf{k}_2) \Psi_b(\mathbf{k}_1, \eta) \Psi_c(\mathbf{k}, \eta), \end{aligned} \quad (\text{C2})$$

where

$$\Omega_{ab} = \begin{pmatrix} 0 & -1/2 \\ -3/2 & 1/2 \end{pmatrix}, \quad (\text{C3})$$

and $\gamma_{abc}^{(s)}$ is a symmetrized vertex matrix given in terms of the functions α [Eq. (A14)] and β [Eq. (A15)]. Finally, the solution to the perturbation equations can be given in terms of an inverse Laplace transform,

$$\begin{aligned} \Psi_a(\mathbf{k}, \eta) = g_{ab}(\eta) \phi(\mathbf{k}) + \int_0^\eta d\eta' g_{ab}(\eta - \eta') \\ \times \gamma_{bcd}^{(s)}(\mathbf{k}, \mathbf{k}_1, \mathbf{k}_2) \Psi_c(\mathbf{k}_1, \eta') \Psi_d(\mathbf{k}, \eta'), \end{aligned} \quad (\text{C4})$$

where g_{ab} is the linear propagator, defined for positive η as

$$g_{ab}(\eta) = \frac{e^\eta}{5} \begin{pmatrix} 3 & 2 \\ 3 & 2 \end{pmatrix} - \frac{e^{-3\eta/2}}{5} \begin{pmatrix} -2 & -2 \\ 3 & -3 \end{pmatrix} \quad (\text{C5})$$

and $g_{ab}(\eta) = 0$ for $\eta < 0$. Analogously to SPT, Eq. (C4) can be solved by a series expansion,

$$\Psi_a(\mathbf{k}, \eta) = \sum_{n=1}^{\infty} \Psi_a^{(n)}(\mathbf{k}, \eta), \quad (\text{C6})$$

where

$$\begin{aligned} \Psi_a^{(n)}(\mathbf{k}, \eta) = \int \delta_D(\mathbf{k} - \mathbf{k}_{1\dots n}) \mathcal{F}_{aa_1\dots a_n}^{(n)}(\mathbf{k}_1, \dots, \mathbf{k}_n; \eta) \\ \times \phi(\mathbf{k}_1) \cdots \phi(\mathbf{k}_n) \end{aligned} \quad (\text{C7})$$

and $\mathbf{k}_{1\dots n} = \mathbf{k}_1 + \dots + \mathbf{k}_n$. The kernel function \mathcal{F} satisfies recurrence relations that are analogous to Eqs. (A20), (A21). In this fashion, the SPT solutions are obtained. However, this approach allows for a simplified formalism, because Feynman diagrams can be used. The basic rules are described in detail in Ref. [70].

Nonlinearities modify the linear propagator into a fully nonlinear one, defined as

$$G_{ab}(k, \eta) \delta_D(\mathbf{k} - \mathbf{k}') = \left\langle \frac{\delta \Psi_a(\mathbf{k}, \eta)}{\delta \phi_b(\mathbf{k}')} \right\rangle. \quad (\text{C8})$$

This represents the response of the final density and velocity fields to variations in initial conditions. Using the series expansion [Eq. (C6)], it can be expressed in terms of the linear propagator:

$$G_{ab}(k, \eta) = g_{ab}(k, \eta) + \sum_{n=2}^{\infty} \left\langle \frac{\delta \Psi_a^{(n)}(\mathbf{k}, \eta)}{\delta \phi_b(\mathbf{k}')} \right\rangle. \quad (\text{C9})$$

Nonlinearities also modify the vertex functions. Thus, the symmetric full vertex function Γ is defined in terms of the fully nonlinear propagator, with the vertex function γ becoming just the first term of a perturbative expansion:

$$\begin{aligned} \left\langle \frac{\delta^2 \Psi_a(\mathbf{k}, \eta)}{\delta \phi_e(\mathbf{k}_1) \delta \phi_f(\mathbf{k}_2)} \right\rangle \\ = 2 \int_0^\eta ds \int_0^s ds_1 \int_0^s ds_2 G_{ab}(\eta - s) \\ \times \Gamma_{bcd}^{(s)}(\mathbf{k}, s; \mathbf{k}_1, s_1; \mathbf{k}_2, s_2) G_{ce}(s_1) G_{df}(s_2). \end{aligned} \quad (\text{C10})$$

Switching again to the Feynman diagram formalism, the nonlinear propagator satisfies Dyson's formula:

$$G_{ab}(\mathbf{k}, \eta) = g_{ab}(\eta) + \int_0^\eta ds_1 \int_0^{s_1} ds_2 g_a c(\eta - s_1) \times \Sigma_{cd}(\mathbf{k}, s_1, s_2) G_{db}(\mathbf{k}, s_2, \eta'), \quad (\text{C11})$$

where Σ represents the sum of the principal path irreducible diagrams (diagrams that cannot be split into disjoint pieces by removing a linear propagator from the principal path).

In the small-scale limit, the infinite series for the propagator can be resummed after a lengthy computation to [71]:

$$G_{ab}(k, a) = g_{ab}(a) \exp\left(-\frac{k^2 \sigma_a^2}{2}\right), \quad (\text{C12})$$

where $\sigma_a^2 = \frac{(a-1)^2}{3} \int \frac{d^3 q P_{\text{lin}}}{2\pi^3 q^2}$.

This method permits the calculation of the n -point correlation function in RPT for an arbitrary number of loops. Explicit expressions for the power spectrum and bispectrum are presented in Ref. [73]. Compared to SPT, this method has the advantage that all the contributions involved are positive and the resummation of the propagator terms gives a well-defined perturbative expansion in the nonlinear regime. However, the expressions involved are computationally demanding—one has to solve numerically a set of integro-differential equations. Moreover, more than one loop is required to obtain an accurate result, even on mildly nonlinear scales.

In order to solve these problems, Refs. [73,75] proposed a method that simplifies the calculation dramatically. The scheme is called MPTBREEZE and in this formalism only the late-time propagator is calculated and hence no time integrations are required. First, the nonlinear propagator is generalized to an arbitrary number of points. The $(n+1)$ -point propagator $\Gamma^{(p)}$ has been defined as

$$\frac{1}{p!} \left\langle \frac{\delta \Psi_a^p(\mathbf{k}, a)}{\delta \phi_{b_1}(\mathbf{k}_1) \cdots \delta \phi_{b_p}(\mathbf{k}_p)} \right\rangle = \delta_D(\mathbf{k} - \mathbf{k}_{1\dots p}) \Gamma_{ab_1\dots b_p}^{(p)}(\mathbf{k}_1, \dots, \mathbf{k}_p, a), \quad (\text{C13})$$

where $\mathbf{k}_{1\dots p} = \mathbf{k}_1 + \cdots + \mathbf{k}_p$. In this framework, the power spectrum can be expressed as

$$P(k, z) = \sum_{r \geq 1} r! \int \delta_D(\mathbf{k} - \mathbf{q}_{1\dots r}) [\Gamma^{(r)}(\mathbf{q}_1, \dots, \mathbf{q}_r, z)]^2 \times P_{\text{lin}}(q_1) \cdots P_{\text{lin}}(q_r) d^3 q_1 \cdots d^3 q_r. \quad (\text{C14})$$

If only the growing mode initial conditions are considered, the growing mode solutions reduce to the following simple expression,

$$\Gamma_\delta^{(n)}(\mathbf{k}_1, \dots, \mathbf{k}_n; z) = D^n(z) F_n^{(s)}(\mathbf{k}_1, \dots, \mathbf{k}_n) \exp[f(k) D^2(z)], \quad (\text{C15})$$

where the function f depends only on the linear power spectrum today:

$$f(k) = \int \frac{d^3 q}{(2\pi)^3} \frac{P_{\text{lin}}(q, z=0)}{504k^3 q^5} \left[6k^7 q - 79k^5 q^3 + 50q^5 k^3 - 21kq^7 + \frac{3}{4}(k^2 - q^2)^3(2k^2 + 7q^2) \log \frac{|k - q|^2}{|k + q|^2} \right]. \quad (\text{C16})$$

The numerical results obtained with this method agree well with the more exact method *RegPT* [154] over the relevant range of scales. Using the simplified MPTBREEZE formalism, the power spectrum up to one loop can be expressed as

$$P_{\text{tree}}^{\text{MPTBreeze}}(k, z) = [\Gamma^{(1)}(k; z)]^2 P_{\text{lin}}(k) \quad (\text{C17})$$

$$P_{1\text{-loop}}^{\text{MPTBreeze}}(k, z) = 2 \int \frac{d^3 q}{(2\pi)^3} [\Gamma^{(2)}(\mathbf{k} - \mathbf{q}, \mathbf{q}; z)]^2 \times P_{\text{lin}}(|\mathbf{k} - \mathbf{q}|) P_{\text{lin}}(q). \quad (\text{C18})$$

Using Eq. (C15) and Eqs. (20) and (23) from Sec. III A and Eq. (97) from Ref. [135], the MPTBREEZE contributions can be expressed in terms of their SPT counterparts as given in Eq. (34).

The bispectrum contributions can be treated in a similar manner [74], and the result up to one loop is given in Eq. (36) in terms of the SPT one-loop contributions [Eqs. (26), (27)].

This prescription allows an easy computation of the power spectrum and bispectrum for this method once the SPT counterparts have been determined, as only one integral function (f) needs to be evaluated, the other terms being calculated in SPT. Unfortunately, applying this theory up to one loop is only expected to give a reliable prediction for the bispectrum up to $k_{\text{max}} = 0.15 h/\text{Mpc}$ at $z = 0$. Therefore, it is desirable to go to two loops in order to increase the range of validity of the model.

