# PHYS 4090: A Study on Cosmological Parity Violation

K. Yue[1, *]

[1] *Department of Physics, The Hong Kong University of Science and Technology, Hong Kong S.A.R., P.R.China*

(Dated: December 9, 2023)

If parity violation exists in the early universe, detectable traces will be left to cosmological observation at later times. One such example is the four-point correlation function (4PCF). It is possible that the cosmological parity violation originates from a Chern-Simons coupling in the axion inflation Lagrangian. Two independent research teams have reported significant evidence for parity violation [**1 and 2**]. To validate their findings, we have developed Python codes capable of computing the parity-odd signal derived from the 4PCF of the Baryon Oscillation Spectroscopic Survey (BOSS). Our program has undergone thorough testing with a toy model consisting of 1500 tetrahedra, and it has been proven to function well. In addition, we have computed the 4PCF estimator for angular momentum multipoles $(1, 1, 1)$ and $(1, 2, 2)$, using a sample of 1000 galaxies from the BOSS CMASS dataset. Currently, we are in the process of developing codes for GPU-acceleration to enable the computation of the 4PCF for larger galaxy samples and finer binning.

## I. INTRODUCTION

The search for parity violation has a long history and remains an active research topic. In the Standard Model (SM), the weak interaction is the only fundamental force that violates parity [**3**], and it played no role in the evolution of the large-scale distribution of matter. Any observation of a cosmological parity violation would therefore imply the existence of a new CP-violating force at inflation. Such a parity-violating force might even be able to explain the observed baryon asymmetry, as envisaged by the Sakharov conditions [**4**]. The detection of a parity violation in cosmological observations would be a significant discovery, providing evidence for physics beyond the SM and potentially shedding light on the nature of dark matter, dark energy, and inflation.

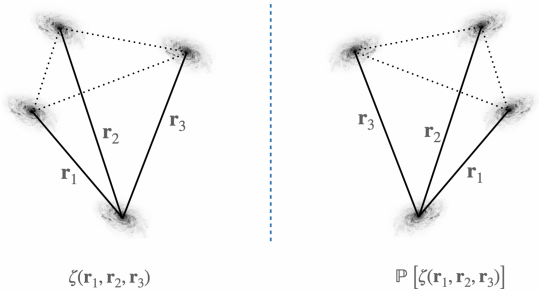


FIG. 1: A tetrahedron and its mirror image cannot be superimposed.

To test for parity violation, parity-sensitive observables are required. The action of the parity operator is equivalent to a rotation, therefore the galaxy's isotropic two-point or three-point correlation function (2PCF or 3PCF) is insensitive to parity. More generally, in a $D$-dimensional space, the $N$-point correlation function (NPCF) is sensitive to parity only if $N > D$ [**5**]. Since our universe is three-dimensional (3D), the simplest statistic to study scalar parity violation in large-scale structure (LSS) surveys is the 4PCF [**6 and 7**]. The 4PCF's sensitivity to parity can also be intuitively understood by the fact that a tetrahedron and its mirror image cannot be superimposed, i.e., $\zeta(\vec{r}_1, \vec{r}_2, \vec{r}_3) \neq \mathbb{P}[\zeta(\vec{r}_1, \vec{r}_2, \vec{r}_3)]$, as shown in FIG. 1 [**1**]. The 4PCF can be used to test for parity violations in the LSS of the universe and potentially constrain theories beyond the SM.

The objective of this research is to validate the findings presented in [**1**] and [**2**]. In the study by [**1**], the authors employed the Baryon Oscillation Spectroscopic Survey (BOSS) CMASS sample to conduct a blind test for parity violation in the 4PCF. They analyzed the data using a non-parametric rank test and a compressed-$\chi^2$ analysis. The rank test resulted in a detection significance of $2.9\sigma$. A similar study was conducted in [**2**], which reported a significance of $3.1\sigma$ in LOWZ and a much higher significance of $7.1\sigma$ in CMASS.

The remaining work is structured as follows. Section II provides an overview of the theoretical concepts related to cosmological parity violation, including the origin of structure from cosmic inflation and the 4PCF. It also covers the necessary corrections to account for non-uniform survey geometry. In Section III, the measurement of the parity-odd 4PCF using toy tetrahedra is illustrated. Section IV presents the results of the measurement of the parity-odd 4PCF using a test sample from Sloan Digital Sky Survey-III (SDSS-III) [**8**], which consists of 1000 galaxies. The limitations of the current research progress and suggestions for future study are discussed in Section V. Finally, Section VI concludes the work. Python codes of our analysis work can be found on https://github.com/ckkyue/4PCF.

* ckkyue@connect.ust.hk

## II. THEORY

### A. Non-Gaussianity

Primordial non-Gaussianity refers to deviations from Gaussian statistics in the initial conditions of the universe, specifically in the distribution of density fluctuations in the cosmic microwave background (CMB) and LSS of the universe. Such deviations, if present, directly translate to signatures of the dynamics and the field content driving inflation [**9 and 10**].

The primordial fluctuations observed in modern cosmological surveys are generated by the quantum fluctuation of the inflaton field, represented as $\phi = \phi_0 + \varphi$. Here, $\phi_0$ represents the background field, while $\varphi$ represents the fluctuation of the field. For simplicity, we consider a massless, real scalar field $\sigma$ in de Sitter space, which is non-interacting and treated as a spectator field. "Spectator" means that the field's back-reaction on the metric has been ignored. The action governing the dynamics of the spectator field is given by

$$S = \int d^4x\, a^3 \left[\frac{1}{2}\dot{\sigma}^2 - \frac{1}{2a^2}(\partial_i\sigma)^2\right]. \tag{1}$$

Converting to Fourier space, eq. (1) becomes [**11**]

$$S = \int dt\, \frac{d^3k}{(2\pi)^3}\, a^3 \left(\frac{1}{2}\dot{\sigma}_{\vec{k}}\dot{\sigma}_{-\vec{k}} - \frac{k^2}{2a^2}\sigma_{\vec{k}}\sigma_{-\vec{k}}\right), \tag{2}$$

which the $\sigma$ field can be quantized as

$$\sigma_{\vec{k}}(t) = v_k(t)a_{\vec{k}} + v_k^*(t)a^\dagger_{-\vec{k}}, \tag{3}$$

and the time-independent creation and annihilation operators satisfy the usual commutation relations:

$$[a_{\vec{k}}, a_{\vec{k}'}] = 0, \quad [a_{\vec{k}}, a^\dagger_{\vec{k}'}] = (2\pi)^3\delta^3(\vec{k} - \vec{k}'). \tag{4}$$

