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: March 31, 2020)

This is the second paper in a series where we propose a method of indirectly measuring large scale structure using information of small scale perturbations. The idea is to use two-point off-diagonal terms of density contrast modes to build a quadratic estimator for long-wavelength modes. We demonstrate in the first paper that our quadratic estimator works well on a dark-matter-only N-body simulation of the snapshot $z = 0$. Here we generalize our theory to the case of a light cone with halo bias taken into consideration. We successfully apply our generalized version of the quadratic estimator to a light cone halo catalog of a 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 nearly impossible right now as pointed out by [1]. In our first work [2], we proposed a method of indirectly measuring large scale structure using small scale information. Physically, long- and short-wavelength modes are correlated in the sense that small scale modes will grow differently depending on the large scale structure they reside in. This phenomena leaves a signature that the Fourier 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 our theory to a halo catalog within a light cone for observational requirements.

First need to account for halo bias [3][4] since eventually we want to apply the estimator to future cosmological surveys [5][6][7]. Halo bias is a term relating halo number density perturbation to matter density perturbation. fill in

Observationally a galaxy catalog will be in a light cone [8] instead of a snapshot. Typical treatment is to cut a light cone into several thin redshift bins need some citations here, can't find a proper one, and analyze the properties within each bin. By doing this we will lose information about long-wavelength modes along line of sight. Thus in this paper we propose a novel method of considering all the halos/galaxies in a light cone together, despite redshift difference, simply by multiplying an extra function proportional to the inverse of linear growth function. Since the universe is linear on sufficiently large scales, two linear order matter density modes at different redshift will still be correlated. With this knowledge we can recover linear power spectrum on large scales from our construction. Assuming a cube volume, we can construct the quadratic estimator for long-wavelength modes using information of non-zero off-diagonal terms like what we did before.

We begin with our treatment of halo bias in a snapshot by constructing the approximated form of matter perturbation field using information of halo positions. We then deal with matter density contrast in a light cone and then build the quadratic estimator. Finally we apply the estimator to different N-body simulations and successfully extract large scale modes both in a snapshot and in a light cone.

II. HALO BIAS

Let's first do a quick review of how we construct the quadratic estimator of a dark-matter-only catalog [2] before we get to a halo catalog. Starting from the perturbative series of the matter density contrast in Fourier space [9][10]:

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

where "m" stands for matter 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. 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}. By computing the off-diagonal term $\langle\delta_{\mathrm{m}}^{(1)}(\vec{k};a)\delta_{\mathrm{m}}^{(2)}(\vec{k};a)\rangle$ we can prove the following relation of the two-point function of two short-wavelength modes:

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

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 satisfiy the squeezed-limit condition $\vec{k}_s + \vec{k}'_s = \vec{k}_l$. f is given by:

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

$P_{\rm m,lin}$ is the linear matter power spectrum; F_2 is a function extremely insensitive to the choice of cosmological parameters in a dark-energy-dominated universe [11] with expression:

$$F_2(\vec{k}_1,\vec{k}_2)=\frac{5}{7}+\frac{2}{7}\frac{(\vec{k}_1\cdot\vec{k}_2)^2}{k_1^2k_2^2}+\frac{\vec{k}_1\cdot\vec{k}_2}{2k_1k_2}\left[\frac{k_1}{k_2}+\frac{k_2}{k_1}\right]. \quad (4)$$

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

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

with $\vec{k}'_s=\vec{k}_l-\vec{k}_s$. A is the normalization factor defined by requiring that $\langle\hat{\delta}^{(1)}_{\rm m}(\vec{k}_l;a)\rangle=\delta^{(1)}_{\rm m}(\vec{k}_l;a)$. g is given by minimizing noise term. They can be expressed as:

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

where $P_{\rm m,nl}$ is the nonlinear matter power spectrum. With this choice of the weighting function g, the Gaussian noise term is equal to the normalization factor $N(\vec{k}_l)=A(\vec{k}_l)$. And the projected detectability of a power spectrum measurement using this quadratic estimator can be written as:

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

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 $\vec{k}_l$.

So far we have finished reviewing the general formalism of the large scale quadratic estimator in a dark-matter-only catalog. Now we generalize it to a halo catalog.

Ignoring higher order halo bias, the density contrast of halos can be related to that of matter at a fixed time a as [4]:

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

here "h" means halo. $n_{\rm h}$ is the halo number density field at a given position and $\bar{n}_{\rm h}$ is the mean number density of halos. b_1 is the linear bias parameter relating halo and matter density contrasts. Linear bias is a function of halo mass and position and has been well approximated using analytical expressions[12][13]. We will use the Tinker halo bias function throughout this paper[14]. Since the linear bias b_1 depends not only on halo mass but also halo redshift, it will be very challenging for us to use $\delta_{\rm h}$ than $\delta_{\rm m}$ within a light cone IV. Therefore we choose to build an approximated matter density contrast out of halo positions following [15]:

$$\delta_{\rm m}(\vec{r};a)\equiv\frac{m_{\rm h}(\vec{r};a)-\bar{m}_{\rm h}(a)}{\bar{n}_{\rm h}(a)} \quad (9)$$