APPENDIX D: RESUMMED LAGRANGIAN PERTURBATION THEORY

Alternatively, perturbation theory can be derived as a function of the Lagrangian coordinates \mathbf{q} , which are related to their Eulerian counterparts \mathbf{x} by the displacement field Ψ :

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + \Psi(\mathbf{q}, t). \quad (\text{D1})$$

Under the assumption that the density perturbations at initial times are negligible, Eulerian and Lagrangian coordinates are related by the continuity equation: $\rho(\mathbf{x}) d^3 x = \bar{\rho} d^3 q$. Using the properties of the Dirac δ_D distribution, this leads to

$$\delta(\mathbf{x}) = \int d^3q \delta_D[\mathbf{x} - \mathbf{q} - \Psi(\mathbf{q})] - 1, \quad (\text{D2})$$

whose Fourier transform is [155]

$$\delta(\mathbf{k}) = \int d^3q e^{-i\mathbf{k}\cdot\mathbf{q}} [e^{-i\mathbf{k}\cdot\Psi(\mathbf{q})} - 1]. \quad (\text{D3})$$

This expression can be used to derive the observable power spectrum in Eulerian space [80,155,156],

$$P(\mathbf{k}) = \int d^3\Delta_{12} e^{-i\mathbf{k}\cdot\Delta_{12}} \{ \langle e^{-i\mathbf{k}\cdot[\Psi(\mathbf{q}_1) - \Psi(\mathbf{q}_2)]} \rangle - 1 \}, \quad (\text{D4})$$

where $\Delta_{ij} \equiv \mathbf{q}_i - \mathbf{q}_j$, and the expectation value only depends on the separation Δ_{12} due to homogeneity. Likewise, the bispectrum can be written as [84]

$$B(k_1, k_2, k_3) = \int d^3\Delta_{12} \int d^3\Delta_{13} e^{-i\mathbf{k}\cdot(\Delta_{12} + \Delta_{13})} \times \{ \langle e^{-i\mathbf{k}_2\cdot[\Psi(\mathbf{q}_1) - \Psi(\mathbf{q}_2)] - i\mathbf{k}_3\cdot[\Psi(\mathbf{q}_1) - \Psi(\mathbf{q}_3)]} \rangle - 1 \}; \quad (\text{D5})$$

also here the expectation value only depends on the separations Δ_{12} , Δ_{13} . Equations (D4), (D5) relate the observable (Eulerian) density polyspectra to the displacement field Ψ . We follow Ref. [80] and use the cumulant expansion theorem:

$$\langle e^{-iX} \rangle = \exp \left[\sum_{N=1}^{\infty} \frac{(-i)^N}{N!} \langle X^N \rangle_c \right], \quad (\text{D6})$$

where $\langle X^N \rangle_c$ represents the cumulant of the random variable X [56]. By applying the cumulant expansion to Eqs. (D4), (D5), and expanding the powers of N with the binomial theorem, two types of terms are obtained: those depending on Ψ at one point, and those depending on Ψ at two different points. In Refs. [80,84] it was demonstrated that, if both sets of terms are expanded to the same order, the LPT results are identical to those obtained in SPT for both the power spectrum and bispectrum. However in Ref. [80] it was found that, for large separations, the terms

depending on Ψ at one point are much larger than those depending on Ψ at two points, so that the first set of terms should be kept as it is, and only the second set should be expanded. This renormalized approach is called RLPT.

In order to derive explicit expressions for the matter power spectrum and bispectrum, we need to expand the displacement field as a function of the matter overdensity δ . The displacement field follows the equation of motion

$$\frac{d^2\Psi}{dt^2} + 2H \frac{d\Psi}{dt} = -\nabla_x \phi[\mathbf{q} + \Psi(\mathbf{q})], \quad (\text{D7})$$

where ϕ is the gravitational potential. The polyspectra of Ψ can be calculated by expanding it as a series of the density field:

$$\Psi^{(n)}(\mathbf{p}) = \frac{iD^n}{n!} \int \frac{d^3p_1}{(2\pi)^3} \dots \frac{d^3p_n}{(2\pi)^3} \delta_D \left(\sum_{j=1}^n \mathbf{p}_j - \mathbf{p} \right) \times \mathbf{L}^{(n)}(\mathbf{p}_1, \dots, \mathbf{p}_n) \delta_1(\mathbf{p}_1) \dots \delta_1(\mathbf{p}_n), \quad (\text{D8})$$

where δ_1 indicates the linear density perturbation at present times, and the perturbative kernels $\mathbf{L}^{(n)}$ are the analogues of the SPT kernels F_n , G_n , and are also obtained from a recursion relation [56].

Using this expansion leads to the RLPT power spectrum of Eq. (37) [80], and to the bispectrum of Eq. (38) [84].

APPENDIX E: GOING TO TWO LOOPS

We outline in this appendix the two-loop calculations for the matter power spectrum and bispectrum in perturbation theory.

1. Two-loop power spectrum in perturbation theories

a. SPT two-loop terms

The two-loop power spectrum in SPT can be expressed as [65,135]

$$P_{2\text{-loop}} = P_{15} + P_{24} + P_{33}^{(I)} + P_{33}^{(II)}, \quad (\text{E1})$$

where the four terms are

$$P_{51}(k) = D^6(z) \int \frac{d^3p}{(2\pi)^3} \int \frac{d^3q}{(2\pi)^3} 30F_5^{(s)}(\mathbf{k}, \mathbf{q}, -\mathbf{q}, \mathbf{p}, -\mathbf{p}) P_{\text{lin}}(k) P_{\text{lin}}(q) P_{\text{lin}}(p) \quad (\text{E2})$$

$$P_{42}(k) = D^6(z) \int \frac{d^3p}{(2\pi)^3} \int \frac{d^3q}{(2\pi)^3} 24F_2^{(s)}(\mathbf{q}, \mathbf{k} - \mathbf{q}) F_4^{(s)}(-\mathbf{q}, \mathbf{q} - \mathbf{k}, \mathbf{p}, -\mathbf{p}) P_{\text{lin}}(q) P_{\text{lin}}(p) P_{\text{lin}}(|\mathbf{k} - \mathbf{q}|) \quad (\text{E3})$$

$$P_{33}^{(I)}(k) = D^6(z) \int \frac{d^3p}{(2\pi)^3} \int \frac{d^3q}{(2\pi)^3} 9F_3^{(s)}(-\mathbf{k}, \mathbf{p}, -\mathbf{p}) F_3^{(s)}(\mathbf{k}, \mathbf{q}, -\mathbf{q}) P_{\text{lin}}(k) P_{\text{lin}}(q) P_{\text{lin}}(p) \quad (\text{E4})$$

$$P_{33}^{(II)}(k) = D^6(z) \int \frac{d^3p}{(2\pi)^3} \int \frac{d^3q}{(2\pi)^3} 6F_3^{(s)}(\mathbf{q}, \mathbf{p}, \mathbf{k} - \mathbf{q} - \mathbf{p}) F_3^{(s)}(-\mathbf{q}, -\mathbf{p}, -\mathbf{k} + \mathbf{q} + \mathbf{p}) P_{\text{lin}}(q) P_{\text{lin}}(p) P_{\text{lin}}(|\mathbf{k} - \mathbf{q} - \mathbf{p}|). \quad (\text{E5})$$

b. Two-loop RPT power spectrum

In the case of RPT, the two-loop calculation is simplified considerably, because only one of the terms from Eqs. (E2)–(E5) appears in this theory [75]. Thus, the two-loop power spectrum can be calculated by modifying the expression of $P_{33}^{(II)}$ as follows:

$$P_{2\text{-loop}}^{\text{MPTbreeze}}(k, z) = P_{33}^{(II)}(k, z) \exp[2f(k)D^2(z)]. \quad (\text{E6})$$

The IR-safe evaluation of this integral has been described in Ref. [136].

2. Two-loop bispectrum in perturbation theories

In SPT, the loop expansion is obtained by considering the expansion from Eq. (A16) up to the relevant order, together with the integral expression (A18) and then using Wick's theorem. For two loops, there are terms up to $F_6^{(s)}$ in the kernels, which make the numerical evaluation cumbersome. In RPT, the number of terms is however drastically reduced.

a. Two-loop RPT calculation

The tree-level and one-loop bispectrum in this theory have been discussed above, and hence we proceed directly to the two-loop terms. The generating function for the RPT bispectrum is given by Eq. (59) of Ref. [73]. At two loops, using the notation from Ref. [73], we need to take $r + s + t = 4$. As only one of these numbers can be 0, there are only three choices for r , s and t (plus permutations) giving nonvanishing contributions, which we will treat in turn:

- (a) $r = 3, s = 1, t = 0$ (+5 perms.)
- (b) $r = 2, s = 2, t = 0$ (+2 perms.)
- (c) $r = 1, s = 1, t = 2$ (+2 perms.)

In all three cases, the expressions involved will depend on the functions $\Gamma^{(n)}$ only up to $\Gamma^{(4)}$, which in turn can be expressed in terms of the corresponding kernel functions F_n^s through Eq. (C15). Even though the expressions that we obtained for the MPTBREEZE two-loop bispectra are relatively simple, they cannot be integrated directly because they have various poles where at least one of the arguments of $F_n^{(s)}$ vanishes. However, we know that the divergences between the various terms must cancel exactly after performing the integration, but numerically this is an issue because the divergent parts are expected to be much bigger than the finite result and thus the numerical result may not be reliable. In order to solve this problem, we will use the methods developed in Refs. [136–138]. Compared to SPT, where some of the terms involve the kernel $F_6^{(s)}$, the expressions appearing in this method represent a significant simplification.