$v_k$ satisfies the classical equation of motion of $\sigma$:

$$\ddot{v}_k + 3H\dot{v}_k + \frac{k^2}{a^2}v_k = 0. \tag{5}$$

The solution to eq. (5) is

$$v_k(\tau) = \frac{H}{\sqrt{2k^3}}(1 + ik\tau)e^{-ik\tau}, \tag{6}$$

which we have switched to conformal time $\tau$ for convenience. In fact, $\sigma$ is the fluctuation $\varphi$ of the inflaton field, of which the Fourier modes have the same solution as $\sigma$:

$$\varphi(\tau, \vec{k}) = \frac{H}{\sqrt{2k^3}}(1 + ik\tau)e^{-ik\tau}. \tag{7}$$

At the limit of conformal time $\tau \to 0$, which corresponds to the end of inflation, the initial conditions of the density fluctuations in our universe are given by $\varphi(0, \vec{k}) = \frac{H}{\sqrt{2k^3}}$. To extract statistical information from these initial conditions, we consider the 2PCF, which gives $\langle\varphi^2\rangle \sim \frac{H}{2k^3}$. Higher-point correlation functions $\langle\varphi^n\rangle$ with $n \geq 3$ provide information about the interactions of the $\varphi$ modes and are referred to as primordial non-Gaussianity [**12**].

Using the spatial flat gauge $\zeta = -\frac{H}{\dot{\phi}_0}\varphi$, we can translate the inflaton fluctuation given by eq. (7) into curvature fluctuation $\zeta$. Expanding in the Fourier space,

$$\zeta(t, \vec{x}) = \int \frac{d^3k}{(2\pi)^3}\zeta(t, \vec{k})e^{i\vec{k}\cdot\vec{x}}. \tag{8}$$

Since $\zeta(t, \vec{x})$ is real, for each $k$-mode we have $\zeta^*(t, \vec{k}) = \zeta(t, -\vec{k})$. Under parity transformation, the NPCF transforms according to

$$\langle\Pi_i^n\zeta(t, \vec{k}_i)\rangle \xrightarrow{\mathcal{P}} \langle\Pi_i^n\zeta(t, -\vec{k}_i)\rangle = \langle\Pi_i^n\zeta(t, \vec{k}_i)\rangle^*. \tag{9}$$

This implies that the imaginary part of the NPCF is parity-odd.

### B. Possible source of parity violation: axion inflation

One possible source of parity violation is a Chern-Simons coupling between an inflaton field and a $U(1)$ gauge field $A_\mu$. The corresponding action is given by [**13**]:

$$S = \int d^4\sqrt{-g}\left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m_A^2A_\mu A^\mu - \frac{1}{4\Lambda}\phi F_{\mu\nu}\tilde{F}^{\mu\nu}\right]. \tag{10}$$

Here, $F_{\mu\nu}$ represents the field strength of $A_\mu$, and $\tilde{F}^{\mu\nu} \equiv \frac{1}{2}\frac{\varepsilon^{\mu\nu\rho\sigma}}{\sqrt{-g}}F_{\rho\sigma}$ represents its Hodge dual. The action of the axion coupling can be isolated as follows:

$$\begin{aligned} S_J &= \int d^4x\sqrt{-g}\left(-\frac{1}{4\Lambda}\phi F_{\mu\nu}\tilde{F}^{\mu\nu}\right) \\ &= -\int d^4x\frac{1}{8\Lambda}\phi\varepsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} \\ &= \int d^4x\phi J(\tau, \vec{x}) \\ &= \int d\tau\frac{d^3k}{(2\pi)^3}\phi J_{\vec{k}}(\tau). \end{aligned} \tag{11}$$

The current is

$$J(\tau, \vec{x}) = -\frac{1}{8\Lambda}\varepsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}, \tag{12}$$

and its Fourier transform is

$$J_{\vec{k}}(\tau) = \frac{a^4}{\Lambda}\int d^3x\, e^{-i\vec{k}\cdot\vec{x}}\vec{E}\cdot\vec{B}, \tag{13}$$

where $\vec{E}$ and $\vec{B}$ represent the electric and magnetic fields associated with $A_\mu$. The term $\vec{E}\cdot\vec{B}$ in the current given in eq. (13) can be viewed as the source that induces parity violation in the correlations between curvature perturbations and density fluctuations. This effect may be observed in the 4PCF of galaxy distributions in the later universe.

In this research, we are particularly interested in the calculation of the 4PCF. The calculation of correlation functions bears resemblance to that of the $S$-matrix in quantum field theory. However, we are looking at the expectation values of field operators at a later time given the "in" state, in the absence of an explicit "out" state. To achieve this, we need to apply the well-known in-in formalism. The detailed and full derivation of the 4PCF in axion inflation is reserved for future work. For now, we only present the essential equations required for the calculation in the in-in formalism.

The master formula in the in-in formalism calculates the time dependent expectation value of an operator:

$$\langle Q(\tau)\rangle = \langle 0|\left[\bar{T}\left\{e^{i\int_{\tau_0}^{\tau} H_I(\tau')d\tau'}\right\}\right] Q^I(\tau) \left[T\left\{e^{-i\int_{\tau_0}^{\tau} H_I(\tau')d\tau'}\right\}\right]|0\rangle\,, \tag{14}$$

where $T$ is the time-ordering operator, $\bar{T}$ is the anti-time ordering operator, and $Q^I(\tau)$ is the operator in the interaction picture. A different and more convenient way to express eq. (14) is through the use of the commutator form:

$$\langle Q(\tau)\rangle = \sum_{n=0}^{\infty}\int_{\tau_0}^{\tau} d\tau_1\cdots\int_{\tau_0}^{\tau_{n-1}} d\tau_n \langle[H_I(\tau_n),[H_I(\tau_{n-1}),\ldots,[H_I(\tau_1),Q^I(\tau)]...]]\rangle. \tag{15}$$