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

$$m_{\rm h}(\vec{r};a)=\sum_i^{N_{\rm h}}\frac{\delta_{\rm D}(\vec{r}-\vec{r}_i)}{b_1(M_i;a_i)}\Big|_{a_i=a} \quad (10)$$

$$n_{\rm h}(\vec{r};a)=\sum_i^{N_{\rm h}}\delta_{\rm D}(\vec{r}-\vec{r}_i)\Big|_a \quad (11)$$

where $\delta_{\rm D}$ is the Dirac delta function; $N_{\rm h}$ is the total number of halos. a_i, M_i and $\vec{r}_i$ is the cosmological growth factor, mass and position of the i-th halo, respectively. In this section we can use the simplification $b_1(M_i,a_i)=b_1(M_i)$ since we consider only a fixed time in this section. We can use Cloud-in-Cell (CIC) or Triangular Shaped Cloud (TSC) scheme [16] to smooth the sharply peaked Dirac delta function numerically.

We consider Eq. (9) as a general formula of expressing matter density using the distribution of halos only, since in real surveys the bias can be expressed as a function of luminosity and color [17]. We will not use halo density contrast $\delta_{\rm h}$ directly, instead we will use $\delta_{\rm m}$ defined in Eq. (9) as the approximated matter density contrast during the calculation of the quadratic estimator Eq. (5) both for a snapshot and for a light cone.

III. DEMONSTRATION WITH N-BODY SIMULATION I

We test the power of the quadratic estimator 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 [18]. This simulation used a flat ΛCDM model with Planck Collaboration XVI (2014) [19] cosmological parameters. Halos were found in this catalog using *Rockstar* code [20]. We will focus on halos with masses between $5\times10^{12}h^{-1}M_\odot<M<10^{13}h^{-1}M_\odot$ for simplicity. And we have similar results in other mass bins as well.

We use the approximated density contrast defined in Eq. (9) and it is able to perfectly capture the nonlinear matter power spectrum. Thus the projected detectability will be the same as our last our, shown in Fig. 1.

We also transform the true Fourier modes $\delta_{\rm m}(\vec{k}_l)$ and the estimated long-wavelength modes $\hat{\delta}_{\rm m}(\vec{k}_l)$ back into real space and compare them in Fig. 2. We can see that this quadratic estimator is able to successfully recapture large scale over/under densities.

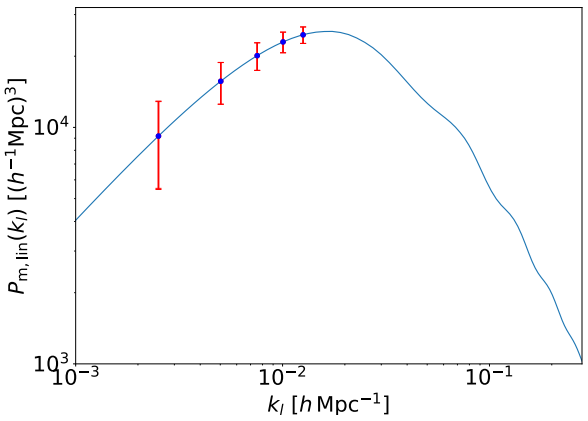

FIG. 1. Long-wavelength power spectrum and its error from Eq. (7) which can be expressed as $P_{\mathrm{m,lin}}(k_l)\sigma(k_l)$. Boxsize of the survey is assumed to be $L = 2.5\,h^{-1}$ Gpc thus volume $V = L^3$ and width $\Delta k = 2\pi/L$. The integration range for $\vec{k}_s$ is from $0.03\,h\,\mathrm{Mpc}^{-1}$ to $0.22\,h\,\mathrm{Mpc}^{-1}$.

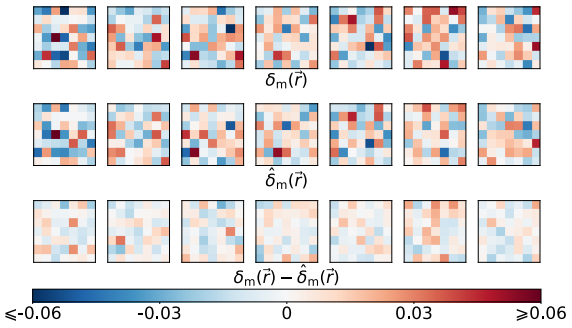

FIG. 2. Comparison of the true density field of the Big-MDPL halo catalog within mass bin $5 \times 10^{12}h^{-1}M_\odot < M < 10^{13}h^{-1}M_\odot$ ($\delta_{\mathrm{m}}(\vec{r})$ computed using the directly measured large scale modes, top row) and the density field from the quadratic estimator ($\hat{\delta}_{\mathrm{m}}(\vec{r})$, middle row). The bottom row shows their difference. Each panel represents a slice through the simulation volume, $2500\,h^{-1}$ 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 $0.22\,h\,\mathrm{Mpc}^{-1}$.