We note that from Eq. (C15) all expressions will have a prefactor,

$$D^8(z) \exp[(f(k_1) + f(k_2) + f(k_3))D^2(z)], \quad (\text{E7})$$

and therefore in the following paragraphs we will omit this factor because it does not affect the calculation. We will denote the three integrals by B_a , B_b and B_c and the integrands with corresponding lowercase letters. Then the final two-loop MPTBREEZE bispectrum is

$$B_{2\text{-loop}}^{\text{MPTbreeze}}(k_1, k_2, k_3, z) = D^8(z) \exp[(f(k_1) + f(k_2) + f(k_3))D^2(z)] [B_a(k_1, k_2, k_3) + B_b(k_1, k_2, k_3) + B_c(k_1, k_2, k_3)]. \quad (\text{E8})$$

The expressions for the three bispectra are as follows:

$$B_a(k_1, k_2, k_3) = 24 \int \frac{d^3 q_1}{(2\pi)^3} \int \frac{d^3 q_2}{(2\pi)^3} F_4^{(s)}(\mathbf{q}_1, \mathbf{q}_2, -\mathbf{k}_3 - \mathbf{q}_1 - \mathbf{q}_2, -\mathbf{k}_2) F_3^{(s)}(-\mathbf{q}_1, -\mathbf{q}_2, \mathbf{k}_3 + \mathbf{q}_1 + \mathbf{q}_2) \times P_{\text{lin}}(q_1) P_{\text{lin}}(q_2) P_{\text{lin}}(k_3) P_{\text{lin}}(|\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{k}_3|) \quad (\text{E9})$$

$$B_b(k_1, k_2, k_3) = 24 \int \frac{d^3 q_1}{(2\pi)^3} \int \frac{d^3 q_2}{(2\pi)^3} F_4^{(s)}(\mathbf{q}_1, \mathbf{k}_3 - \mathbf{q}_1, \mathbf{q}_2, -\mathbf{k}_2 - \mathbf{q}_2) F_2^{(s)}(-\mathbf{q}_2, \mathbf{k}_2 + \mathbf{q}_2) F_2^{(s)}(-\mathbf{q}_1, \mathbf{k}_3 + \mathbf{q}_1) \times P_{\text{lin}}(q_1) P_{\text{lin}}(q_2) P_{\text{lin}}(|\mathbf{q}_1 + \mathbf{k}_3|) P_{\text{lin}}(|\mathbf{q}_2 + \mathbf{k}_2|) \quad (\text{E10})$$

$$B_c(k_1, k_2, k_3) = 36 \int \frac{d^3 q_1}{(2\pi)^3} \int \frac{d^3 q_2}{(2\pi)^3} F_3^{(s)}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2) F_2^{(s)}(-\mathbf{k}_1 + \mathbf{q}_1 + \mathbf{q}_2, -\mathbf{k}_3 - \mathbf{q}_1 - \mathbf{q}_2) \times F_3^{(s)}(\mathbf{k}_3 + \mathbf{q}_1 + \mathbf{q}_2, -\mathbf{q}_1, -\mathbf{q}_2) P_{\text{lin}}(q_1) P_{\text{lin}}(q_2) P_{\text{lin}}(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|) P_{\text{lin}}(|\mathbf{k}_3 + \mathbf{q}_1 + \mathbf{q}_2|). \quad (\text{E11})$$

We will treat each of them in turn and show how to remove the singularities before the integration.

1. B_a

The expression for B_a has singularities when $q_1 = 0$, $q_2 = 0$ and $\mathbf{q}_1 + \mathbf{q}_2 = -\mathbf{k}_3$. By considering the variable $\mathbf{q}_3 = -\mathbf{q}_1 - \mathbf{q}_2 - \mathbf{k}_3$, B_a can be reexpressed in terms of a triple integral by adding a Dirac-delta function:

$$B_a(k_1, k_2, k_3) = 24 \int \frac{d^3 q_1}{(2\pi)^3} \int \frac{d^3 q_2}{(2\pi)^3} \int \frac{d^3 q_3}{(2\pi)^3} F_4^{(s)}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) F_3^{(s)}(-\mathbf{q}_1, -\mathbf{q}_2, -\mathbf{q}_3) \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3 + \mathbf{k}_3) \times P_{\text{lin}}(q_1) P_{\text{lin}}(q_2) P_{\text{lin}}(k_3) P_{\text{lin}}(q_3). \quad (\text{E12})$$

This expression is now completely symmetric in $\mathbf{q}_1 \leftrightarrow \mathbf{q}_2 \leftrightarrow \mathbf{q}_3$ and hence all ordering of the magnitudes of these three wave vectors are equivalent after a suitable relabeling of the variables. As there are six possible permutations of q_1 , q_2 and q_3 ,

$$B_a = \int \int \int \frac{d^3 q_1 d^3 q_2 d^3 q_3}{(2\pi)^9} b_b 6\Theta(q_3 - q_2)\Theta(q_2 - q_1). \quad (\text{E13})$$

Hence, the delta function and \mathbf{q}_3 can now be eliminated and the final expression is obtained:

$$B_a(k_1, k_2, k_3) = 24 \int \frac{d^3 q_1}{(2\pi)^3} \int \frac{d^3 q_2}{(2\pi)^3} F_4^{(s)}(\mathbf{q}_1, \mathbf{q}_2, -\mathbf{k}_3 - \mathbf{q}_1 - \mathbf{q}_2, -\mathbf{k}_2) F_3^{(s)}(-\mathbf{q}_1, -\mathbf{q}_2, \mathbf{k}_3 + \mathbf{q}_1 + \mathbf{q}_2) \times P_{\text{lin}}(q_1) P_{\text{lin}}(q_2) P_{\text{lin}}(k_3) P_{\text{lin}}(|\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{k}_3|) \times 6\Theta(|\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{k}_3| - q_2)\Theta(q_2 - q_1). \quad (\text{E14})$$

The expression of Eq. (E14) has a leading divergence when $q_1 = q_2 = 0$ and a subleading divergence when $q_1 = 0$, \mathbf{q}_2 fixed. The two divergences corresponding to $\mathbf{q}_1 + \mathbf{q}_2 = -\mathbf{k}_3$ and $q_2 = 0$, at fixed \mathbf{q}_1 , have disappeared because the Heaviside functions evaluate to 0 in those limits. In order to eliminate all divergences at the integrand level, we can also symmetrize in $\mathbf{q}_{1,2} \leftrightarrow -\mathbf{q}_{1,2}$:

$$b_a(\mathbf{q}_1, \mathbf{q}_2) \rightarrow \frac{1}{4} [b_a(\mathbf{q}_1, \mathbf{q}_2) + b_a(-\mathbf{q}_1, \mathbf{q}_2) + b_a(\mathbf{q}_1, -\mathbf{q}_2) + b_a(-\mathbf{q}_1, -\mathbf{q}_2)]. \quad (\text{E15})$$

We will use this symmetrization for the b and c terms as well.

2. B_b

The B_b term has divergences for $q_1 = 0$, $\mathbf{q}_1 = -\mathbf{k}_3$, $q_2 = 0$ and $\mathbf{q}_2 = -\mathbf{k}_1$. We note that b_b is symmetric under the transformations $\mathbf{q}_1 \leftrightarrow -\mathbf{k}_3 - \mathbf{q}_1$ and $\mathbf{q}_2 \leftrightarrow -\mathbf{k}_2 - \mathbf{q}_2$. We can exploit the three symmetries that now appear in the integrand by restricting the integration region to $q_1 < |\mathbf{k}_3 + \mathbf{q}_1|$ and $q_2 < |\mathbf{k}_1 + \mathbf{q}_2|$ and introducing two Heaviside functions and a factor of 2^2 :

$$B_b(k_1, k_2, k_3) = 24 \int \frac{d^3 q_1}{(2\pi)^3} \int \frac{d^3 q_2}{(2\pi)^3} F_4^{(s)}(\mathbf{q}_1, \mathbf{k}_3 - \mathbf{q}_1, \mathbf{q}_2, -\mathbf{k}_2 - \mathbf{q}_2) F_2^{(s)}(-\mathbf{q}_2, \mathbf{k}_2 + \mathbf{q}_2) F_2^{(s)}(-\mathbf{q}_1, \mathbf{k}_3 + \mathbf{q}_1) \times P_{\text{lin}}(q_1) P_{\text{lin}}(q_2) P_{\text{lin}}(|\mathbf{q}_1 + \mathbf{k}_3|) P_{\text{lin}}(|\mathbf{q}_2 + \mathbf{k}_2|) \times 4\Theta(|\mathbf{k}_2 + \mathbf{q}_2| - q_2)\Theta(|\mathbf{k}_3 + \mathbf{q}_1| - q_1). \quad (\text{E16})$$

This expression is not symmetric in $\mathbf{q}_1 \leftrightarrow \mathbf{q}_2$, but we can symmetrize it by symmetrizing the whole integrand (including the delta functions):

$$b_b(\mathbf{q}_1, \mathbf{q}_2) \rightarrow \frac{1}{2} [b_b(\mathbf{q}_1, \mathbf{q}_2) + b_b(\mathbf{q}_2, \mathbf{q}_1)]. \quad (\text{E17})$$

After the symmetrization, we aim to restrict the integration range to $q_1 < q_2$, and we achieve this by adding an additional Θ -function, thus obtaining the final answer:

$$\begin{aligned}
B_b(k_1, k_2, k_3) = & 24 \int \frac{d^3 q_1}{(2\pi)^3} \int \frac{d^3 q_2}{(2\pi)^3} [F_4^{(s)}(\mathbf{q}_1, \mathbf{k}_3 - \mathbf{q}_1, \mathbf{q}_2, -\mathbf{k}_2 - \mathbf{q}_2) F_2^{(s)}(-\mathbf{q}_2, \mathbf{k}_2 + \mathbf{q}_2) F_2^{(s)}(-\mathbf{q}_1, \mathbf{k}_3 + \mathbf{q}_1) \\
& \times P_{\text{lin}}(q_1) P_{\text{lin}}(q_2) P_{\text{lin}}(|\mathbf{q}_1 + \mathbf{k}_3|) P_{\text{lin}}(|\mathbf{q}_2 + \mathbf{k}_2|) \times 4\Theta(|\mathbf{k}_2 + \mathbf{q}_2| - q_2) \Theta(|\mathbf{k}_3 + \mathbf{q}_1| - q_1) \\
& + F_4^{(s)}(\mathbf{q}_2, \mathbf{k}_3 - \mathbf{q}_2, \mathbf{q}_1, -\mathbf{k}_2 - \mathbf{q}_1) F_2^{(s)}(-\mathbf{q}_1, \mathbf{k}_2 + \mathbf{q}_1) F_2^{(s)}(-\mathbf{q}_2, \mathbf{k}_3 + \mathbf{q}_2) \\
& \times P_{\text{lin}}(q_1) P_{\text{lin}}(q_2) P_{\text{lin}}(|\mathbf{q}_2 + \mathbf{k}_3|) P_{\text{lin}}(|\mathbf{q}_1 + \mathbf{k}_2|) \times 4\Theta(|\mathbf{k}_2 + \mathbf{q}_1| - q_1) \Theta(|\mathbf{k}_3 + \mathbf{q}_2| - q_2)] \Theta(q_2 - q_1).
\end{aligned} \tag{E18}$$

Hence all the leading and subleading divergences have been moved to $q_1 = q_2 = 0$ and $q_1 = 0$, at \mathbf{q}_2 fixed. For all the other poles in the kernels, the Heaviside functions vanish.

