

## Include the Extreme Faint Galaxy in Shear Measurement

HEKUN LI<sup>1</sup> AND JUN ZHANG<sup>\*1</sup>

<sup>1</sup>*Department of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China*

Submitted to ApJ

### ABSTRACT

*Keywords:* gravitational lensing: weak-methods: data analysis

#### 1. INTRODUCTION

noise bias, SNR threshold  
precision achieved  
cite the previous paper, cut and calibration to avoid bias  
introduce some aspects of Fourier\_Quad and PDF\_SYM  
recently, we found small bias on the faint end  
multiplicative & additive bias of these methods

#### 2. APPROACH THE SUB-PERCENT LEVEL ACCURACY

##### 2.1. The Fourier\_Quad method

The Fourier\_Quad is a model-independent shear measurement method which estimates the shape of galaxy based on its 2D power spectrum (Zhang 2008; Zhang et al. 2015, 2017). The shear estimators are defined as:

$$\begin{aligned} G_1 &= - \int d^2\vec{k} (k_x^2 - k_y^2) T(\vec{k}) M(\vec{k}) \\ G_2 &= -2 \int d^2\vec{k} k_x k_y T(\vec{k}) M(\vec{k}) \\ N &= 2 \int d^2\vec{k} \left[ k^2 - \frac{\beta^2}{2} k^4 \right] T(\vec{k}) M(\vec{k}), \\ U &= -\beta^2 \int d^2\vec{k} (k_x^4 - 6k_x^2 k_y^2 + k_y^4) T(\vec{k}) M(\vec{k}), \\ V &= -4\beta^2 \int d^2\vec{k} (k_x^3 k_y - k_x k_y^3) T(\vec{k}) M(\vec{k}). \end{aligned} \quad (1)$$

The  $M(\vec{k})$  is the power spectrum of the galaxy image from which the background and Poisson noise have been subtracted through:

$$M(\vec{k}) = \left| \tilde{f}_S(\vec{k}) \right|^2 - P_S - \left| \tilde{f}_B(\vec{k}) \right|^2 + P_B \quad (2)$$

$$P_S = \frac{\int_{|\vec{k}| > k_c} d^2\vec{k} \left| \tilde{f}_S(\vec{k}) \right|^2}{\int_{|\vec{k}| > k_c} d^2\vec{k}}, \quad P_B = \frac{\int_{|\vec{k}| > k_c} d^2\vec{k} \left| \tilde{f}_B(\vec{k}) \right|^2}{\int_{|\vec{k}| > k_c} d^2\vec{k}},$$

where  $\tilde{f}_S(\vec{k})$  and  $\tilde{f}_B(\vec{k})$  are the Fourier transformations of the galaxy image and a neighboring image of background noise respectively.  $P_B$  and  $P_S$  are the estimates of the Poisson noise power spectrum of the background noise and source images respectively. A critical wavelength,  $k_c$ , is required to avoid the contamination of the source power (Zhang et al. 2015). The factor  $T(\vec{k})$  is the ratio of power spectrum of the target isotropic Gaussian PSF (point spread function) to that of the original PSF,  $T(\vec{k}) = |\tilde{W}_\beta(\vec{k})|^2 / |\tilde{W}_P(\vec{k})|^2$ . It is designed to convert the original PSF,  $W_P(\vec{x})$ , to an isotropic Gaussian PSF,  $W_\beta(\vec{x})$ , so that the effect of PSF can be corrected rigorously and model-independently. The scale radius,  $\beta$ , of  $W_\beta(\vec{x})$  should be chosen slightly larger than that of  $W_P(\vec{x})$  to avoid the singularities in the factor  $T(\vec{k})$ .

Fourier\_Quad method provides two statistical approaches for shear recovery from the estimators, the average approach and the PDF\_SYM approach. The former approach recovers the shear signal by taking ensemble averages of the shear estimators:

$$\frac{1}{2} \frac{\langle G_1 \rangle}{\langle N \rangle} = g_1 + O(g_{1,2}^3), \quad \frac{\langle G_2 \rangle}{\langle N \rangle} = g_2 + O(g_{1,2}^3). \quad (3)$$

It is precise to the second order of shear in accuracy (Zhang et al. 2015). Zhang et al. (2015) have shown the appealing advantage of Fourier\_Quad that including the extreme faint sources or even point sources would not bias the shear measurement. However, an appropriate weight should be considered in Eq.3 because the bright galaxies will dominate in the ensemble average and enlarge the estimate of error bar.

