Mon. Not. R. Astron. Soc. **388,** 1819–1830 (2008) doi:10.1111/j.1365-2966.2008.13524.x

Precision of diffuse 21-cm lensing

Tingting Lu[1]⋆ and Ue-Li Pen[2]⋆

[1]*Department of Astronomy and Astrophysics, University of Toronto, Canada M5S 3H4*
[2]*Canadian Institute for Theoretical Astrophysics, University of Toronto, Canada M5S 3H8*

Accepted 2008 May 29. Received 2008 April 16; in original form 2007 October 4

ABSTRACT

We study the limits of accuracy for weak lensing maps of dark matter using diffuse 21-cm radiation from the pre-reionization epoch using simulations. We improve on previous 'optimal' quadratic lensing estimators by using shear and convergence instead of deflection angles. This is a generalization of the deflection estimator, and is more optimal for non-Gaussian sources. The cross-power spectrum of shear and convergence is an unbiased estimator of lensing power spectrum which does not require knowledge of the source four-point function. We find that non-Gaussianity provides a limit to the accuracy of weak lensing reconstruction, even if instrumental noise is reduced to zero. The best reconstruction result is equivalent to Gaussian sources with effective independent cell of side length $2.0\,h^{-1}$ Mpc. Using a source full map from $z = 10$ to 20, this limiting sensitivity allows mapping of dark matter at a signal-to-noise ratio greater than 1 out to $l \lesssim 6000$, which is better than any other proposed technique for large-area weak lensing mapping.

Key words: gravitational lensing – methods: *N*-body simulations – cosmology: observations – cosmology: theory – dark matter – large-scale structure of Universe.

1 INTRODUCTION

Gravitational lensing has been the most successful tool to directly map the distribution of dark matter, and is an important component of modern precision cosmology. Weak gravitational lensing has developed rapidly in recent years, which allows the measurement of the projected dark matter density along arbitrary lines of sight using galaxies as sources. Recently, Smith, Zahn & Doré (2007) have demonstrated the first cosmic microwave background (CMB) lensing detection. The goal is now to achieve high-precision cosmological measurements through lensing, at better than 1 per cent accuracy.

Galaxies are plentiful on the sky, but their intrinsic properties are not understood from first principles, and must be measured from the data. Future surveys may map as many as 10^{10} source objects. Using galaxies as lensing sources has several potential limits (Hirata & Seljak 2004), including the need to calibrate redshift-space distributions and point spread function (PSF) corrections, to be better than the desired accuracy, say 1 per cent. This will be challenging for the next generation of experiments.

Some sources, such as the CMB, are in principle very clean, since its redshift and statistical properties are well understood. Unfortunately, there is only one 2D CMB sky with an exponential damping at $l \gg 1000$, which limits the number of source modes to $\sim 10^6$.

⋆E-mail: ttlu@cita.utoronto.ca (TL); pen@cita.utoronto.ca (U-LP)

The potential of detecting the 21-cm background from the dark ages will open a new window for cosmological detections. Studying the 21-cm background as high-redshift lensing source, as well as the physics of the 21-cm background itself, will provide rich and valuable information to the evolution of Universe. The number of modes on the sky is potentially very large, with numbers of 10^{16} or more. For this reason, 21-cm lensing has recently attracted attention. However, most of the reconstruction methods are based on a Gaussian assumption (Cooray 2004; Pen 2004; Zahn & Zaldarriaga 2006; Metcalf & White 2007; Hilbert, Metcalf & White 2007). In contrast to CMB lensing, where the Gaussian assumption works well, non-Gaussianity in 21-cm lensing may affect the results. Non-linear gravitational clustering leads to non-Gaussianity, and ultimately to reionization. In this paper, we will address the problem of the lensing of pre-reionization gas.

21-cm emission is similar to CMB: both are diffuse backgrounds. It is natural to apply the techniques used in CMB lensing. Hu & Okamoto (2002) expand the CMB lensing field in terms of the gravitational potential (or deflection angles), and construct a trispectrum-based quadratic estimator of potential with maximum signal-to-noise ratio (S/N). However, unlike CMB, the 21-cm background has a 3D distribution and is intrinsically non-Gaussian. A fully 3D analysis is explored in Zahn & Zaldarriaga (2006), where they generalize the 2D quadratic estimator of CMB lensing (Hu & Okamoto 2002) to the 3D Optimal Quadratic Deflection Estimator (OQDE).

A local estimator was proposed in Pen (2004), which assumed a power-law density power spectrum. In this paper, we

will design localized estimators for the lensing fields under the Gaussian assumption, and apply the derived reconstruction technique to Gaussian and non-Gaussian sources. The influence of non-Gaussianity can be measured by comparing the numerical results between the Gaussian and non-Gaussian sources.

Quadratic lensing reconstruction is a two-point function of the lensed brightness temperature field of the 21-cm emission. In this paper, 3D quadratic estimators are constructed for the convergence (κ), as well as the shear (γ). Our method recovers the κ and γ directly instead of gravitational potential or deflection angles. Our estimators have in principle the same form as the OQDE, consisting of the covariance of two filtered temperature maps. The OQDE reconstructs the deflection angle, while our estimators reconstruct the κ and γ fields. Our filtering process can be written as a convolution of the observed fields. As presented in Appendix A and Section 4, our combined estimator is unbiased, and equally optimal as the OQDE for Gaussian sources. Our estimator has better performance for non-Gaussian sources, and recovers three extra (constant) modes for finite fields.

Other authors also developed reconstruction methods from alternative approaches. Metcalf & White (2007) give an estimator for shear. They choose 2D slices at different redshifts, and then treat these slices as independent source samples for the same lensing structure. They neglect the information between these slices, so the efficiency of reconstruction depends on the width of the slices. Cooray (2004) expands the lensed field to higher order of the gravitational potential, and investigates the resulting corrections to the lensed power spectrum.

This paper is organized as follows. The basic framework of lensing and the reconstruction method is introduced in Section 2. The numerical methods are presented in Section 3. The results are discussed in Section 4. We conclude in Section 5.

2 LENSING AND RECONSTRUCTION

Photons are deflected by the gravitational pull of clumpy matter when they propagate from the source to the observer. This effect can be used to map the mass distribution if we can measure the distortion of an image. In this section, we will first review the lensing theory, which serves to define our notation. We then develop an optimal quadratic estimator using a maximum likelihood method. The reconstruction depends on the power spectrum of the source. The noise and normalization of the reconstruction are calculated in Appendix A.

2.1 Lensing

The Jacobian matrix describing the mapping between the source and image planes is defined as

$$\mathbf{J}(\boldsymbol{\theta}, \chi) = \frac{1}{f_K(\chi)} \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\theta}}. \tag{1}$$

Here χ is the radial coordinate and $f_K(\chi)$ is the comoving angular diameter distance. We consider a ray bundle intersecting at the observer and denote $\boldsymbol{x}(\boldsymbol{\theta}, \chi)$ as the comoving transverse coordinate of a ray.

In the lensing literature, the physical quantities frequently used to describe a lensing field are convergence κ and shear γ, which are given by

$$\mathbf{J}(\boldsymbol{\theta}, \chi) = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}.$$

Equivalently, the convergence and shear can also be written as $\kappa = (\Phi_{,11} + \Phi_{,22})/2\,; \gamma_1 = (\Phi_{,11} - \Phi_{,22})/2\,; \gamma_2 = \Phi_{,12}$. Φ is the projected 2D potential:

$$\Phi = \frac{2}{c^2} \int_0^\chi \mathrm{d}\chi' \frac{f_K(\chi') f_K(\chi - \chi')}{f_K(\chi)} \phi[f_K(\chi')\boldsymbol{\theta}(\chi'), \chi']. \tag{2}$$

Here subscripts '1' and '2' refer to the derivative to the two perpendicular transverse coordinates, and ϕ is the 3D Newtonian gravitational potential. Note that the integral should be computed along the actual perturbed path of each photon. In the Born approximation, the deflection is approximated by an integral along the unperturbed path.

In the small-angle approximation (Limber 1954), $\nabla_\perp^2$ can be replaced by ∇^2 in the integral. We get the Limber equation

$$\kappa = \frac{3H_0^2}{2}\, \Omega_\mathrm{m} \int_0^\chi \mathrm{d}\chi' g(\chi', \chi)\, \frac{\delta}{a(\chi')}, \tag{3}$$

where $g(\chi', \chi) = f_K(\chi') f_K(\chi - \chi')/f_K(\chi)$. Ω_m is the mass density parameter, H_0 is the current Hubble constant, a is the scalefactor and δ the overdensity.

Kaiser (1992) derived the Fourier-space version of the Limber equation

$$P_\kappa(l) = \frac{9}{4} \left(\frac{H_0}{c}\right)^4 \Omega_\mathrm{m}^2 \int_0^{\chi_\mathrm{H}} \mathrm{d}\chi \frac{g^2(\chi)}{a^2(\chi)} P\left(\frac{l}{f_K(\chi)}, \chi\right). \tag{4}$$

Here $P_\kappa(l)$ is the 2D power spectrum of the κ field, $P(l/f_K(\chi), \chi)$ is the 3D power spectrum of matter and χ_H is the comoving distance to the Hubble horizon. The equation is valid when the power spectrum P_κ evolves slowly over time corresponding to the scales of fluctuation of interest, and these fluctuation scales are smaller than the horizon scale.

2.2 Reconstruction of large-scale structure

We first heuristically review quadratic lensing estimation in two dimensions. Then we will proceed with a generalization to 3D with a quantitative derivation.