3. B_c

The expression for B_c has only one direct symmetry $\mathbf{q}_1 \leftrightarrow \mathbf{q}_2$, but this is not enough. Therefore we introduce the following notation,

$$\mathbf{q}_3 = \mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2 \tag{E19}$$

$$\mathbf{q}_4 = \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2, \tag{E20}$$

and we then introduce two additional integrations and two Dirac delta functions. The integral becomes

$$\begin{aligned}
B_c(k_1, k_2, k_3) = & 36 \int \int \int \int \frac{d^3 q_1 d^3 q_2 d^3 q_3 d^3 q_4}{(2\pi)^{12}} F_3^{(s)}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) F_2^{(s)}(-\mathbf{q}_3, \mathbf{q}_4) F_3^{(s)}(-\mathbf{q}_4, -\mathbf{q}_1, -\mathbf{q}_2) \\
& \times P_{\text{lin}}(q_1) P_{\text{lin}}(q_2) P_{\text{lin}}(q_3) P_{\text{lin}}(q_4) \delta_D(\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2 - \mathbf{q}_3) \delta_D(\mathbf{k}_2 + \mathbf{q}_3 - \mathbf{q}_4).
\end{aligned} \tag{E21}$$

This expression is already symmetric in $\mathbf{q}_1 \leftrightarrow \mathbf{q}_2$, and we symmetrize it in all the other variables, obtaining 12 possible permutations and a fully symmetric expression. We can now introduce an ordering of the four variables in terms of their magnitude (e.g. $q_4 \geq q_3 \geq q_2 \geq q_1$), knowing that all the other orderings can be obtained by a suitable relabeling of the variables. There are $4! = 24$ permutations of the four variables and, keeping only one of the permutations, we need to multiply it by the following product of Heaviside functions:

$$24\Theta(q_4 - q_3)\Theta(q_3 - q_2)\Theta(q_2 - q_1). \tag{E22}$$

There are now 12 summands, and each of them involves integrals over \mathbf{q}_1 , \mathbf{q}_2 , \mathbf{q}_3 and \mathbf{q}_4 , three Heaviside functions and two delta functions. The aim is now to perform two of the integrations, in order to eliminate the delta functions. In ten of the terms, it turns out that it is possible integrate over \mathbf{q}_3 and \mathbf{q}_4 . In the other two, \mathbf{q}_3 and \mathbf{q}_4 appear in the same combination in both delta functions. For those terms we integrate over \mathbf{q}_2 and \mathbf{q}_4 , and then relabel $\mathbf{q}_3 \rightarrow \mathbf{q}_2$. The final expression that we thus obtain only has divergences for $q_1 = q_2 = 0$ and $q_1 = 0$, with \mathbf{q}_2 fixed as required:

$$\begin{aligned}
b_c(k_1, k_2, k_3) = & 36(2F_2^{(s)}(\mathbf{k}_2 - \mathbf{q}_1, \mathbf{q}_1) F_3^{(s)}(-\mathbf{q}_1, -\mathbf{q}_2, -\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2) F_3^{(s)}(\mathbf{q}_1 - \mathbf{k}_2, \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2) \\
& \times P_{\text{lin}}(q_1) P_{\text{lin}}(|\mathbf{q}_1 - \mathbf{k}_2|) P_{\text{lin}}(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2|) P_{\text{lin}}(q_2) \\
& \times \Theta(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2| - |\mathbf{q}_1 - \mathbf{k}_2|) \Theta(|\mathbf{q}_1 - \mathbf{k}_2| - q_2) \Theta(q_2 - q_1) \\
& + 2F_2^{(s)}(-\mathbf{q}_1, \mathbf{k}_2 + \mathbf{q}_1) F_3^{(s)}(-\mathbf{k}_2 - \mathbf{q}_1, -\mathbf{q}_2, -\mathbf{k}_1 + \mathbf{q}_1 + \mathbf{q}_2) F_3^{(s)}(\mathbf{q}_1, \mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2) P_{\text{lin}}(q_1) P_{\text{lin}}(|\mathbf{k}_2 + \mathbf{q}_1|) \\
& \times P_{\text{lin}}(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|) P_{\text{lin}}(q_2) \Theta(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2| - |\mathbf{k}_2 + \mathbf{q}_1|) \Theta(|\mathbf{k}_2 + \mathbf{q}_1| - q_2) \Theta(q_2 - q_1) \\
& + 2F_2^{(s)}(-\mathbf{q}_1, \mathbf{k}_2 + \mathbf{q}_1) F_3^{(s)}(-\mathbf{k}_2 - \mathbf{q}_1, -\mathbf{q}_2, -\mathbf{k}_1 + \mathbf{q}_1 + \mathbf{q}_2) F_3^{(s)}(\mathbf{q}_1, \mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2) \\
& \times P_{\text{lin}}(q_1) P_{\text{lin}}(|\mathbf{k}_2 + \mathbf{q}_1|) P_{\text{lin}}(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|) P_{\text{lin}}(q_2) \Theta(|\mathbf{k}_2 + \mathbf{q}_1| - |\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|) \Theta(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2| - q_2) \Theta(q_2 - q_1)
\end{aligned}$$

$$\begin{aligned}
& + 2F_2^{(s)}(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2, -\mathbf{k}_1 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(-\mathbf{q}_1, -\mathbf{q}_2, -\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(\mathbf{q}_1, \mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2) \\
& \times P_{\text{lin}}(q_1)P_{\text{lin}}(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(q_2) \\
& \times \Theta(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2| - |\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|)\Theta(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2| - q_2)\Theta(q_2 - q_1) \\
& + 2F_2^{(s)}(\mathbf{k}_2 - \mathbf{q}_1, \mathbf{q}_1)F_3^{(s)}(-\mathbf{q}_1, -\mathbf{q}_2, -\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(\mathbf{q}_1 - \mathbf{k}_2, \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2) \\
& \times P_{\text{lin}}(q_1)P_{\text{lin}}(|\mathbf{q}_1 - \mathbf{k}_2|)P_{\text{lin}}(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(q_2) \\
& \times \Theta(|\mathbf{q}_1 - \mathbf{k}_2| - |\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2|)\Theta(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2| - q_2)\Theta(q_2 - q_1) \\
& + 2F_2^{(s)}(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2, -\mathbf{k}_1 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(-\mathbf{q}_1, -\mathbf{q}_2, -\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(\mathbf{q}_1, \mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2) \\
& \times P_{\text{lin}}(q_1)P_{\text{lin}}(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(q_2) \\
& \times \Theta(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2| - |\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2|)\Theta(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2| - q_2)\Theta(q_2 - q_1) \\
& + 2F_2^{(s)}(\mathbf{k}_2 - \mathbf{q}_2, \mathbf{q}_2)F_3^{(s)}(-\mathbf{q}_1, -\mathbf{q}_2, -\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(\mathbf{q}_1, \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2 - \mathbf{k}_2) \\
& \times P_{\text{lin}}(q_1)P_{\text{lin}}(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(q_2)P_{\text{lin}}(|\mathbf{q}_2 - \mathbf{k}_2|) \\
& \times \Theta(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2| - q_2)\Theta(|\mathbf{q}_2 - \mathbf{k}_2| - |\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2|)\Theta(q_2 - q_1) \\
& + 2F_2^{(s)}(\mathbf{k}_2 - \mathbf{q}_2, \mathbf{q}_2)F_3^{(s)}(-\mathbf{q}_1, -\mathbf{q}_2, -\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(\mathbf{q}_1, \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2 - \mathbf{k}_2)P_{\text{lin}}(q_1) \\
& \times P_{\text{lin}}(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(q_2)P_{\text{lin}}(|\mathbf{q}_2 - \mathbf{k}_2|)\Theta(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2| - |\mathbf{q}_2 - \mathbf{k}_2|)\Theta(|\mathbf{q}_2 - \mathbf{k}_2| - q_2)\Theta(q_2 - q_1) \\
& + 2F_2^{(s)}(-\mathbf{q}_2, \mathbf{k}_2 + \mathbf{q}_2)F_3^{(s)}(-\mathbf{q}_1, -\mathbf{k}_2 - \mathbf{q}_2, -\mathbf{k}_1 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(\mathbf{q}_1, \mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2)P_{\text{lin}}(q_1) \\
& \times P_{\text{lin}}(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(q_2)P_{\text{lin}}(|\mathbf{k}_2 + \mathbf{q}_2|)\Theta(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2| - q_2)\Theta(|\mathbf{k}_2 + \mathbf{q}_2| - |\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|)\Theta(q_2 - q_1) \\
& + 2F_2^{(s)}(-\mathbf{q}_2, \mathbf{k}_2 + \mathbf{q}_2)F_3^{(s)}(-\mathbf{q}_1, -\mathbf{k}_2 - \mathbf{q}_2, -\mathbf{k}_1 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(\mathbf{q}_1, \mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2) \\
& \times P_{\text{lin}}(q_1)P_{\text{lin}}(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(q_2)P_{\text{lin}}(|\mathbf{k}_2 + \mathbf{q}_2|)\Theta(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2| - |\mathbf{k}_2 + \mathbf{q}_2|)\Theta(|\mathbf{k}_2 + \mathbf{q}_2| - q_2)\Theta(q_2 - q_1) \\
& + 2F_2^{(s)}(\mathbf{k}_2 - \mathbf{q}_1, \mathbf{q}_1)F_3^{(s)}(-\mathbf{q}_1, -\mathbf{q}_2, -\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(\mathbf{q}_1 - \mathbf{k}_2, \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2)P_{\text{lin}}(q_1) \\
& \times P_{\text{lin}}(|\mathbf{q}_1 - \mathbf{k}_2|)P_{\text{lin}}(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(q_2)\Theta(|\mathbf{q}_1 - \mathbf{k}_2| - q_1)\Theta(|\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1 - \mathbf{q}_2| - q_2)\Theta(q_2 - |\mathbf{q}_1 - \mathbf{k}_2|) \\
& + 2F_2^{(s)}(-\mathbf{q}_1, \mathbf{k}_2 + \mathbf{q}_1)F_3^{(s)}(-\mathbf{k}_2 - \mathbf{q}_1, -\mathbf{q}_2, -\mathbf{k}_1 + \mathbf{q}_1 + \mathbf{q}_2)F_3^{(s)}(\mathbf{q}_1, \mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2, \mathbf{q}_2) \\
& \times P_{\text{lin}}(q_1)P_{\text{lin}}(|\mathbf{k}_2 + \mathbf{q}_1|)P_{\text{lin}}(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2|)P_{\text{lin}}(q_2)\Theta(|\mathbf{k}_2 + \mathbf{q}_1| - q_1)\Theta(|\mathbf{k}_1 - \mathbf{q}_1 - \mathbf{q}_2| - q_2)\Theta(q_2 - |\mathbf{k}_2 + \mathbf{q}_1|).
\end{aligned} \tag{E23}$$