To approach the lower statistical error bound, Zhang et al. (2017) propose a new shear recovery method in the

Fourier\_Quad framework which is called the PDF\_SYM approach. It recovers the shear signal by symmetrizing the probability distribution function (PDF) of the shear estimators. Zhang et al. (2017) show that the PDF of  $\hat{G}_{1/2}$  would be maximally symmetrical respect to zero if the shear estimate,  $\hat{g}_{1/2}$  is the true one (see the Ref. for more details).

$$\hat{G}_1 = G_1 - \hat{g}_1(N + U) = \int d^2\vec{k} P_k T(\vec{k}) M(\vec{k}), \quad (4)$$

$$\hat{G}_2 = G_2 - \hat{g}_2(N - U) = \int d^2\vec{k} Q_k T(\vec{k}) M(\vec{k}),$$

$$P_k = k_x^2 - k_y^2 + \hat{g}_1 [2k^2 - \beta^2 (k^4 + k_x^4 - 6k_x^2 k_y^2 + k_y^4)],$$

$$Q_k = 2k_x k_y + \hat{g}_2 [2k^2 - \beta^2 (k^4 - k_x^4 + 6k_x^2 k_y^2 - k_y^4)].$$

It needs two additional shear estimators,  $U$  and  $V$ . The  $V$  term is kept for the rotation with  $U$ . The PDF\_SYM approach estimates each component of shear ( $g_1$  or  $g_2$ ) independently. Therefore, when estimating one component, it is convenient to ignore the other one or set it zero.

Unlike the weighted sum of shear estimators used by the other methods, the PDF\_SYM approach includes the full PDF of the shear estimators and do not weight down the faint sources. Therefore, it could make full use of the additional information from faint sources to achieve a lower statistical error. The PDF\_SYM approach would be a promising method for shear recovery in the Fourier\_Quad framework.

Recently, we find the PDF\_SYM approach would be biased by the very faint galaxies. Figure 1 shows the multiplicative and additive bias from the sample of different SNR. The multiplicative bias would increase to about 1% level when the SNR is about 10. If the PSF has some anisotropy ( $e \sim 0.1$ ), we could also find additive bias ( $\sim 6 \times 10^{-4}$ ) at the same SNR level. However, the average approach is immune to this bias even applied to the extreme noisy sample.

## 2.2. Bias due to the source and noise mix

To reveal the origin of the bias, let us start from the components of noisy galaxy image which could be written as a sum of source and noise image, *i.e.*  $f_S(\vec{x}) = f_G(\vec{x}) + f_N(\vec{x})$ . Therefore, the modified power spectrum of galaxy image,  $M(\vec{k})$ , can be rewritten as

$$M(\vec{k}) = \left| \tilde{f}_G(\vec{k}) \right|^2 + C(\vec{k}) + \Delta N(\vec{k}). \quad (5)$$

The first term is the power spectrum of the noise free galaxy image.  $C(\vec{k})$  is the mixture term of the Fourier transform of source and that of the background noise.  $\Delta N(\vec{k})$  is the noise power spectrum residual.

$$C(\vec{k}) = \tilde{f}_G^*(\vec{k}) \tilde{f}_N(\vec{k}) + \tilde{f}_G(\vec{k}) \tilde{f}_N^*(\vec{k}), \quad (6)$$

$$\Delta N(\vec{k}) = \left| \tilde{f}_N(\vec{k}) \right|^2 - \left| \tilde{f}_B(\vec{k}) \right|^2.$$

We have dropped the estimate of the Poisson noise power spectrum,  $P_S$  and  $P_B$ , because they are not import here.

To understand the bias mechanism, we should consider the contribution of  $C(\vec{k})$  to the shear estimators in Eq.(1). For the simplicity in calculation, let  $F_k$  and  $N_k$  denotes the Fourier transform of noise free galaxy image and that of the background noise image respectively. Assuming that the PDF\_SYM approach has been applied, we obtain

$$\hat{G}_1 = \int d^2\vec{k} P_k T(\vec{k}) \left( |F_k|^2 + C(\vec{k}) \right) \quad (7)$$

$$= \Delta^2 \sum_k A_k \left( |F_k|^2 + 2F_k^{Re} N_k^{Re} + 2F_k^{Im} N_k^{Im} \right),$$

$$\hat{G}_2 = \int d^2\vec{k} Q_k T(\vec{k}) \left( |F_k|^2 + C(\vec{k}) \right)$$

$$= \Delta^2 \sum_k B_k \left( |F_k|^2 + 2F_k^{Re} N_k^{Re} + 2F_k^{Im} N_k^{Im} \right).$$

Here, we have dropped the notation of integral variation for simplicity and rewritten the integration into the discrete form.