IV. LIGHT CONE FORMALISM

We first define the modified matter density contrast in a light cone as:

$$d_{\mathrm{m}}^{\mathrm{LC}}(\vec{r}) := d_{\mathrm{m}}(\vec{r})\frac{D_{\mathrm{ini}}}{D_1(a(r))} \tag{12}$$

where "LC" stands for "light cone". We set $a(\vec{r}=0)$ to be the origin of the light cone thus $a(\vec{r}) = a(r)$ will not depend of the direction of position $\vec{r}$. The goal of the inverse of linear growth function is to try to set the large scale amplitude the same for the whole light cone. $d_{\mathrm{m}}(\vec{r})$ is a field defined in the light cone as:

$$d_{\mathrm{m}}(\vec{r}) = \frac{\rho_{\mathrm{m}}(\vec{r}) - \bar{\rho}_{\mathrm{m}}(a(r))}{\bar{\rho}_{\mathrm{m}}(a(r))} \tag{13}$$

$$= \frac{n_{\mathrm{h}}(\vec{r}) - \bar{n}_{\mathrm{h}}(a(r))}{\bar{n}_{\mathrm{h}}(a(r))} \tag{14}$$

The first row Eq. (13) is the most accurate definition using matter density. $\rho_{\mathrm{m}}(\vec{r})$ is the matter density at position $\vec{r}$, $\bar{\rho}_{\mathrm{m}}(a(r))$ is the mean matter density at snapshot $a(r)$.

However in real cosmic surveys we cannot gain the information of matter density directly, so we generalize Eq. (9) to approximate this purely theoretical expression when applying to a halo/galaxy catalog in a light cone in the second row Eq. (14). $n_{\mathrm{h}}(\vec{r})$ is defined similarly as Eq. (10). Now we consider all halos in the light cone each with different scale factors, so there is no a dependence in the field n_{h}. $\bar{n}_{\mathrm{h}}(a(r))$ and $\bar{n}_{\mathrm{h}}(a(r))$ is the mean density of $n_{\mathrm{h}}(\vec{r})$ and $n_{\mathrm{h}}(\vec{r})$ at snapshot $a(r)$, respectively.

Notice this modified light cone matter density contrast $d_{\mathrm{m}}^{\mathrm{LC}}(\vec{r})$ can be fully determined observationally since we get to know $n_{\mathrm{h}}(\vec{r})$ and $n_{\mathrm{h}}(\vec{r})$ in cosmic surveys. And if we cut the light cone into several thin slices we are able to get a interpolation function for each value of $n_{\mathrm{h}}(a(r))$ and $\bar{n}_{\mathrm{h}}(a(r))$. We continue our derivation using the definition from matter density field Eq. (13) for the $d_{\mathrm{m}}(\vec{r})$ field. It's easy to check that all our results apply to the $d_{\mathrm{m}}(\vec{r})$ field from halo catalog Eq. (14), using the approximated matter density contrast formula in a snapshot - Eq. (9).

Each point $\vec{r}$ in the light cone corresponds to a snapshot with cosmological scale factor $a(r)$. In that snapshot we have the usual matter density contrast $\delta_{\mathrm{m}}(\vec{x}; a(r))$ and its Fourier transform $\delta_{\mathrm{m}}(\vec{k}; a(r))$ satisfying Eq. (1). Both $\delta_{\mathrm{m}}(\vec{x}; a(r))$ and $\delta_{\mathrm{m}}(\vec{k}; a(r))$ are defined at a **fixed** time $a(r)$, and we cannot measure them neither in surveys nor light cone simulations. We take advantage of their well-known analytical properties (such as Eq. (1)) and use them during analytical derivations. Luckily they will not show up in our final results thus our formalism can be applied in light cone simulations, and eventually surveys. $\delta_{\mathrm{m}}(\vec{x}; a(r))$ can be explicitly written as:

$$\delta_{\mathrm{m}}(\vec{x}; a(r)) = \frac{\rho_{\mathrm{m}}(\vec{x}; a(r)) - \bar{\rho}_{\mathrm{m}}(a(r))}{\bar{\rho}_{\mathrm{m}}(a(r))} \tag{15}$$

We can relate $d_{\mathrm{m}}(\vec{r})$ and $\delta_{\mathrm{m}}(\vec{x}; a)$ by noticing that:

$$\rho_{\mathrm{m}}(\vec{r}) = \rho_{\mathrm{m}}(\vec{x}; a(r))|_{\vec{x}=\vec{r}} = \rho_{\mathrm{m}}(\vec{r}; a(r)) \tag{16}$$

this equality holds obviously and we further get:

$$d_{\mathrm{m}}(\vec{r}) = \delta_{\mathrm{m}}(\vec{r}; a(r)) = \int \frac{d^3\vec{k}}{(2\pi)^3} e^{i\vec{k}\cdot\vec{r}}\delta_{\mathrm{m}}(\vec{k}; a(r)) \tag{17}$$

In the second step, we explicitly write down the definition of Fourier transform of matter density contrast in a snapshot with $a(r)$.

Eq. (17) is of great importance since it relates an observable $d_{\rm m}(\vec{r})$ (thus $d_{\rm m}^{\rm LC}(\vec{r})$) in a light cone to some unmeasurable quantity $\delta_{\rm m}(\vec{k};a(r))$ in a snapshot, but we know the properties of $\delta_{\rm m}(\vec{k};a(r))$ quite well theoretically. We can use this relation Eq. (17) to further perform analytical calculations.

