Large Scale Structure Reconstruction with Short-Wavelength Modes: Halo Bias and Light Cone Formalism

Peikai Li,[1, 2] Rupert A. C. Croft,[1, 2] and Scott Dodelson[1, 2]

[1]*Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[2]*McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

(Dated: June 25, 2020)

This is the second paper in a series where we propose a method of indirectly measuring large scale structure using information from small scale perturbations. The idea is to build a quadratic estimator from small scale modes that provides a map of structure on large scales. We demonstrated in the first paper that the quadratic estimator works well on a dark-matter-only N-body simulation at a snapshot of $z = 0$. Here we generalize the theory to the case of a light-cone with halo bias and redshift space distortions taken into consideration. We successfully apply the generalized version of the quadratic estimator to a light cone halo catalog of an N-body simulation of size $\sim 5.6\,(h^{-1}\,\mathrm{Gpc})^3$. The most distant point in the light cone is at a redshift of 1.4, which indicates that we might be able to apply our method to next generation galaxy surveys.

I. INTRODUCTION

Directly measuring the distribution of matter on large scales is extremely difficult right now as pointed out, e.g., by [1]. The attempt to use small scale perturbations to infer large scale information has been frequently discussed in recent years [2][3][4][5][6]. In our first work [7], we proposed a method of indirectly measuring large scale structure using the small scale density contrast. Physically, long- and short-wavelength modes are correlated because small scale modes will grow differently depending on the large scale structure they reside in. This phenomenon leaves a signature in Fourier space: the two-point statistics of short-wavelength matter density modes will have non-zero off-diagonal terms proportional to long-wavelength modes. This is our starting point for constructing the quadratic estimator for long-wavelength modes. We tested the power of the quadratic estimator using a dark-matter-only catalog from an N-body simulation in the first paper. In this work, we generalize Ref. [7] to account for three effects that must be accounted for before applying the techniques to upcoming surveys [8][9][10]: (i) we observe galaxies, not the dark matter field; (ii) we observe on a light cone not a snapshot; and (iii) peculiar velocities lead to redshift space distortions.

First we need to account for halo bias [11][12]. Halo bias is a term relating the halo number density contrast to the matter density contrast. We consider only the leading order of halo bias in this work, as done in recent treatments of galaxy surveys [13][14][15]. One of the main difficulties with this generalization is that the linear halo bias depends on both redshift and halo mass. We will account for this and construct an estimator of the matter density contrast from halo information.

Observationally a galaxy catalog will be in a light cone [16] instead of a snapshot. The typical treatment is to cut a light cone into several thin redshift bins [?] [Scott: This reference is not a good example since it is a photometric survey and this paper is only cosmic shear, not galaxy clustering. Can you find a Sloan or eboss reference?] and analyze the properties within each bin. Doing this, though, loses information about the long-wavelength modes along line of sight. Thus in this paper we propose a method of considering all the halos/galaxies in a light cone together, accounting for the redshift difference by multiplying by an extra factor proportional to the inverse of the linear growth function. Since the universe is linear at sufficiently large scales, linear matter density modes at different redshift will still be correlated. With this knowledge we can recover the linear power spectrum on large scales. Using a cubic volume, we construct the quadratic estimator for long-wavelength modes using information from non-zero off-diagonal terms as in Ref. .[7].

Another observational effect we need to take into account is redshift space distortion [17]. This effect can be straightforwardly included by combining the Kaiser formalism with the perturbative expansion developed in real space.

We begin with a brief review of the formalism developed in Ref. [7], then present our treatment of halo bias in a snapshot by constructing the approximate form of the matter perturbation field using the information from a halo catalog. We then deal with the matter density contrast in a light cone and then build the quadratic estimator, finally adding in redshift space distortions. Finally we apply the estimator to different N-body simulations and successfully extract large scale modes accounting for all three effects. In table I we list important symbols and notation used in this work.

Name	Symbol	Defining Relation
Matter Density Contrast	$\delta_{\mathrm{m}}(\vec{r};a)$	$\delta_{\mathrm{m}}(\vec{r};a) \equiv (\rho_{\mathrm{m}}(\vec{r};a) - \bar{\rho}_{\mathrm{m}}(a))/\bar{\rho}_{\mathrm{m}}(a)$
Initial Matter Density Contrast	$\delta_{\mathrm{m,ini}}^{(1)}(\vec{r})$	$\delta_{\mathrm{m,ini}}^{(1)}(\vec{r}) = \delta_{\mathrm{m}}^{(1)}(\vec{r};a_{\mathrm{ini}})$
Halo (Number) Density Field	$n_{\mathrm{h}}(\vec{r};a)$	$n_{\mathrm{h}}(\vec{r};a) = \sum_{i=1}^{N_{\mathrm{h}}} \delta_{\mathrm{D}}(\vec{r} - \vec{r}_i)$
Halo (Number) Density Contrast	$\delta_{\mathrm{h}}(\vec{r};a)$	Eq. (8); $\delta_{\mathrm{h}}(\vec{r};a) \equiv (n_{\mathrm{h}}(\vec{r};a) - \bar{n}_{\mathrm{h}}(a))/\bar{n}_{\mathrm{h}}(a)$
Weighted Halo (Number) Density Field	$\eta_{\mathrm{h}}(\vec{r};a)$	Eq. (10); $\eta_{\mathrm{h}}(\vec{r};a) = \sum_{i=1}^{N_{\mathrm{h}}}(\delta_{\mathrm{D}}(\vec{r} - \vec{r}_i)/b_1(M_i;a_i))$
Approximate Matter Density Contrast	$\delta_{\mathrm{m}}^{\mathrm{h}}(\vec{r};a)$	Eq. (9); $\delta_{\mathrm{m}}^{\mathrm{h}}(\vec{r};a) \equiv (\eta_{\mathrm{h}}(\vec{r};a) - \bar{\eta}_{\mathrm{h}}(a))/\bar{n}_{\mathrm{h}}(a)$
Modified Matter Density Contrast	$d_{\mathrm{m}}(\vec{r})$	Eq. (12); $d_{\mathrm{m}}(\vec{r}) \equiv \delta_{\mathrm{m}}(\vec{r};a(r))D_{\mathrm{ini}}/D_1(a(r))$
Approximate Modified Matter Density Contrast	$d_{\mathrm{m}}^{\mathrm{h}}(\vec{r})$	$d_{\mathrm{m}}^{\mathrm{h}}(\vec{r}) \equiv \delta_{\mathrm{m}}^{\mathrm{h}}(\vec{r};a(r))D_{\mathrm{ini}}/D_1(a(r))$
Linear Matter Power Spectrum	$P_{\mathrm{m,lin}}(k;a)$	$\langle \delta_{\mathrm{m}}^{(1)}(\vec{k};a)\delta_{\mathrm{m}}^{(1)}(\vec{k}';a)\rangle \equiv (2\pi)^3 \delta_{\mathrm{D}}(\vec{k}+\vec{k}')P_{\mathrm{m,lin}}(k;a)$
Nonlinear Matter Power Spectrum	$P_{\mathrm{m,nl}}(k;a)$	$\langle \delta_{\mathrm{m}}(\vec{k};a)\delta_{\mathrm{m}}(\vec{k}';a)\rangle \equiv (2\pi)^3 \delta_{\mathrm{D}}(\vec{k}+\vec{k}')P_{\mathrm{m,nl}}(k;a)$
Initial Matter Power Spectrum	$P_{\mathrm{m,ini}}(k)$	$P_{\mathrm{m,ini}}(k) = (D_{\mathrm{ini}}/D_1(a))^2 P_{\mathrm{m,lin}}(k;a)$ or Eq. (19)
Light Cone Matter Power Spectrum	$\mathcal{P}_{\mathrm{m}}(k)$	Eq. (20); $\langle d_{\mathrm{m}}(\vec{k})d_{\mathrm{m}}(\vec{k}')\rangle \equiv (2\pi)^3 \delta_{\mathrm{D}}(\vec{k}+\vec{k}')\mathcal{P}_{\mathrm{m}}(k)$
Approximate Light Cone Matter Power Spectrum	$\mathcal{P}_{\mathrm{m}}^{\mathrm{h}}(k)$	$\langle d_{\mathrm{m}}^{\mathrm{h}}(\vec{k})d_{\mathrm{m}}^{\mathrm{h}}(\vec{k}')\rangle \equiv (2\pi)^3 \delta_{\mathrm{D}}(\vec{k}+\vec{k}')\mathcal{P}_{\mathrm{m}}^{\mathrm{h}}(k)$

TABLE I. Table of symbols and notation. Extra subscript "rs" stands for the corresponding quantity in redshift space.

II. REVIEW OF QUADRATIC ESTIMATOR

We first review the construction of a quadratic estimator of a dark-matter-only catalog [7] before moving to a halo catalog. Starting from the perturbative series of the matter density contrast in Fourier space [18][19]:

$$\delta_{\mathrm{m}}(\vec{k};a) = \delta_{\mathrm{m}}^{(1)}(\vec{k};a) + \delta_{\mathrm{m}}^{(2)}(\vec{k};a) + \cdots$$
$$= \frac{D_1(a)}{D_{\mathrm{ini}}}\delta_{\mathrm{m}}^{(1)}(\vec{k};a_{\mathrm{ini}}) + \left[\frac{D_1(a)}{D_{\mathrm{ini}}}\right]^2 \delta_{\mathrm{m}}^{(2)}(\vec{k};a_{\mathrm{ini}}) + \cdots \text{(1)}$$

where "m" stands for matter, the superscript $i = 1,2,\cdots$ corresponds to the i-th order term of the expansion, D_1 is the linear growth function while $D_{\mathrm{ini}} = D_1(a = a_{\mathrm{ini}})$ is the value of D_1 at some initial time a_{ini}, and $\delta_{\mathrm{m}}(\vec{k};a)$ is the full Fourier space matter density contrast in a snapshot when the scale factor was equal to a. Thus, $\delta_{\mathrm{m}}^{(1)}$ is the linear density contrast, so that the first term on the right in the first line above evolves as shown by the first term on the right in the second line.