Assuming that the noise is mainly white noise, the PDF of  $N_k^{Re}$  and  $N_k^{Im}$  should be a normal distribution which is independent of  $\vec{k}$ . The standard deviation is assume to be  $\sigma$ . The PDF of  $G_1$  and  $G_2$  is given by<sup>1</sup>:

$$P_G(G_1, G_2) = \frac{1}{(2\pi)^2} \frac{\pi}{\sqrt{B}} \exp \left[ -\frac{A}{4B} \right] \quad (8)$$

where

$$A = (G_1 - \Gamma_1)^2 U_3 + (G_2 - \Gamma_2)^2 U_1 \quad (9)$$

$$-2(G_1 - \Gamma_1)(G_2 - \Gamma_2)U_2,$$

$$B = U_1 U_3 - U_2^2,$$

$$(\Gamma_1, \Gamma_2) = \Delta^2 \sum_k |F_k|^2 (A_k, B_k)$$

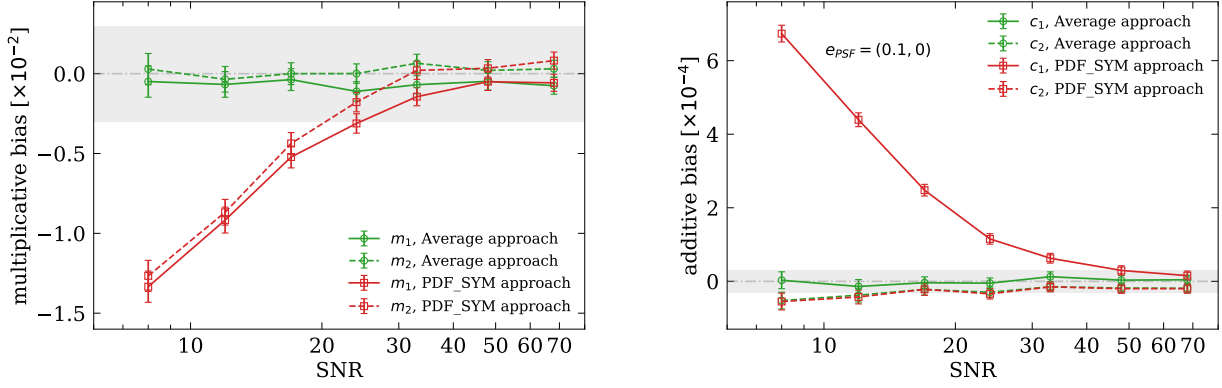
$$(U_1, U_2, U_3) = 2\sigma^2 \Delta^4 \sum_k |F_k|^2 (A_k^2, A_k B_k, B_k^2)$$

The 1-D PDF of  $G_1(G_2)$  is needed by the PDF\_SYM approach for  $g_1(g_2)$  estimating. Therefore, for example, let us consider the PDF of  $G_1$ . By integrating the  $G_2$ , we would have

$$\int_{-\infty}^{\infty} dG_2 P_G(G_1, G_2) \quad (10)$$

$$= \frac{1}{(2\pi)^2} \frac{\pi\sqrt{\pi}}{\sqrt{U_1}} \exp \left\{ -\frac{(G_1 - \Gamma_1)^2}{4U_1} \right\}$$

<sup>1</sup> The details of calculation in this section can be found in Appendix A.