To derive eq. (15) from eq. (14), we make use of the Baker–Campbell–Hausdorff (BCH) formula:

$$e^X Y e^{-X} = \frac{[(X)^n,Y]}{n!}, \tag{16}$$

where $X$ and $Y$ represent operators that may not commute. For simplicity, we neglect the time-ordering operator $T$ for the time being, and choose $X = i\int_{\tau_0}^{\tau} H_I(\tau')d\tau'$ and $Y = Q^I(\tau)$. We then proceed to calculate

$$\begin{aligned}\langle 0|e^X Y e^{-X}|0\rangle &= \sum_{n=0}^{\infty}\langle 0|\,\frac{[(X)^n,Y]}{n!}\,|0\rangle \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\,\langle 0|\,[i\int_{\tau_0}^{\tau_{\sigma_1}} H_I(\tau')d\tau', \\ &\qquad [\ldots,[i\int_{\tau_0}^{\tau_{\sigma_n}} H_I(\tau')d\tau', Q^I(\tau)]...]\,|0\rangle \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\int_{\tau_0}^{\tau_{\sigma_1}} d\tau'\cdots\int_{\tau_0}^{\tau_{\sigma_n}} d\tau'\,\langle 0|\,[H_I(\tau_{\sigma_1}), \\ &\qquad [\ldots,[H_I(\tau_{\sigma_n}),Q^I(\tau)]\ldots]\,|0\rangle\,,\end{aligned} \tag{17}$$

where $\sigma_i$ represents the permutation of $\tau_1,\ldots,\tau_n$. If we reintroduce the time-ordering operator $T$, the integrals in eq. (17) will be arranged in the same order as in eq. (15), while also absorbing the factor $\frac{1}{n!}$ that appears in front.

### C. 4PCF estimator

The fractional density fluctuation is $\delta(\vec{x}) \equiv \frac{\rho(\vec{x})}{\bar{\rho}} - 1$, where $\bar{\rho}$ is the average density and $\rho(\vec{x})$ is the density field at some $\vec{x}$ from the origin. The Planck and Wilkinson Microwave Anisotropy Probe (WMAP) results have provided the most stringent limits on the non-Gaussianity of the distribution of density fluctuations in the early universe **[14 and 15]**.

The 4PCF is a measure of the correlation between four points in space. It quantifies the excessive tendency of groups of four galaxies to cluster together compared to random distribution, and is mathematically defined as

$$\hat{\zeta}(\vec{r}_1,\vec{r}_2,\vec{r}_3) \equiv \langle\delta(\vec{x})\delta(\vec{x}+\vec{r}_1)(\vec{x}+\vec{r}_2)(\vec{x}+\vec{r}_3)\rangle, \tag{18}$$

where $\langle\rangle$ indicates an ensemble average of the density fluctuation field. Applying the ergodic principle, the ensemble average can be replaced by an integral of spatial position $\vec{x}$ over the volume $V$ of the survey, rewriting eq. (18) into

$$\hat{\zeta}(\vec{r}_1,\vec{r}_2,\vec{r}_3) = \frac{1}{V}\int d\vec{x}\delta(\vec{x})\delta(\vec{x}+\vec{r}_1)(\vec{x}+\vec{r}_2)(\vec{x}+\vec{r}_3). \tag{19}$$

From eq. (19), the 4PCF estimated from cosmological survey is

$$\hat{\zeta}(\vec{r}_1,\vec{r}_2,\vec{r}_3) = \frac{1}{V}\int_{\vec{r}_1,\vec{r}_2,\vec{r}_3\in\text{bins}} d\vec{x}\delta(\vec{x}) \times\delta(\vec{x}+\vec{r}_1)\delta(\vec{x}+\vec{r}_2)\delta(\vec{x}+\vec{r}_3), \tag{20}$$

which we have included the finite-size radial bins. As a remark, we choose $r_1 < r_2 < r_3$ to avoid over-counting

the tetrahedra. It implies that measuring the 4PCF requires counting of quadruplets of galaxies (see FIG. 2 [**16**]), and the computational complexity $\sim \mathcal{O}(N_g^4)$ with $N_g$ being the number of galaxies. Considering the substantial number of galaxies, approximately $10^6$, in the BOSS CMASS [**17**], it is imperative to improve our algorithm to finish the computation within a reasonable amount of time. In the following, we outline a method that reduces the complexity to $\mathcal{O}(N_g^2)$ to address this issue.

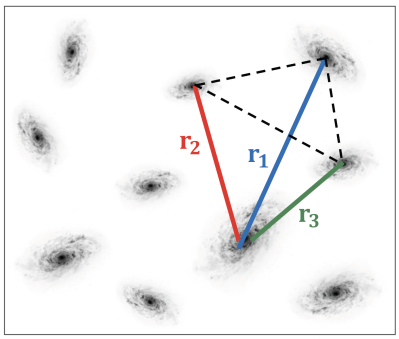


FIG. 2: A quadruplet of galaxy. Three vectors $\vec{r}_1$, $\vec{r}_2$, and $\vec{r}_3$ extend from the primary vertex.

Ignoring redshift space distortions, 3D distribution of galaxies is assumed to be isotropic on cosmological scales, i.e. invariant under simultaneous rotations of all coordinates. The isotropic basis is hence an efficient method to extract cosmological information systematically. Expanding the 4PCF in the basis of isotropic functions [**1**],

$$\hat{\zeta}(\vec{r}_1,\vec{r}_2,\vec{r}_3) = \sum_{l_1 l_2 l_3} \hat{\zeta}_{l_1 l_2 l_3}(r_1,r_2,r_3)\mathcal{P}_{l_1 l_2 l_3}(\hat{\vec{r}}_1,\hat{\vec{r}}_2,\hat{\vec{r}}_3). \tag{21}$$

The first part of eq. (21) is the coefficients, and the second part is the isotropic basis functions [**17**], which can be decomposed into spherical harmonics as

$$\mathcal{P}_{l_1 l_2 l_3}(\hat{\vec{r}}_1,\hat{\vec{r}}_2,\hat{\vec{r}}_3) \equiv (-1)^{l_1+l_2+l_3} \sum_{m_1 m_2 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \times Y_{l_1 m_1}(\hat{\vec{r}}_1) Y_{l_2 m_2}(\hat{\vec{r}}_2) Y_{l_3 m_3}(\hat{\vec{r}}_3), \tag{22}$$

which the Wigner 3-$j$ symbol

$$\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \equiv \frac{(-1)^{l_1-l_2-m_3}}{\sqrt{2l_3+1}} \langle l_1 m_1 l_2 m_2 | l_3(-m_3)\rangle \tag{23}$$

is included, and the summation includes $m_i \in [-l_i, l_i]$.