When evaluating the two-point function of the full density contrast, cross-terms appear. For example, $\langle \delta_{\mathrm{m}}^{(1)}(\vec{k};a)\delta_{\mathrm{m}}^{(2)}(\vec{k}';a)\rangle$ is proportional to $\delta_{\mathrm{m}}^{(1)}(\vec{k}+\vec{k}',a)$ if both $\vec{k}$ and $\vec{k}'$ correspond to short wavelengths but their sum is small (long wavelength). Explicitly, keeping terms up to second order,

$$\langle \delta_{\mathrm{m}}(\vec{k}_s;a)\delta_{\mathrm{m}}(\vec{k}_s';a)\rangle = f(\vec{k}_s, \vec{k}_s')\delta_{\mathrm{m}}^{(1)}(\vec{k}_l;a). \quad \text{(2)}$$

Here $\vec{k}_s$ and $\vec{k}_s'$ are two short-wavelength modes and $\vec{k}_l$ is a long-wavelength mode ($\vec{k}_s, \vec{k}_s' \gg \vec{k}_l$). They satisfy the squeezed-limit condition $\vec{k}_s + \vec{k}_s' = \vec{k}_l$ and f is given by:

$$f(\vec{k}_s, \vec{k}_s') = 2F_2(-\vec{k}_s, \vec{k}_s + \vec{k}_s')P_{\mathrm{m,lin}}(k_s;a)$$
$$+ 2F_2(-\vec{k}_s', \vec{k}_s + \vec{k}_s')P_{\mathrm{m,lin}}(k_s';a). \quad \text{(3)}$$

Here $P_{\mathrm{m,lin}}$ is the linear matter power spectrum and F_2 is a function particularly insensitive to the choice of cosmological parameters in a dark-energy-dominated universe [20]:

$$F_2(\vec{k}_1, \vec{k}_2) = \frac{5}{7} + \frac{2}{7}\frac{(\vec{k}_1 \cdot \vec{k}_2)^2}{k_1^2 k_2^2} + \frac{\vec{k}_1 \cdot \vec{k}_2}{2k_1 k_2}\left[\frac{k_1}{k_2} + \frac{k_2}{k_1}\right]. \quad \text{(4)}$$

Eq. (2) indicates that we can estimate the long-wavelength modes using small scale information with the following minimum variance quadratic estimator:

$$\hat{\delta}_{\mathrm{m}}^{(1)}(\vec{k}_l;a) = A(\vec{k}_l) \int \frac{d^3\vec{k}_s}{(2\pi)^3} g(\vec{k}_s, \vec{k}_s')\delta_{\mathrm{m}}(\vec{k}_s;a)\delta_{\mathrm{m}}(\vec{k}_s';a) \quad \text{(5)}$$

with $\vec{k}_s' = \vec{k}_l - \vec{k}_s$. The normalization factor A is defined by requiring that $\langle \hat{\delta}_{\mathrm{m}}^{(1)}(\vec{k}_l;a)\rangle = \delta_{\mathrm{m}}^{(1)}(\vec{k}_l;a)$, and the weighting function g is obtained by minimizing the noise. They can be expressed as:

$$A(\vec{k}_l) = \left[\int \frac{d^3\vec{k}_s}{(2\pi)^3} g(\vec{k}_s, \vec{k}_s')f(\vec{k}_s, \vec{k}_s')\right]^{-1}$$
$$g(\vec{k}_s, \vec{k}_s') = \frac{f(\vec{k}_s, \vec{k}_s')}{2P_{\mathrm{m,nl}}(k_s)P_{\mathrm{m,nl}}(k_s')} \quad \text{(6)}$$

where $P_{\mathrm{m,nl}}$ is the nonlinear matter power spectrum. With this choice of the weighting function g, the noise on the estimator $N(\vec{k}_l) = A(\vec{k}_l)$ if non-gaussian terms in the four-point function are neglected. Therefore, the projected detectability of a power spectrum measurement using this quadratic estimator can be written as:

$$\frac{1}{\sigma^2(k_l)} = \frac{Vk_l^2\Delta k}{(2\pi)^2}\left[\frac{P_{\mathrm{m,lin}}(k_l)}{P_{\mathrm{m,lin}}(k_l) + A(k_l)}\right]^2, \quad \text{(7)}$$

where V is the volume of a survey and Δk is the width of long-wavelength mode bins. We also take advantage of the fact that $A(\vec{k}_l)$ does not depend on the direction of the long mode $\vec{k}_l$.

III. HALO BIAS

Ignoring higher order halo bias, the number density contrast of halos can be related to the matter density contrast at a fixed time a as [12]:

$$\delta_{\mathrm{h}}(\vec{r}; a) \equiv \frac{n_{\mathrm{h}}(\vec{r}; a) - \bar{n}_{\mathrm{h}}(a)}{\bar{n}_{\mathrm{h}}(a)} = b_1 \delta_{\mathrm{m}}(\vec{r}; a) \qquad (8)$$

here "h" denotes halo, n_{h} is the halo number density field at a given position, $\bar{n}_{\mathrm{h}}$ is the mean number density of halos and b_1 is the linear bias parameter relating halo and matter density contrasts. The linear bias b_1 is a function of halo mass and position and has been well approximated using analytical expressions [21][22]. We will use the Tinker halo bias function throughout this paper [23]. Since the linear bias b_1 depends not only on halo mass but also halo redshift, it will be more challenging for us to use δ_{h} than δ_{m} within a light cone unless we use a very narrow mass bin. Therefore we choose to build an approximate matter density contrast out of halo positions in a snapshot following [24]. To distinguish this from the standard matter contrast, we add an "h" to the superscript. Then,

$$\delta_{\mathrm{m}}^{\mathrm{h}}(\vec{r}; a) \equiv \frac{\eta_{\mathrm{h}}(\vec{r}; a) - \bar{\eta}_{\mathrm{h}}(a)}{\bar{n}_{\mathrm{h}}(a)} \qquad (9)$$

where $\eta_{\mathrm{h}}(\vec{r}; a)$ is the usual halo density field with each halo weighted by the inverse of its own halo bias, and $\bar{\eta}_{\mathrm{h}}(a)$ is its mean value. They can be explicitly written as:

$$\eta_{\mathrm{h}}(\vec{r}; a) = \sum_{i=1}^{N_{\mathrm{h}}} \left. \frac{\delta_{\mathrm{D}}(\vec{r} - \vec{r}_i)}{b_1(M_i; a_i)} \right|_{a_i = a} \qquad (10)$$

$$\bar{\eta}_{\mathrm{h}}(a) = \langle \eta_{\mathrm{h}}(\vec{r}; a) \rangle |_a \qquad (11)$$

where δ_{D} is the Dirac delta function, N_{h} is the total number of halos and a_i, M_i and $\vec{r}_i$ are the cosmological scale factor, mass and position of the i-th halo, respectively. [Scott: I changed the second equation above and propose to add this sentence:] In the second line, the brackets denote an average over all $\vec{r}$ in the shell defined by $a(|\vec{r}|) = a$. Since we consider only a fixed time in this section, we can use the simplification $b_1(M_i, a_i) = b_1(M_i)$. We can use the Cloud-in-Cell (CIC) or the Triangular Shaped Cloud (TSC) scheme [25] to smooth the sharply peaked Dirac delta function numerically[1].
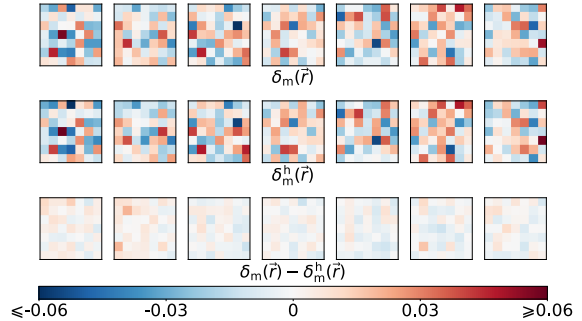
---

[1] We use the TSC scheme in this work.

FIG. 1. Comparison between the real matter density contrast $\delta_{\mathrm{m}}(\vec{r})$ and the approximate matter density contrast using halo information $\delta_{\mathrm{m}}^{\mathrm{h}}(\vec{r})$ at a snapshot when $z = 0$. We use the halo catalog from BIGMDPL.

We consider Eq. (9) to be a practical and general way of estimating the matter density using the positions of galaxies, since in real surveys the bias can be expressed as a function of galaxy luminosity and color [26]. We will not use the halo density contrast δ_{h} directly, instead we will use $\delta_{\mathrm{m}}^{\mathrm{h}}$ defined in Eq. (9) as the approximate matter density contrast during the calculation of the quadratic estimator Eq. (5) both in a snapshot and in a light cone. We test this using a halo catalog from a cosmological N-body simulation. We use the $z = 0$ snapshot from BigMDPL, one of the Multi-Dark cosmological simulations [27]. This simulation used a flat ΛCDM model with Planck Collaboration XVI (2014) [28] cosmological parameters. Halos were found in this catalog using the *Rockstar* code [29]. We will focus on halos with masses between $2.2 \times 10^{12} h^{-1} M_{\odot} < M < 10^{14} h^{-1} M_{\odot}$ for simplicity (although we have found similar results in other mass bins).

The top panel of Fig. 1 shows the actual matter density δ_{m} in a set of seven slices, each broken up into 49 square regions on volume [Scott: add number here]. The second panel shows the matter density inferred via Eq. (9) and the bottom panel shows the difference between them. The differences are small, showing the $\delta_{\mathrm{m}}^{\mathrm{h}}$ is a good estimator for the underlying matter density.

We use the approximate density contrast defined in Eq. (9) and compute its power spectrum; as shown in Fig. 2 this is equal to the true matter power spectrum within ten percent. Note that the plot covers the scales of interest, those that will be used to construct the quadratic estimator. We therefore expect the projected detectability will be the same as when applied to the dark matter field itself, as in [7].