**Figure 1.** The multiplicative and additive bias measured from the galaxy of different SNR level. The shaded region in the left (right) panel is between  $\pm 3 \times 10^{-3}$  ( $\pm 3 \times 10^{-5}$ ) which is the fluctuation allowed in the Fourier\_Quad method.

where

$$\begin{aligned} \Gamma_1 &= \int d^2 \vec{k} (k_x^2 - k_y^2) \exp(-\beta^2 k^2) |F^S(\vec{k})|^2, \\ U_1 &= 2\sigma^2 \Delta^2 \int d^2 \vec{k} (k_x^2 - k_y^2)^2 \exp(-2\beta^2 k^2) \frac{|F^S(\vec{k})|^2}{|W_P(M\vec{k})|^2}, \\ M &= \begin{bmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{bmatrix} \end{aligned} \quad (11)$$

We have assumed that the 1-D PDF of  $G_1$  has been symmetrized by the true shear signal. Therefore, any asymmetry part, which should vanish, in the PDF will cause the bias. Assuming the ellipticity of the PSF is  $(e_1, e_2)$ , and suppose the components are small. For convenience in the rest of our discussion, we set  $g_2 = 0$ , and the ellipticity  $e_2$  of the PSF is also zero. Then, we have

$$|W_P(M\vec{k})|^{-2} = W^{-2} W' [2(e_1 + g_1)(k_x^2 - k_y^2)] + W^{-1} \quad (12)$$

Therefore, we can rewrite the functions of  $\Gamma_1$  and  $U_1$  as:

$$\begin{aligned} U_1 &= U_{10} + \delta U, \\ \Gamma_1 &= \int k dk \int d\theta \exp(-\beta^2 k^2) |F^S(k, \theta)|^2 k^2 \cos(2\theta) \end{aligned} \quad (13)$$

where

$$\begin{aligned} U_1 &= 2\sigma^2 \Delta^2 \int k dk \int d\theta \exp(-2\beta^2 k^2) \\ &\quad \times |F^S(k, \theta)|^2 W^{-1} k^4 \cos^2(2\theta) \\ \delta U &= 2\sigma^2 \Delta^2 \int k dk \int d\theta \exp(-2\beta^2 k^2) \\ &\quad \times |F^S(k, \theta)|^2 [2(e_1 + g_1) W^{-2} W' k^6 \cos^3(2\theta)] \end{aligned} \quad (14)$$

For the 1-D PDF of  $G_1$ , we have:

$$\int_{-\infty}^{\infty} dG_2 P_G(G_1, G_2) \quad (15)$$

$$\begin{aligned} &= \frac{1}{(2\pi)^2} \frac{\pi \sqrt{\pi}}{\sqrt{U_1}} \exp \left\{ -\frac{(G_1 - \Gamma_1)^2}{4U_1} \right\} \\ &\approx \frac{1}{(2\pi)^2} \frac{\pi \sqrt{\pi}}{\sqrt{U_{10}}} \left[ 1 - \frac{\delta U}{2U_{10}} + \frac{(G_1 - \Gamma_1)^2}{4U_{10}} \frac{\delta U}{U_{10}} \right] \\ &\quad \times \exp \left\{ -\frac{(G_1 - \Gamma_1)^2}{4U_{10}} \right\} \end{aligned}$$

The terms proportional to  $\delta U$ , when averaged over galaxies, generate the asymmetry of  $P_G$ .

It should be noted that  $U_1$  is derived from the  $C(\vec{k})$  which only contains  $\tilde{f}_G(\vec{k})$  not the power spectrum  $|\tilde{f}_G(\vec{k})|^2$ . Therefore, the PSF deconvolution (the factor  $T(\vec{k}) = |\tilde{W}_\beta(\vec{k})|^2 / |\tilde{W}_P(\vec{k})|^2$ ) will give rise to an additional factor,  $1/|W_P(M\vec{k})|^2$ , in  $U_1$ . However, the PSF is appropriately removed from the  $\Gamma_1$  which is derived from the power spectrum of galaxy image. So the appropriate PSF deconvolution for  $C(\vec{k})$ , the mixture of Fourier transform of the galaxy image and background image, is the origin of bias. See Appendix A for more details.

### 2.3. Numerical test

To test the calculation against data, we simulate galaxy images of different SNR. We use the random walk method to generate galaxies.

#### 2.3.1. Image simulation

The real galaxies consist of millions of stars which are the point sources. Each galaxy is made of a number of points. In this method, we let a random walker wanders in a circular region from image center. The points of galaxy are the collection of positions of each step the walker has reached. The merit of random walk method includes precise shape distortion, less computational cost and efficient PSF convolution. We use the

Moffat profile for PSF in the simulation.

$$I(r) \propto \left[ 1 + \left( \frac{r}{r_d} \right)^2 \right]^{-3.5} H(r_c - r). \quad (16)$$

$H(r_c - r)$  is the Heaviside step function,  $r_d$  is the scale length and  $r_c$  is set to 4 times  $r_d$ . The stamp size is  $44 \times 44$ . 40 shear pairs random choice on the  $g_1 - g_2$  plane are used.  $10^7$  galaxies are generated for each shear point.