Lensing changes the spatial distribution of a temperature field. Lensing estimation relies on statistical changes to quadratic quantities in the source plane temperature field. We use a tilde to denote a lensed quantity. All estimators work by convolving the temperature field with a window,

$$\tilde{T}_1(\boldsymbol{x}) = \int \mathrm{d}^2x' \tilde{T}(\boldsymbol{x}') W_1(\boldsymbol{x} - \boldsymbol{x}'), \tag{5}$$

and a second window

$$\tilde{T}_2(\boldsymbol{x}) = \int \mathrm{d}^2x' \tilde{T}(\boldsymbol{x}') W_2(\boldsymbol{x} - \boldsymbol{x}'). \tag{6}$$

The quadratic estimator is simply the product of the two convolved temperature fields,

$$E(\boldsymbol{x}) \equiv \tilde{T}_1(\boldsymbol{x}) \tilde{T}_2(\boldsymbol{x}). \tag{7}$$

In the weak lensing case, the estimator is a linear function of the weak lensing parameters (κ, γ). The simplest case is two equal, azimuthally symmetric window functions $W_1 = W_2 = f(r)$. We first consider the limit where κ is a constant value, then the estimator is linearly proportional to κ:

$$\langle E \rangle \propto \kappa + V, \tag{8}$$

and V is a constant related to the mean covariance. Here $\langle \cdots \rangle$ means ensemble average. For a stochastic random field, the ensemble average can be calculated by the volume average if the volume is big

enough. We can absorb V as well as the normalization coefficient into E for convenience, i.e. $E(\boldsymbol{x}) \equiv \tilde{T}_1(\boldsymbol{x})\tilde{T}_2(\boldsymbol{x}) - V$. When κ is spatially variable, E needs to be normalized by a scale-dependent factor $b(k)$. This corresponds to a convolution of κ with a kernel:

$$\langle E(\boldsymbol{x})\rangle = \int \mathrm{d}^2x'\kappa(\boldsymbol{x}')b(\boldsymbol{x}-\boldsymbol{x}'), \tag{9}$$

where kernel b is the Fourier transform of the normalization factor. Therefore $E(\boldsymbol{k}) = b(\boldsymbol{k})\hat{\kappa}(\boldsymbol{k})$, here $\hat{\kappa}(\boldsymbol{k})$ is the Fourier transform of the estimator $\hat{\kappa}(\boldsymbol{x})$.

One can optimize the functions to minimize the error on the lensing variables. In this paper we will compare various forms of the smoothing windows, which include as special case the traditional OQDE. The simplest case is a constant value of κ, for which one can compute its variance:

$$\langle\hat{\kappa}^2\rangle = \left\langle(\tilde{T}_1(\boldsymbol{x})^2)^2\right\rangle. \tag{10}$$

Lensing is a small perturbation of the variance, therefore we can calculate the variance from the unlensed source field, i.e. $\langle(T_1(\boldsymbol{x})^2)^2\rangle \approx \langle(\tilde{T}_1(\boldsymbol{x})^2)^2\rangle$. Performing a variation to minimize the variance, one can find the optimal window function. Generally, one would think a fixed window function might not be optimal for all scales, i.e. $\langle\hat{\kappa}(\boldsymbol{k})^2\rangle$ might not be minimized at the same time for all $\boldsymbol{k}$ modes. Fortunately, it turns out that the optimal window functions do not depend on the spatial structure of the lensing field. Only the normalization factor b in equation (9) is scale dependent. We solve the optimal window function at scales $k = |\boldsymbol{k}| \lesssim k_a$, where the constant κ approximation works well. Here the characteristic scale k_a is determined by the smallest scale k_c of 21-cm sources resolved in a given experiment, as well as the non-Gaussianity of sources. We will explain the details later in Section 3. We expect the resulting window function to be optimal for all scales, and will verify this at Section 4.

Shear and deflection angles are tensorial and vectorial quantities and require anisotropic or vectorial choices of the window function

$$\begin{aligned}\boldsymbol{E}_\gamma &= \tilde{\boldsymbol{T}}_1\tilde{\boldsymbol{T}}_2,\ \tilde{\boldsymbol{T}}_1 = \int \mathrm{d}^2\theta'\tilde{T}(\boldsymbol{\theta}')\boldsymbol{W_1}(\boldsymbol{\theta}-\boldsymbol{\theta}'),\ \tilde{\boldsymbol{T}}_2\\ &= \int \mathrm{d}^2\theta'\tilde{T}(\boldsymbol{\theta}')\boldsymbol{W_2}(\boldsymbol{\theta}-\boldsymbol{\theta}'),\end{aligned} \tag{11}$$

$$\begin{aligned}\boldsymbol{E}_\mathrm{d} &= \tilde{\boldsymbol{T}}_1\tilde{T}_2,\ \tilde{\boldsymbol{T}}_1 = \int \mathrm{d}^2\theta'\tilde{T}(\boldsymbol{\theta}')\boldsymbol{W_1}(\boldsymbol{\theta}-\boldsymbol{\theta}'),\ \tilde{T}_2\\ &= \int \mathrm{d}^2\theta'\tilde{T}(\boldsymbol{\theta}')W_2(\boldsymbol{\theta}-\boldsymbol{\theta}').\end{aligned} \tag{12}$$

This will be explained in detail in Sections 2.2.2 and 2.3.

The source is usually treated as a Gaussian random field in the literature on reconstruction methods. While this is valid for CMB on large angular scales, 21-cm background sources are not necessarily Gaussian. In this paper we attempt to understand the influence of this non-Gaussianity. Optimal estimators for Gaussian sources are not necessarily optimal for non-Gaussian sources. Here, we will construct the convergence and shear field directly, instead of following the deflection angles or potential field reconstruction in CMB lensing. There are three reasons to do this. First, the strength of lensing is evident through the magnitude of κ or γ since they are dimensionless quantities. The rms deflection angle of photons from 21-cm emission is at the magnitude of a few arcmin, which is comparable to the lensing scales we are resolving. Some authors argued that perturbation theory on the deflection angle will break down at these scales (Cooray 2004; Mandel & Zaldarriaga 2006). However, κ and γ are still small and can still work with perturbation calculations without ambiguity. Secondly, κ and γ have well-defined limits as they approach a constant, while only spatially variable deflection angles or potentials can be measured. This significantly simplifies the derivations. Finally, κ and γ are standard variables to use in broader lensing studies, such as strong lensing and cosmic shear. Using the same convention in different subfields will help to generalize the underlying physics of lensing.

The estimators are unbiased, as shown in Appendix A. Furthermore, we confirm that our combined estimators from κ and γ have the same optimality as the OQDE for Gaussian sources. When the sources are non-Gaussian, our new estimators are more optimal.

2.2.1 *Maximum likelihood estimator of* $\boldsymbol{\kappa}$

We now derive the quantitative window functions for 21-cm lensing reconstruction. Due to their similarity, it is helpful to quickly review the reconstruction in CMB lensing. The early work by Zaldarriaga & Seljak (1999) used the quadratic combination of the derivatives of the CMB field to reconstruct the lens distribution, originally also using a κ and γ formulation. Since the CMB has an intrinsic Gaussian distribution, the optimal quadratic estimator (Hu 2001) can also be applied to lensing reconstruction with CMB polarization (Hu & Okamoto 2002). Zahn & Zaldarriaga (2006) generalized the optimal quadratic estimator of CMB lensing to 21-cm lensing.

We will construct estimators for κ and γ with the 21-cm brightness temperature fields, starting from a maximum likelihood derivation. We will show that the OQDE and our approach are the same if the sources are Gaussian; however, the problem is simplified in a intuitive way by using the limit that κ and γ vary slowly in small scales. We will show later the optimal window functions have the same form when κ and γ vary rapidly.

The magnification is

$$\mu = \frac{1}{(1-\kappa)^2-\gamma^2} \sim 1+2\kappa. \tag{13}$$

The last approximation is valid since both κ and γ are much smaller than 1 in the weak lensing regime.

We use Bayesian statistics and assume the prior distribution of parameter κ to be flat. For an M-pixel map on the sky, the posterior likelihood function of the source field has a Gaussian distribution, and can be written as

$$\mathcal{P}(\tilde{T}(\boldsymbol{k})) = (2\pi)^{-M/2}\det(\mathbf{C}_{\tilde{T}\tilde{T}})^{-1/2}\mathrm{e}^{-1/2\tilde{T}^\dagger\mathbf{C}_{\tilde{T}\tilde{T}}{}^{-1}\tilde{T}}. \tag{14}$$

Here $\tilde{T} = \tilde{T}_\mathrm{b} + n$ is the brightness temperature of the diffuse 21-cm emission lensed by the large-scale structure plus measurement noise. To simplify the algebra, we use the negative logarithm $\mathcal{L}$ of the likelihood function in our calculation,

$$\mathcal{L} = -\ln\mathcal{P} = \frac{1}{2}\tilde{T}^\dagger\mathbf{C}^{-1}_{\tilde{T}\tilde{T}}\tilde{T} + \frac{1}{2}\ln\det\mathbf{C}_{\tilde{T}\tilde{T}}. \tag{15}$$

Here $\tilde{T}$ is the 3D discrete Fourier transform of measured temperature. $\mathbf{C}_{\tilde{T}\tilde{T}} = \mathbf{C}_\mathrm{S} + \mathbf{C}_\mathrm{N}$ is the covariance matrix, and the signal contribution $\mathbf{C}_\mathrm{S}$ and noise contribution $\mathbf{C}_\mathrm{N}$ are both diagonal in Fourier space and uncorrelated to each other. In the continuum limit, the likelihood function can be written as

$$\mathcal{L} = \frac{1}{4\pi^2}\left[\int \mathrm{d}^3k\ln\tilde{P}^\mathrm{tot}_\mathrm{3D}(\boldsymbol{k}) + \int \mathrm{d}^3k\frac{|\tilde{T}(\boldsymbol{k})|^2}{\tilde{P}^\mathrm{tot}_\mathrm{3D}(\boldsymbol{k})}\right]. \tag{16}$$

We use $\tilde{P}^\mathrm{tot}_\mathrm{3D} = \tilde{P}_\mathrm{3D}(\boldsymbol{k}) + P_\mathrm{N}(\boldsymbol{k})$ to represent the signal plus noise power spectrum in the following text, where $\tilde{P}_\mathrm{3D}(\boldsymbol{k})$ is 3D power spectrum of the distorted 21-cm field, and $P_\mathrm{N}(\boldsymbol{k})$ is the noise power spectrum. The dimensionless power spectrum of the 3D 21-cm gas

can be written as

$$\Delta^2_{\rm 3D}(k) = \frac{k^3}{2\pi^2} P_{\rm 3D}(k), \tag{17}$$

where $k = |\boldsymbol{k}|$ since the gas is statistically isotropic.