We apply the quadratic estimator to the field defined in Eq. (9). We transform the true Fourier modes $\delta_{\mathrm{m}}^{\mathrm{h}}(\vec{k}_l)$

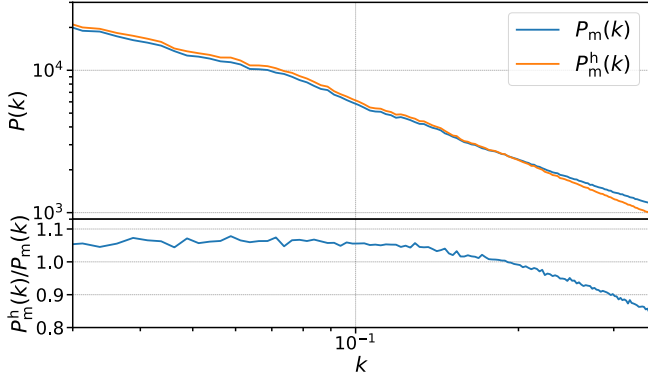FIG. 2. $P_{\mathrm{m}}^{\mathrm{h}}(k)$ is the approximate matter power spectrum using the halo catalog from BIGMDPL at a snapshot $z = 0$ and Eq. (9), while $P_m(k)$ is the true matter spectrum; the bottom panel is the ratio of the two, showing that over the scales of interest, $\delta_{\mathrm{m}}^{\mathrm{h}}$ is a good estimator for the density contrast.
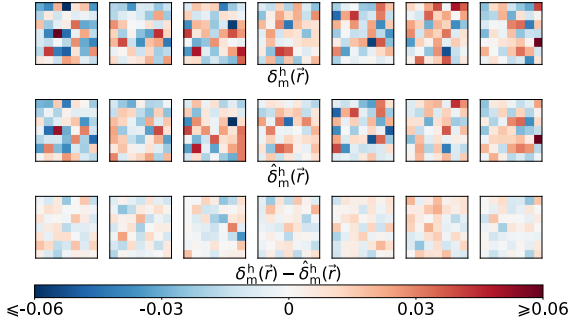

FIG. 3. Comparison of the true density field of the Big-MDPL halo catalog within mass bin $2.2 \times 10^{12} h^{-1} M_\odot < M < 10^{14} h^{-1} M_\odot$ ($\delta_{\mathrm{m}}^{\mathrm{h}}(\vec{r})$ computed using the directly measured large scale modes, top row) and the density field from the quadratic estimator ($\hat{\delta}_{\mathrm{m}}^{\mathrm{h}}(\vec{r})$, middle row). The bottom row shows their difference. Each panel represents a slice through the simulation volume, $2500 h^{-1}$ Mpc wide, and one cell $\sim 357 h^{-1}$ Mpc thick. The integration range of $\vec{k}_s$ is from $0.03 h\,\mathrm{Mpc}^{-1}$ to $0.22 h\,\mathrm{Mpc}^{-1}$.

[Scott: Shouldn't we show δ_m here?] and the long-wavelength modes estimated using the quadratic estimator, $\hat{\delta}_{\mathrm{m}}^{\mathrm{h}}(\vec{k}_l)$, back into real space and compare them in Fig. 3. We can see that this quadratic estimator is able to successfully recapture large-scale over/under densities. As further evidence of this, Fig. 4 shows the large-scale power spectrum obtained using the quadratic estimator on the halos and compares it with the theoretical matter power spectrum [Scott: is this right? is the spectrum

linear?]. Also shown are the error bars on the inferred spectrum using Eq. (7). Using halos to infer the matter power spectrum appears to work well in this case when the halo masses are well-determined.
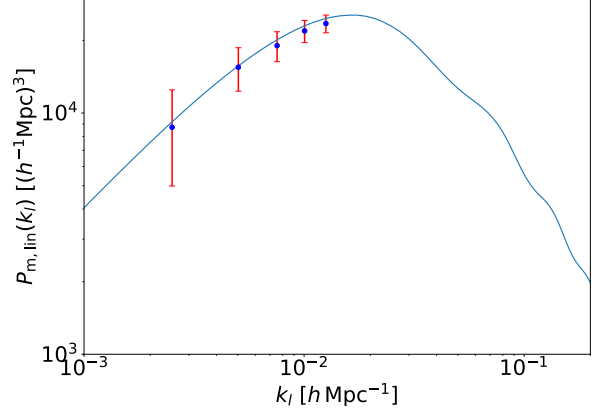

FIG. 4. Long-wavelength power spectrum using the Big-MDPL halo catalog, and its error from Eq. (7) which can be expressed as $P_{\mathrm{m,lin}}(k_l)\sigma(k_l)$. The box size of the survey is be $L = 2.5 h^{-1}$ Gpc thus volume $V = L^3$ and width $\Delta k = 2\pi/L$. The integration range for $\vec{k}_s$ is from $0.03 h\,\mathrm{Mpc}^{-1}$ to $0.22 h\,\mathrm{Mpc}^{-1}$. [Scott: Is this using halos?]yes [Scott: and the quadratic estimator? Also, explain what the solid curve is.]

IV. LIGHT CONE FORMALISM

The difficulty with applying this formalism to real data is that we are able to observe a galaxy a distance r from us only as it was when the scale factor was equal to $a(r)$. Here $a = 1$ when $r = 0$ and decreases as the distance from us increases. Therefore, we do not have access to the full $\delta_{\mathrm{m}}^{\mathrm{h}}(\vec{r}; a)$ for all values of $\vec{r}$, only for that thin shell from which light emitted when the scale factor was equal to a would have reached us by now.

To deal with this, we first define a quantity – the modified matter density contrast that can be observed:

$$d_{\mathrm{m}}(\vec{r}) \equiv \delta_{\mathrm{m}}(\vec{r}; a(r)) \frac{D_{\mathrm{ini}}}{D_1(a(r))} \qquad (12)$$

Note that – even apart from the ratio of growth factors – $d_{\mathrm{m}}(\vec{r})$ is very different than $\delta_{\mathrm{m}}(\vec{r}; a)$. The latter is defined at a fixed value of the scale factor and requires knowledge of the density at all positions at that value of a, knowledge that we cannot obtain. The former, d_{m}, is the density at the position $\vec{r}$ at the time at which we observe it $a(|\vec{r}|)$. That is, this quantity is defined within a light cone. We multiply this observable by the ratio of growth factors so that the large scale mode amplitudes at different times will share the same normalization.

We can take advantage of this definition to generalize the quadratic estimator to the real-life situation of light-cone observations. Consider the Fourier transform

$$d_{\rm m}(\vec{k}) = \int_V d^3\vec{r}\, d_{\rm m}(\vec{r}) e^{-i\vec{k}\cdot\vec{r}}$$
$$= \int_V d^3\vec{r}\, \delta_{\rm m}(\vec{r}; a(r)) \frac{D_{\rm ini}}{D_1(a(r))} e^{-i\vec{k}\cdot\vec{r}} \quad (13)$$

where V is the volume of the light cone. We can make progress by writing

$$\delta_{\rm m}(\vec{r}; a(r)) = \int \frac{d^3\vec{k}}{(2\pi)^3} e^{i\vec{k}\cdot\vec{r}} \delta_{\rm m}(\vec{k}; a(r)) \quad (14)$$

so that

$$d_{\rm m}(\vec{k}) = \int \frac{d^3\vec{k}'}{(2\pi)^3} \int_V d^3\vec{r}\, e^{i(\vec{k}'-\vec{k})\cdot\vec{r}}$$
$$\times \left[\frac{D_{\rm ini}}{D_1(a(r))} \delta_{\rm m}(\vec{k}'; a(r)) \right]. \quad (15)$$

In the case of the snapshot the term in brackets had no $\vec{r}$ dependence, so the integral over space led to a Dirac delta function, and then the integral over $\vec{k}'$ to a tautology. Now, though, we have an expression for an observable $d_{\rm m}(\vec{k})$ in terms of a theoretical quantity $\delta_{\rm m}(\vec{k}'; a(r))$ that cannot be observed but whose behavior is governed by perturbation theory as outlined in Eq. (1). Therefore, perturbation theory predicts that

$$d_{\rm m}(\vec{k}) = d_{\rm m}^{(1)}(\vec{k}) + d_{\rm m}^{(2)}(\vec{k}) + \cdots \quad (16)$$

with each order given by:

$$d_{\rm m}^{(i)}(\vec{k}) = \int \frac{d^3\vec{k}'}{(2\pi)^3} \int_V d^3\vec{r}\, e^{i(\vec{k}'-\vec{k})\cdot\vec{r}}$$
$$\times \left[\frac{D_{\rm ini}}{D_1(a(r))} \delta_{\rm m}^{(i)}(\vec{k}'; a(r)) \right]. \quad (17)$$

Theoretically, we know that

$$\langle \delta_{\rm m}^{(1)}(\vec{k}; a) \delta_{\rm m}^{(1)}(\vec{k}'; a') \rangle$$
$$= \frac{D_1(a)}{D_{\rm ini}} \frac{D_1(a')}{D_{\rm ini}} \langle \delta_{\rm m,ini}^{(1)}(\vec{k}_1) \delta_{\rm m,ini}^{(1)}(\vec{k}_2) \rangle$$
$$= \frac{D_1(a)}{D_{\rm ini}} \frac{D_1(a')}{D_{\rm ini}} (2\pi)^3 \delta_{\rm D}(\vec{k}+\vec{k}') P_{\rm m,ini}(k) \quad (18)$$

where $\delta_{\rm m,ini}^{(1)}(\vec{k}) \equiv \delta_{\rm m}^{(1)}(\vec{k}; a_{\rm ini})$ and where $P_{\rm m,ini}$ is the linear matter power spectrum at the initial time $a_{\rm ini}$. Equating the second and third lines shows that the first order term of $d_{\rm m}^{(1)}$ satisfies [Scott: why is this approximate? Isn't is exact?]

$$\langle d_{\rm m}^{(1)}(\vec{k}) d_{\rm m}^{(1)}(\vec{k}') \rangle \equiv (2\pi)^3 \delta_{\rm D}(\vec{k}+\vec{k}') P_{\rm m,ini}(k). \quad (19)$$

This relation Eq. (19) tells us that the first order term of the modified light cone matter density contrast $d_{\rm m}(\vec{k})$

(which corresponds to long-wavelength modes) still characterizes the linear evolution information of the light cone, since the redshift difference in the light cone has been cancelled by the extra factor $D_{\rm ini}/D_1(a(r))$. We define the light cone matter power spectrum $\mathcal{P}_{\rm m}(k)$ as:

$$\langle d_{\rm m}(\vec{k}) d_{\rm m}(\vec{k}') \rangle \equiv (2\pi)^3 \delta_{\rm D}(\vec{k}+\vec{k}') \mathcal{P}_{\rm m}(k) \quad (20)$$