The three long expressions can be added together with their corresponding permutations to obtain the final two-loop result, which is then free of any divergences before the integration.

APPENDIX F: NINE-PARAMETER MODEL

The tree-level prediction is the simplest model for the bispectrum. As its simpler counterpart, the linear power spectrum, it is only accurate for very low values of the wave number. A simple improvement over the tree level would be to substitute the linear with the nonlinear power spectrum in Eq. (24), e.g. as calculated with the HALOFIT method [106,131]; this result can then be tuned further by modifying the kernel $F_2^{(s)}$ in order to better fit simulations. This idea has been proposed in Ref. [130] and here we discuss a more elaborate version of it, which fits N -body simulation better, introduced by Ref. [133]. Each of the three terms of the kernel

[Eq. (16)] is modified by a multiplicative function, as follows:

$$\begin{aligned}
F_2^{\text{eff}}(\mathbf{q}_1, \mathbf{q}_2) &= \frac{5}{7}a(n_1, q_1)a(n_2, q_2) \\
&+ \frac{1}{2} \frac{\mathbf{q}_1 \cdot \mathbf{q}_2}{q_1 q_2} \left(\frac{q_1}{q_2} + \frac{q_2}{q_1} \right) b(n_1, q_1)b(n_2, q_2) \\
&+ \frac{2}{7} \frac{(\mathbf{q}_1 \cdot \mathbf{q}_2)^2}{q_1^2 q_2^2} c(n_1, q_1)c(n_2, q_2), \tag{F1}
\end{aligned}$$

such that the bispectrum can be expressed as

$$\begin{aligned}
B(k_1, k_2, k_3) &= F_2^{\text{eff}}(\mathbf{k}_1, \mathbf{k}_2)P_{\text{Halofit}}(k_1)P_{\text{Halofit}}(k_2) \\
&+ 2 \text{ perms.} \tag{F2}
\end{aligned}$$

The functions $a(n, k)$, $b(n, k)$, $c(n, k)$ can be expressed in terms of nine coefficients that are determined numerically (a_1, \dots, a_n) by fitting N -body simulations:

$$a(n, k) = \frac{1 + \sigma_8^{a_6}(z)[0.7Q_3(n)]^{1/2}(qa_1)^{n+a_2}}{1 + (qa_1)^{n+a_2}} \quad (\text{F3})$$

$$b(n, k) = \frac{1 + 0.2a_3(n+3)(qa_7)^{n+3+a_8}}{1 + (qa_7)^{n+3.5+a_8}} \quad (\text{F4})$$

$$c(n, k) = \frac{1 + 4.5a_4/[1.5 + (n+3)^4](qa_5)^{n+3+a_9}}{1 + (qa_5)^{n+3.5+a_9}}. \quad (\text{F5})$$

The functions $n(k)$, $Q_3(n)$ and q are defined as

$$n(k) = \frac{d \log P_{\text{lin}}(k)}{d \log k} \quad (\text{F6})$$

$$Q_3(n) = \frac{4 - 2^n}{1 + 2^{n+1}} \quad (\text{F7})$$

$$q = \frac{k}{k_{\text{NL}}}, \quad (\text{F8})$$

where k_{NL} is the nonlinear scale defined as the solution to the equation

$$\frac{k_{\text{NL}}^3 P_{\text{lin}}(k_{\text{NL}})}{2\pi^2} = 1. \quad (\text{F9})$$

The parameters have been calibrated to give a maximum of 10% error in the matter bispectrum for $z \in [0, 1.5]$ and $k \leq 0.4 h/\text{Mpc}$ and they are $a_1 = 0.484$, $a_2 = 3.740$, $a_3 = -0.849$, $a_4 = 0.392$, $a_5 = 1.013$, $a_6 = -0.575$, $a_7 = 0.128$, $a_8 = -0.722$ and $a_9 = -0.926$.

APPENDIX G: INGREDIENTS OF THE HALO MODEL

We assume that dark matter halos are virialized spheres of mass $m = \frac{4\pi}{3} R_v^3 \Delta_v \bar{\rho}$, where R_v is the virial radius. The virial overdensity is $\Delta_v = 18\pi^2 \approx 180$ in matter domination and it depends weakly on cosmology; we fix $\Delta_v = 200$ in our model in order to match the assumptions of the numerical fits for the mass function and concentration. The initial overdensity of spherically collapsed objects, extrapolated to the present time using linear theory, is $\delta_c = \frac{3}{5} (\frac{3\pi}{2})^{2/3} \approx 1.686$ [157].

1. Halo profile

We use the Navarro-Frenk-White (NFW) profile [158], which can be expressed in terms of two parameters, $\rho_s(m)$ and $r_s(m)$, describing the scaling radius and associated density, where the profile slope changes:

$$\rho(r|m) = \frac{\rho_s}{r/r_s(1+r/r_s)^2}. \quad (\text{G1})$$

The NFW profile can be recast in terms of the concentration $c \equiv R_v/r_s$, which can be calibrated from N -body simulations [159]. Hence,

$$\rho_s = \frac{\Delta_v \bar{\rho}}{3} \frac{c^3}{\log(1+c) - \frac{1}{1+c}}; \quad r_s = \left(\frac{3m}{4\pi c^3 \Delta_v \bar{\rho}} \right)^{1/3}. \quad (\text{G2})$$

Substituting these definitions into Eq. (G1) and applying a Fourier transformation, we obtain [160]

$$u(k|m) = \frac{4\pi\rho_s r_s^3}{m} \left\{ \sin(kr_s) [\text{Si}((1+c)kr_s) - \text{Si}(kr_s)] - \frac{\sin(kr_s)}{(1+c)kr_s} + \cos(kr_s) [\text{Ci}((1+c)kr_s) - \text{Ci}(kr_s)] \right\}, \quad (\text{G3})$$

where $\text{Si}(x)$ and $\text{Ci}(x)$ are the sine and cosine integral functions. We use the fitting function to the concentration obtained from the Bolshoi simulation [161],

$$c(m, z) = 9.2\kappa(z)D(z)^{1.3} \left(\frac{m}{10^{12}h^{-1}M} \right)^{-0.09} \times \left[1 + 0.013 \left(\frac{m}{10^{12}h^{-1}M} D(z)^{-\frac{1.3}{0.09}} \right)^{0.25} \right] \quad (\text{G4})$$

with $\kappa(z) = 1.26$ at $z = 0$ and $\kappa(z) = 0.96$ at $z \geq 1$, as in Ref. [162].

2. Halo mass function

The number density of halos of mass m and redshift z is given by $n(m, z)$, which can be written as [163]

$$\frac{m}{\bar{\rho}} n(m, z) dm = f(\nu) d\nu. \quad (\text{G5})$$

Here the peak height $\nu \equiv \delta_c^2/\sigma^2(m, z)$ is obtained from the variance of the linear density field filtered with a top-hat function in Fourier space $W_f(x) = (3/x^3)[\sin(x) - x\cos(x)]$ on the scale $R_f = [3m/(4\pi\bar{\rho})]^{1/3}$:

$$\sigma^2(m, z) = \frac{D^2(z)}{2\pi^2} \int_0^\infty dk k^2 W_f^2(kR_f) P_{\text{lin}}(k), \quad (\text{G6})$$

where $P_{\text{lin}}(k)$ is the linear matter power spectrum and $D(z)$ is the linear growth function.