The geometry of the 21-cm field will be changed by lensing:

$$\begin{aligned}\tilde{T}_{\rm b}(\boldsymbol{k}_\perp, k_\parallel) &= \int {\rm d}^3x \tilde{T}_{\rm b}(\boldsymbol{x}) {\rm e}^{-{\rm i}\boldsymbol{k}\cdot\boldsymbol{x}} \\ &= \int {\rm d}^2x_\perp \int {\rm d}x_\parallel T_{\rm b}((1-\kappa)\boldsymbol{x}_\perp, x_\parallel) {\rm e}^{-{\rm i}(\boldsymbol{k}_\perp\cdot\boldsymbol{x}_\perp + k_\parallel x_\parallel)} \\ &= \frac{1}{(1-\kappa)^2} T_{\rm b}\left(\frac{\boldsymbol{k}_\perp}{1-\kappa}, k_\parallel\right), \end{aligned} \tag{18}$$

where '$\perp$' and '$\parallel$' means the perpendicular and parallel direction to the line of sight, respectively. We ignore the contribution of shear first. Then the length-scale is magnified on the transverse plane by a factor κ. Isotropy is broken in 3D but is still conserved on the 2D cross-section. The statistical properties of the 21-cm field will be changed by lensing, i.e. the power spectrum will also change:

$$\begin{aligned}\left\langle \tilde{T}^*_{\rm b}(\boldsymbol{k}_\perp, k_\parallel) \tilde{T}_{\rm b}(\boldsymbol{k}'_\perp, k'_\parallel) \right\rangle &= (2\pi)^2 \delta^{\rm 2D}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp)(2\pi) \\ &\quad \times \delta^{\rm D}(k_\parallel - k'_\parallel) \tilde{P}_{\rm 3D}(\boldsymbol{k}_\perp, k_\parallel). \end{aligned} \tag{19}$$

The delta function has the property

$$\delta^{\rm 2D}\left(\frac{\boldsymbol{k}_\perp}{1-\kappa} - \frac{\boldsymbol{k}'_\perp}{1-\kappa}\right) = (1-\kappa)^2 \delta^{\rm 2D}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp). \tag{20}$$

Therefore the relationship between the unlensed and lensed power spectrum is

$$\begin{aligned}\tilde{P}_{\rm 3D}(\boldsymbol{k}_\perp, k_\parallel) &= \frac{1}{(1-\kappa)^2} P_{\rm 3D}\left(\frac{\boldsymbol{k}_\perp}{1-\kappa}, k_\parallel\right) \\ &= \frac{1}{(1-\kappa)^2} P_{\rm 3D}\left[\sqrt{\frac{k_\perp^2}{(1-\kappa)^2} + k_\parallel^2}\right]. \end{aligned} \tag{21}$$

The second equivalence is due to the statistical isotropy of the unlensed power spectrum.

The Taylor expansion of the lensed power spectrum is

$$\tilde{P}^{\rm tot}_{\rm 3D} = P^{\rm tot}_{\rm 3D} + \left.\frac{\partial \tilde{P}^{\rm tot}_{\rm 3D}}{\partial\kappa}\right|_{\kappa=0} \kappa + \frac{1}{2} \left.\frac{\partial^2 \tilde{P}^{\rm tot}_{\rm 3D}}{\partial\kappa^2}\right|_{\kappa=\kappa_\star} \kappa^2, \tag{22}$$

where $\kappa_\star$ in the residual term lies between 0 and κ. The first-order derivative at $\kappa = 0$ is $2P_{\rm 3D} + \Delta P_{\rm 3D}$, where $\Delta P_{\rm 3D} = P'_{\rm 3D}\, k(k_\perp^2/k^2)$ and $P'_{\rm 3D}(k) = {\rm d}P_{\rm 3D}(k)/{\rm d}k$. In this paper, we will only consider the first-order perturbation. Now we can verify the accuracy of the approximation $\Delta\kappa_{\rm Taylor}/\kappa$, which is the ratio of the residual term to the second term in equation (22). If we write the first-order derivative of power spectrum to κ as $G(\kappa) = \partial\tilde{P}^{\rm tot}_{\rm 3D}/\partial\kappa$, and the second-order derivative as $R(\kappa) = \partial^2\tilde{P}^{\rm tot}_{\rm 3D}/\partial\kappa^2$, then

$$\frac{\Delta\kappa_{\rm Taylor}}{\kappa} = \frac{1}{2}\frac{R(\kappa=\kappa_\star)}{R(\kappa=0)}\frac{R(\kappa=0)}{G(\kappa=0)}\kappa. \tag{23}$$

In most cases, power spectra have an approximately power-law shape $P_{\rm 3D} = P_0 k^n$. For dark matter like power spectrum, $\Delta^2_{\rm 3D}(k) \propto k$, there is $n = -2$. $R(\kappa=0)/G(\kappa=0) = 3 - 4(k_\perp/k)^2 \in [-1, 3]$. $R(\kappa=\kappa_\star)/R(\kappa=0) = 1/(1-\kappa)^4$ when $k_\perp/k = 0$; $R(\kappa=\kappa_\star)/R(\kappa=0) \to 22/(1-\kappa)^2 - 21$ when $k_\perp/k \to 1$. κ to the epoch of reionization is about at the 5 per cent level, therefore

$$\left|\frac{\Delta\kappa_{\rm Taylor}}{\kappa}\right| \le 1.8\kappa. \tag{24}$$

The precision of a first-order κ reconstruction would not be better than the magnitude of κ itself, i.e. the quadratic reconstruction would not be better than a few per cent level. We will calculate to first-order accuracy in κ throughout this paper; however, we need to keep in mind that the first-order perturbation approximation will contribute several per cent error in our calculation.

The maximum likelihood condition requires

$$\frac{\delta\mathcal{L}}{\delta\kappa} \approx \frac{1}{2}L^3 \int \frac{{\rm d}^3k}{(2\pi)^3} \frac{\left(\tilde{P}^{\rm tot}_{\rm 3D} - |\tilde{T}|^2 L^{-3}\right)}{\tilde{P}^{\rm tot2}_{\rm 3D}} \frac{\delta\tilde{P}_{\rm 3D}}{\delta\kappa} = 0. \tag{25}$$

Since we calculate to first-order accuracy in κ, a further simplification is $1/\tilde{P}^{\rm tot2}_{\rm 3D} \approx [1 - 2\kappa(2P_{\rm 3D} + \Delta P_{\rm 3D})/P_{\rm 3D}]/P^2_{\rm 3D}$. The first-order solution is

$$E_\kappa = \int \frac{{\rm d}^3k}{(2\pi)^3} (|\tilde{T}|^2 L^{-3}) \mathcal{F}^\kappa(\boldsymbol{k}) - V_\kappa. \tag{26}$$

To simplify the problem, we assume the source is a cube with physical length L in each dimension. The offset constant $V_\kappa = \langle\sigma^2\rangle = \int {\rm d}^3k/(2\pi)^3 P^{\rm tot}_{\rm 3D}(\boldsymbol{k})\mathcal{F}^\kappa(\boldsymbol{k})$, and the optimal filter $\mathcal{F}^\kappa$ is

$$\mathcal{F}^\kappa(\boldsymbol{k}) = \frac{2P_{\rm 3D}(\boldsymbol{k}) + \Delta P_{\rm 3D}(\boldsymbol{k})}{P^{\rm tot2}_{\rm 3D}(\boldsymbol{k}) Q_\kappa}, \tag{27}$$

with $Q_\kappa = \int {\rm d}^3k/(2\pi)^3 (2P_{\rm 3D} + \Delta P_{\rm 3D})^2(\boldsymbol{k})/(P^{\rm tot}_{\rm 3D})^2(\boldsymbol{k})$.

From Parseval's theorem, we can rewrite equation (26) in the form of a convolution of the density field and a window function in real space:

$$\begin{aligned}\int \frac{{\rm d}^3k}{(2\pi)^3} \tilde{T}^*(\boldsymbol{k})\tilde{T}(\boldsymbol{k})\mathcal{F}^\kappa(\boldsymbol{k}) &= \int {\rm d}^3x \tilde{T}^\kappa_{\rm w_1}(\boldsymbol{x}) \tilde{T}^\kappa_{\rm w_2}(\boldsymbol{x}) \\ &= L^2 \int {\rm d}x_\parallel \tilde{T}^\kappa_{\rm w_1}(\boldsymbol{x}_\perp, x_\parallel) \tilde{T}^\kappa_{\rm w_2}(\boldsymbol{x}_\perp, x_\parallel). \end{aligned} \tag{28}$$

In equation (28) the two window functions are the decomposition of the optimal filter $W^\kappa_1(\boldsymbol{k}) W^\kappa_2(\boldsymbol{k}) = \mathcal{F}^\kappa(\boldsymbol{k})$. The last '=' in equation (28) holds when κ is constant. One can choose $W^\kappa_1(\boldsymbol{k}) = W^\kappa_2(\boldsymbol{k}) = \sqrt{\mathcal{F}_\kappa}$. If $\mathcal{F}_\kappa < 0$, we choose $W^\kappa_1 = -W^\kappa_2 = \sqrt{|\mathcal{F}_\kappa|}$. The convergence field is equivalent to the covariance of the measured maps smoothed by two windows. In the slowly spatially varying κ limit, all decomposition into two windows are equivalent. As we will show later, the shear construction can also be represented in the form of the covariance of two filtered temperature maps. These maps will have symmetric probability density functions (PDFs), which can reduce the non-Gaussianity of the maps so that a better S/N level can be achieved, when the shear window functions are chosen properly. The last two steps in equation (28) assume that the fluctuation of the convergence field is slow compared to the filter. Then we can apply the estimator to each beam in the map:

$$E_\kappa(\boldsymbol{x}_\perp) = L^{-1} \int {\rm d}x_\parallel \tilde{T}^\kappa_{\rm w_1}(\boldsymbol{x}) \tilde{T}^\kappa_{\rm w_2}(\boldsymbol{x}) - V_\kappa, \tag{29}$$

where $\tilde{T}^\kappa_{\rm w_1}$ and $\tilde{T}^\kappa_{\rm w_2}$ are the convolution of $\tilde{T}$ and window function $W^\kappa_1(\boldsymbol{x})$ and $W^\kappa_2(\boldsymbol{x})$, respectively, which are the real-space version of $W^\kappa_1(\boldsymbol{k})$ and $W^\kappa_2(\boldsymbol{k})$. The reconstruction of κ is dominated by the gradient of the power spectrum ${\rm d}\ln\Delta^2/{\rm d}\ln k$, which follows from the expression of our estimator in equation (26).