Fig. 5 shows that this power spectrum is approximately equal to the matter power spectrum on large scales. There are also off-diagonal terms in Eq. (20) which contains extremely valuable information, and we will look into them in next section V.
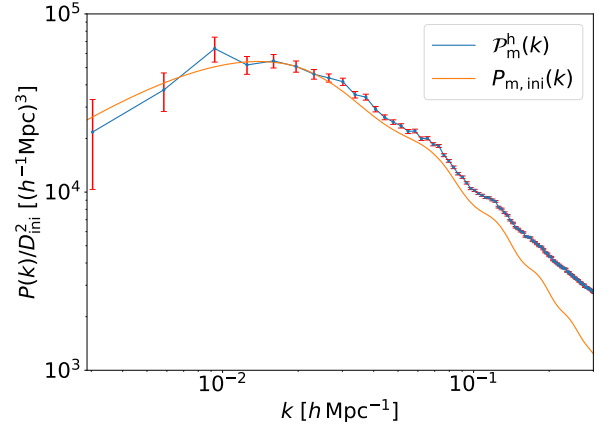

FIG. 5. [Scott: I think it would be simplest to remove the line connecting the points with error bars.] Comparison of the light cone power spectrum $\mathcal{P}_{\rm m}^{\rm h}(k)$ (the superscript "h" means $\mathcal{P}$ is computed using halo information as was $P_{\rm m}^{\rm h}$ in Fig. 2) with the initial power spectrum $P_{\rm m,ini}(k)$, both scaled by $1/D_{\rm ini}^2$. Here $P_{\rm m,ini}(k)$ is plotted according to the? theory predicted linear power spectrum, and $\mathcal{P}_{\rm m}^{\rm h}(k)$ is computed using the? halo catalog of MICE Grand Challenge light cone simulation [30][31]. Halo mass bin is chosen to be $[2.2 \times 10^{12} h^{-1} M_\odot, 10^{14} h^{-1} M_\odot]$. They match each other pretty well on large scales ($k \lesssim 0.02\, h\,\mathrm{Mpc}^{-1}$) despite tiny systematic error. The error bars shows errors due to limited number of independent modes in each bin.

[Scott: I think there is enough complicated notation with 4 or so different delta's and a similar number of different power spectra that it makes sense to have a table that explains/defines each of them.]

V. QUADRATIC ESTIMATOR

¡¡¡¡¡¡¡ HEAD Using the expressions of Eq. (17), we can compute the two-point correlations of two short-wavelength Fourier modes $d_{\rm m}(\vec{k}_s)$ and $d_{\rm m}(\vec{k}'_s)$ up to second order. Again we assume the squeezed limit $\vec{k}_l = \vec{k}_s + \vec{k}'_s$: ======= Using Eq. (17), we can compute the

two-point correlations of two short-wavelength Fourier modes $d_{\rm m}(\vec{k}_s)$ and $d_{\rm m}(\vec{k}'_s)$ where the sum of the two is chosen to be a long wavelength mode, $\vec{k}_l = \vec{k}_s + \vec{k}'_s$:
¿¿¿¿¿¿¿ 2b3be61f28790afbee962f03cb038d37f9539495

$$\langle d_{\rm m}(\vec{k}_s)d_{\rm m}(\vec{k}'_s)\rangle|_{\vec{k}_s+\vec{k}'_s=\vec{k}_l}$$
$$=\langle d_{\rm m}^{(1)}(\vec{k}_s)d_{\rm m}^{(2)}(\vec{k}'_s)\rangle + \langle d_{\rm m}^{(2)}(\vec{k}_s)d_{\rm m}^{(1)}(\vec{k}'_s)\rangle. \qquad (21)$$

Substituting Eq. (17) into Eq. (21) and evaluating the first bracket as an example:

$$\langle d_{\rm m}^{(1)}(\vec{k}_s)d_{\rm m}^{(2)}(\vec{k}'_s)\rangle$$
$$=\int \frac{d^3\vec{k}}{(2\pi)^3}\int \frac{d^3\vec{k}'}{(2\pi)^3}\int_V d^3\vec{r}\int_V d^3\vec{r}'$$
$$e^{i(\vec{k}\cdot\vec{r}+\vec{k}'\cdot\vec{r}')-i(\vec{k}_s\cdot\vec{r}+\vec{k}'_s\cdot\vec{r}')}\frac{D_{\rm ini}}{D_1(a(r))}\frac{D_{\rm ini}}{D_1(a(r'))}$$
$$\times \langle \delta_{\rm m}^{(1)}(\vec{k};a(r))\delta_{\rm m}^{(2)}(\vec{k}';a(r'))\rangle \qquad (22)$$

We have computed $\langle \delta_{\rm m,ini}^{(1)}(\vec{k})\delta_{\rm m,ini}^{(2)}(\vec{k}')\rangle$ in our previous work [7], the result gives:

$$\langle \delta_{\rm m,ini}^{(1)}(\vec{k})\delta_{\rm m,ini}^{(2)}(\vec{k}')\rangle = 2F_2(-\vec{k},\vec{k}+\vec{k}')P_{\rm m,ini}(k)\delta_{\rm m,ini}^{(1)}(\vec{k}+\vec{k}') \qquad (23)$$

We can use this result to further determine the value of the bracket in Eq. (22):

$$\langle \delta_{\rm m}^{(1)}(\vec{k};a(r))\delta_{\rm m}^{(2)}(\vec{k}';a(r'))\rangle$$
$$=2\frac{D_1(a(r))}{D_{\rm ini}}\left[\frac{D_1(a(r'))}{D_{\rm ini}}\right]^2\langle \delta_{\rm m,ini}^{(1)}(\vec{k})\delta_{\rm m,ini}^{(2)}(\vec{k}')\rangle$$
$$=2\left[\frac{D_1(a(r'))}{D_{\rm ini}}\right]^2 F_2(-\vec{k},\vec{k}+\vec{k}')P_{\rm m,ini}(k)\delta_{\rm m}^{(1)}(\vec{k}+\vec{k}';a(r))$$
$$(24)$$

where we make use of the definition of the the linear growth factor $D_1(a(r))$. After plugging Eq. (24) into Eq. (22), the $\vec{r}'$ integral becomes

$$\int_V d^3\vec{r}'e^{-i(\vec{k}'-\vec{k}'_s)\vec{r}'}\frac{D_1(a(r'))}{D_{\rm ini}} \simeq C\,(2\pi)^3\delta_{\rm D}(\vec{k}'-\vec{k}'_s) \qquad (25)$$

since $D_1(a(r'))$ is a slowly varying function. See appendix A for a more detailed analysis. The constant C can be further determined via integrating over $\vec{k}'$ on both sides of Eq. (25), which gives:

$$C = \frac{D_1(a=1)}{D_{\rm ini}}. \qquad (26)$$

Therefore, [Scott: I think the second equality is exact, just redefining k, is that right? Also, changed the argument of δ_m in the 3rd equality. I also reworded after this.

Please check.]

$$\langle d_{\rm m}^{(1)}(\vec{k}_s)d_{\rm m}^{(2)}(\vec{k}'_s)\rangle$$
$$\simeq 2C\int_V d^3\vec{r}\int \frac{d^3\vec{k}}{(2\pi)^3}F_2(\vec{k},-\vec{k}+\vec{k}'_s)P_{\rm m,ini}(|\vec{k}-\vec{k}'_s|)$$
$$\times \frac{D_{\rm ini}}{D_1(a(r))}e^{-i(\vec{k}-\vec{k}_s-\vec{k}'_s)\vec{r}}\delta_{\rm m}^{(1)}(\vec{k};a(r))$$
$$=2C\,F_2(-\vec{k}_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k_s)\int \frac{d^3\vec{k}}{(2\pi)^3}$$
$$\times \int_V d^3\vec{r}\frac{D_{\rm ini}}{D_1(a(r))}e^{-i(\vec{k}-\vec{k}_s-\vec{k}'_s)\vec{r}}\delta_{\rm m}^{(1)}(\vec{k}+\vec{k}'_s;a(r))$$
$$=2C\,F_2(-\vec{k}_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k_s)d_{\rm m}^{(1)}(\vec{k}_s+\vec{k}'_s) \qquad (27)$$

where in the second step, we redefined the dummy variable from $\vec{k} \to \vec{k}-\vec{k}'_s$. Then the integral over d^3r is a Dirac delta function, leaving the final result.

With this definition of $d_{\rm m}$, then, we can obtain the long-wavelength modes from the off-diagonal two-point functions of short-wavelength modes:

$$\langle d_{\rm m}(\vec{k}_s)d_{\rm m}(\vec{k}'_s)\rangle|_{\vec{k}_s+\vec{k}'_s=\vec{k}_l} = f(\vec{k}_s,\vec{k}'_s)d_{\rm m}^{(1)}(\vec{k}_l) \qquad (28)$$

with

$$f(\vec{k}_s,\vec{k}'_s) = 2C\,F_2(-\vec{k}_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k_s)$$
$$+ 2C\,F_2(-\vec{k}'_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k'_s). \qquad (29)$$

Note that f here differs from the f defined in Eq. (3) by a factor of C only.