Different choices for the mass function $f(\nu)$ are possible. The simplest form was derived by Press and Schechter [163] analytically assuming spherical collapse, finding

$$f^{\text{PS}}(\nu) = \sqrt{2\nu/\pi} e^{-\nu/2}. \quad (\text{G7})$$

This simple model only matches the results from N -body simulations within a factor of two (see e.g. Ref. [164]); more elaborated models calibrated on N -body simulations

include those by Refs. [165,166]. We use here the numerical fit by Tinker *et al.* [167], where

$$f^T(\nu) = \alpha[1 + (\beta\nu)^{-2\phi}]\nu^{2\eta} \exp(-\gamma\nu^2/2); \quad (\text{G8})$$

the coefficients β , γ , η and ϕ have the following redshift dependence,

$$\begin{aligned} \beta &= \beta_0(1+z)^{0.20}; & \gamma &= \gamma_0(1+z)^{-0.01} \\ \eta &= \eta_0(1+z)^{0.27}; & \phi &= \phi_0(1+z)^{-0.08} \end{aligned} \quad (\text{G9})$$

with $\alpha = 0.368$, $\beta_0 = 0.589$, $\gamma_0 = 0.864$, $\eta_0 = -0.243$ and $\phi_0 = -0.729$.

3. Halo clustering

Finally, we need a model for the clustering of the dark matter halo centers, i.e. P_h and B_h . Under the simplest assumption of local deterministic bias, the halo overdensity in real space can be expressed as a Taylor expansion of the dark matter overdensity field, where the coefficients are the bias parameters [168], assuming both fields have been smoothed on a relatively large scale R . This expression also holds in Fourier space, where the first constant term b_0 is relegated to $k = 0$ and is thus irrelevant.

If we calculate the matter power spectra using SPT at tree level, we obtain [168]

$$\begin{aligned} P_h(k|m_1, m_2) &= b_1(m_1)b_1(m_2)D^2(z)P_0(k), \\ B_h(k_1, k_2, k_3|m_1, m_2, m_3, z) &= b_1(m_1)b_1(m_2)b_1(m_3)D^6(z)B_0(k_1, k_2, k_3) \\ &+ [b_1(m_1)b_1(m_2)b_2(m_3)D^4(z)P_{\text{lin}}(k_1)P_{\text{lin}}(k_2) + 2 \text{cyc}], \end{aligned} \quad (\text{G10})$$

where we only need the first two bias parameters b_1 , b_2 . They can be derived from the halo mass function using the peak-background split technique [160,165,169–171]; this method consists of dividing the Lagrangian density perturbations into short and long wavelength modes, and assuming that halo collapse happening on the short scales is enhanced or suppressed by the long-scale modulations in the dark-matter perturbations, which effectively alter the collapse threshold. After transformation from Lagrangian to Eulerian space assuming spherical collapse, the first two bias coefficients for the Press-Schechter mass function are [170]

$$b_1^{\text{PS}}(\nu) = 1 + \frac{\nu - 1}{\delta_c}, \quad (\text{G11})$$

$$b_2^{\text{PS}}(\nu) = \frac{8}{21} \frac{\nu - 1}{\delta_c} + \frac{\nu^4 - 3\nu^2}{\delta_c^2}. \quad (\text{G12})$$

While using the Tinker *et al.* mass function we find

$$b_1^T(\nu) = \frac{2\phi}{\delta_c[(\beta\nu)^{2\phi} + 1]} + \frac{\gamma\nu^2 + \delta_c - 2\eta - 1}{\delta_c}, \quad (\text{G13})$$

$$\begin{aligned} b_2^T(\nu) &= \frac{2(42\gamma\nu^2\phi + 8\delta_c\phi - 84\eta\phi + 42\phi^2 - 21\phi)}{21\delta_c^2[(\beta\nu)^{2\phi} + 1]} \\ &+ \frac{21\gamma^2\nu^4 + 8\gamma\delta_c\nu^2 - 84\gamma\eta\nu^2 - 63\gamma\nu^2}{21\delta_c^2} \\ &+ \frac{-16\delta_c\eta - 8\delta_c + 84\eta^2 + 42\eta}{21\delta_c^2}. \end{aligned} \quad (\text{G14})$$

In order to enforce consistency with the definition of matter overdensity, and to recover linear theory for $k \rightarrow 0$, we must finally impose the following conditions [172]:

$$\int_0^\infty dm \frac{m}{\bar{\rho}} n(m) = 1 \quad (\text{G15})$$

$$\int_0^\infty dm \frac{m}{\bar{\rho}} n(m)b_1(m) = 1, \quad (\text{G16})$$

$$\int_0^\infty dm \frac{m}{\bar{\rho}} n(m)b_i(m) = 0, \quad \forall i > 1. \quad (\text{G17})$$

APPENDIX H: COMBINED HALO-PT MODEL

The derivation of this model can be summarized as follows. The probability that a particle at Lagrangian position \mathbf{q}_1 belongs to a halo with mass in $[m, m + dm]$ is $dF = f(\nu)d\nu$. The probability that a particle at position \mathbf{q}_2 , at a distance $q = |\mathbf{q}_2 - \mathbf{q}_1|$ is situated in the same halo can be expressed as

$$F_m(q) = \frac{(2q_m - q)^2(4q_m + q)}{16q_m^3}, \quad \text{if } 0 \leq q \leq 2q_m, \quad (\text{H1})$$

and $F_m(q) = 0$ if $q > q_m$. Then the probability that the pair $[\mathbf{q}_1, \mathbf{q}_2]$ belongs to one (or two) halos is respectively

$$F_{1h}(q) = \int_{\nu_{q/2}}^\infty d\nu f(\nu)F_m(q), \quad (\text{H2})$$

$$F_{2h}(q) = 1 - F_{1h}(q). \quad (\text{H3})$$

In order to derive an analytic expression in Eulerian space, the function $F_{2h}(q)$ at perturbative level is further approximated by its value at $F_{2h}(q \sim 1/k)$.

In terms of the Eulerian particle positions $\mathbf{x}(\mathbf{q}, t)$, the matter power spectrum can be expressed as [155,173]

$$P(k) = \int d^3q \langle e^{i\mathbf{k} \cdot \Delta \mathbf{x}} - e^{i\mathbf{k} \cdot \mathbf{q}} \rangle, \quad (\text{H4})$$

where $\Delta \mathbf{x} = \mathbf{x}(\mathbf{q}) - \mathbf{x}(0)$ and $\langle \dots \rangle$ represents statistical average. The term $e^{i\mathbf{k}\cdot\mathbf{q}}$ is normally neglected in perturbation theory as it only produces a nonzero contribution at $k = 0$, but it is important in the halo model [100]. The power spectrum can be split between the contributions coming from pairs in one (or two) halos as

$$P_{1h(2h)}(k) = \int_0^\infty d^3q F_{1h(2h)}(q) \langle e^{i\mathbf{k}\cdot\Delta\mathbf{x}} - e^{i\mathbf{k}\cdot\mathbf{q}} \rangle_{1h(2h)}, \quad (\text{H5})$$

where in this case the averages are conditional on the set of pairs being in exactly one of the terms. The terms described in Eq. (H5) correspond to the one- and two-halo terms from the halo model of Sec. IV.

We can then split the power spectra further between perturbative and nonperturbative regimes. Considering the

perturbative case and the expected physical behavior of the two terms, it must hold $F_{1h} \equiv 0$ at all levels of perturbation theory. Hence $F_{2h} \equiv 1$, and the two-halo contribution is fully perturbative. However, the two-halo power spectrum can be obtained more easily by replacing the conditional average of Eq. (H5) with the full average given by perturbation theory, and by weighting instead the results with F_{2h} from Eq. (H3). This yields the result of Eq. (48).

For the one-halo contribution, it is assumed that the halos are fully virialized, and hence,

$$\langle e^{i\mathbf{k}\cdot\Delta\mathbf{x}} \rangle_m = u^2(k|m). \quad (\text{H6})$$

Substituting the result back into Eq. (H5) and changing the order of integration yields the result of Eq. (47).