We first demonstrate the rotational invariance of the isotropic basis functions. The isotropy of these functions implies that they should not be dependent on specific components along any axis, i.e., $\sum_{i=1}^3 m_i = 0$. This condition is automatically satisfied since the Clebsch–Gordan (CG) coefficient $\langle l_1 m_1 l_2 m_2 | l_3(-m_3)\rangle$ in the Wigner 3-$j$ symbol is non-zero only if $\sum_{i=1}^3 m_i = 0$. Moreover, these basis functions are orthonormal:

$$\int d\hat{\vec{r}}_1 d\hat{\vec{r}}_2 d\hat{\vec{r}}_3 \mathcal{P}_{l_1 l_2 l_3}(\hat{\vec{r}}_1,\hat{\vec{r}}_2,\hat{\vec{r}}_3) \times \mathcal{P}^*_{l_1' l_2' l_3'}(\hat{\vec{r}}_1,\hat{\vec{r}}_2,\hat{\vec{r}}_3) = \delta_{l_1 l_1'}\delta_{l_2 l_2'}\delta_{l_3 l_3'}. \tag{24}$$

Next, we show the coefficients part in eq. (21) is separable in $r_i$. Consider the integral

$$\begin{aligned} &\int d\hat{\vec{r}}_1 d\hat{\vec{r}}_2 d\hat{\vec{r}}_3 \mathcal{P}^*_{l_1 l_2 l_3}(\hat{\vec{r}}_1,\hat{\vec{r}}_2,\hat{\vec{r}}_3)\hat{\zeta}(\hat{\vec{r}}_1,\hat{\vec{r}}_2,\hat{\vec{r}}_3) \\ &= \int d\hat{\vec{r}}_1 d\hat{\vec{r}}_2 d\hat{\vec{r}}_3 \mathcal{P}^*_{l_1 l_2 l_3}(\hat{\vec{r}}_1,\hat{\vec{r}}_2,\hat{\vec{r}}_3) \\ &\quad \times \sum_{l_1' l_2' l_3'} \hat{\zeta}_{l_1' l_2' l_3'}(r_1,r_2,r_3)\mathcal{P}_{l_1' l_2' l_3'}(\hat{\vec{r}}_1,\hat{\vec{r}}_2,\hat{\vec{r}}_3) \\ &= \sum_{l_1' l_2' l_3'} \int d\hat{\vec{r}}_1 d\hat{\vec{r}}_2 d\hat{\vec{r}}_3 \mathcal{P}^*_{l_1 l_2 l_3}(\hat{\vec{r}}_1,\hat{\vec{r}}_2,\hat{\vec{r}}_3) \\ &\quad \times \mathcal{P}_{l_1' l_2' l_3'}(\hat{\vec{r}}_1,\hat{\vec{r}}_2,\hat{\vec{r}}_3)\hat{\zeta}_{l_1' l_2' l_3'}(r_1,r_2,r_3) \\ &= \hat{\zeta}_{l_1 l_2 l_3}(r_1,r_2,r_3), \end{aligned} \tag{25}$$

which we have applied the orthonormal relation given in eq. (24) to arrive at the last line. Define harmonic coefficients as

$$a_{lm}(x;\vec{r}_i) \equiv \int d\hat{\vec{r}}_i \delta(\vec{x}+\vec{r}_i) Y_{lm}(\hat{\vec{r}}_i), \tag{26}$$

eq. (25) turns into

$$\begin{aligned} &\hat{\zeta}_{l_1 l_2 l_3}(r_1,r_2,r_3) \\ &= \int d\hat{\vec{r}}_1 d\hat{\vec{r}}_2 d\hat{\vec{r}}_3 \sum_{m_1 m_2 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \\ &\quad \times Y_{l_1 m_1}(\hat{\vec{r}}_1) Y_{l_2 m_2}(\hat{\vec{r}}_2) Y_{l_3 m_3}(\hat{\vec{r}}_3) \\ &\quad \times \frac{1}{V}\int d\vec{x} d\hat{\vec{r}}_1 d\hat{\vec{r}}_2 d\hat{\vec{r}}_3 \delta(\vec{x})\delta(\vec{x}+\vec{r}_1)\delta(\vec{x}+\vec{r}_2)\delta(\vec{x}+\vec{r}_3) \\ &= \frac{1}{V}\sum_{m_1 m_2 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \\ &\quad \times \int d\vec{x} d\hat{\vec{r}}_1 d\hat{\vec{r}}_2 d\hat{\vec{r}}_3 \delta(\vec{x})\delta(\vec{x}+\vec{r}_1)\delta(\vec{x}+\vec{r}_2)\delta(\vec{x}+\vec{r}_3) \\ &\quad \times Y_{l_1 m_1}(\hat{\vec{r}}_1) Y_{l_2 m_2}(\hat{\vec{r}}_2) Y_{l_3 m_3}(\hat{\vec{r}}_3) \\ &= \frac{1}{V}\sum_{m_1 m_2 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \int d\vec{x}\delta(\vec{x}) \prod_{i=1}^3 a_{l_i m_i}(\vec{x};r_i). \end{aligned} \tag{27}$$

Using eq. (27), the computational complexity of the 4PCF estimator is reduced to $\mathcal{O}(N_g^2)$.