The quadratic estimator can be similarly formed as:

$$\hat{d}_{\rm m}^{(1)}(\vec{k}_l) = \mathcal{A}(\vec{k}_l)\int \frac{d^3\vec{k}_s}{(2\pi)^3}g(\vec{k}_s,\vec{k}'_s)d_{\rm m}(\vec{k}_s)d_{\rm m}(\vec{k}'_s) \qquad (30)$$

with $\vec{k}'_s = \vec{k}_l - \vec{k}_s$ and g being the weighting function. By requiring that $\langle\hat{d}^{(1)}(\vec{k}_l)\rangle = d^{(1)}(\vec{k}_l)$ we can similarly determine the normalization function $\mathcal{A}$:

$$\mathcal{A}(\vec{k}_l) = \left[\int \frac{d^3\vec{k}_s}{(2\pi)^3}g(\vec{k}_s,\vec{k}'_s)f(\vec{k}_s,\vec{k}'_s)\right]^{-1} \qquad (31)$$

Similar to our last work, by minimizing the noise we get the expression for the weighting function g:

$$g(\vec{k}_s,\vec{k}'_s) = \frac{f(\vec{k}_s,\vec{k}'_s)}{2\mathcal{P}_{\rm m}(k_s)\mathcal{P}_{\rm m}(k'_s)}$$
$$=C\frac{F_2(-\vec{k}_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k_s)+F_2(-\vec{k}'_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k'_s)}{\mathcal{P}_{\rm m}(k_s)\mathcal{P}_{\rm m}(k'_s)}$$
$$(32)$$

with this choice of g the noise term $\mathcal{N}$ is identical to the normalization factor $\mathcal{A}$. And the projected detectability is defined similarly as Eq. (7):

$$\frac{1}{\sigma(k_l)^2} = \frac{Vk_l^2\Delta k}{(2\pi)^2}\left[\frac{P_{\rm m,ini}(k_l)}{P_{\rm m,ini}(k_l)+\mathcal{A}(k_l)}\right]^2, \qquad (33)$$

Using the quadratic estimator Eq. (30) we can use small scale information of the whole light cone (although below we consider only a cube-shaped sub-volume) to infer the large scale field of the modified matter density contrast $d_{\mathrm{m}}(\vec{r})$. According to Eq. (19), this is equivalent to the large scale linear matter perturbations in the light cone.

VI. DEMONSTRATION WITH N-BODY SIMULATION II

A. Real Space

We use the MICE Grand Challenge light cone N-body simulation (MICE-GC) to demonstrate the power of the estimator in a light cone. The catalog contains one octant of the full sky up to $z = 1.4$ (comoving distance $3072\,h^{-1}\,\mathrm{Mpc}$) without simulation box repetition. We use the largest cube that can be fit into the octant as the region V of integration in Eq. (13), as shown in Fig. 6. This simulation used a flat ΛCDM model with cosmological parameters $\Omega_{\mathrm{m}} = 0.25$, $\sigma_8 = 0.8$, $n_{\mathrm{s}} = 0.95$, $\Omega_{\mathrm{b}} = 0.044$, $\Omega_{\Lambda} = 0.75$, $h = 0.7$. We also assume [Scott: Change "assume" to "normalize" and drop the footnote? I don't think this is an assumption, just a normalization choice.] $D_{\mathrm{ini}} = 1$ throughout this section[2].
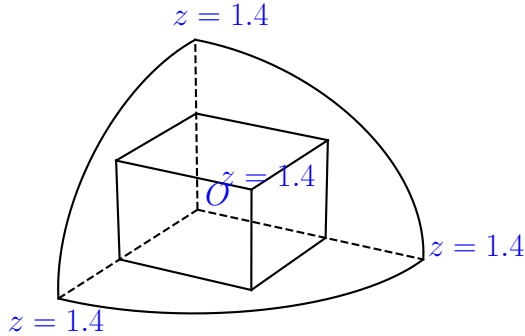
FIG. 6. Boxsize of the cube is $L = 3072/\sqrt{3}\,h^{-1}\,\mathrm{Mpc} \sim 1774\,h^{-1}\,\mathrm{Mpc}$. Volume is $V = L^3 \sim 5.6\,(h^{-1}\,\mathrm{Mpc})^3$. Only one point in the cube can reach the redshift of 1.4, and $z = 0$ is at the origin O of the octant.

---

[2] $D_{\mathrm{ini}} = 1$ is physically ill-defined, because the maximum value of D_{ini} is about 0.997 for a flat universe with $\Omega_{\mathrm{m}} = 0.25$. But D_{ini} is simply a normalization constant and the choice of its value does not affect the feature of the final result.

We first extract 1 in 700 matter particles' positions in the full light cone of the MICE-GC simulation and then focus on the largest cube within it. This corresponds to $\sim 5.14 \times 10^7$ particles in total each with $2.9 \times 10^{10} h^{-1} M_{\odot}$ particle mass. Thus we have a number density of 3.4×10^{-3} particles/(Mpc/h)^3, which is similar to Dark Energy Survey (DES) [32] full sample of galaxies. We use Nbodykit [33] to compute the modified matter density contrast field d_m in Fourier space and get the estimated one $\hat{d}_m$ using Eq. (30). Than we transform them back into real space and compare them it Fig. 7. The cube with volume $\sim 5.6\,(h^{-1}\,\mathrm{Gpc})^3$ is divided into 7^3 cells. Each row in Fig. 7 contains 7 panels; each panel is a slice of the cube. We see that the quadratic estimator does an excellent job of extracting the large scale density - it reproduces almost every cell with large over- or under-densities, and the difference is much smaller than the density field itself.

In Fig. 7, the left top cell of each row corresponds to the cell containing the origin ($z = 0$) in it; and the right bottom cell corresponds to the farthest cell from the origin ($z \sim 1.4$). We can use Fig. 8 to get a better view of this in a light cone, where in each row we still show the usual directly measure field $d_{\mathrm{m}}(\vec{r})$, estimated field $\hat{d}_{\mathrm{m}}(\vec{r})$ and their difference $(d_{\mathrm{m}} - \hat{d}_{\mathrm{m}})(\vec{r})$ respectively. We have two columns here, the left column shows cells near the origin and the right column shows cells far from the origin. While all the information in this figure is contained in Fig. 7, it still is a more straightforward way of expressing our result in a light cone.
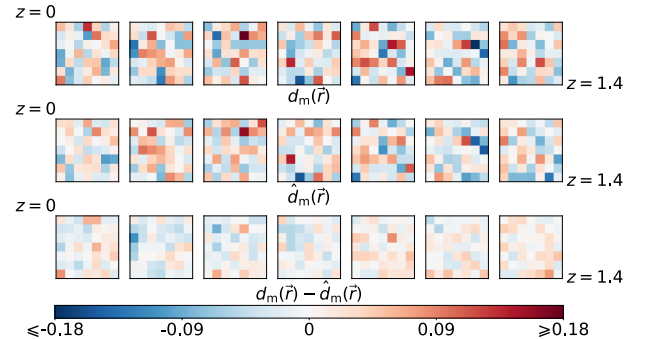

FIG. 7. Comparison of the true density field of the 1 in 700 matter particles of the Mice-GC simulation ($\delta_{\mathrm{m}}(\vec{r})$, top row) and the density field from the quadratic estimator ($\hat{\delta}_{\mathrm{m}}(\vec{r})$, middle row). The bottom row shows their difference. Each panel represents a slice through the simulation volume, $1774\,h^{-1}\,\mathrm{Mpc}$ wide, and one cell $\sim 253\,h^{-1}\,\mathrm{Mpc}$ thick. The integration range of $\vec{k}_s$ is from $0.03\,h\,\mathrm{Mpc}^{-1}$ to $0.22\,h\,\mathrm{Mpc}^{-1}$.

Then we consider the halo catalog in the same light cone with halo masses between $2.2 \times 10^{12} h^{-1} M_{\odot} < M <$
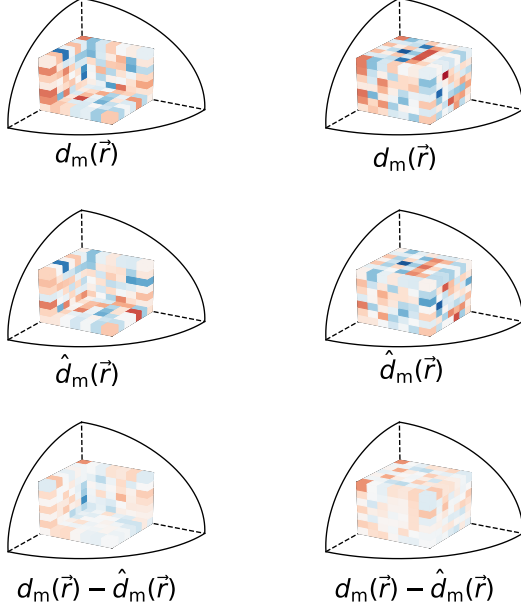
FIG. 8. Comparison of the true density field and the estimated density field of dark-matter-only catalog in a light cone with the same colorbar as Fig. 7. First row shows the modified matter density contrast near the origin (inner surfaces of the cube, $0 \leqslant z \lesssim 0.69$); second row shows the modified matter density contrast far from the origin (outer surfaces of the cube, $0.69 \lesssim z \lesssim 1.4$).

$10^{14}h^{-1}M_{\odot}$. We obtained similar results in other mass bins as well. We use Eq. (9) as the approximate matter density contrast, $n_{\mathrm{h}}(\vec{r})$ [Scott: What is n_h? How does it differ from n_h?] can be easily computed using halo positions and the Tinker bias function. We cut the whole light cone (the octant) into several slices and compute the mean density field $\bar{n}_{\mathrm{h}}$ and $\bar{n}_{\mathrm{h}}$ of each slice, then get the mean field $\bar{n}_{\mathrm{h}}(a(r))$ and $\bar{n}_{\mathrm{h}}(a(r))$ at each position by interpolation. We compute the directly measured modified matter contrast $d_{\mathrm{m}}^{\mathrm{h}}(\vec{r})$ and also the power spectrum in Fig. 5. The power spectrum agrees well with the linear matter power spectrum on large scales. We use the quadratic estimator Eq. (30) to get the reconstructed modified field and transform them back into real space. Again we have two similar plots Fig. 9 and Fig. 10. From Fig. 9 we can see that our quadratic estimator is still able to extract large scale information, especially large over- or under-density cells in the first few low-redshift panels. The difference seems larger when we go to higher redshift and is worst on the very right panel. We observe the same feature in Fig. 10 where the cells near the origin match each other better than cells far from the origin. The main reason for this effect is that nonlinear bias (b_2, $b_3 \cdots$) becomes more important when we go deeper into the light cone [34] [Scott: Why is this? I would have
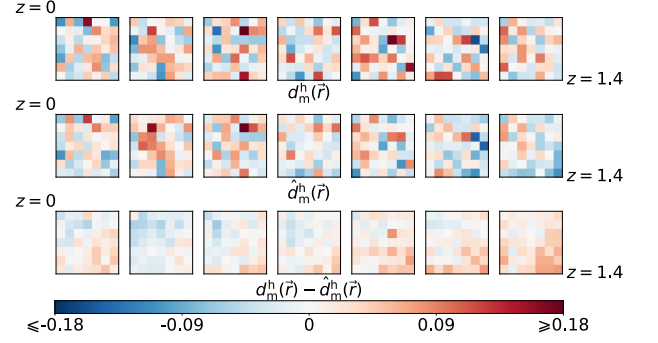
thought higher redshift was simpler?].