We can generalize the estimator to a spatially varying lensing field. In Appendix A we show

$$\int {\rm d}^2x'_\perp \kappa(\boldsymbol{x}'_\perp) b_\kappa(\boldsymbol{x}_\perp - \boldsymbol{x}'_\perp) = \langle E_\kappa(\boldsymbol{x}_\perp)\rangle. \tag{30}$$

Equivalently, for smaller scales, we will need to normalize the reconstructed lensing field by a scale-dependent factor in Fourier space, which is calculated in Appendix A.

$$\hat{\kappa}(\boldsymbol{l}) = b_\kappa^{-1}(\boldsymbol{l}) E_\kappa(\boldsymbol{l}) = \kappa(\boldsymbol{l}) + n(\boldsymbol{l}), \tag{31}$$

where $\boldsymbol{l} = \boldsymbol{k}_\perp \chi(z_s)$, and z_s is the redshift of the source. Here $b_\kappa(l)$ is the normalization factor ($\lim_{l\to 0} b_\kappa(l) = 1$) and $n(l)$ is the noise, since different Fourier modes are independent. They do not depend on direction because variables related to κ are isotropic on the transverse plane. In Appendix A, we show that the normalization factor is unity at small l when Q_κ has the form as $Q_\kappa = \int \mathrm{d}^3k/(2\pi)^3 (2P_{3D} + \Delta P_{3D})^2(\boldsymbol{k})/(P^{tot}_{3D})^2(\boldsymbol{k})$.

Amblard, Vale & White (2004) have pointed out that the reconstructed κ could be biased due to the non-Gaussianity of lenses. One possible reason is that there is only one CMB sky behind each patch of lensing field. For 21 cm, there are many source planes, and we expect there is no such bias effect by averaging over these planes.

2.2.2 *Estimator of shear*

When shear is taken into account, not only the scale but the directions of the coordinates are changed. We will start the derivation from the constant shear case. In analogy to κ, the optimal window function will be the same when γ is spatially variable:

$$\begin{aligned}\tilde{T}_b(\boldsymbol{k}_\perp, k_\parallel) &= \int \mathrm{d}^3x \tilde{T}_b(\boldsymbol{x}) \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \\ &= \int \mathrm{d}^2x_\perp \int \mathrm{d}x_\parallel T_b(\mathbf{J}\boldsymbol{x}_\perp, x_\parallel) \mathrm{e}^{-\mathrm{i}(\boldsymbol{k}_\perp\cdot\boldsymbol{x}_\perp + k_\parallel x_\parallel)} \\ &= |\mathbf{J}|^{-1} \int \mathrm{d}^2x'_\perp \int \mathrm{d}x_\parallel T_b(\boldsymbol{x}'_\perp, x_\parallel) \mathrm{e}^{-\mathrm{i}(\boldsymbol{k}'_\perp\cdot\boldsymbol{x}'_\perp + k_\parallel x_\parallel)} \\ &= |\mathbf{J}|^{-1} T_b(\mathbf{J}^{-1}\boldsymbol{k}_\perp, k_\parallel), \end{aligned} \tag{32}$$

here $\mathrm{d}^2x'_\perp = |\mathbf{J}|\mathrm{d}^2x_\perp$, $\boldsymbol{k}'_\perp = \mathbf{J}^{-1}\boldsymbol{k}_\perp$. Now the symmetry is broken even on the transverse plane due to the anisotropic distortion caused by the shear.

Since $\delta^{2D}(\mathbf{J}^{-1}\boldsymbol{k}) = |\mathbf{J}|\delta^{2D}(\boldsymbol{k})$, equation (19) implies

$$\begin{aligned}\tilde{P}_{3D}(\boldsymbol{k}_\perp, k_\parallel) = |\mathbf{J}|^{-1} P_{3D}(\mathbf{J}^{-1}\boldsymbol{k}_\perp, k_\parallel) \approx (1+2\kappa)[P_{3D}(k) + \Delta P_{3D}(\boldsymbol{k}) \\ \times (\kappa + \gamma_1 \cos 2\theta_{\boldsymbol{k}_\perp} + \gamma_2 \sin 2\theta_{\boldsymbol{k}_\perp})], \end{aligned} \tag{33}$$

where $\theta_{\boldsymbol{k}_\perp}$ is the angle between $\boldsymbol{k}_\perp$ and the transverse coordinate. Note that in the calculation of κ, we have ignored the contribution from shear. The reason is that the overall contribution of shear would be zero to first-order magnitude of κ due to the angular dependence of shear on $\theta_{\boldsymbol{k}_\perp}$.

Maximum likelihood requires $\delta\mathcal{L}/\delta\gamma_1 = 0$ and $\delta\mathcal{L}/\delta\gamma_2 = 0$. The maximum likelihood shear estimators can be written as a tensor $\boldsymbol{E}_\gamma$:

$$E_{\gamma ij} = L^{-1} \int \mathrm{d}x_\parallel \tilde{T}^\gamma_{w_i} \tilde{T}^\gamma_{w_j}, \tag{34}$$

where $\tilde{T}^\gamma_{w_i}$ is convolution of the temperature field with W^γ_i, and $W^\gamma_i(\boldsymbol{k}) = (2\Delta P_{3D}/P^2_{3D}Q_\gamma)^{1/2} k_i/k_\perp$, $k_i (i, j = 1, 2)$ is one of the two components of $\boldsymbol{k}_\perp$ on the transverse plane. When $\Delta P < 0$, we can choose $W^\gamma_1 = |2\Delta P_{3D}/P^2_{3D}Q_\gamma|^{1/2} k_1/k_\perp$, $W^\gamma_2 = -|2\Delta P_{3D}/P^2_{3D}Q_\gamma|^{1/2} k_2/k_\perp$. The normalization factor $Q_\gamma = \int \mathrm{d}^3k/(2\pi)^3 \Delta P_{3D}(\boldsymbol{k})(2k_1k_2/k^2_\perp)^2$. The two components of shear are now

$$\hat{\gamma}_1 = E_{\gamma 12} = E_{\gamma 21},\ \hat{\gamma}_2 = \frac{E_{\gamma 11} - E_{\gamma 22}}{2}. \tag{35}$$

Note that there is a difference between the reconstruction for convergence and shear. Shear reconstruction depends on the gradient of $P(k)$, while convergence reconstruction depends on the gradient of $\Delta^2(k)$ in a 2D analogue. To test our method, we can generate a Gaussian source field with power-law power spectrum $P(k) = k^\beta$. In the 2D analogue case, the convergence field cannot be measured if $\beta = -2$, because the variance is conserved. However in three dimensions, when $\beta = -3$, the convergence field can still be measured, which is due to the more complicated shape of the window function in 3D. When $\beta = 0$, the shear cannot be measured in either two or three dimensions.

In analogy to κ reconstruction, we can calculate the normalization factors b_{γ_1} and b_{γ_2}. The calculations for the normalization factors and noise are presented in Appendix A.

2.3 The combined estimator and the OQDE

The combined estimator of κ can be written as

$$\hat{\kappa}_{comb}(\boldsymbol{l}) = c_1\hat{\kappa}(\boldsymbol{l}) + c_2\hat{\gamma}_E(\boldsymbol{l}), \tag{36}$$

where $\hat{\gamma}_E$ is the convergence constructed from shear field,

$$\hat{\gamma}_E(\boldsymbol{l}) = \hat{\gamma}_1(\boldsymbol{l})\cos 2\boldsymbol{\theta}_l + \hat{\gamma}_2(\boldsymbol{l})\sin 2\boldsymbol{\theta}_l, \tag{37}$$

and $\boldsymbol{\theta}_l$ is the angle of $\boldsymbol{l}$. c_1 and c_2 are the weights of two components. The optimal c_1 and c_2 can be calculated from the covariance matrix of the two estimators

$$\mathbf{C}_\kappa = \begin{pmatrix} \langle\hat{\kappa}(\boldsymbol{l})^2\rangle - \langle\hat{\kappa}(\boldsymbol{l})\rangle^2 & \langle\hat{\kappa}(\boldsymbol{l})\hat{\gamma}_E(\boldsymbol{l})\rangle - \langle\hat{\kappa}(\boldsymbol{l})\rangle\langle\hat{\gamma}_E(\boldsymbol{l})\rangle \\ \langle\hat{\kappa}(\boldsymbol{l})\hat{\gamma}_E(\boldsymbol{l})\rangle - \langle\hat{\kappa}(\boldsymbol{l})\rangle\langle\hat{\gamma}_E(\boldsymbol{l})\rangle & \langle\hat{\gamma}_E(\boldsymbol{l})^2\rangle - \langle\hat{\gamma}_E(\boldsymbol{l})\rangle^2 \end{pmatrix}.$$

To minimize the variance of $\hat{\kappa}_{comb}(\boldsymbol{l})$, c_1 and c_2 are sum of the first and second row components, respectively, in $\mathbf{C}^{-1}_\kappa$, the inverse matrix of $\mathbf{C}_\kappa$. When $l \lesssim l_a = k_a\chi(z_s)$, $\hat{\kappa}(\boldsymbol{l})$ and $\hat{\gamma}_E(\boldsymbol{l})$ are uncorrelated and the covariance matrix is diagonal. The weights are simply proportional to the reciprocal of the noise of two individual estimators $c_1 \propto N_\kappa(\boldsymbol{l})^{-1}$ and $c_2 \propto N_{\gamma_E}(\boldsymbol{l})^{-1}$.

The lensing power spectrum P_κ is measured by taking the observed power spectrum $P_{\hat{\kappa}}$ and subtracting the computable power spectrum of the noise N_κ

$$P_\kappa(\boldsymbol{l}) = P_{\hat{\kappa}}(\boldsymbol{l}) - N_\kappa(\boldsymbol{l}). \tag{38}$$

If the source is Gaussian and its power spectrum is known, the noise power spectrum of estimator $N_\kappa(\boldsymbol{l})$ can be calculated using the method described in Appendix A. If source is non-Gaussian, $N_\kappa(\boldsymbol{l})$ is not derivable from the two-point function, the lensing power spectrum could be biased. For scales $l \lesssim l_a = k_a\chi(z_s)$, $N_{\kappa\gamma_E}(\boldsymbol{l}) \approx 0$. The $\hat{\kappa} - \hat{\gamma}_E$ cross-power spectrum does not contain the noise term and does not depend on the source four-point function. It is less optimal, but more robust. Therefore the cross-power spectrum of convergence and shear will be an unbiased estimator of lensing power spectrum even when we do not know the unlensed source four-point function, and the noise of the two estimators themselves.