Consider the Fourier transform of this modified density contrast within the light cone:

$$d_{\rm m}^{\rm LC}(\vec{k}) = \int_V d^3\vec{r}\, d_{\rm m}(\vec{r}) \frac{D_{\rm ini}}{D_1(a(r))} e^{-i\vec{k}\cdot\vec{r}} \tag{18}$$

V is the whole volume of the light cone. Similar to our previous work, we still want to compute the off-diagonal term of $d_{\rm m}^{\rm LC}(\vec{k})$ up to second order. With small scale perturbations of $d_{\rm m}^{\rm LC}(\vec{k})$ which can be measured observationally, we can estimate long-wavelength modes of $d_{\rm m}^{\rm LC}(\vec{k})$. The first and second order term of $d_{\rm m}^{\rm LC}(\vec{k})$ will be:

$$d_{\rm m}^{\rm LC,(i)}(\vec{k}) = \int_V d^3\vec{r} \int \frac{d^3\vec{k}}{(2\pi)^3} e^{i(\vec{k}-\vec{k})\cdot\vec{r}} \delta_{\rm m}^{(\rm i)}(\vec{k};a(r)) \frac{D_{\rm ini}}{D_1(a(r))} \tag{19}$$

with $i = 1, 2$. These are defined simply by substituting the first and second order term of Eq. (17) into Eq. (18). It's easy to prove that the first order term of $\delta_{\rm m}^{\rm LC}$ satisfies the following relation is cube volume necessary here?:

$$\langle d_{\rm m}^{\rm LC,(1)}(\vec{k}) d_{\rm m}^{\rm LC,(1)}(\vec{k}')\rangle = (2\pi)^3 \delta_{\rm D}(\vec{k}+\vec{k}') P_{\rm m,ini}(k) \tag{20}$$

$P_{\rm m,ini}$ is the linear matter power spectrum at some initial time $a_{\rm ini}$. This relation Eq. (20) tells us that the first order term of the modified light cone matter density contrast $d_{\rm m}^{\rm LC}(\vec{k})$ (which corresponds to long-wavelength modes) still characterizes the linear evolution information of the light cone, since redshift difference in the light cone has been cancelled by the extra factor $D_{\rm ini}/D_1(a(r))$ of the integrand. Or equivalently speaking, we have $P_{\rm m}^{\rm LC}(k) \simeq P_{\rm m,ini}(k)$ for long-wavelength modes (when $k \to 0$), as shown in Fig. 3. $P_{\rm m}^{\rm LC}(k)$ is the power spectrum of density contrast Eq. (12) defined as the diagonal term of the two-point correlation:

$$\langle d_{\rm m}^{\rm LC}(\vec{k}) d_{\rm m}^{\rm LC}(\vec{k}')\rangle \simeq (2\pi)^3 \delta_{\rm D}(\vec{k}+\vec{k}') P_{\rm m}^{\rm LC}(k) \tag{21}$$

There are also off-diagonal terms in Eq. (21) which contains extremely valuable information, and we will look into them in next section V.

V. QUADRATIC ESTIMATOR

Using expressions of Eq. (19), we can compute the two-point correlations of two short-wavelength Fourier modes $d_{\rm m}^{\rm LC}(\vec{k}_s)$ and $d_{\rm m}^{\rm LC}(\vec{k}'_s)$. Again we take the squeezed limit

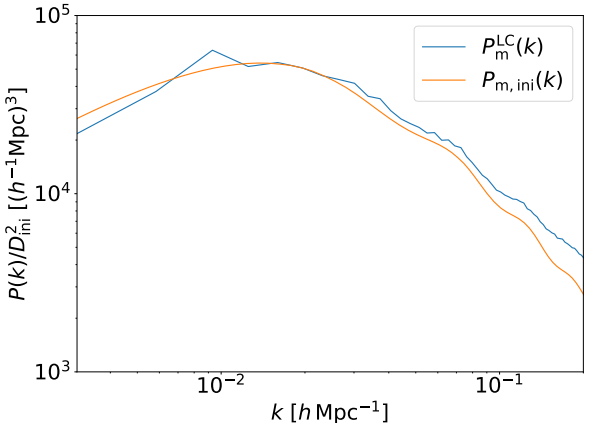

FIG. 3. Comparison of the light cone power spectrum $P_{\rm m}^{\rm LC}(k)$ with the initial power spectrum $P_{\rm m,ini}(k)$, both scaled by $1/D_{\rm ini}^2$. $P_{\rm m,ini}(k)$ is plotted according to theory predicted linear power spectrum, and $P_{\rm m}^{\rm LC}(k)$ is computed using halo catalog of MICE Grand Challenge light cone simulation [21][22]. Halo mass bin is chosen to be $[2.2\times10^{12}h^{-1}M_\odot, 10^{14}h^{-1}M_\odot]$. They match each other pretty well on large scales ($k \lesssim 0.02\,h\,{\rm Mpc}^{-1}$) despite tiny systematic error.