FIG. 9. Comparison of the true density field of the halo number density contrast in mass bin $2.2 \times 10^{12}h^{-1}M_{\odot} < M < 10^{14}h^{-1}M_{\odot}$ of the Mice-GC simulation ($\delta_{\mathrm{m}}(\vec{r})$, top row) and the density field from the quadratic estimator ($\hat{\delta}_{\mathrm{m}}(\vec{r})$, middle row) in the same light cone as Fig. 7, and their difference (bottom row). The integration range of $\vec{k}_s$ is from $0.03\,h\,\mathrm{Mpc}^{-1}$ to $0.35\,h\,\mathrm{Mpc}^{-1}$. We use a larger upper limit to slightly reduce the noise term.

B. Redshift Space

In a spectroscopic survey, the observed distribution of halos (and galaxies) is distorted and squashed when we use their redshift as an indicator of their radial distance due to halos' peculiar velocity. In the plane-parallel approximation, the mapping from real space to redshift space is given by (e.g., see [19]):

$$\vec{r}_{\mathrm{rs}} = \vec{r} + \frac{(\vec{u} \cdot \hat{n})\hat{n}}{aH(a)} \tag{34}$$

where $\vec{r}_{\mathrm{rs}}$ denotes redshift space coordinates, $\vec{u}$ is the peculiar velocity field, $\hat{n}$ is the direction of the line of sight, and $H(a)$ is the Hubble parameter. As before, the matter density contrast in the redshift space can be written as a perturbative series:

$$\delta_{\mathrm{m,rs}}(\vec{k};a) = \sum_{n=1}^{\infty} \int \frac{d^3\vec{k}_1}{(2\pi)^3} \cdots \int \frac{d^3\vec{k}_n}{(2\pi)^3} \delta_{\mathrm{D}}(\vec{k} - \vec{k}_1 - \cdots - \vec{k}_n)$$
$$\times Z_n(\vec{k}_1, \cdots, \vec{k}_n; a)\delta_{\mathrm{m}}^{(1)}(\vec{k}_1;a) \cdots \delta_{\mathrm{m}}^{(1)}(\vec{k}_n;a) \tag{35}$$

where now the window functions are

$$Z_1(\vec{k};a) = 1 + f\mu^2 \tag{36}$$
$$Z_2(\vec{k}_1,\vec{k}_2;a) = F_2(\vec{k}_1,\vec{k}_2) + f\mu^2 G_2(\vec{k}_1,\vec{k}_2)$$
$$+ \frac{f\mu k}{2}\left[\frac{\mu_1}{k_1}(1 + f\mu_2^2) + \frac{\mu_2}{k_2}(1 + f\mu_1^2)\right] \tag{37}$$
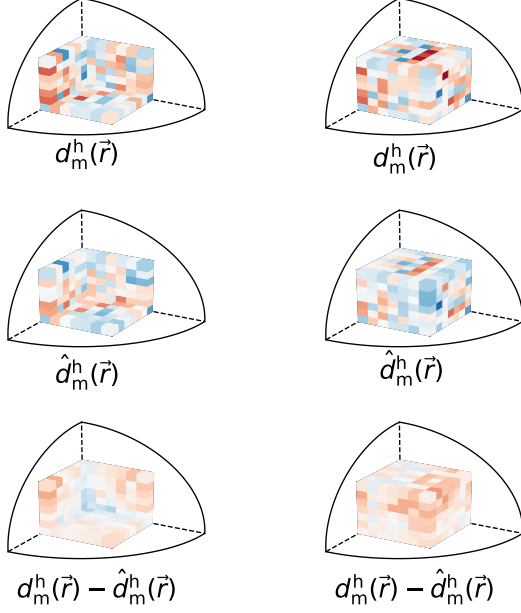
FIG. 10. Comparison of the true density field and the estimated density field of halo catalog in a light cone with the same colorbar as Fig. 7 (or Fig. 9). First row shows the modified matter density contrast near the origin (inner surfaces of the cube, $0 \leqslant z \lesssim 0.69$); second row shows the modified matter density contrast far from the origin (outer surfaces of the cube, $0.69 \lesssim z \lesssim 1.4$).

with $\mu \equiv \vec{k} \cdot \hat{n}/k$ and $\mu_{1,2} \equiv \vec{k}_{1,2} \cdot \hat{n}/k_{1,2}$. The window functions now depend on the growth rate

$$f(a) \equiv \frac{d \ln D_1(a)}{d \ln a} \tag{38}$$

and on

$$G_2(\vec{k}_1, \vec{k}_2) = \frac{3}{7} + \frac{4}{7} \frac{(\vec{k}_1 \cdot \vec{k}_2)^2}{k_1^2 k_2^2} + \frac{\vec{k}_1 \cdot \vec{k}_2}{2 k_1 k_2} \left[\frac{k_1}{k_2} + \frac{k_2}{k_1} \right]. \tag{39}$$

The modified matter density contrast in the redshift space of a light cone can then be expressed as:

$$d_{\mathrm{m,rs}}(\vec{k}) = \int_V d^3\vec{r} \int \frac{d^3\vec{k}'}{(2\pi)^3} e^{i(\vec{k}'-\vec{k})\cdot\vec{r}} \delta_{\mathrm{m,rs}}(\vec{k}'; a(r)) \frac{D_{\mathrm{ini}}}{D_1(a(r))} \tag{40}$$

One can prove that this definition still satisfies the relation (see appendix B for a detailed derivation):

$$\langle d_{\mathrm{m,rs}}(\vec{k}_s) d_{\mathrm{m,rs}}(\vec{k}'_s) \rangle|_{\vec{k}_s+\vec{k}'_s=\vec{k}_l} = f_{\mathrm{rs}}(\vec{k}_s, \vec{k}'_s) d_{\mathrm{m,rs}}^{(1)}(\vec{k}_l) \tag{41}$$

with

$$f_{\mathrm{rs}}(\vec{k}_s, \vec{k}'_s) = 2C \frac{Z_1(\vec{k}_s; a(r \simeq L/2))}{Z_1(\vec{k}_s + \vec{k}'_s; a(r \simeq L/2))}$$
$$\times \left[Z_2(-\vec{k}_s, \vec{k}_s + \vec{k}'_s; a = 1) P_{\mathrm{m,ini}}(k_s) \right.$$
$$\left. + Z_2(-\vec{k}'_s, \vec{k}_s + \vec{k}'_s; a = 1) P_{\mathrm{m,ini}}(k'_s) \right] \tag{42}$$

where $L = 3072 \, h^{-1}$ Mpc when we use MICE-GC simulation which is the depth of the simulation. It comes from the approximation when we take the ratio term out of the integrand. The redshift dependence of the f function is solved as in Eq. (25).

In Fig. 11 we use the halo catalog of MICE-GC simulation to reconstruct the monopole moment of the redshift space matter power spectrum.
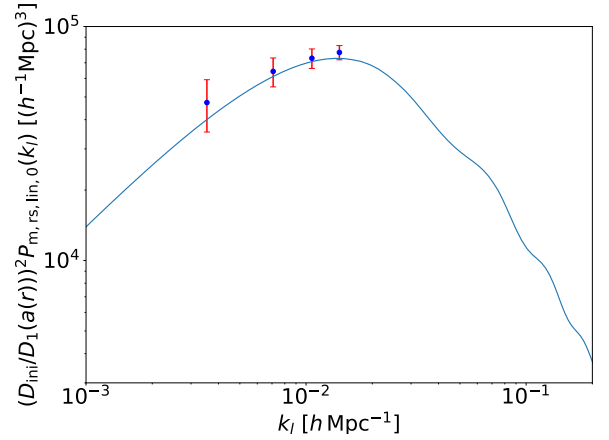

FIG. 11. Solid curve is the theoretical curve for the monopole moment of the linear redshift space matter power spectrum from Kaiser formula (detailed discussion in appendix B). Blue dots and their corresponding error bars are measured using the quadratic estimator in redshift space. The integration range for $\vec{k}_s$ is from $0.03 \, h \, \mathrm{Mpc}^{-1}$ to $0.35 \, h \, \mathrm{Mpc}^{-1}$.

VII. CONCLUSION

A. Summary

In prior work [7] we have shown that the amplitude and phase of large scale density fluctuations can be recovered by applying a quadratic estimator to measurements of small scale Fourier modes and their correlations. In this paper we extend that work (which was limited to a matter density field at a single instance in cosmic time) in the following three ways in order to make it applicable to observational data. All extensions are tested on

appropriate mock survey datasets derived from N-body simulations.

1. Halos bias. We allow for the effect of clustering bias between galaxies and dark matter by including an appropriate weighting scheme applied to individual halos.

2. Light cone effect. The redshift difference between parts of large volume surveys can be considerable. We prove that using a suitable weighting involving the linear growth factor, the quadratic estimator can correctly recover large scale modes from small wavelength modes measured on the light cone.

3. Redshift space distortions. We prove that the relevant relationship between small and large scale modes applies in redshift space, and show how redshift space distortions can be included in our formalism.

B. Discussion

Our formalism includes the major effects that are relevant for an application to observational data. There are some minor aspects however which will need to be dealt with when this occurs. One is the fact that we have tested on homogeneous mock surveys, when real observations will include masked data (to account for bright stars for example), and a potentially complex window function. We have also used a cubical volume excised from the full light cone in our mock surveys in order to compute Fourier space clustering. This simplification means that a significant quantity of data is not used for analysis. It will therefore be useful to develop an approach to account for the effect of convolution with the window function in Fourier space. Because we are concerned with the direct measurement of small scale modes this should not be such a large effect, particularly for surveys with significant sky coverage.