The 2D OQDE in CMB lensing can be written as product of two filtered temperature fields (Hu 2001; Lewis & Challinor 2006). Furthermore, the 3D OQDE can be written in the same form as equation (12), though it is not explicit (private communication with Oliver Zahn):

$$\boldsymbol{E}_d(\boldsymbol{\theta}) = L^{-1} \int \mathrm{d}x_\parallel \boldsymbol{T}_1(\boldsymbol{\theta}, x_\parallel) T_2(\boldsymbol{\theta}, x_\parallel) \tag{39}$$

and

$$\int \mathrm{d}^2\theta' \boldsymbol{d}(\boldsymbol{\theta}') b_d(\boldsymbol{\theta} - \boldsymbol{\theta}') = \langle \boldsymbol{E}_d(\boldsymbol{\theta})\rangle. \tag{40}$$

Here b_d is a normalization factor, $\boldsymbol{T}_1 = \int \mathrm{d}^2\theta' T(\boldsymbol{\theta}')\boldsymbol{W}_1(\boldsymbol{\theta} - \boldsymbol{\theta}')$ and $T_2 = \int \mathrm{d}^2\theta' T(\boldsymbol{\theta}') W_2(\boldsymbol{\theta} - \boldsymbol{\theta}')$ are convolved temperature fields, where the window functions are Fourier transforms of

$$\begin{aligned}\boldsymbol{W}_1(\boldsymbol{l}, k_\parallel) &= \frac{-\mathrm{i}\boldsymbol{l} P_{3D}(\boldsymbol{l}, k_\parallel)}{\tilde{P}^{tot}_{3D}(\boldsymbol{l}, k_\parallel)}, \\ W_2(\boldsymbol{l}, k_\parallel) &= \frac{1}{\tilde{P}^{tot}_{3D}(\boldsymbol{l}, k_\parallel)}. \end{aligned} \tag{41}$$

We note that the OQDE and our estimators have the same form.

The contribution from lensing in equation (10) is secondary, and the noise of reconstruction is mainly determined by the unlensed terms. Therefore we can measure the numerical reconstruction noise without lensing the sources.

3 NUMERICAL METHODS

3.1 Simulation

The fluctuation in the 21-cm brightness temperature may depend on many factors, such as the gas density, temperature, neutral fraction, radial velocity gradient and Ly α flux (Barkana & Loeb 2005). In our paper, we do not consider the redshift-space distortion effect caused by the non-zero radial peculiar velocity gradient, and simply assume that the brightness temperature is proportional to the density of the neutral gas:

$$T_{\rm b} \approx (27\,{\rm mK})\left(\frac{1+z}{10}\right)^{1/2}\frac{T_{\rm s}-T_{\rm CMB}}{T_{\rm s}}(1+\delta_{\rm HI}), \tag{42}$$

where $T_{\rm b}$ is the brightness temperature increment respective to CMB, $T_{\rm s}$ is the spin temperature, which is expected to be much bigger than $T_{\rm CMB}$ once structure are non-linear, and $\delta_{\rm HI}$ is the overdensity of the neutral hydrogen.

Our paper mainly focuses on the non-Gaussian aspect and 3D properties of the reconstruction, and these effects also exist in a pure dark matter distribution. The neutral gas will trace the total mass distribution, which is dominated by the dark matter haloes. A simplification is to use the dark matter as the source directly. Even though this will bring some bias at small scales, the approximation is valid at large scales (Trac & Pen 2004). The dark matter distributions are generated using the PMFAST code (Merz, Pen & Trac 2005).

The high-resolution PMFAST simulation was performed on a 1456^3 fine mesh with 3.9×10^8 particles. The production platform was the IA-64 'lobster' cluster at CITA, which consists of eight nodes. One of them was upgraded, so we used the remaining seven nodes. Each node contains four 733 MHz Itanium-1 processors and 64 GB RAM. The simulation started at an initial redshift $z_i = 100$ and ran for 63 steps with comoving box size $L = 50\,h^{-1}$ Mpc. The initial condition was generated using the Zeldovich approximation, and the matter transfer function was calculated using CMBFAST (Seljak & Zaldarriaga 1996). The cosmological parameters were chosen in accordance with the *Wilkinson Microwave Anisotropy Probe* (*WMAP*) result (Spergel 2003): $\Omega_{\rm m} = 0.27, \Omega_\Lambda = 0.73$, $\Omega_{\rm b} = 0.044, n = 1.0, \sigma_8 = 0.84$ and $h_0 = 0.71$. 20 independent boxes were generated. We had 3D data at $z = 7$ at hand, and used them in our numerical tests for convenience.

3.2 Convergence and shear map construction

The dimensionless power spectrum, which is the contribution to the variance of overdensity per logarithmic interval in spatial wavenumber, can be measured from the source data in the periodic simulation box.

To reduce the computation time, our numeric results on the reconstruction used a resampled distribution. We generate 20 independent sources, each on 512^3 grids, to investigate the statistics. The total comoving length along the line of sight of 20 simulation boxes is $1\,h^{-1}$ Gpc, which is about the same size as the observable 21-cm region distributed between redshifts 10 and 20. The correlation between the boxes can be ignored since the box size is much larger than the non-linear length-scale, and the number of neglected modes is small. The evolution of structure over this redshift is significant. Our simulations were all measured at the same redshift, so we anticipate the real effects of non-Gaussianity to be smaller. On the other hand, large-scale power is generated by reionization bubbles, which may contribute to non-Gaussianity as well. In Fig. 1, the solid line is the average power spectrum of the resampled sources. To measure the dependence of non-Gaussianity on scale, we will compare the results with different scales of experimental noise cut-off.

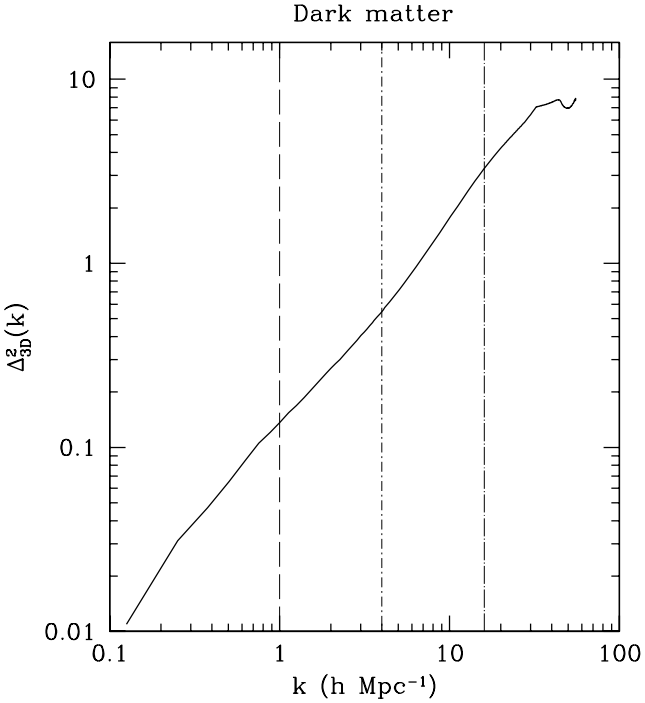

Figure 1. The average dimensionless power spectrum of the resampled dark matter from the 1456^3 N-body simulation in three dimensions are given. The solid line is the power spectrum on the 512^3 grids. The resampled sources keep the non-linearity and the non-Gaussianity of the structures up to $k \sim 30\,h\,{\rm Mpc}^{-1}$. Three different experimental noise cut-offs are shown with $k_{\rm c} = 1, 4, 16\,h\,{\rm Mpc}^{-1}$, which represent the linear, quasi-linear and non-linear scales.

We simply assume the noise to be zero above a cut-off and infinity below the cut-off scale. This is a reasonable approximation for a filled aperture experiment, which has good brightness sensitivity, and an exponentially growing noise at small scales. Three cut-offs were chosen at $k_{\rm c} = 1, 4, 16\,h\,{\rm Mpc}^{-1}$, which represent the linear, quasi-linear and non-linear scales. Three different experimental noise levels are shown as vertical lines in Fig. 1.

In principle, the convergence map is the variance (or covariance when the filter $\mathcal{F}^\kappa$ has negative value) of the overdensity field after a specified filtering process. Shear is the covariance of the two maps, since the anisotropic filter cannot be factored into a perfect square. We need to smooth the maps to extract the lensing signal with maximum S/N. The window function used to smooth the lensed map, which is isotropic in the transverse directions to the line of sight, can be calculated using equation (27). The gradient of the power spectrum becomes negative at small scales; which comes from the limited resolution of the N-body simulation and is unphysical. The experimental noise will put a natural cut-off at small scales.

As mentioned in Section 2.2.2, the reconstruction of κ will depend on $2P + \Delta P$. In two dimensions, this is equivalent to the gradient

of the 2D version of $\Delta^2_{2D} = k^2 P_{2D}(k)/2\pi$. In three dimensions, it is more complicated since $\Delta P(\boldsymbol{k})$ is not isotropic. The optimal window functions have two parts W_1 and W_2, the choice of which is not unique. One might expect a symmetric decomposition to have the best S/N. The optimal filter of κ is positive except at a few modes, and can be decomposed into two equivalent parts (one part needs to contain a minus sign for those negative value of the filter). In contrast to κ, the shear construction needs to use the covariance between two different windowed temperature fields, since there is a sin or cos component in the window function. The window is a function of the transverse and parallel components of $\boldsymbol{k}$.

We can calculate the mean covariance of the two smoothed maps along the redshift axis for each pixel. From equation (29) we can construct the convergence map. Shear maps are reconstructed in the same way, except different optimal window functions are used. The anisotropic part $\cos 2\theta_{\boldsymbol{k}_\perp}$ can be decomposed into $\cos\theta_{\boldsymbol{k}_\perp} - \sin\theta_{\boldsymbol{k}_\perp}$ and $\cos\theta_{\boldsymbol{k}_\perp} + \sin\theta_{\boldsymbol{k}_\perp}$. Both windows can generate a field with even PDF so that the distribution is less non-Gaussian. This is consistent with the numerical results as shown in Fig. 6. Using these two maps, we construct the γ_1 map with their covariance, as shown in equation (35). Similarly we can get the γ_2 map.

4 NUMERICAL RESULTS AND DISCUSSION

Cooray (2004) claims that the variance will not vary considerably and is not an ideal measurement of the lensing signal. Even though the κ field itself is only a few per cent, the integrated effect from the 3D images will reduce the noise ratio significantly to uncover the signal. Zahn & Zaldarriaga (2006) solve the problem from an alternative approach by generalizing the minimum variance quadratic estimator (Hu & Okamoto 2002) in CMB lensing to 3D.

Related work was done in Metcalf & White (2007), where they also construct quadratic estimators of shear and convergence in real space, even though they did not include the correlation between the 2D slices along the line of sight and they did not choose the estimator with minimized noise.

4.1 Non-Gaussianity

The dark matter distribution is linear at large scales, and can be treated as Gaussian. In the non-linear scales, when the amplitude of density fluctuations is big, the structure becomes highly non-Gaussian. Reference Gaussian sources with identical power spectrum to the dark matter are generated.