### D. Edge-correction

To mitigate the effects of non-uniform survey geometry, the 4PCF estimator in eq. (27) needs to be modified to the generalized Landy-Szalay form **[18 and 19]**:

$$\hat{\zeta}(\vec{r}_1,\vec{r}_2,\vec{r}_3)=\frac{\mathcal{N}(\vec{r}_1,\vec{r}_2,\vec{r}_3)}{\mathcal{R}(\vec{r}_1,\vec{r}_2,\vec{r}_3)}, \tag{28}$$

where $\mathcal{N}$ and $\mathcal{R}$ are the 4PCF estimates obtained from "data-minus-random" and random catalogs respectively. The edge-corrected 4PCF multiplets are then given by

$$\begin{aligned}\hat{\zeta}_{l_1l_2l_3}(r_1,r_2,r_3)=&\sum_{l_1'l_2'l_3'}[M^{-1}]^{l_1'l_2'l_3'}_{l_1l_2l_3}(r_1,r_2,r_3)\\&\times\frac{\mathcal{N}_{l_1'l_2'l_3'}(r_1,r_2,r_3)}{\mathcal{R}_{000}(r_1,r_2,r_3)}.\end{aligned} \tag{29}$$

The coupling matrix is defined as

$$\begin{aligned}M^{l_1'l_2'l_3'}_{l_1l_2l_3}(r_1,r_2,r_3)=&\frac{(-1)^{l_1'+l_2'+l_3'}}{(4\pi)^{\frac{3}{2}}}\times\sum_{L_1L_2L_3}\frac{\mathcal{R}_{L_1L_2L_3}(r_1,r_2,r_3)}{\mathcal{R}_{000}(r_1,r_2,r_3)}\\&\left[\prod_{i=1}^{3}\sqrt{(2l_i+1)(2L_i+1)(2l_i'+1)}\right]\begin{Bmatrix}l_1&L_1&l_1'\\l_2&L_2&l_2'\\l_3&L_3&l_3'\end{Bmatrix}\\&\times\begin{pmatrix}l_1&L_1&l_1'\\0&0&0\end{pmatrix}\begin{pmatrix}l_2&L_2&l_2'\\0&0&0\end{pmatrix}\begin{pmatrix}l_3&L_3&l_3'\\0&0&0\end{pmatrix},\end{aligned} \tag{30}$$

and the curly bracket denotes the Wigner 9-$j$ symbol.

To sum up, we first expand the 4PCF in basis of isotropic functions. This expansion allows us to separate the coefficients into decoupled integrals. By doing so, the estimator is factorized into independent components corresponding to pairs of galaxies. This factorization significantly reduces the computational complexity to approximately $\mathcal{O}(N_g^2)$, enabling the calculation of the 4PCF for galaxy surveys in a feasible manner.

## III. TOY MODEL OF TETRAHEDRA

In line with the approach in **[2]**, a toy model consisting of 1500 tetrahedra within a cubic box with a side length of $L=1000$ has been constructed. The purpose of this toy model is to assess the functionality of our 4PCF estimator codes. Each tetrahedron in the model has three sides extending from the primary vertex, with lengths $r_1\sim 10$, $r_2\sim 20$, and $r_3\sim 30$. In order to enhance the realism of each tetrahedron, a random rotation is introduced to every vertex, where the rotation angle $\theta$ is selected from the range of $[-180°, 180°]$ for each of the three Cartesian coordinates. Additionally, random distances $\Delta r_1\in[0,1]$, $\Delta r_2\in[-2,2]$, and $\Delta r_3\in[-1,0]$ are incorporated to modify the values of $r_1$, $r_2$, and $r_3$ respectively. To achieve minimal overlap between tetrahedra, a minimum separation of 60 (roughly twice the longest side $r_3$ of the tetrahedron) between any two primary vertices is imposed.

For a given tetrahedron, we select one point as the origin and assign subscripts to three vectors originating from that point, ensuring $r_1<r_2<r_3$. To determine the handedness, we observe the directions when looking down from the origin along each vector $\vec{r}_i$. The direction from $\vec{r}_1$ to $\vec{r}_3$ defines the handedness, as illustrated in FIG. 3. FIG. 4 shows an example cubic box filled with 1500 counterclockwise tetrahedra.

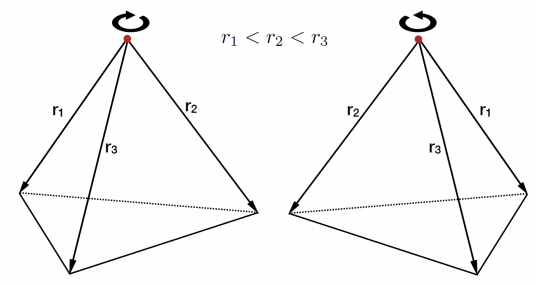


FIG. 3: We define the handedness of a tetrahedron by the direction of $\vec{r}_1$ to $\vec{r}_3$ when viewed from the origin. **[1]**

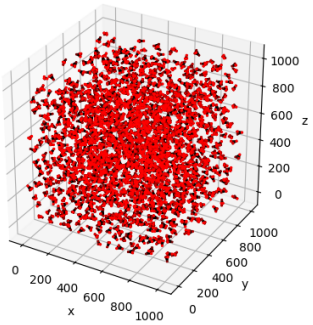


FIG. 4: A cubic box with a side length of $L=1000$ is depicted, and it contains 1500 counterclockwise tetrahedra. Each tetrahedron has a primary vertex, from which three sides extend with lengths $r_1\sim 10$, $r_2\sim 20$, and $r_3\sim 30$.

To verify the functionality of our codes, we proceed to calculate the 4PCF estimator for each tetrahedron, considering both counterclockwise and clockwise tetrahedra. The computation is performed within a specific radial range: $r_1\in[5,15]$, $r_2\in[15,25]$, and $r_3\in[25,35]$. This restriction ensures that only the vertex with three sides of lengths approximately $r_1\sim 10$, $r_2\sim 20$, and $r_3\sim 30$ is considered as the primary vertex. The purpose of this selection is to illustrate the connection between the handedness of a tetrahedron and the parity-odd 4PCF. The obtained results are presented in FIG. 5 and 6, using angular momentum multipoles $(l_1,l_2,l_3)=(1,1,1)$ and $(1,2,2)$, respectively.