At present the large-scale limitations on direct measurement of clustering are observational systematics (e.g., [35]). These include angular variations in obscuration, seeing, sky brightness, colors, extinction and magnitude errors. Because these result in relatively small modulations of the measured galaxy density, they will affect large scale modes most importantly, hence the utility of our indirect measurements of clustering on these scales. Quantification of these effects on the scales for which we do measure clustering will still be needed though. It will be also be instructive to apply large-scale low amplitude modulations to our mock surveys in order to test how well the quadratic estimator works with imperfect data. Even small scale issues with clustering, such as fiber collisions [36] could affect our reconstruction, depending on how their effects propagate through the quadratic estimator.

Observational datasets exist at present which could be used to carry out measurements using our methods.

These include the SDSS surveys BOSS [37] and eBOSS [38] (both luminous red galaxies and emission line galaxies). Substantial extent in both angular coordinates and redshift are necessary, so that deep but narrow surveys such as VIMOS [39] or DEEP2 [40] would not be suitable. In the near future, the available useful data will increase rapidly with the advent of WEAVE [41] and DESI [10]. Space based redshift surveys with EUCLID [42] and WFIRST [9] will expand the redshift range, and SPHEREx [43], due for launch even earlier will offer maximum sky coverage, and likely the largest volume of all.

In order to model what is expected from all these datasets, the effective range of wavelengths used in the reconstruction of large-scale modes should be considered. Surveys covering large volumes but with low galaxy number density will have large shot noise contributions to density fluctuations, and this will limit the range of scales that can be used. For example, in our present work we have successfully tested number densities of $\sim 3 \times 10^{-3}$ galaxies per $(Mpc/h)^3$. Surveys such as the eBOSS quasar redshift survey [44] with a number density ~ 100 times lower will not be useful, for example.

Once an indirect measurement of large scale modes has been made from an observational dataset, there are many different potential applications. We can break these up into two groups, involving the power spectrum itself, and the map (and statistics beyond $P_m(k)$) which can be derived from it.

First, because of the effect of observational systematics mentioned above, and the fact the our indirect estimate of clustering is sensitive to fluctuations beyond the survey boundaries itself, then it is likely that the measurement we propose would correspond to the largest scale estimate of three dimensional matter clustering yet made. This would in itself be an exciting test of theories, for example probing the power spectrum beyond the matter-radiation equality turnover, and allowing access to the Harrison-Zeldovich portion. There has been much work analyzing large-scale anomalies in the clustering measured from the CMB [45][46][47], and it would be extremely useful to see if anything comparable is seen from galaxy large-scale structure data. On smaller scales, one could use the matter-radiation equality turnover as a cosmic ruler [48], and this would allow comparison to measurements based on BAO [49].

Second, there will be much information in the reconstructed maps of the large-scale densities (such as Fig. 9). One could look at statistics beyond the power spectrum, such as counts-in-cells [50], or the bispectrum, and see how consistent they are with model expectations. One can also compare to the directly measured density field and obtain information on the large-scale systematic effects which are modulating the latter. Cross-correlation of the maps with those of different tracers can also be carried out. For example the large-scale potential field inferred can be used in conjunction with CMB observations to constrain the Integrated Sachs Wolfe effect[51].

In general, as we will be looking at large scale fluc-

tuations beyond current limits by perhaps an order of magnitude in scale or more, one may expect to find interesting constraints on new physics. For example evidence for the ΛCDM model was seen in the first reliable measurements of large-scale galaxy clustering on scales greater than $10\,h^{-1}$Mpc (e.g., [52]). Moving to wavelengths beyond $2\pi/(k = 0.02) \sim 300$ Mpc may yet lead to more surprises.

ACKNOWLEDGMENTS

Appendix A: Approximation in Eq. (25)

In Fig. 12 we use a simple 1D example to demonstrate that the approximation we made in Eq. (25) is valid:

$$\int_V d^3\vec{r}' e^{-i\vec{q}\cdot\vec{r}'} D_1(a(r')) \simeq D_1(a=1)(2\pi)^3 \delta_D(\vec{q}) \quad \text{(A1)}$$

This 1D result can be also be applied to 3D since...

The blue curve corresponds to the Dirac delta function in a finite volume multiplied by $D_1(a=1)$ (right-hand-side of the above equation), and the orange curve is the left-hand-side. We can see that the orange curve is still sharply peaked at $q=0$.

For a relatively large scale, e.g., $L = 3072\,h^{-1}\mathrm{Mpc}$, we have $\Delta k = 2\pi/L \approx 0.002\,h\,\mathrm{Mpc}^{-1}$. Generally we want to further evaluate the following integral:

$$\int \frac{d^3\vec{q}}{(2\pi)^3} f(\vec{q}) \int_V d^3\vec{r}' e^{-i\vec{q}\cdot\vec{r}'} D_1(a(r')) \quad \text{(A2)}$$

Unless $f(\vec{q})$ is a rapidly fluctuating function at scales $q \approx 0.002$, Eq. (25) remains a good approximation. Recall that $f(\vec{q})$ in our case is a combination of F_2 and δ_m, it is slowly varying compared to the scale $q \approx 0.002$. So it is a convincing approximation based on our analysis.
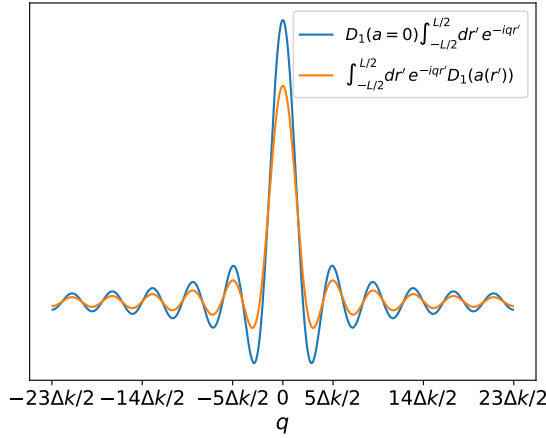

FIG. 12. 1D demonstration of the approximation made in Eq. (25). Here $\Delta k = 2\pi/L$.

Appendix B: Proof of Eq. (41) in Redshift Space

One can prove that the leading order term of of $d_\mathrm{m}(\vec{k})$ still characterizes the linear matter power spectrum in redshift space:

$$\langle d_\mathrm{m}^{(1)}(\vec{k}) d_\mathrm{m}^{(1)}(\vec{k}')\rangle$$
$$= (2\pi)^3 \delta_D(\vec{k}+\vec{k}')(1 + f(a=1)\mu^2)^2 P_{\mathrm{m,ini}}(k)$$
$$= (2\pi)^3 \delta_D(\vec{k}+\vec{k}')\left[\frac{D_\mathrm{ini}}{D_1(a=1)}\right]^2 P_{\mathrm{m,rs,lin}}(\vec{k};a=1) \quad \text{(B1)}$$

where

$$P_{\mathrm{m,rs,lin}}(\vec{k};a) = (1+f(a)\mu^2)^2 P_{\mathrm{m,lin}}(k;a) \quad \text{(B2)}$$

is the leading order redshift space matter power spectrum first derived by N. Kaiser [17]. We can decompose its direction-dependence by Legendre polynomials expansion as:

$$P_{\mathrm{m,rs},\ell}(k;a) \equiv \frac{2\ell+1}{2} \int_{-1}^1 d\mu P_{\mathrm{m,rs}}(\vec{k};a)\mathcal{L}_\ell(\mu) \quad \text{(B3)}$$

The monopole ($\ell = 0$) and the quadrupole ($\ell = 2$) moments have been measured in recent surveys [53]. Now we start our proof of Eq. (41), first we have:

$$\langle \delta_{\mathrm{m,rs}}^{(1)}(\vec{k};a(r)) \delta_{\mathrm{m,rs}}^{(2)}(\vec{k}';a(r'))\rangle$$
$$= Z_1(\vec{k};a(r))\langle \delta_\mathrm{m}^{(1)}(\vec{k};a(r)) \delta_{\mathrm{m,rs}}^{(2)}(\vec{k}';a(r'))\rangle$$
$$= 2Z_1(\vec{k};a(r))\left[\frac{D_1(a(r'))}{D_\mathrm{ini}}\right]^2 Z_2(-\vec{k},\vec{k}+\vec{k}';a(r'))$$
$$\times P_{\mathrm{m,ini}}(k)\delta_\mathrm{m}^{(1)}(\vec{k}+\vec{k}';a(r))$$
$$= \frac{Z_1(\vec{k};a(r))}{Z_1(\vec{k}+\vec{k}';a(r))}\left\{2\left[\frac{D_1(a(r'))}{D_\mathrm{ini}}\right]^2 Z_2(-\vec{k},\vec{k}+\vec{k}';a(r'))\right.$$
$$\left.\times P_{\mathrm{m,ini}}(k)\delta_{\mathrm{m,rs}}^{(1)}(\vec{k}+\vec{k}';a(r))\right\} \quad \text{(B4)}$$

where in the step we use the first order expression of $\delta_\mathrm{m,rs}$ in Eq. (35); and second step can be similarly derived as Eq. (24) with F_2 replaced by Z_2; and in the last step we use the expression again:

$$\delta_{\mathrm{m,rs}}^{(1)}(\vec{k};a(r)) = Z_1(\vec{k};a(r))\delta_\mathrm{m}^{(1)}(\vec{k};a(r)) \quad \text{(B5)}$$

in order to get the similar expression as Eq. (24) in the braces. So when we evaluate $\langle d_\mathrm{m,rs}^{(1)}(\vec{k}_s) d_\mathrm{m,rs}^{(2)}(\vec{k}_s')\rangle$ we get an expression close to Eq. (27):

$$\langle d_\mathrm{m,rs}^{(1)}(\vec{k}_s) d_\mathrm{m,rs}^{(2)}(\vec{k}_s')\rangle$$
$$\simeq 2C\,Z_2(-\vec{k}_s,\vec{k}_s+\vec{k}_s';a=1)P_{\mathrm{m,ini}}(k_s)$$
$$\times \int_V d^3\vec{r} \int \frac{d^3\vec{k}}{(2\pi)^3} \frac{Z_1(\vec{k}-\vec{k}_s';a(r))}{Z_1(\vec{k};a(r))} \frac{D_\mathrm{ini}}{D_1(a(r))}$$
$$e^{-i(\vec{k}-\vec{k}_s-\vec{k}_s')\vec{r}}\delta_\mathrm{m}^{(1)}(\vec{k};a(r))$$
$$\simeq \frac{Z_1(\vec{k}_s;a(r\simeq L/2))}{Z_1(\vec{k}_s+\vec{k}_s';a(r\simeq L/2))}$$
$$\times 2C\,Z_2(-\vec{k}_s,\vec{k}_s+\vec{k}_s';a=1)P_{\mathrm{m,ini}}(k_s)\delta_\mathrm{m}^{(1)}(\vec{k}_s+\vec{k}_s')$$
$$\quad \text{(B6)}$$

Z_2 is fixed at $a = 1$ when we use a similar approximation as Eq. (25). In order to recover $d_{\rm m}^{(1)}(\vec{k}_s + \vec{k}'_s)$ from the integral, we also take the ratio $Z_1(\vec{k} - \vec{k}'_s; a(r))/Z_1(\vec{k}; a(r))$ out of the integrand by assuming that $\vec{k} = \vec{k}_s + \vec{k}'_s$ and $r = L/2$ where L is the depth of the survey. By performing these approximations we have proved Eq. (41).