$\vec{k}_l = \vec{k}_s + \vec{k}'_s$:

$$\begin{aligned}&\langle d_{\rm m}^{\rm LC}(\vec{k}_s) d_{\rm m}^{\rm LC}(\vec{k}'_s)\rangle|_{\vec{k}_s+\vec{k}'_s=\vec{k}_l}\\ =&\langle d_{\rm m}^{\rm LC,(1)}(\vec{k}_s) d_{\rm m}^{\rm LC,(2)}(\vec{k}'_s)\rangle + \langle d_{\rm m}^{\rm LC,(2)}(\vec{k}_s) d_{\rm m}^{\rm LC,(1)}(\vec{k}'_s)\rangle\end{aligned} \tag{22}$$

Substituting Eq. (19) into Eq. (22) and evaluate the first bracket as an example:

$$\begin{aligned}&\langle d_{\rm m}^{\rm LC,(1)}(\vec{k}_s) d_{\rm m}^{\rm LC,(2)}(\vec{k}'_s)\rangle\\ =&\int_V d^3\vec{r} \int_V d^3\vec{r}' \int \frac{d^3\vec{k}}{(2\pi)^3} \int \frac{d^3\vec{k}'}{(2\pi)^3}\\ &e^{i(\vec{k}\cdot\vec{r}+\vec{k}'\cdot\vec{r}')-i(\vec{k}_s\cdot\vec{r}+\vec{k}'_s\cdot\vec{r}')} \frac{D_{\rm ini}}{D_1(a(r))} \frac{D_{\rm ini}}{D_1(a(r'))}\\ \times&\langle \delta_{\rm m}^{(1)}(\vec{k};a(r)) \delta_{\rm m}^{(2)}(\vec{k}';a(r'))\rangle\end{aligned} \tag{23}$$

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

$$\langle \delta_{\rm m,ini}^{(1)}(\vec{k}) \delta_{\rm m,ini}^{(2)}(\vec{k}')\rangle = 2F_2(-\vec{k}, \vec{k}+\vec{k}') P_{\rm m,ini}(k) \delta_{\rm m,ini}^{(1)}(\vec{k}+\vec{k}') \tag{24}$$

We can use this result to further determine the value of the bracket in Eq. (23):

$$\begin{aligned}&\langle \delta_{\rm m}^{(1)}(\vec{k};a(r)) \delta_{\rm m}^{(2)}(\vec{k}';a(r'))\rangle\\ =&\, 2\frac{D_1(a(r))}{D_{\rm ini}} \left[\frac{D_1(a(r'))}{D_{\rm ini}}\right]^2 \langle \delta_{\rm m,ini}^{(1)}(\vec{k}) \delta_{\rm m,ini}^{(2)}(\vec{k}')\rangle\\ =&\, 2\left[\frac{D_1(a(r'))}{D_{\rm ini}}\right]^2 F_2(-\vec{k}, \vec{k}+\vec{k}') P_{\rm m,ini}(k) \delta_{\rm m}^{(1)}(\vec{k}+\vec{k}';a(r))\end{aligned} \tag{25}$$

where we make use of the definition of the the linear growth factor $D_1(a(r))$. Plugging Eq. (25) into Eq. (23), we can see that the only $\vec{r}'$ dependent integral can be written as (here we simply choose a cube volume):

$$\int_V d^3\vec{r}' e^{-i(\vec{k}'-\vec{k}'_s)\vec{r}'} \frac{D_1(a(r'))}{D_{\rm ini}} \simeq C\,(2\pi)^3\delta_{\rm D}(\vec{k}'-\vec{k}'_s) \quad (26)$$

since $D_1(a(r'))$ is a slowly varying function. C is a constant and can be further determined via integrating over $\vec{k}$ on both sides of Eq. (26), which gives:

$$C = \frac{D_1(a=1)}{D_{\rm ini}} \quad (27)$$

thus:

$$\begin{aligned}
&\langle d_{\rm m}^{{\rm LC},(1)}(\vec{k}_s) d_{\rm m}^{{\rm LC},(2)}(\vec{k}'_s)\rangle \\
&= 2C\int_V d^3\vec{r}\int\frac{d^3\vec{k}}{(2\pi)^3}F_2(\vec{k},-\vec{k}+\vec{k}'_s)P_{\rm m,ini}(|\vec{k}-\vec{k}'_s|) \\
&\quad\times\frac{D_{\rm ini}}{D_1(a(r))}e^{-i(\vec{k}-\vec{k}_s-\vec{k}'_s)\vec{r}}\delta_{\rm m}^{(1)}(\vec{k};a(r)) \\
&\simeq 2C\,F_2(-\vec{k}_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k_s) \\
&\quad\times\int_V d^3\vec{r}\int\frac{d^3\vec{k}}{(2\pi)^3}\frac{D_{\rm ini}}{D_1(a(r))}e^{-i(\vec{k}-\vec{k}_s-\vec{k}'_s)\vec{r}}\delta_{\rm m}^{(1)}(\vec{k};a(r)) \\
&= 2C\,F_2(-\vec{k}_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k_s)d_{\rm m}^{{\rm LC},(1)}(\vec{k}_s+\vec{k}'_s) \quad (28)
\end{aligned}$$

where in the first step, we perform a redefinition of integration dummy variable. And we use an approximation in the second step in order to get the first order term $d_{\rm m}^{{\rm LC},(1)}(\vec{k}_l)$ out of the integral.