### 2.3.2. Test result

In the image simulation, we could separate the components of  $M(\vec{k})$  for measurement independently. We test different combinations of the components from  $M(\vec{k})$ . We find that the noise power spectrum residual,  $\Delta N(\vec{k})$ , would not bias the measurement. While, the result will be biased if the  $C(\vec{k})$  term is included in the measurement. Furthermore, we find that both multiplicative and additive bias will vanish when  $|\widetilde{W}_P(\vec{k})|$  rather than  $|\widetilde{W}_P(\vec{k})|^2$  is used in the PSF deconvolution ( $T(\vec{k})$  factor) for the  $C(\vec{k})$ . We present the comparison in Figure 2.

### 2.4. Improvement to the sub-percent precision

Although an appropriate PSF deconvolution for the  $C(\vec{k})$  term would correct the bias, image components separation could not be achieved in the real data. It should be noted that what we have for measurement are the noisy images and the shear estimators measured from that. Among the components of  $M(\vec{k})$ , the  $C(\vec{k})$  part is the one that easy to estimate. We propose a new method that utilizes the shear estimators from the estimated  $C(\vec{k})$  to correct the bias.

Recall that the original measurement, Eq.(2), employs the power spectrum of a noise image to subtract the contamination to shear estimators from the background noise. However, this operation leaves an annoying term,  $C(\vec{k})$ , to the shear measurement. In our new method,

we need a second noise image,  $f_{B'}(\vec{x})$ , to estimate this component. Then, we would get the shear estimators from this component by applying the shear measurement to it. There are three steps:

- Add the second noise image to the original noisy galaxy image,  $f_S(\vec{x})$ , to get a new image,  $f_{S'}(\vec{x})$ .

$$f_{S'}(\vec{x}) = f_G(\vec{x}) + f_N(\vec{x}) + f_{B'}(\vec{x}). \quad (17)$$

- Subtract the power spectrum of the new noise image and that of the original noisy galaxy image from the power spectrum of the new image above.

$$C'(\vec{k}) = \left| \widetilde{f}_{S'}(\vec{x}) \right|^2 - \left| \widetilde{f}_S(\vec{x}) \right|^2 - \left| \widetilde{f}_B(\vec{x}) \right|^2 \quad (18)$$

$$= \widetilde{f}_G^*(\vec{k}) \widetilde{f}_B(\vec{k}) + \widetilde{f}_G(\vec{k}) \widetilde{f}_B^*(\vec{k}).$$

The noise-noise power mixture,  $\widetilde{f}_N^*(\vec{k}) \widetilde{f}_B(\vec{k}) + \widetilde{f}_N(\vec{k}) \widetilde{f}_B^*(\vec{k})$ , has been neglected because we find it would not change the result in our new method.

- Apply the shear measurement to  $C'$  to get the new shear estimator:  $G_1^C$ ,  $G_2^C$ ,  $N^C$ ,  $U^C$ , and  $V^C$ .

The calculation of Section 2.2 indicates that the noise part in  $C(\vec{k})$  is not import because it vanishes at last. Therefore, utilizing  $C'(\vec{k})$  is a promising way to solve this problem.

use the formulae to show the way of correction by new crossing term

introduce the method to obtain the estimated crossing term & apply to the simulated data

figure of the results of correction, random walk galaxies, SNR.1 ~ SNR.n

figure of the same results from the Galsim galaxies, SNR.1 ~ SNR.n

figure of a more realistic simulation, galaxies form faint to bright, separate the bright and faint sample