We treat the $1\,h^{-1}$ Gpc region at $z = 10$–20 as 20 independent sources. Structures at these redshifts are similar to those at $z = 7$ used by us, though less non-linear. We can expect to see similar non-Gaussianity effects in the reconstruction with the $1\,h^{-1}$ Gpc space except that the non-linear scale is smaller. We compare the reconstruction noise with three different experimental noise as well as the lensing signal in Figs 2–4. The thick solid line is the lensing power spectrum, which is calculated with the Limber integral of the 3D power spectra of dark matter using equation (4). We use the publicly available code HALOFIT.F (Smith et al. 2003) to generate the non-linear dark matter power spectra. The code provides both their fitting results, and the results using the Peacock–Dodds formula (Peacock & Dodds 1996, PD96 hereafter). The HALOFIT code fits the power spectrum at low redshift to Virgo and GIF CDM simulations, which used the transfer function of Efstathiou, Bond & White (1992). At higher redshifts, the code does not operate. We use a combination of the two: HALOFIT power spectra are used for redshifts lower than $z = 3.0$, and PD96 power spectra are used for higher redshifts.

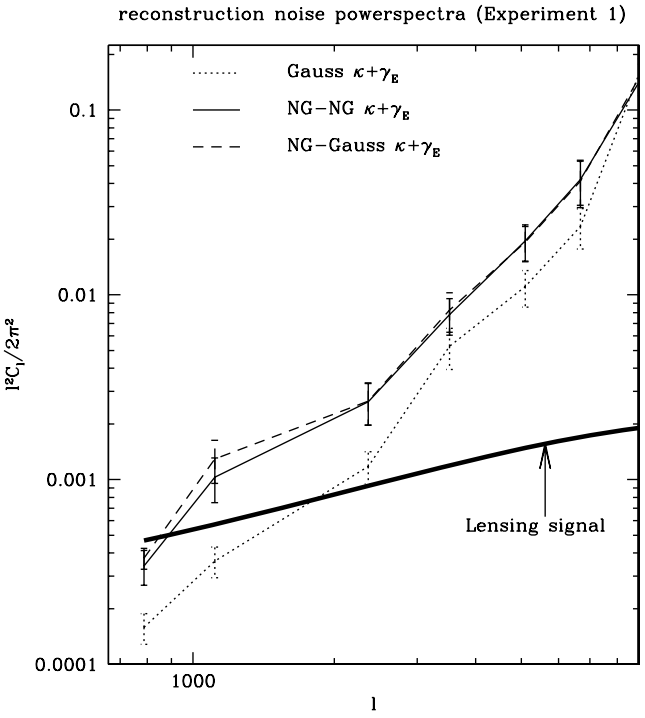

Figure 2. The noise of lensing maps from different estimators using experimental noise 1, which cuts off at $k_c = 1\,h\,\mathrm{Mpc}^{-1}$. We treat the $1\,h^{-1}$ Gpc rectangle of gas at $z = 10$–20 as 20 independent sources each is a $50\,h^{-1}$ Mpc box size cube. Structures at these redshifts are similar to those at $z = 7$ used by us, though less non-linear. We can expect to see qualitatively similar non-Gaussianity effects in the reconstruction with the $1\,h^{-1}$ Gpc space except that the non-Gaussianity of sources may be smaller, but reionization may change that, too. The curves are truncated at $\sqrt{2}k_c$, where the noise goes to infinity. The thick solid line is the expected lensing signal. The dotted line is the lensing reconstruction noise for a simulated Gaussian source with the same power spectrum. The dashed curve is the noise from the *N*-body simulation using the Gaussian estimator, which increases modestly compared to the Gaussian source. It is identical for the optimal κ, γ reconstruction as it is for the deflection angle. The thin solid line is noise when shear and convergence are reweighted by their non-Gaussian variances.

Since the reconstruction noise of κ is isotropic, one can always choose the direction of the lensing mode $\boldsymbol{l}$ to be parallel with a coordinate axis. In this direction, $\gamma_1(\boldsymbol{l}) = \kappa(\boldsymbol{l}), \gamma_2(\boldsymbol{l}) = 0$ and $\gamma_E = \gamma_1$, which simplifies the numerical calculation. As shown in Section 2.3, the combined estimator becomes the sum of κ and γ_E with weights. The optimal weights are calculated from the inverse matrix of $\mathbf{C}_\kappa$, covariance between the two estimators. The matrix is close to diagonal in small l approximation, so for simplicity we take the diagonal terms. For Gaussian sources, the weights are the reciprocal of the noise of the two estimators, respectively. For non-Gaussian sources, we could use the same weights as the Gaussian sources, i.e. using the reciprocal of the noise measured from Gaussian sources, or weight them by calculating their respective measured noise from non-Gaussian sources. We will show that the combined estimator with Gaussian noise derived weights has the same noise as the OQDE for both Gaussian and non-Gaussian sources. Figs 2–4 are results using noise cut-offs from experiment 1, 2 and 3. The curves are truncated at $\sqrt{2}k_c$. The non-Gaussianity increased the noise of all estimators. The first cut-off falls in the linear regime, where the

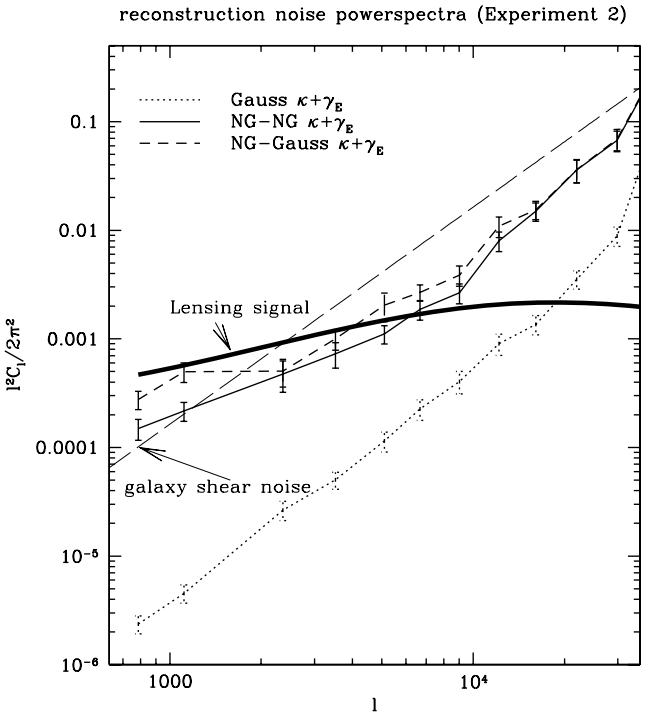

Figure 3. Same of Fig. 2, but with cut-off at the quasi-linear scales $k_c = 4\,h\,\mathrm{Mpc}^{-1}$. The effect of non-Gaussianity of sources is more pronounced. We can compare the S/N with a fiducial cosmic shear survey of sources in the same $10 < z < 20$ redshift range, which reconstructs the lensing from the shape of galaxies, with a surface number density of $14\,\mathrm{arcmin}^{-2}$. To map the lensing to the same S/N with redshift $z \sim 1$ sources requires a density of $56\,\mathrm{arcmin}^{-2}$ (Hu & White 2001) with rms ellipticity of 0.4. We see that proposed optical lensing surveys are unlikely to outperform 21-cm sources.

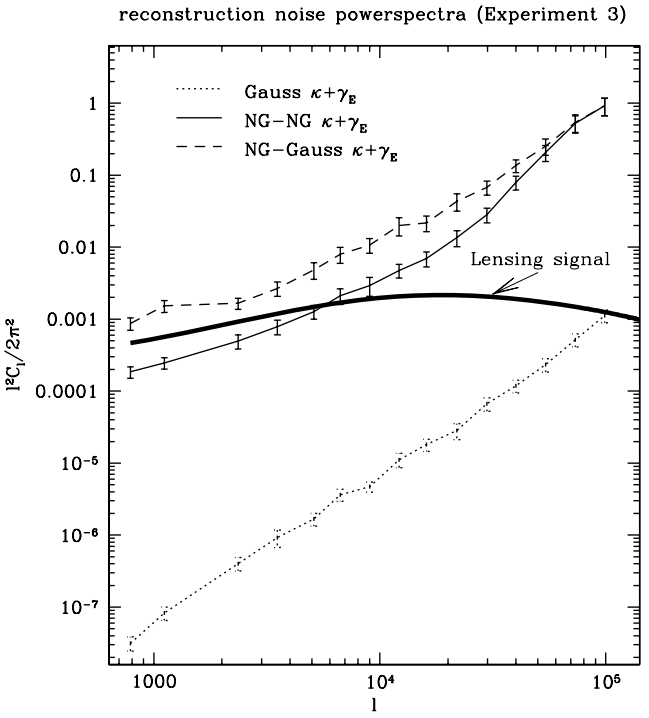

Figure 4. Same of Fig. 2, but with cut-off at the non-linear scales $k_c = 16\,h\,\mathrm{Mpc}^{-1}$. At the highly non-linear scales, the non-Gaussian noise variance is about 3–4 mag higher than the Gaussian noise. The combined reweighted estimator (NG–NG $\kappa + \gamma_E$) has noise about half an order of magnitude lower than the OQDE.

non-Gaussianity only has a modest effect on the noise. The second cut-off is at the quasi-linear scale. Here the non-Gaussianity increases the noise of the OQDE by about one to two orders of magnitude. At the highly non-linear scales, the non-Gaussian noise is about 3–4 mag higher than the Gaussian noise, and in fact higher than that of the more noisy experiment.

Our estimators were derived in the limit that κ and γ are constant. For spatially variable lens, we solve for the required normalization factors. In the OQDE, the windows do not depend on the scale of the lens, so one might guess the same ansatz to hold for the (κ, γ) estimators. We verify this numerically in Fig. 5. The solid and dotted lines are for the Gaussian and non-Gaussian sources, respectively. The differences are less than a few per cent, and consistent with integration errors from the tabulated power spectrum, and most importantly, independent of scale, as we had expected. The vertical line shows the characteristic scale $l_a = \chi(z_s)k_a \sim \chi(z_s)k_c/2$, or below which the window functions are optimized. As shown in Fig. 7, the reconstruction noise is proportional to k^{-3} if sources are Gaussian, and most of the contribution to the reconstruction is from scales near k_c. Therefore our approximation holds at scales $l < \chi(z_s)k_c/2$ with small deviance. For non-Gaussian sources, l_a will also be affected by the non-Gaussianity of sources. While the optimality is only proved at low l, we find the combined estimator and OQDE equally optimal for Gaussian sources at all scales. For non-Gaussian sources, they also have the same results. We do note that for a finite size survey, the (κ, γ) recover the constant mode, which is lost in the OQDE. Three more numbers are recovered.