-
- [1] Planck Collaboration, [arXiv:1502.01582](#).
 - [2] Planck Collaboration, [arXiv:1502.01589](#).
 - [3] A. H. Guth, *Phys. Rev. D* **23**, 347 (1981).
 - [4] V. F. Mukhanov and G. V. Chibisov, *JETP Lett.* **33**, 532 (1981).
 - [5] S. Hawking, *Phys. Lett.* **115B**, 295 (1982).
 - [6] S. Kachru, R. Kallosh, A. Linde, J. Maldacena, L. McAllister, and S. P. Trivedi, *J. Cosmol. Astropart. Phys.* **10** (2003) 013.
 - [7] N. Turok, *Classical Quantum Gravity* **19**, 3449 (2002).
 - [8] D. Baumann and L. McAllister, *Inflation and String Theory* (Cambridge University Press, Cambridge, England, 2015).
 - [9] J. Martin, C. Ringeval, and V. Vennin, *Phys. Dark Univ.* **5–6**, 75 (2014).
 - [10] Planck Collaboration, [arXiv:1502.02114](#).
 - [11] X. Chen, *Adv. Astron.* **2010**, 1 (2010).
 - [12] Planck Collaboration, [arXiv:1502.01592](#).
 - [13] J. Carrasco, M. Hertzberg, and L. Senatore, *J. High Energy Phys.* **09** (2012) 82.
 - [14] M. Alvarez *et al.*, [arXiv:1412.4671](#).
 - [15] J. Kwan, K. Heitmann, S. Habib, N. Padmanabhan, H. Finkel, N. Frontiere, and A. Pope, *Astrophys. J.* **810**, 35 (2015).
 - [16] S. A. Rodríguez-Torres *et al.*, [arXiv:1509.06404](#).
 - [17] D. G. York *et al.*, *Astron. J.* **120**, 1579 (2000).
 - [18] D. J. Eisenstein *et al.*, *Astron. J.* **142**, 72 (2011).
 - [19] K. S. Dawson, D. J. Schlegel *et al.*, *Astron. J.* **145**, 10 (2013).
 - [20] Dark Energy Survey Collaboration, [arXiv:astro-ph/0510346](#).
 - [21] H. T. Diehl *et al.* (DES), *Proc. SPIE Int. Soc. Opt. Eng.* **9149**, 91490V (2014).
 - [22] Z. Ivezic, J. A. Tyson, B. Abel *et al.*, [arXiv:0805.2366](#).
 - [23] R. Laureijs, J. Amiaux, S. Arduini, J. Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet *et al.*, [arXiv:1110.3193](#).
 - [24] B. Flaugher and C. Bebek, The Dark Energy Spectroscopic Instrument (DESI), in *Proceedings of SPIE, Ground-based and Airborne Instrumentation for Astronomy V* Vol. 9147 (2014), p. 91470S.
 - [25] D. Spergel, N. Gehrels, C. Baltay *et al.*, [arXiv:1503.03757](#).
 - [26] O. Doré *et al.*, [arXiv:1412.4872](#).
 - [27] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, *Phys. Rev. D* **77**, 123514 (2008).
 - [28] S. Matarrese and L. Verde, *Astrophys. J. Lett.* **677**, L77 (2008).
 - [29] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, *J. Cosmol. Astropart. Phys.* **08** (2008) 031.
 - [30] J.-Q. Xia, A. Bonaldi, C. Baccigalupi, G. De Zotti, S. Matarrese, L. Verde, and M. Viel, *J. Cosmol. Astropart. Phys.* **08** (2010) 013.
 - [31] A. Mana, T. Giannantonio, J. Weller, B. Hoyle, G. Hütsi, and B. Sartoris, *Mon. Not. R. Astron. Soc.* **434**, 684 (2013).
 - [32] A. J. Ross *et al.*, *Mon. Not. R. Astron. Soc.* **428**, 1116 (2013).
 - [33] T. Giannantonio, A. J. Ross, W. J. Percival, R. Crittenden, D. Bacher, M. Kilbinger, R. Nichol, and J. Weller, *Phys. Rev. D* **89**, 023511 (2014).
 - [34] T. Giannantonio and W. J. Percival, *Mon. Not. R. Astron. Soc.* **441**, L16 (2014).
 - [35] L. Verde, L. Wang, A. F. Heavens, and M. Kamionkowski, *Mon. Not. R. Astron. Soc.* **313**, 141 (2000).
 - [36] R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga, *Phys. Rev. D* **69**, 103513 (2004).
 - [37] N. Bartolo, S. Matarrese, and A. Riotto, *J. Cosmol. Astropart. Phys.* **10** (2005) 010.
 - [38] D. Jeong and E. Komatsu, *Astrophys. J.* **703**, 1230 (2009).
 - [39] T. Nishimichi, A. Taruya, K. Koyama, and C. Sabiu, *J. Cosmol. Astropart. Phys.* **07** (2010) 002.
 - [40] M. Liguori, E. Sefusatti, J. R. Fergusson, and E. P. S. Shellard, *Adv. Astron.* **2010**, 1 (2010).

- [41] T. Baldauf, U. Seljak, and L. Senatore, *J. Cosmol. Astropart. Phys.* **04** (2011) 006.
- [42] J. N. Fry, *Phys. Rev. Lett.* **73**, 215 (1994).
- [43] E. Sefusatti, M. Crocce, S. Pueblas, and R. Scoccimarro, *Phys. Rev. D* **74**, 023522 (2006).
- [44] S. Matarrese, L. Verde, and A. F. Heavens, *Mon. Not. R. Astron. Soc.* **290**, 651 (1997).
- [45] R. Scoccimarro, S. Colombi, J. N. Fry, J. A. Frieman, E. Hivon, and A. Melott, *Astrophys. J.* **496**, 586 (1998).
- [46] A. Shirata, Y. Suto, C. Hikage, T. Shiromizu, and N. Yoshida, *Phys. Rev. D* **76**, 044026 (2007).
- [47] L. Verde, A. F. Heavens, S. Matarrese, and L. Moscardini, *Mon. Not. R. Astron. Soc.* **300**, 747 (1998).
- [48] R. Scoccimarro, H. M. P. Couchman, and J. A. Frieman, *Astrophys. J.* **517**, 531 (1999).
- [49] E. Sefusatti and E. Komatsu, *Phys. Rev. D* **76**, 083004 (2007).
- [50] H. Guo and Y. P. Jing, *Astrophys. J.* **702**, 425 (2009).
- [51] J. E. Pollack, R. E. Smith, and C. Porciani, *Mon. Not. R. Astron. Soc.* **420**, 3469 (2012).
- [52] S. Saito, T. Baldauf, Z. Vlah, U. Seljak, T. Okumura, and P. McDonald, *Phys. Rev. D* **90**, 123522 (2014).
- [53] J. N. Fry, *Astrophys. J.* **279**, 499 (1984).
- [54] M. H. Goroff, B. Grinstein, S.-J. Rey, and M. B. Wise, *Astrophys. J.* **311**, 6 (1986).
- [55] B. Jain and E. Bertschinger, *Astrophys. J.* **431**, 495 (1994).
- [56] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, *Phys. Rep.* **367**, 1 (2002).
- [57] T. Buchert, *Astron. Astrophys.* **223**, 9 (1989).
- [58] F. Moutarde, J.-M. Alimi, F. R. Bouchet, R. Pellat, and A. Ramani, *Astrophys. J.* **382**, 377 (1991).
- [59] T. Buchert, *Mon. Not. R. Astron. Soc.* **254**, 729 (1992).
- [60] T. Buchert, *Mon. Not. R. Astron. Soc.* **267**, 811 (1994).
- [61] F. R. Bouchet, S. Colombi, E. Hivon, and R. Juszkiewicz, *Astron. Astrophys.* **296**, 575 (1995).
- [62] J. Ehlers and T. Buchert, *Gen. Relativ. Gravit.* **29**, 733 (1997).
- [63] T. Tatekawa, *Recent Res. Dev. Phys. Chem.* **2**, 1 (2005).
- [64] M. White, *Mon. Not. R. Astron. Soc.* **439**, 3630 (2014).
- [65] J. Carlson, M. White, and N. Padmanabhan, *Phys. Rev. D* **80**, 043531 (2009).
- [66] P. Valageas, *Phys. Rev. D* **88**, 083524 (2013).
- [67] S. Tassev, *J. Cosmol. Astropart. Phys.* **06** (2014) 008.
- [68] Z. Vlah, U. Seljak, and T. Baldauf, *Phys. Rev. D* **91**, 023508 (2015).
- [69] R. Scoccimarro, *Ann. N.Y. Acad. Sci.* **927**, 13 (2001).
- [70] M. Crocce and R. Scoccimarro, *Phys. Rev. D* **73**, 063519 (2006).
- [71] M. Crocce and R. Scoccimarro, *Phys. Rev. D* **73**, 063520 (2006).
- [72] M. Crocce and R. Scoccimarro, *Phys. Rev. D* **77**, 023533 (2008).
- [73] F. Bernardeau, M. Crocce, and R. Scoccimarro, *Phys. Rev. D* **78**, 103521 (2008).
- [74] F. Bernardeau, M. Crocce, and R. Scoccimarro, *Phys. Rev. D* **85**, 123519 (2012).
- [75] M. Crocce, R. Scoccimarro, and F. Bernardeau, *Mon. Not. R. Astron. Soc.* **427**, 2537 (2012).
- [76] P. Valageas, *Astron. Astrophys.* **465**, 725 (2007).
- [77] A. Taruya and T. Hiramatsu, *Astrophys. J.* **674**, 617 (2008).
- [78] P. McDonald, *Phys. Rev. D* **75**, 043514 (2007).
- [79] S. Matarrese and M. Pietroni, *J. Cosmol. Astropart. Phys.* **06** (2007) 026.
- [80] T. Matsubara, *Phys. Rev. D* **77**, 063530 (2008).
- [81] T. Okamura, A. Taruya, and T. Matsubara, *J. Cosmol. Astropart. Phys.* **08** (2011) 012.
- [82] J. Carlson, B. Reid, and M. White, *Mon. Not. R. Astron. Soc.* **429**, 1674 (2013).
- [83] C. Rampf and T. Buchert, *J. Cosmol. Astropart. Phys.* **06** (2012) 021.
- [84] C. Rampf and Y. Y. Y. Wong, *J. Cosmol. Astropart. Phys.* **06** (2012) 018.
- [85] P. Valageas, T. Nishimichi, and A. Taruya, *Phys. Rev. D* **87**, 083522 (2013).
- [86] T. Matsubara, *Phys. Rev. D* **90**, 043537 (2014).
- [87] N. S. Sugiyama, *Astrophys. J.* **788**, 63 (2014).
- [88] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, *J. Cosmol. Astropart. Phys.* **07** (2012) 051.
- [89] M. P. Hertzberg, *Phys. Rev. D* **89**, 043521 (2014).
- [90] J. J. M. Carrasco, S. Foreman, D. Green, and L. Senatore, *J. Cosmol. Astropart. Phys.* **07** (2014) 056.
- [91] J. J. M. Carrasco, S. Foreman, D. Green, and L. Senatore, *J. Cosmol. Astropart. Phys.* **07** (2014) 057.
- [92] E. Pajer and M. Zaldarriaga, *J. Cosmol. Astropart. Phys.* **08** (2013) 037.
- [93] S. M. Carroll, S. Leichenauer, and J. Pollack, *Phys. Rev. D* **90**, 023518 (2014).
- [94] R. A. Porto, L. Senatore, and M. Zaldarriaga, *J. Cosmol. Astropart. Phys.* **05** (2014) 022.
- [95] L. Senatore and M. Zaldarriaga, *J. Cosmol. Astropart. Phys.* **02** (2015) 013.
- [96] U. Seljak, *Mon. Not. R. Astron. Soc.* **318**, 203 (2000).
- [97] C.-P. Ma and J. N. Fry, *Astrophys. J.* **543**, 503 (2000).
- [98] J. A. Peacock and R. E. Smith, *Mon. Not. R. Astron. Soc.* **318**, 1144 (2000).
- [99] A. Cooray and R. Sheth, *Phys. Rep.* **372**, 1 (2002).
- [100] P. Valageas and T. Nishimichi, *Astron. Astrophys.* **527**, A87 (2011).
- [101] P. Valageas and T. Nishimichi, *Astron. Astrophys.* **532**, A4 (2011).
- [102] A. Mead, J. Peacock, C. Heymans, S. Joudaki, and A. Heavens, *Mon. Not. R. Astron. Soc.* **454**, 1958 (2015).
- [103] I. Mohammed and U. Seljak, *Mon. Not. R. Astron. Soc.* **445**, 3382 (2014).
- [104] U. Seljak and Z. Vlah, *Phys. Rev. D* **91**, 123516 (2015).
- [105] 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).
- [106] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri, *Astrophys. J.* **761**, 152 (2012).
- [107] K. Heitmann, M. White, C. Wagner, S. Habib, and D. Higdon, *Astrophys. J.* **715**, 104 (2010).
- [108] E. Sefusatti, M. Crocce, and V. Desjacques, *Mon. Not. R. Astron. Soc.* **406**, 1014 (2010).
- [109] E. Sefusatti, M. Crocce, and V. Desjacques, *Mon. Not. R. Astron. Soc.* **425**, 2903 (2012).
- [110] M. M. Schmittfull, D. M. Regan, and E. P. S. Shellard, *Phys. Rev. D* **88**, 063512 (2013).
- [111] M. Schmittfull, T. Baldauf, and U. Seljak, *Phys. Rev. D* **91**, 043530 (2015).