## 3. CONCLUSION

## APPENDIX

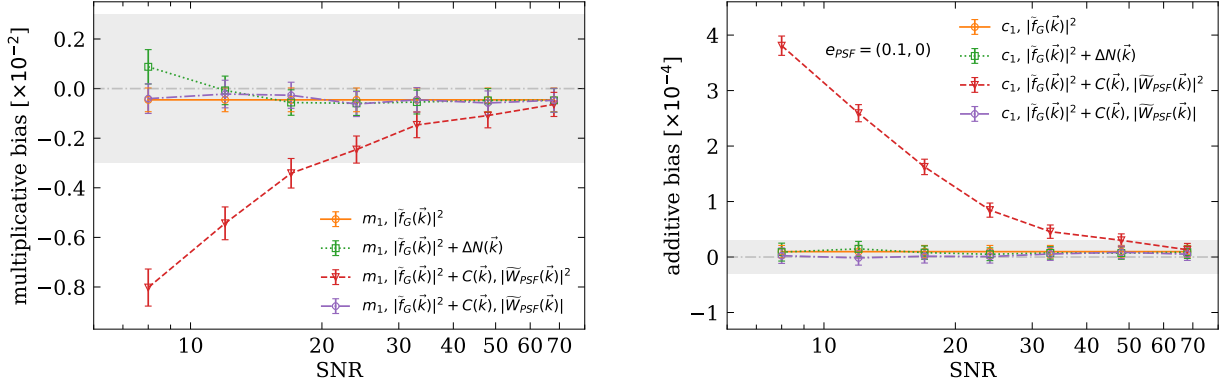
### A. APPENDIX A

$$P_N(N_k^{Re}, N_k^{Im}) = \frac{1}{2\pi\sigma^2} \exp \left[ -\frac{(N_k^{Re})^2 + (N_k^{Im})^2}{2\sigma^2} \right] \quad (A1)$$

$$P_G(G_1, G_2) = \frac{1}{(2\pi)^2} \frac{\pi}{\sqrt{C}} \exp \left[ -\frac{D}{4C} \right]. \quad (A2)$$

The parameters are defined as:

$$C = U_1 U_3 - U_2^2, \quad (A3)$$



**Figure 2.** The multiplicative and additive bias measured from different components of the power spectrum of noisy galaxy image. Because the curves of  $m_1$  and  $m_2$  are very similar, only  $m_1$  curves are presented. Only  $c_1$ 's are shown because all  $c_2$ 's are consistent with zero ( $e_2$  of PSF is zero)). Orange curves: Measured from the power spectrum of noise free galaxy image. Green curves: Measured from the combination of noise free part and the noise power spectrum residual,  $\Delta N(\vec{k})$ . Red curves: Measured from the combination of noise free galaxy image power spectrum and the Fourier transform mixture,  $C(\vec{k})$ . The PSF power spectrum,  $|\widetilde{W}_{PSF}(\vec{k})|^2$ , is used for deconvolution. Purple curves: Same components are used for measurement as the red curves. However,  $|\widetilde{W}_{PSF}(\vec{k})|$  is used (in  $T(\vec{k})$ ) in PSF deconvolution for  $C(\vec{k})$ .

$$\begin{aligned}
 D &= (G_1 - \Gamma_1)^2 U_3 - 2(G_1 - \Gamma_1)(G_2 - \Gamma_2) U_2 \\
 &\quad + (G_2 - \Gamma_2)^2 U_1, \\
 (\Gamma_1, \Gamma_2) &= \Delta^2 \sum_k |F_k|^2 (A_k, B_k) \\
 (U_1, U_2, U_3) &= 2\sigma^2 \Delta^4 \sum_k |F_k|^2 (A_k^2, A_k B_k, B_k^2)
 \end{aligned}$$

$$\begin{aligned}
 \Gamma_1 &= \int d^2 \vec{k} P'_k \exp(-\beta^2 k^2) |F_S(\vec{k})|^2 \\
 U_1 &= 2\sigma^2 \Delta^2 \int d^2 \vec{k} Q'_k \exp(-2\beta^2 k^2) \frac{|F_S(\vec{k})|^2}{|W_{PSF}(\vec{k})|^2}, \\
 P'_k &= (1 + 4\beta^2 g_2 k_x k_y) (k_x^2 - k_y^2), \\
 Q'_k &= (1 + 8\beta^2 g_2 k_x k_y) (k_x^2 - k_y^2)^2.
 \end{aligned} \tag{A4}$$

## B. APPENDIX B

## REFERENCES

- Rowe B, Jarvis M, Mandelbaum R, et al. 2015, A&C, 10, 121  
 Zhang J. 2008, MNRAS, 383, 113  
 Zhang J, Luo W & Sebastien F. 2015 JCAP, 1, 24  
 Zhang J, Zhang P, Luo W. 2017, ApJ, 834, 8