The combined estimator with κ and γ_E weighted by using the non-Gaussian noise is more optimal than weighted by using Gaussian noise, therefore has lower noise than the OQDE. In fact, the non-Gaussian noise of γ_E is much smaller than κ. To investigate the origin of this change, we first investigate the cause of the increased noise in non-Gaussian sources for κ. This could be because either the non-Gaussianity leads to a high kurtosis in κ, which boosts the errors; or the non-Gaussianity may lead to correlations between modes, resulting in a smaller number of independent modes, and thus a larger error.

In Fig. 6, the PDF of maps smoothed with the κ window are shown. The top, middle and bottom panel show the results with experimental noise cut-offs 1, 2 and 3. The solid line is the PDF for maps smoothed with κ window (T_1^κ, T_2^κ in Section 2.2.1). Because the window functions are almost symmetric, we plot only one PDF. To see the full dynamic range on the x-axis, we plot $\pm|T|^{1/4}$ as x-axis and PDF $(|T|^{1/4})|T|^{15/4}$ as the y-axis. The integral of the x-axis weighted by the y-axis will give $\langle T^4\rangle$, which is basically a estimation of the point-wise non-Gaussian reconstruction noise. Here PDF $(|T|^{1/4})$ is the PDF of $|T|^{1/4}$. To compare with a Gaussian distribution, dotted lines are also plotted. The contributions to the $\langle T^4\rangle$ in experiment 1 mainly come from small-fluctuation regions. In experiment 2, the large outliers play a more important role but one can still expect the curve to converge. In experiment 3, most contributions come from rare regions with high fluctuations. Caution should be exercised in the interpretation of the most non-linear scales, since a larger number of source samples may result

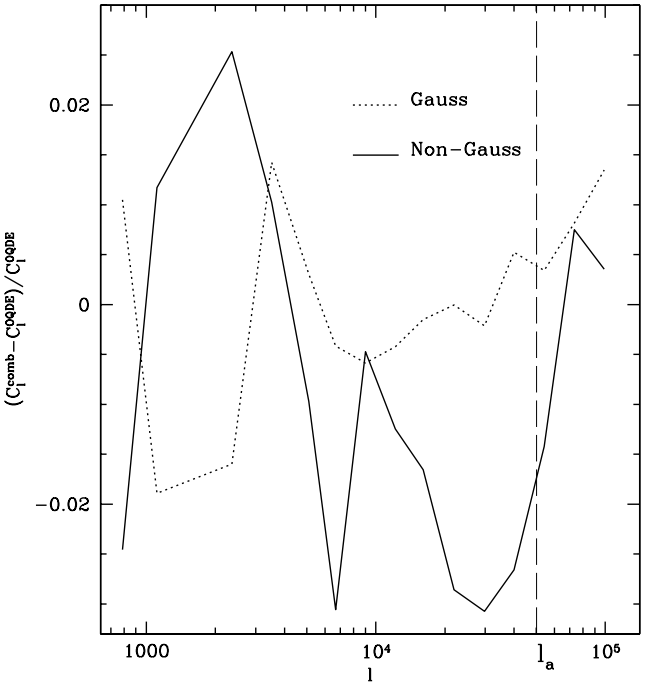

Figure 5. The comparison of reconstruction noise from the combined (κ, γ) estimator and the OQDE. As shown in Section 2.3, the combined estimator becomes the sum of κ and γ_E with weights. For Gaussian sources, the weights are the reciprocal of the noise of the two estimators, respectively. To compare with the OQDE, we could still use the same weights for non-Gaussian sources, i.e. using the reciprocal of the noise measured from Gaussian sources. Note that the combined estimator with these weights will be less optimal than using weights calculated from the noise of non-Gaussian sources. The vertical line shows the characteristic scale l_a below which the window functions are optimized. While the optimality is only proved at low l, we find the combined estimator and OQDE equally optimal for Gaussian sources at all scales. For non-Gaussian sources, they also have the same results. The scatter is consistent with numerical integration errors from the tabulated power spectrum. Similarly, the noise for non-Gaussian sources is consistent between the two estimators. We conclude that the combined estimator is numerically equivalent to OQDE if the weights are optimized for Gaussian sources.

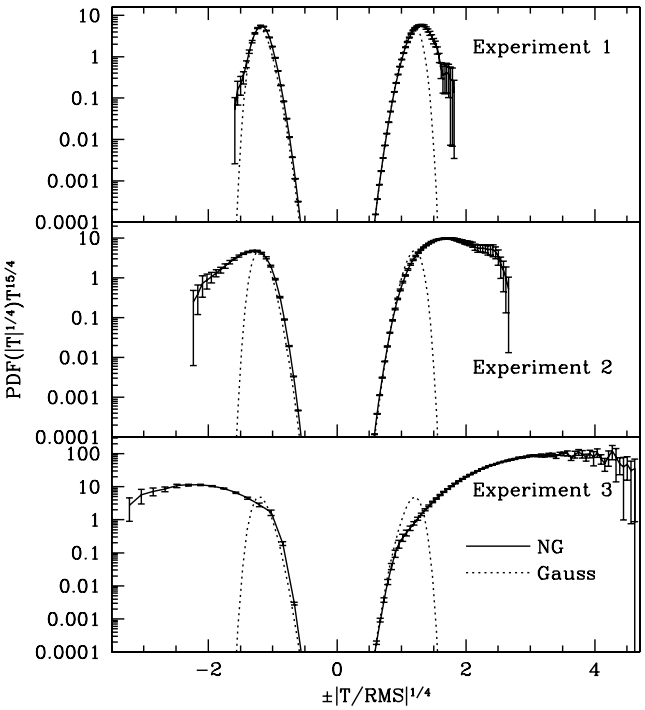

Figure 6. The PDF of maps smoothed with a κ window are shown. The top, middle and bottom panel show the results with experimental noise cut-offs 1, 2 and 3. The solid line is the PDF for maps smoothed with the κ window (T^κ in Section 2.2.1). To see the full dynamic range on the x-axis, we plot the curve with $\pm|T|^{1/4}$ as x-axis and PDF $(|T|^{1/4})|T|^{15/4}$ as the y-axis. The integral of the x-axis weighted by the y-axis will give $\langle T^4\rangle$, which is basically an estimation of the reconstruction noise. The error bars are estimated from the 20 simulations. To compare with a Gaussian distribution, dotted lines are also plotted. The contributions to the $\langle T^4\rangle$ in experiment 1 mainly come from small-fluctuation regions. In experiment 2, the large outliers play a more important role but one can still expect the curve to converge. In experiment 3, most contributions come from rare regions with high fluctuations. Caution should be exercised in the interpretation of the most non-linear scales, since a larger number of source samples may result in a different error. It is clear, however, that the noise has increased dramatically.

in a different error. It is clear, however, that the noise has increased dramatically.

The kurtosis of κ is $\langle(T_1^\kappa)^4\rangle/\langle(T_1^\kappa)^2\rangle^2-3$, and an analogous quantity can be defined by $\langle(T_1^\gamma\ T_2^\gamma)^2\rangle/(\langle(T_1^\gamma)^2\rangle\langle(T_2^\gamma)^2\rangle)-1$ for shear. $T_1^\kappa\approx T_2^\kappa$, and T_1^γ is uncorrelated with T_2^γ. The noise of κ and γ is determined by both kurtosis and number of independent cells. For experimental noise 1, the kurtosis of T^κ and T^γ are 1.2 and 0.29, respectively. The effectively independent cube cells for κ and γ have side length 4.8 and $4.6\,h^{-1}$ Mpc, respectively. The corresponding Gaussian sources with the same cut-off have effective cell size 3.0 and $3.5\,h^{-1}$ Mpc. For experimental noise 2, the kurtosis of T^κ and T^γ are 18 and 5.7, respectively. The effective cell size for κ and γ are 1.8 and $1.5\,h^{-1}$ Mpc, respectively. The corresponding Gaussian sources with the same cut-off have effective cell size 1.0 and $1.1\,h^{-1}$ Mpc. For experimental noise 3, the kurtosis for T^κ and T^γ are 1.6×10^3 and 3.5×10^2, respectively. The effective cell size for κ and γ are 540 and $310\,h^{-1}$ kpc, respectively. The corresponding Gaussian sources with the same cut-off have effective cell size 240 and $290\,h^{-1}$ kpc. We conclude that the shear measurements have lower non-Gaussian noise both because of a smaller point-wise kurtosis and less correlation between modes.

We will see later that experiment 2 has the largest S/N, which is larger than unity for $l\lesssim6000$. We can compare the S/N with cosmic shear surveys, which reconstruct lensing from the shape of galaxies. The noise can be estimated by $\langle\gamma^2\rangle/n_{\rm eff}$ (Hu & White 2001; Hoekstra et al. 2006), where we use $\langle\gamma^2\rangle^{1/2}\approx0.4$ as the rms intrinsic ellipticity, and $n_{\rm eff}$ is the effective number density of galaxies. We plot the shear noise from a survey of sources in the same redshift range $10<z<20$ in Fig. 3, with a surface number density of 14 arcmin^{-2}. For more realistic source redshifts $z\sim1$ in proposed optical surveys (Hu & White 2001), this corresponds to a surface density of 56 arcmin^{-2} to achieve the same fidelity of dark matter reconstruction. In the CFHTLS wide survey the source galaxies are distributed at redshifts lower than 3, and their effective number density is $\sim$12 galaxies arcmin^{-2} (Hoekstra et al. 2006). This noise is larger still. Even though non-Gaussian 21-cm lensing saturates lensing reconstruction, it still measures more modes than current proposed optical surveys.