- [112] J. R. Fergusson, D. M. Regan, and E. P. S. Shellard, *Phys. Rev. D* **86**, 063511 (2012).
- [113] J. R. Fergusson, M. Liguori, and E. P. S. Shellard, *Phys. Rev. D* **82**, 023502 (2010).
- [114] Y. P. Jing and G. Börner, *Astrophys. J.* **503**, 37 (1998).
- [115] H. A. Feldman, J. A. Frieman, J. N. Fry, and R. Scoccimarro, *Phys. Rev. Lett.* **86**, 1434 (2001).
- [116] R. Scoccimarro, H. A. Feldman, J. N. Fry, and J. A. Frieman, *Astrophys. J.* **546**, 652 (2001).
- [117] L. Verde *et al.*, *Mon. Not. R. Astron. Soc.* **335**, 432 (2002).
- [118] Y. P. Jing and G. Börner, *Astrophys. J.* **607**, 140 (2004).
- [119] Y. Wang, X. Yang, H. J. Mo, F. C. van den Bosch, and Y. Chu, *Mon. Not. R. Astron. Soc.* **353**, 287 (2004).
- [120] F. Marín, *Astrophys. J.* **737**, 97 (2011).
- [121] F. A. Marín *et al.*, *Mon. Not. R. Astron. Soc.* **432**, 2654 (2013).
- [122] H. Gil-Marín, J. Noreña, L. Verde, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider, *Mon. Not. R. Astron. Soc.* **451**, 539 (2015).
- [123] H. Gil-Marín, L. Verde, J. Noreña, A. J. Cuesta, L. Samushia, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider, *Mon. Not. R. Astron. Soc.* **452**, 1914 (2015).
- [124] D. Babich, P. Creminelli, and M. Zaldarriaga, *J. Cosmol. Astropart. Phys.* **08** (2004) 009.
- [125] T. Baldauf, U. Seljak, V. Desjacques, and P. McDonald, *Phys. Rev. D* **86**, 083540 (2012).
- [126] J. R. Fergusson and E. P. S. Shellard, *Phys. Rev. D* **80**, 043510 (2009).
- [127] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, *J. Cosmol. Astropart. Phys.* **05** (2006) 004.
- [128] P. D. Meerburg, J. P. van der Schaar, and P. S. Corasaniti, *J. Cosmol. Astropart. Phys.* **05** (2009) 018.
- [129] X. Chen, *Adv. Astron.* **2010**, 1 (2010).
- [130] R. Scoccimarro and H. M. P. Couchman, *Mon. Not. R. Astron. Soc.* **325**, 1312 (2001).
- [131] 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).
- [132] A. Lewis, A. Challinor, and A. Lasenby, *Astrophys. J.* **538**, 473 (2000).
- [133] H. Gil-Marín, C. Wagner, F. Fragkoudi, R. Jimenez, and L. Verde, *J. Cosmol. Astropart. Phys.* **02** (2012) 047.
- [134] R. Scoccimarro, S. Colombi, J. N. Fry, J. A. Frieman, E. Hivon, and A. Melott, *Astrophys. J.* **496**, 586 (1998).
- [135] J. J. M. Carrasco, S. Foreman, D. Green, and L. Senatore, *J. Cosmol. Astropart. Phys.* **07** (2014) 057.
- [136] J. J. M. Carrasco, S. Foreman, D. Green, and L. Senatore, *J. Cosmol. Astropart. Phys.* **07** (2014) 056.
- [137] T. Baldauf, L. Mercolli, M. Mirbabayi, and E. Pajer, *J. Cosmol. Astropart. Phys.* **05** (2015) 007.
- [138] R. E. Angulo, S. Foreman, M. Schmittfull, and L. Senatore, *J. Cosmol. Astropart. Phys.* **10** (2015) 039.
- [139] D. Figueroa, E. Sefusatti, A. Riotto, and F. Vernizzi, *J. Cosmol. Astropart. Phys.* **08** (2012) 036.
- [140] R. E. Smith, R. K. Sheth, and R. Scoccimarro, *Phys. Rev. D* **78**, 023523 (2008).
- [141] V. Springel, N. Yoshida, and S. D. White, *New Astron.* **6**, 79 (2001).
- [142] V. Springel, *Mon. Not. R. Astron. Soc.* **364**, 1105 (2005).
- [143] R. Scoccimarro, *Mon. Not. R. Astron. Soc.* **299**, 1097 (1998).
- [144] M. Crocce, S. Pueblas, and R. Scoccimarro, *Mon. Not. R. Astron. Soc.* **373**, 369 (2006).
- [145] N. McCullagh, D. Jeong, and A. S. Szalay, *Mon. Not. R. Astron. Soc.* **455**, 2945 (2016).
- [146] E. Komatsu *et al.*, *Astrophys. J. Suppl. Ser.* **192**, 18 (2011).
- [147] D. M. Regan, M. M. Schmittfull, E. P. S. Shellard, and J. R. Fergusson, *Phys. Rev. D* **86**, 123524 (2012).
- [148] R. E. Smith, V. Desjacques, and L. Mariani, *Phys. Rev. D* **83**, 043526 (2011).
- [149] C.-T. Chiang, C. Wagner, F. Schmidt, and E. Komatsu, *J. Cosmol. Astropart. Phys.* **05** (2014) 048.
- [150] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics: Volume 6* (Butterworth-Heinemann, Washington, DC, 1987).
- [151] M. Peloso and M. Pietroni, *J. Cosmol. Astropart. Phys.* **05** (2013) 031.
- [152] A. Kehagias and A. Riotto, *Nucl. Phys.* **B873**, 514 (2013).
- [153] T. Hahn, *Comput. Phys. Commun.* **168**, 78 (2005).
- [154] A. Taruya, F. Bernardeau, T. Nishimichi, and S. Codis, *Phys. Rev. D* **86**, 103528 (2012).
- [155] A. N. Taylor and A. J. S. Hamilton, *Mon. Not. R. Astron. Soc.* **282**, 767 (1996).
- [156] K. B. Fisher and A. Nusser, *Mon. Not. R. Astron. Soc.* **279**, L1 (1996).
- [157] J. E. Gunn and J. R. Gott, III, *Astrophys. J.* **176**, 1 (1972).
- [158] J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **462**, 563 (1996).
- [159] J. S. Bullock, T. S. Kolatt, Y. Sigad, R. S. Somerville, A. V. Kravtsov, A. A. Klypin, J. R. Primack, and A. Dekel, *Mon. Not. R. Astron. Soc.* **321**, 559 (2001).
- [160] R. Scoccimarro, R. K. Sheth, L. Hui, and B. Jain, *Astrophys. J.* **546**, 20 (2001).
- [161] A. A. Klypin, S. Trujillo-Gomez, and J. Primack, *Astrophys. J.* **740**, 102 (2011).
- [162] T. Giannantonio, C. Porciani, J. Carron, A. Amara, and A. Pillepich, *Mon. Not. R. Astron. Soc.* **422**, 2854 (2012).
- [163] W. H. Press and P. Schechter, *Astrophys. J.* **187**, 425 (1974).
- [164] T. Giannantonio and C. Porciani, *Phys. Rev. D* **81**, 063530 (2010).
- [165] R. K. Sheth and G. Tormen, *Mon. Not. R. Astron. Soc.* **308**, 119 (1999).
- [166] J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottlöber, and D. E. Holz, *Astrophys. J.* **688**, 709 (2008).
- [167] J. L. Tinker, B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes, and S. Gottlöber, *Astrophys. J.* **724**, 878 (2010).
- [168] J. N. Fry and E. Gaztanaga, *Astrophys. J.* **413**, 447 (1993).
- [169] S. Cole and N. Kaiser, *Mon. Not. R. Astron. Soc.* **237**, 1127 (1989).
- [170] H. J. Mo and S. D. M. White, *Mon. Not. R. Astron. Soc.* **282**, 347 (1996).
- [171] H. J. Mo, Y. P. Jing, and S. D. M. White, *Mon. Not. R. Astron. Soc.* **284**, 189 (1997).
- [172] A. Refregier and R. Teyssier, *Phys. Rev. D* **66**, 043002 (2002).
- [173] P. Schneider and M. Bartelmann, *Mon. Not. R. Astron. Soc.* **273**, 475 (1995).