Another interesting thing to notice in the redshift space construction is that, the Gaussian noise term $\mathcal{A}_{\rm rs}$ is direction-dependent following:

$$\mathcal{A}_{\rm rs}(\vec{k}_l) \propto (1 + f(a = 1)\mu_{\vec{k}_l}^2) \tag{B7}$$

One benefit we can take from this numerical result is that the projected detectability in redshift space still will not depend on the direction of $\vec{k}_l$. So the error bars will be the same for both real space and redshift space.

[1] C. Modi, M. White, A. Slosar, and E. Castorina, J. Cosmol. Astropart. P. **11**, 023 (2019), arXiv:1907.02330 [astro-ph.CO].

[2] T. Baldauf, U. Seljak, L. Senatore, and M. Zaldarriaga, J. Cosmol. Astropart. P. **10**, 031 (2011), arXiv:1106.5507 [astro-ph.CO].

[3] D. Jeong and M. Kamionkowski, Phys. Rev. Lett. **108**, 251301 (2012), arXiv:1203.0302 [astro-ph.CO].

[4] Y. Li, W. Hu, and M. Takada, Phys. Rev. D **90**, 103530 (2014), arXiv:1408.1081 [astro-ph.CO].

[5] H.-M. Zhu, U.-L. Pen, Y. Yu, X. Er, and X. Chen, Phys. Rev. D **93**, 103504 (2016), arXiv:1511.04680 [astro-ph.CO].

[6] A. Barreira and F. Schimidt, J. Cosmol. Astropart. P. **06**, 03 (2017), arXiv:1703.09212 [astro-ph.CO].

[7] P. Li, S. Dodelson, and R. A. C. Croft, Phys. Rev. D **101**, 083510 (2020), arXiv:2001.02780 [astro-ph.CO].

[8] LSST Daek Energy Science Collaboration, arXiv:1211.0310 [astro-ph.CO].

[9] WFIRST Science Definition Team, arXiv:1208.4012 [astro-ph.IM].

[10] DESI Collaboration, arXiv:1907.10688 [astro-ph.IM].

[11] A. V. Kravtsov and A. Klypin, Anatoly, Astrophys. J. **520**, 437 (1999), arXiv:astro-ph/9812311 [astro-ph].

[12] V. Desjacques, D. Jeong, and F. Schmidt, Phys. Rept. **733**, 1 (2018), arXiv:1611.09787 [astro-ph.CO].

[13] U. Seljak, A. Makarov, R. Mandelbaum, C. M. Hirata, N. Padmanabhan, P. McDonald, M. R. Blanton, M. Tegmark, N. A. Bahcall, and J. Brinkmann (SDSS Collaboration), Phys. Rev. D **71**, 043511 (2005), arXiv:astro-ph/0406594.

[14] C. Chang et al. (DES Collaboration), Mon. Not. Roy. Astron. Soc. **459**, 3203 (2016), arXiv:1601.00405 [astro-ph.CO].

[15] J. Prat et al. (DES Collaboration), Mon. Not. Roy. Astron. Soc. **473**, 1667 (2018), arXiv:1609.08167 [astro-ph.CO].

[16] S. M. Carroll, arXiv:gr-qc/9712019 [gr-qc].

[17] N. Kaiser, Mon. Not. Roy. Astron. Soc. **227**, 1 (1987).

[18] B. Jain and E. Bertschinger, Astrophys. J. **431**, 495 (1994), arXiv:astro-ph/9311070 [astro-ph].

[19] F. Bernardeau, S. Colombi, E. Gaztañaga, and Scoccimarro, Phys. Rept. **367**, 1 (2012), arXiv:astro-ph/0112551 [astro-ph].

[20] R. Takahashi, Prog. Theor. Phys. **120**, 549–559 (2008), arXiv:0806.1437 [astro-ph.CO].

[21] U. Seljak and M. S. Warre, Mon. Not. Roy. Astron. Soc. **355**, 129 (2004), arXiv:astro-ph/0403698 [astro-ph].

[22] S. Bhattacharya, K. Heitmann, M. White, Z. Lukić, C. Wagner, and S. Habib, Astrophys. J. **732**, 122 (2011), arXiv:1005.2239 [astro-ph.CO].

[23] J. L. Tinker, B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yeps, and S. Gottlober, Astrophys. J. **724**, 878 (2010), arXiv:1001.3162 [astro-ph.CO].

[24] W. J. Pervical et al., Astrophys. J. **657**, 645 (2007), arXiv:astro-ph/0608636 [astro-ph].

[25] P. Fosalba, M. Crocce, E. Gaztañaga, and F. J. Castander, Mon. Not. Roy. Astron. Soc. **460**, 3624–3636 (2015), arXiv:1512.07295 [astro-ph.CO].

[26] J. G. Cresswell and W. J. Pervical, Mon. Not. Roy. Astron. Soc. **392**, 682 (2008), arXiv:0808.1101 [astro-ph].

[27] A. Klypin, G. Yepes, S. Gottlober, F. Prada, and S. Hess, Mon. Not. Roy. Astron. Soc. **457**, 4340 (2014), arXiv:1411.4001 [astro-ph.CO].

[28] Planck Collaboration, Astron. Astrophys. **571**, 66 (2014), arXiv:1303.5076 [astro-ph.CO].

[29] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, Astrophys. J. **762**, 109 (2013), arXiv:1110.4372 [astro-ph.CO].

[30] P. Fosalba, M. Crocce, E. Gaztañaga, and F. J. Castander, Mon. Not. Roy. Astron. Soc. **448**, 2987 (2015), arXiv:1312.1707 [astro-ph.CO].

[31] M. Crocce, F. J. Castander, E. Gaztañaga, P. Fosalba, and J. Carretero, Mon. Not. Roy. Astron. Soc. **453**, 1513 (2015), arXiv:1312.2013 [astro-ph.CO].

[32] Dark Energy Survey Collaboration, Mon. Not. Roy. Astron. Soc. **460**, 1270–1299 (2016), arXiv:1601.00329 [astro-ph.CO].

[33] N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak, and Z. Slepian, Astrophys. J. **156**, 160 (2018), arXiv:1712.05834 [astro-ph.CO].

[34] T. Lazeyras, C. Wagner, T. Baldauf, and F. Schmidt, J. Cosmol. Astropart. P. **02**, 018 (2016), arXiv:1511.01096 [astro-ph.CO].

[35] S. Ho et al., Astrophys. J. **761**, 24 (2012), arXiv:1201.2137 [astro-ph.CO].

[36] C. Hahn, R. Scoccimarro, M. R. Blanton, J. L. Tinker, and S. A. Rodríguez-Torres, Mon. Not. Roy. Astron. Soc. **467**, 1940 (2017), arXiv:1609.01714 [astro-ph.CO].

[37] K. S. Dawson et al., Astron. J. **145**, 10 (2013), arXiv:1208.0022 [astro-ph.CO].

[38] K. S. Dawson et al., Astron. J. **151**, 44 (2016), arXiv:1508.04473 [astro-ph.CO].

[39] O. Le Fèvre et al., Astron. Astrophys. **576**, A79 (2015), arXiv:1403.3938 [astro-ph.CO].

[40] A. L. Coil, J. A. Newman, M. C. Cooper, M. Davis, S. Faber, D. C. Koo, and C. N. Willmer, Astrophys. J. **644**, 671 (2006), arXiv:astro-ph/0512233.

[41] G. Dalton et al., Proceedings of the SPIE **9147** (2014), 10.1117/12.2055132, arXiv:1412.0843 [astro-ph.CO].

[42] J. Amiaux et al., Proceedings of the SPIE **8842** (2021),

10.1117/12.926513, arXiv:1209.2228 [astro-ph.CO].

[43] O. Doré *et al.*, (2014), arXiv:1412.4872 [astro-ph.CO].

[44] M. Ata *et al.*, Mon. Not. Roy. Astron. Soc. **473**, 4773 (2018), arXiv:1705.06373 [astro-ph.CO].

[45] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, Adv. Astron. **2010**, 847541 (2010), arXiv:1004.5602 [astro-ph.CO].

[46] A. Rassat, J.-L. Starck, P. Paykari, F. Sureau, and J. Bobin, JCAP **08**, 006 (2014), arXiv:1405.1844 [astro-ph.CO].

[47] D. J. Schwarz, C. J. Copi, D. Huterer, and G. D. Starkman, Class. Quant. Grav. **33**, 184001 (2016),

arXiv:1510.07929 [astro-ph.CO].

[48] J. Hasenkamp and J. Kersten, JCAP **08**, 024 (2013), arXiv:1212.4160 [hep-ph].

[49] R. Lazkoz, S. Nesseris, and L. Perivolaropoulos, JCAP **07**, 012 (2008), arXiv:0712.1232 [astro-ph].

[50] A. Yang and W. C. Saslaw, Astrophys. J. **729**, 123 (2011), arXiv:1009.0013 [astro-ph.CO].

[51] A. J. Nishizawa, PTEP **2014**, 06B110 (2014), arXiv:1404.5102 [astro-ph.CO].

[52] G. Efstathiou, W. J. Sutherland, and S. J. Maddox, Nature **348**, 705 (1990).

[53] H. Gil-Marín *et al.*, Mon. Not. Roy. Astron. Soc. **460**, 4188 (2016), arXiv:1509.06386 [astro-ph.CO].