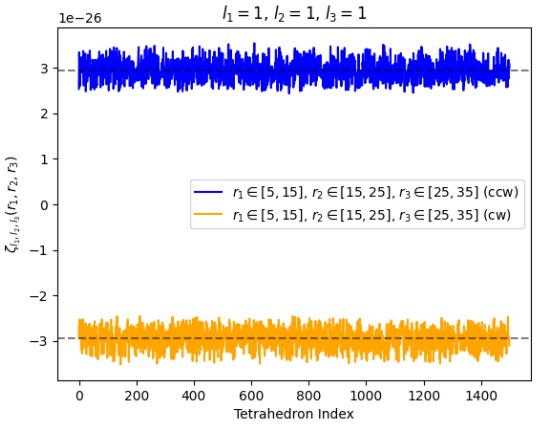


FIG. 5: 4PCF estimators of 1500 counterclockwise (ccw) and clockwise (cw) tetrahedra, using angular momentum multipole $(l_1, l_2, l_3) = (1, 1, 1)$. The dotted line represents the mean value of the 1500 tetrahedra. Clockwise tetrahedra yield the exact same value as their counterclockwise counterparts, but with the opposite sign.

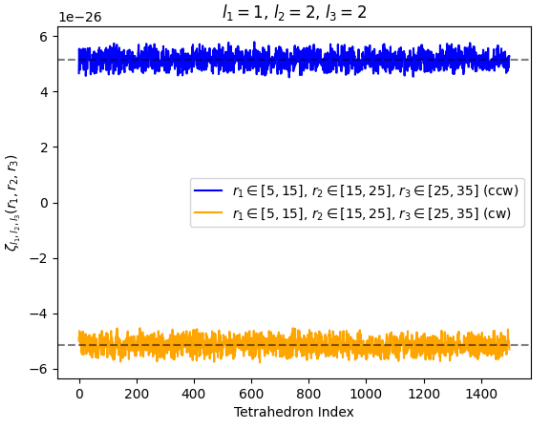


FIG. 6: 4PCF estimators of 1500 counterclockwise (ccw) and clockwise (cw) tetrahedra, using angular momentum multipole $(l_1, l_2, l_3) = (1, 2, 2)$. The dotted line represents the mean value of the 1500 tetrahedra. Clockwise tetrahedra yield the exact same value as their counterclockwise counterparts, but with the opposite sign.

The lowest-lying parity-odd isotropic basis function corresponds to the subscripts $l_1 = l_2 = l_3 = 1$. Put this to eq. (22),

$$\mathcal{P}_{111}(\hat{\vec{r}}_1, \hat{\vec{r}}_2, \hat{\vec{r}}_3) = -\frac{3i}{\sqrt{2}(4\pi)^{\frac{3}{2}}} \hat{\vec{r}}_1 \cdot (\hat{\vec{r}}_2 \times \hat{\vec{r}}_3). \tag{31}$$

As anticipated, the counterclockwise tetrahedra give a positive projection on this function, while the clockwise tetrahedra gives a negative projection (see FIG. 5).

In a given tetrahedron, any of its four vertices can be chosen as the primary vertex. Among these vertices, some will be considered "counterclockwise" while others will be deemed "clockwise." When the 4PCF estimators are placed in the same radial bin and angular momentum multipole $(l_1, l_2, l_3)$, there is a possibility of internal cancellation occurring between positive and negative values. This cancellation leads to a diminished parity-odd signal. To eliminate the cancellation, it is preferable to use finer binning, which reduces the internal cancellation. The effect of using finer binning is demonstrated in Section IV.

## IV. DATA

Our test sample consists of 1000 galaxies extracted from the twelfth data release (DR12) **[20]** of the SDSS-III **[8]** as part of the BOSS. Two samples, namely CMASS and LOWZ, are included in the survey, and we use the former for our study. Luminous Red Galaxies (LRGs) account for approximately 74% in the CMASS sample, while the rest are late-type spirals **[21]**. To analyze the data, we divide the radial range $r_i \in [20, 160]h^{-1}$Mpc into 10 evenly spaced bins, with a width of $\Delta r = 14h^{-1}$Mpc. Furthermore, we impose the conditions $r_2 > r_1 + \Delta r$ and $r_3 > r_2 + \Delta r_1$, resulting in a total of 56 radial bins. The impact of finer binning is also investigated by employing 18 evenly spaced bins covering a range of $r_i \in [20, 164]h^{-1}$Mpc, giving a total of 816 radial bins. Due to time constraints, we only expand the parity-odd 4PCF in angular momentum multipole $(1, 1, 1)$ and $(1, 2, 2)$. Our codes have not yet incorporated edge-correction.

To convert angles and redshifts to Cartesian coordinates, we employ a fiducial cosmology with parameters $\Omega_m = 0.31$, $\Omega_\Lambda = 0$, $\Omega_r h^2 = 4.165 \times 10^{-5}$, $\Omega_k = 1 - (\Omega_m + \Omega_r + \Omega_\Lambda)$, and $h = 0.676$. In practice, the integral presented in eq. (27) needs to be transformed into a sum over discrete galaxies, where each galaxy is assigned a weight based on

$$w = w_{\rm fkp} w_{\rm sys} (w_{\rm noz} + w_{\rm cp} - 1), \tag{32}$$

which $w_{\rm fkp}$ represents the FKP weight **[22]**, $w_{\rm sys}$ is the systematic weight that combines stellar density and seeing, $w_{\rm noz}$ accounts for the redshift failure weight, and $w_{\rm cp}$ represents the fiber collision weight. The redshift failure weight $w_{\rm noz}$ is a factor that accounts for galaxies in the sample whose redshift measurements were unsuccessful or unreliable. This weight is used to downweight or exclude such galaxies from the analysis to prevent their influence on the results. On the other hand, the fiber collision weight $w_{\rm cp}$ is applied to account for pairs of galaxies that have overlapping fiber collision zones in the spectroscopic observations. These collisions occur when two galaxies are close enough that only one of them can be assigned a spectrum because of the limited fiber size. The fiber collision weight adjusts the weights of these galaxies to eliminate the effects of incomplete sampling

caused by the collisions. Eq. (27) is then rewritten as

$$\hat{\zeta}_{l_1 l_2 l_3}(r_1, r_2, r_3) = \sum_{i=1}^{N_g} \sum_{m_1 m_2 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \times w_i a_{l_1 m_1}(\vec{x}_i; r_1) a_{l_2 m_2}(\vec{x}_i; r_2) a_{l_3 m_3}(\vec{x}_i; r_3). \tag{33}$$

The computed 4PCF estimators are displayed in FIG. 7, 8, 9, and 10. FIG. 7 and 8 demonstrate the results using angular momentum multipole $(1, 1, 1)$ and $(1, 2, 2)$ respectively, with $\Delta r = 14h^{-1}$Mpc being the bin width. Their sizes are comparable, both with an order of $\sim 10^{-17}$. FIG. 9 shows $\hat{\zeta}_{111}$ with a finer binning of $\Delta r = 8h^{-1}$Mpc, which gives an order of roughly $10^2$ bigger than the measurement using $\Delta r = 14h^{-1}$Mpc, which is consistent with our expectation. For visibility, we rescale the 4PCF in FIG. 9 by multiplying a factor of $r_1 r_2 r_3$, and the result is illustrated in FIG. 10.