Again, we have proved that with this construction in Eq. (12), we can get long-wavelength modes from off-diagonal terms of short-wavelength modes:

$$\langle d_{\rm m}^{\rm LC}(\vec{k}_s)d_{\rm m}^{\rm LC}(\vec{k}'_s)\rangle|_{\vec{k}_s+\vec{k}'_s=\vec{k}_l} = f(\vec{k}_s,\vec{k}'_s)d_{\rm m}^{{\rm LC},(1)}(\vec{k}_l) \quad (29)$$

with

$$\begin{aligned}
f(\vec{k}_s,\vec{k}'_s) &= 2C\,F_2(-\vec{k}_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k_s) \\
&\quad + 2C\,F_2(-\vec{k}'_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k'_s) \quad (30)
\end{aligned}$$

The quadratic estimator can similarly be formed as:

$$\hat{d}_{\rm m}^{{\rm LC},(1)}(\vec{k}_l) = \mathcal{A}(\vec{k}_l)\int\frac{d^3\vec{k}_s}{(2\pi)^3}g(\vec{k}_s,\vec{k}'_s)d_{\rm m}^{\rm LC}(\vec{k}_s)d_{\rm m}^{\rm LC}(\vec{k}'_s) \quad (31)$$

with $\vec{k}'_s = \vec{k}_l - \vec{k}_s$ and g being a weighting function. $\mathcal{A}$ is the normalization function determined similarly by requiring that $\langle\hat{d}^{{\rm LC},(1)}(\vec{k}_l)\rangle = d^{{\rm LC},(1)}(\vec{k}_l)$:

$$\mathcal{A}(\vec{k}_l) = \left[\int\frac{d^3\vec{k}_s}{(2\pi)^3}g(\vec{k}_s,\vec{k}'_s)f(\vec{k}_s,\vec{k}'_s)\right]^{-1} \quad (32)$$

Similar to our last work, g can be calculated by minimizing the noise term and the result is:

$$\begin{aligned}
g(\vec{k}_s,\vec{k}'_s) &= \frac{f(\vec{k}_s,\vec{k}'_s)}{2P_{\rm m}^{\rm LC}(k_s)P_{\rm m}^{\rm LC}(k'_s)} \\
&= C\frac{F_2(-\vec{k}_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k_s)+F_2(-\vec{k}'_s,\vec{k}_s+\vec{k}'_s)P_{\rm m,ini}(k'_s)}{P_{\rm m}^{\rm LC}(k_s)P_{\rm m}^{\rm LC}(k'_s)}
\end{aligned} \quad (33)$$

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):

$$\frac{1}{\sigma^{\rm LC}(k_l)^2} = \frac{Vk_l^2\Delta k}{(2\pi)^2}\left[\frac{P_{\rm m,ini}(k_l)}{P_{\rm m,ini}(k_l)+\mathcal{A}(k_l)}\right]^2, \quad (34)$$

In Fig. 4, we show the projected error of measuring the initial power spectrum using this quadratic estimator Eq. (31) in a light cone. The difference of the error bar in this plot from the error bar of the cosmic variance ($\mathcal{A}=0$) is negligible.

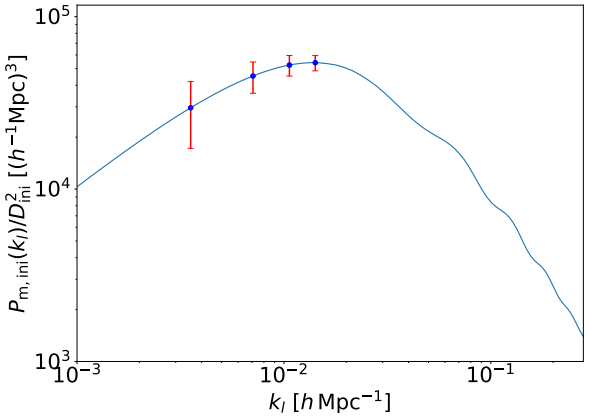

FIG. 4. Initial power spectrum of a light cone and its error from Eq. (34) which can be expressed as $P_{\rm m,ini}(k_l)\sigma_{\rm LC}(k_l)$. Boxsize of the survey is assumed to be $L = 1.773\,h^{-1}\,{\rm Gpc}$ thus volume $V = L^3$ and width $\Delta k = 2\pi/L$. The integration range for $\vec{k}_s$ is from $0.03\,h\,{\rm Mpc}^{-1}$ to $0.22\,h\,{\rm Mpc}^{-1}$.

Using the quadratic estimator Eq. (31) we can use small scale information of the whole light cone (for now we only consider a cube-shaped volume) to infer large scale field of the modified matter density contrast $d_{\rm m}^{\rm LC}(\vec{r})$. According to Eq. (20), this is equivalent to large scale linear matter perturbations in the light cone.

VI. DEMONSTRATION WITH N-BODY SIMULATION II

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. (18), as shown in Fig. 5. 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 $D_\mathrm{ini} = 1$ throughout this section[1].

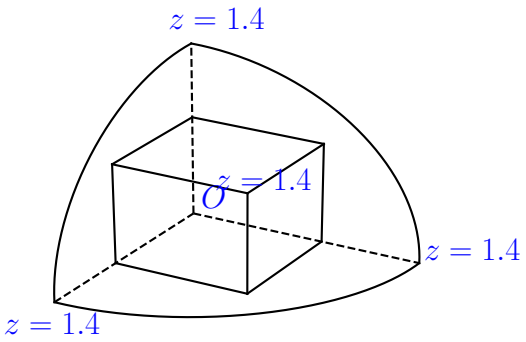

FIG. 5. Boxsize of the cube is $L = 3072/\sqrt{3}\,h^{-1}\,\mathrm{Mpc} \sim 1774\,\mathrm{h}^{-1}\,\mathrm{Mpc}$. Volume is $V = L^3 \sim 5.6\,(h^{-1}\,\mathrm{Mpc})^3$. Only one point in the cube can reach the redshift of 1.4, and $z = 0$ is at the origin O of the octant.

We first extract 1 in 700 matter particles' positions in the full the light cone of 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) [23] full sample of galaxies. We use Nbodykit [24] to compute the modified matter density contrast field in Fourier space and get the estimated one using Eq. (31). Than we transform them back into real space and compare them it Fig. 6. The cube with volume $\sim 5.6\,(h^{-1}\,\mathrm{Gpc})^3$ is divided into 7^3 cells. Each row in Fig. 6 contains 7 panels; each panel is a slice of the cube. We can see that this estimator does an amazing job of extracting the large scale density job - we are able to reproduce almost every cell with large over- or under-densities, and the difference is apparently much smaller than the density field itself.

[1] $D_\mathrm{ini} = 1$ 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.

In Fig. 6, 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. 7 to get a better view of this in a light cone, where in each row we still show the usual directly measure field $d_\mathrm{m}^\mathrm{LC}(\vec{r})$, estimated field $\hat{d}_\mathrm{m}^\mathrm{LC}(\vec{r})$ and their difference $(d_\mathrm{m}^\mathrm{LC} - \hat{d}_\mathrm{m}^\mathrm{LC})(\vec{r})$ 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. 6, it still is a more straightforward way of expressing our result in a light cone.

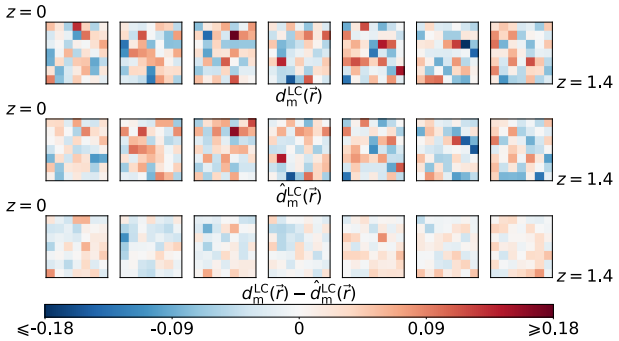

FIG. 6. Comparison of the true density field of the 1 in 700 matter particles of the Mice-GC simulation ($\delta_\mathrm{m}^\mathrm{LC}(\vec{r})$, top row) and the density field from the quadratic estimator ($\hat{\delta}_\mathrm{m}^\mathrm{LC}(\vec{r})$, middle row). The bottom row shows their difference. 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 $2.2 \times 10^{12} h^{-1} M_\odot < M < 10^{14} h^{-1} M_\odot$ for brevity. We have similar results in other mass bins as well. We use Eq. (14) as the approximated matter density contrast, $\pi_\mathrm{h}(\vec{r})$ 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 $\bar{\pi}_\mathrm{h}$ of each slice, then get the mean field $\bar{n}_\mathrm{h}(a(r))$ and $\bar{\pi}_\mathrm{h}(a(r))$ at each position by interpolation. We compute the directly measured modified matter contrast $d_\mathrm{m}^\mathrm{LC}(\vec{r})$ and also the power spectrum in Fig. 3. The power spectrum perfectly characterizes linear matter power spectrum on large scales. We use the quadratic estimator Eq. (31) to get the reconstructed modified field and transform them back into real space. Again we have two similar plots Fig. 8 and Fig. 9. From Fig. 8 we can see that our quadratic estimator is still able to extract large scale information, especially