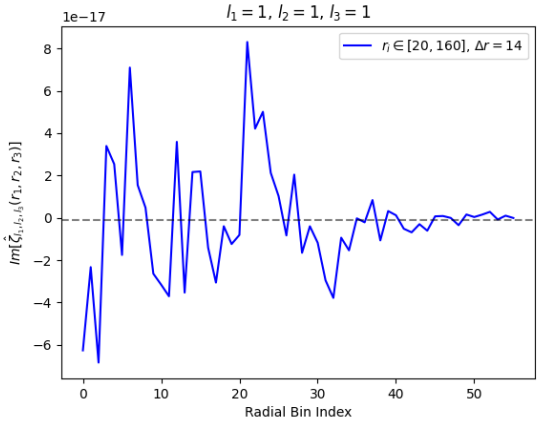


FIG. 7: Measurement of $\hat{\zeta}_{111}$ from the test sample. The measurement was performed using 10 evenly spaced bins covering a range of $r_i \in [20, 160]h^{-1}$Mpc. The dotted line represents the mean value across all radial bins.

## V. OUTLOOK

To assess the statistical significance of the parity-odd signal in the 4PCF, it is necessary to compare the signal observed in the BOSS data with that obtained from mock catalogs. This approach follows the methodology described in **[1]** and **[2]**, which involves conducting a non-parametric rank test. The rank test is a commonly used method for analyzing data when the underlying likelihood is unknown, which we examine the null hypothesis of zero parity-odd 4PCF and calculate the test statistic as follows:

$$\tilde{\chi}^2 = \zeta^T \tilde{C}^{-1} \zeta. \tag{34}$$

In eq. (34), $\zeta$ is a vector representing the set of mea-

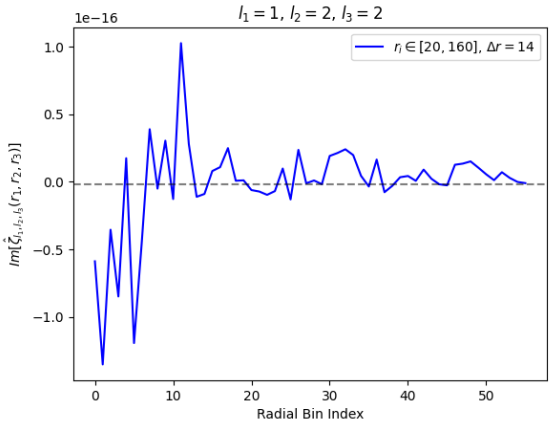


FIG. 8: Measurement of $\hat{\zeta}_{122}$ from the test sample. The measurement was performed using 10 evenly spaced bins covering a range of $r_i \in [20, 160]h^{-1}$Mpc. The dotted line represents the mean value across all radial bins.

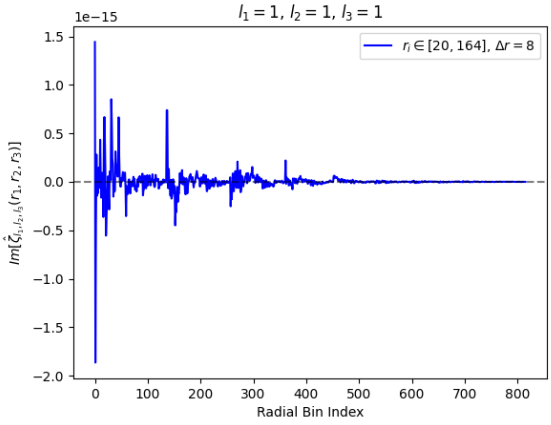


FIG. 9: Measurement of $\hat{\zeta}_{111}$ from the test sample. The measurement was performed using 18 evenly spaced bins covering a range of $r_i \in [20, 164]h^{-1}$Mpc. The dotted line represents the mean value across all radial bins.

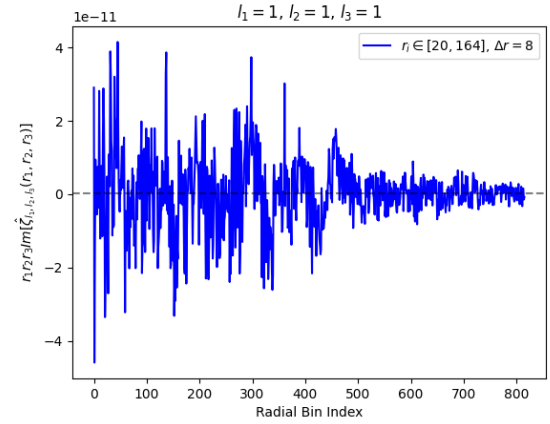


FIG. 10: Same as FIG. 9, but rescaled by a factor of $r_1 r_2 r_3$.

sured parity-odd 4PCF estimators, and $\tilde{C}$ is the theoretical covariance matrix. To conduct the test, we compute $\tilde{\chi}^2$ for both the BOSS data and each of the mock catalogs. The confidence level is upper-constrained at $(1-1/N_{\rm mocks})$, where $N_{\rm mocks}$ is the total number of mock catalogs available.

Before proceeding with the non-parametric rank test, a challenge has been identified in the research progress. Although the computational complexity of the 4PCF has been effectively reduced from $\mathcal{O}(N_g^4)$ to $\mathcal{O}(N_g^2)$, the computation time still remains significant. TABLE I provides a summary of the computation time required for calculating parity-odd 4PCF estimators on my local computer.