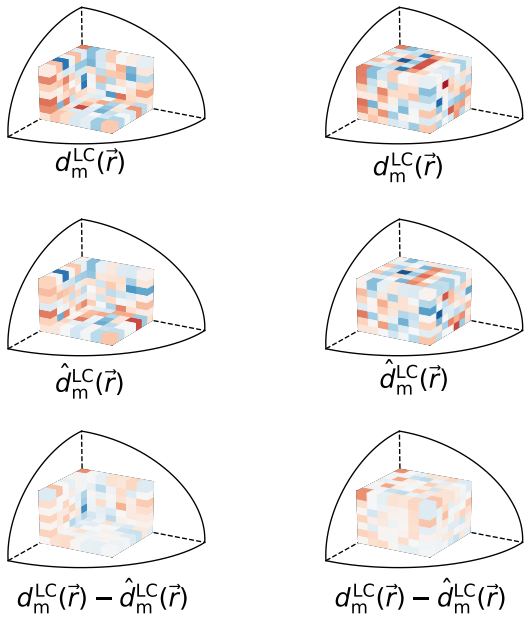

FIG. 7. Comparison of the true density field and the estimated density field of dark-matter-only catalog in a light cone with the same colorbar as Fig. 6. First row shows the modified matter density contrast near the origin (inner surfaces of the cube, $0 \leqslant z \lesssim 0.69$); second row shows the modified matter density contrast far from the origin (outer surfaces of the cube, $0.69 \lesssim z \lesssim 1.4$).

large over- or under-density cells on first few low-redshift panels. The difference seems larger when we go higher into redshift, and we have the worst performance on the very right panel. We observe the same feature in Fig. 9 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 [25].

VII. CONCLUSION

Future work: RSD; higher order perturbations; higher order bias term; full volume of the light cone.

ACKNOWLEDGMENTS

We thank Duncan Campbell for resourceful discussions. We also thank Enrique Gaztanaga for providing us with the 1 in 700 matter particles' positions of Mice-GC simulation. This work is supp orted by U.S. Dept. of Energy contract DE-SC0019248 and NSF AST-1909193. The BigMDPL simulation was performed at LRZ Munich within the PRACE project pr86bu. The CosmoSim

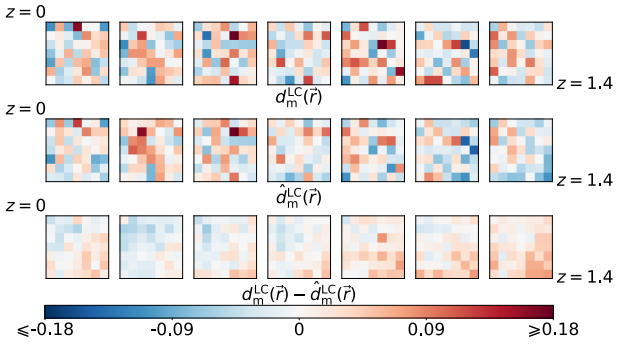

FIG. 8. Comparison of the true density field of the halo number density contrast in mass bin $2.2 \times 10^{12} h^{-1} M_\odot < M < 10^{14} h^{-1} M_\odot$ of the Mice-GC simulation ($\delta_{\rm m}^{\rm LC}(\vec{r})$, top row) and the density field from the quadratic estimator ($\hat{\delta}_{\rm m}^{\rm LC}(\vec{r})$, middle row) in the same light cone as Fig. 6, and their difference (bottom row). The integration range of $\vec{k}_s$ is from $0.03\,h\,{\rm Mpc}^{-1}$ to $0.35\,h\,{\rm Mpc}^{-1}$. We use a larger upper limit to reduce the noise term.

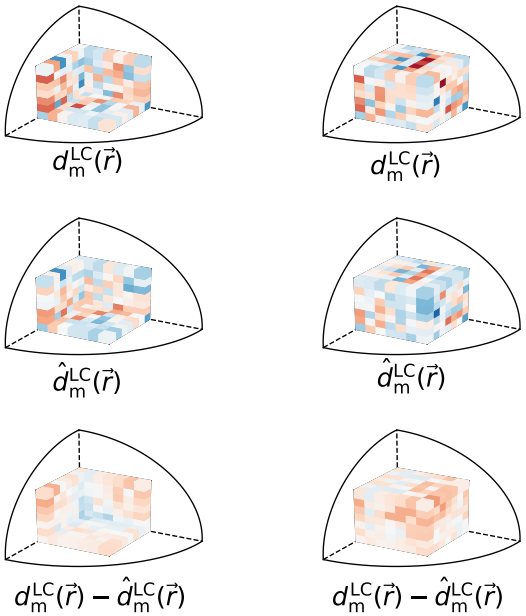

FIG. 9. Comparison of the true density field and the estimated density field of halo catalog in a light cone with the same colorbar as Fig. 6 (or Fig. 8). First row shows the modified matter density contrast near the origin (inner surfaces of the cube, $0 \leqslant z \lesssim 0.69$); second row shows the modified matter density contrast far from the origin (outer surfaces of the cube, $0.69 \lesssim z \lesssim 1.4$).

database (www.cosmosim.org) providing the file access is a service by the Leibniz-Institute for Astrophysics Potsdam (AIP). This work has made use of CosmoHub. CosmoHub has been developed by the Port d'Informació Científica (PIC), maintained through a collaboration of the Institut de Física d'Altes Energies (IFAE) and the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), and was partially funded by the "Plan Estatal de Investigación Científica y Técnica y de Innovación" program of the Spanish government.

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

[2] P. Li, S. Dodelson, and R. A. C. Croft, arXiv:2001.02780 [astro-ph.CO].

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

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

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

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

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

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

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

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

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

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

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

[14] 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].

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

[16] 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].

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

[18] 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].

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

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

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

[22] 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].

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

[24] 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].

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