TABLE I: Computation time of parity-odd 4PCF estimators. $N_r$ denotes the number of radial bins, and $t_{l_1l_2l_3N_r}$ represents the computation time in unit of hours.

| $N_r$ | $(l_1, l_2, l_3)$ | $t_{l_1l_2l_3N_r}$ (hr) |
|---|---|---|
| 56 | $(1, 1, 1)$ | $\sim 0.5$ |
| 56 | $(1, 2, 2)$ | $\sim 1.5$ |
| 816 | $(1, 1, 1)$ | $\sim 9$ |

When computing the parity-odd 4PCF, the summation in eq. (33) involves counting pairs of galaxies. Consequently, we expect that the computation time will scale as $N_g^2$. The complexity increases linearly with the number of radial bin $N_r$ as well. Given the substantial computation time required, as exemplified in TABLE I, conducting such an analysis is currently infeasible. To resolve this issue, we are developing codes with the aim of reducing computation time through GPU-acceleration.

With the abundance of spectroscopic data expected over the next decade from surveys like DESI [**23**], Euclid [**24**] and the Rubin Observatory [**25**], galaxy surveys seem to be ideal for hunting for evidence of parity-violating interactions. These large-scale surveys could help confirm whether the observed parity-odd 4PCF actually exist in our universe.

## VI. CONCLUSION

The code for computing the 4PCF estimator of the BOSS galaxy sample has been successfully established in this work. The functionality of the code has been demonstrated using a toy model of 1500 tetrahedra. $\hat{\zeta}_{111}$ and $\hat{\zeta}_{122}$ of the 4PCF estimator have been computed using 1000 galaxies from the BOSS CMASS galaxy sample. Two different sets of radial bins are employed, with a binning width of $\Delta r = 8h^{-1}$Mpc and $\Delta r = 14h^{-1}$Mpc, respectively. For the former, the 4PCF components corresponding to angular momentum multipoles $(1, 1, 1)$ and $(1, 2, 2)$ are computed, both giving a size of $\sim 10^{-17}$. For multipole $(1, 1, 1)$, we have also applied a finer binning of $\Delta r = 8h^{-1}$Mpc, yielding a result roughly $10^2$ times larger than with $\Delta r = 14h^{-1}$Mpc. On account of the significant computation time required, conducting a non-parametric rank test to extract the statistical significance is currently impractical. To address this issue, GPU-accelerated computation will be utilized to speed up the calculations. This work will be continued as my SCIE 4500 project.

## ACKNOWLEDGMENTS

I would like to express my gratitude to Prof. Tao Liu, Sun Wenrong, and Miao Liyang for the insightful discussions. I would also like to thank Fong Ching for providing the derivation of the commutator form of in-in formalism.

[1] O. H. E. Philcox, Phys. Rev. D **106**, 063501 (2022).
[2] J. Hou, Z. Slepian, and R. N. Cahn, Monthly Notices of the Royal Astronomical Society **522**, 5701 (2023).
[3] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, Phys. Rev. **105**, 1413 (1957).
[4] A. D. Sakharov, in *In The Intermissions... Collected Works on Research into the Essentials of Theoretical Physics in Russian Federal Nuclear Center, Arzamas-16* (World Scientific, 1998) pp. 84–87.
[5] O. H. Philcox and Z. Slepian, Proceedings of the National Academy of Sciences **119**, e2111366119 (2022).
[6] M. Shiraishi, C. Hikage, R. Namba, T. Namikawa, and M. Hazumi, Phys. Rev. D **94**, 043506 (2016).
[7] D. Jeong and M. Kamionkowski, Phys. Rev. Lett. **108**, 251301 (2012).
[8] D. J. Eisenstein, D. H. Weinberg, E. Agol, *et al.*, The Astronomical Journal **142**, 72 (2011).
[9] J. Maldacena, Journal of High Energy Physics **2003**, 013 (2003).
[10] N. Bartolo, S. Matarrese, and A. Riotto, Phys. Rev. D **65**, 103505 (2002).
[11] Y. Wang, Communications in Theoretical Physics **62**, 109 (2014).
[12] T. Liu, X. Tong, Y. Wang, and Z.-Z. Xianyu, Journal of High Energy Physics **2020**, 1 (2020).
[13] X. Niu, M. H. Rahat, K. Srinivasan, and W. Xue, Journal of Cosmology and Astroparticle Physics **2023** (05), 018.
[14] P. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Banday, R. Barreiro, J. Bartlett, *et al.*, A13 **16**, 10.1051/0004-6361/201525830 (2016).
[15] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. Bennett, J. Dunkley, M. Nolta, M. Halpern, R. Hill, N. Odegard, *et al.*, The Astrophysical Journal Supplement Series **208**, 19 (2013).
[16] O. H. E. Philcox, Physics Division Seminar, Presentation (2021), Lawrence Berkeley National Laboratory (LBNL).
[17] R. N. Cahn, Z. Slepian, and J. Hou, A Test for Cosmological Parity Violation Using the 3D Distribution of

Galaxies (2021), arXiv:2110.12004 [astro-ph.CO].
[18] I. Szapudi and A. S. Szalay, The Astrophysical Journal **494**, L41 (1998).
[19] S. D. Landy and A. S. Szalay, Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 412, no. 1, p. 64-71. **412**, 64 (1993).
[20] S. Alam, F. D. Albareti, C. A. Prieto, *et al.*, The Astrophysical Journal Supplement Series **219**, 12 (2015).
[21] K. L. Masters, C. Maraston, R. C. Nichol, *et al.*, Monthly Notices of the Royal Astronomical Society **418**, 1055 (2011), https://academic.oup.com/mnras/article-pdf/418/2/1055/17331036/mnras0418-1055.pdf.
[22] H. A. Feldman, N. Kaiser, and J. A. Peacock, The Astrophysical Journal **426**, 23 (1994).
[23] D. Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, *et al.*, The DESI Experiment Part I: Science, Targeting, and Survey Design (2016), arXiv:1611.00036 [astro-ph.IM].
[24] R. Laureijs, J. Amiaux, S. Arduini, *et al.*, Euclid Definition Study Report (2011), arXiv:1110.3193 [astro-ph.CO].
[25] L. S. Collaboration, P. A. Abell, J. Allison, S. F. Anderson, *et al.*, LSST Science Book, Version 2.0 (2009), arXiv:0912.0201 [astro-ph.IM].