In Fig. 7, we show the reconstruction noise at two different l versus various experimental noise cut-off k_c. The top panel is for the fundamental mode in the box, $l_1=2\pi/L=783$ and the bottom panel is for $l_2=6l_1=4715$. As shown in the plot, it is clear that

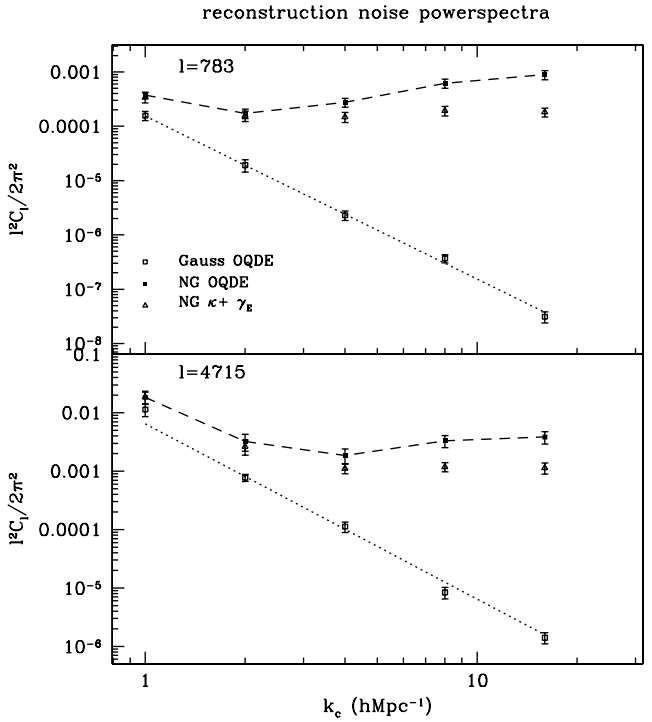

Figure 7. The reconstruction noise versus the cut-off in the experimental noise. The top panel is for $l_1 = 2\pi/L = 783$ and the bottom panel is for $l_2 = 6\,l_1 = 4715$. The noise of Gaussian sources decreases as k_c increases, because of the increasing number of independent modes. The dotted lines are a least-squares fitting power law $N_0\,k_c^{-3}$ to the Gaussian noises, and $N_0 = 3.1 \times 10^{-2}, 1.3 \times 10^{-1}$ for top and bottom panels, respectively. The dashed lines connect the non-Gaussian noise of the OQDE. The triangles are the reconstruction noise for the combined estimator, which is equal to the OQDE at larger scale k_c and about half an order of magnitude lower at large k_c. The noise of the non-Gaussian sources changes slowly and saturates or even increases at small scales.

the noise of a Gaussian source decreases as k_c increases, because of the increasing number of independent modes. The dotted lines are a least-squares fitting power law $N_0\,k_c^{-3}$ to the Gaussian noises, and $N_0 = 3.1 \times 10^{-2}, 1.3 \times 10^{-1}$ for top and bottom panels, respectively. This comes from counting the number of available source modes. The dashed lines connect the non-Gaussian noises of the OQDE. The triangles are the reconstruction noise for the combination estimator, which is equal to the OQDE at larger scale k_c and about half an order of magnitude lower at smaller scales of k_c. From this plot, we can see that experiments with lower noise do not necessarily decrease the reconstruction noise of the OQDE for non-Gaussian sources. And the experimental noise has a limit around the quasi-linear scale where the OQDE achieves its best S/N. The S/N achieves its maximum around $k_c^{\rm NG} \approx 4\,h\,{\rm Mpc}^{-1}$. This cut-off with maximum S/N varies only slowly with l.

If one wants to estimate the effective number of available lensing modes, we can derive an effective cut-off of a Gaussian field which gives the same S/N as the optimal non-Gaussian source estimator. This is $k_c^{\rm G} \approx 2\,h\,{\rm Mpc}^{-1}$, where the power spectrum of source is $\Delta^2 \approx 0.2$. The size of the effectively independent cells is $2.0\,h^{-1}$ Mpc. A simple equivalent Gaussian noise estimate counts all modes up to $\Delta^2(k) < 0.2$, which is perhaps surprisingly low.

For our noise estimates, we stacked simulations all at redshift $z = 7$. While the angular diameter distance does not change much to $z \sim 20$, the structure does evolve. We do not have access to the higher redshift outputs to test this effect, but one would expect a smaller non-linear scale to result in a smaller reconstruction noise.

4.2 Future directions

A possible way to find the optimal window functions for non-Gaussian sources is to divide the window into N frequency bins $W_1(\boldsymbol{k}_1, \boldsymbol{k}_2, \ldots, \boldsymbol{k}_N)$, and apply a numerical variation to those bins. The noise can be measured numerically by applying the estimator to the simulated sources. The process of searching for a optimal filter is equivalent to look for a minimum of reconstruction noise in N-dimensional space $\boldsymbol{k}_1, \boldsymbol{k}_2, \ldots, \boldsymbol{k}_N$. In fact, the optimal window function can be constructed by the inverse of covariance matrix of the source power spectra. We will present details of the method in a future paper. In this paper we only considered the class of windows which are identical to the optimal Gaussian estimators with a hard cut-off, as well as two weightings for shear and convergence.

One can also try to Gaussianize the sources by modifying the PDF of all the sources to be Gaussian. The physical explanation and details of Gaussianization can be found in Weinberg (1992). The basic idea is that every pixel should preserve its rank in the whole field during the Gaussianization process. During structure formation, the non-linear evolution at small scales should not destroy most of the information on the peaks and dips of the linear field. However, this Gaussianization process will change the power spectra of sources, and the reconstructed lensing field will be biased. This is not a linear process, and the variation of power spectrum does not have analytical solution, and can only be measured numerically with simulated sources.

Recently it has been proposed that one could economically achieve brightness mapping of 21-cm emission at lower redshifts (Chang et al. 2008), potentially even with existing telescopes. If individual galaxies are not resolved, one can again ask the question of how one could reconstruct a lensing signal. This is very similar to the problem studied in this paper. We will discuss this with optimal non-Gaussian estimators in a future paper.

5 CONCLUSION

In this paper, we developed the maximum likelihood estimator for the large-scale structure from the 21-cm emission of the neutral gas before the epoch of reionization. The convergence and shears can be constructed separately. They are independent for $l \lesssim l_a$. The cross-power spectrum of convergence and shear is more robust unbiased estimator of lensing power spectrum, which does not require knowledge of the unlensed source four-point function. To test the effects of non-Gaussianity, we applied our estimators to simulated data. The sources were generated by N-body simulations, because gas is expected to trace the total mass distribution. To investigate the influence of non-Gaussianity, we also use Gaussian sources which have the same power spectra as the simulated sources. We applied our estimator and the OQDE on both the Gaussian and non-Gaussian sources. Though our estimators are derived in the simplified case of a constant convergence, the noise of our combined estimator of convergence and shear are the same as the OQDE for Gaussian sources for spatially variable lenses. For a finite survey area, three extra constant modes can be recovered.

The non-Gaussian nature of the source can increase the error bar by orders of magnitude, depending on the experimental cut-off scale. Shear construction is affected less by non-Gaussianity than the convergence field, and the combined estimator with non-Gaussian noise weights is a better choice than reconstructing with

the OQDE. S/N cannot be boosted infinitely by reducing the experimental noise, and achieves its maximum for a cut-off around $k_c^{NG} \approx 4\,h\,\mathrm{Mpc}^{-1}$. Below that scale the S/N starts to saturate or even decrease. The maximum S/N for non-Gaussian sources is equal to Gaussian sources with $k_c^{G} \approx 2\,h\,\mathrm{Mpc}^{-1}$, where the power spectrum of sources is $\Delta^2 \approx 0.2$ and the side length of the effectively independent cells is $2.0\,h^{-1}$ Mpc. The maximum S/N is greater than unity for $l \lesssim 6000$, which makes 21-cm lensing very competitive compared to optical approaches.

ACKNOWLEDGMENTS

We thank Oliver Zahn, Chris Hirata, Brice Ménard and Mike Kesden for helpful discussions. TL thanks Pengjie Zhang, Zhiqi Huang, Hy Trac and Hugh Merz for help in the early stage of the work.

REFERENCES

Amblard A., Vale C., White M., 2004, New Astron., 9, 687
Barkana R., Loeb A., 2005, ApJ, 624, L65
Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303
Cooray A., 2004, New Astron., 9, 173
Efstathiou G., Bond J. R., White S. D. M., 1992, MNRAS, 258, 1P
Hilbert S., Metcalf R., White S. D. M., 2007, MNRAS, 382, 1494
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526
Hoekstra H. et al., 2006, ApJ, 647, 116
Hu W., 2001, ApJ, 557, L79
Hu W., Okamoto T., 2002, ApJ, 574, 566
Hu W., White M., 2001, ApJ, 554, 67
Kaiser N., 1992, ApJ, 388, 272
Lewis A., Challinor A., 2006, Phys. Rep., 429, 1
Limber D. N., 1954, ApJ, 119, 655
Mandel K. S., Zaldarriaga M., 2006, ApJ, 647, 719
Merz H., Pen U., Trac H., 2005, New Astron., 10, 393
Metcalf R. B., White S. D. M., 2007, MNRAS, 381, 447
Peacock J. A., Dodds S. J., 1996, MNRAS, 280, L19 (PD96)
Pen U., 2004, New Astron., 9, 417
Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437
Smith K. M., Zahn O., Doré O., 2007, Phys. Rev. D, 76, 043510
Smith R. E. et al., 2003, MNRAS, 341, 1311
Spergel D. N. et al., 2003, ApJS, 148, 175
Trac H., Pen U., 2004, New Astron., 9, 443
Weinberg D. H., 1992, MNRAS, 254, 315
Zahn O., Zaldarriaga M., 2006, ApJ, 653, 922
Zaldarriaga M., Seljak U., 1999, Phys. Rev. D, 59, 123507

APPENDIX A: NORMALIZATION AND NOISE OF THE ESTIMATOR

In the end of Section 4, the numerical results of the noise of the estimators are shown. Here we will develop the analytical expression for

$$E_\kappa(\boldsymbol{k}_\perp) = b_\kappa(\boldsymbol{k}_\perp)[\kappa(\boldsymbol{k}_\perp) + n(\boldsymbol{k}_\perp)]. \tag{A1}$$

For shear, a similar relationship holds even though b and n are not isotropic.
When κ is spatially variable,

$$\tilde{T}_\mathrm{b}(\boldsymbol{x}) = T_\mathrm{b}(\boldsymbol{x}_\perp - \boldsymbol{D}_0 - \boldsymbol{\delta D}(\boldsymbol{x}_\perp), x_\parallel) = T_\mathrm{b}(\boldsymbol{x}_\perp - \boldsymbol{D}_0, x_\parallel) - \boldsymbol{\nabla}_\perp T_\mathrm{b}(\boldsymbol{x}_\perp - \boldsymbol{D}_0, x_\parallel)\cdot\boldsymbol{\delta D}(\boldsymbol{x}_\perp), \tag{A2}$$

where $\boldsymbol{D}(\boldsymbol{x}_\perp) = \boldsymbol{d}(\boldsymbol{x}_\perp)\,\chi(z_\mathrm{s})$, and $\boldsymbol{d}(\boldsymbol{x}_\perp)$ is the deflection angle. Therefore $\kappa = \boldsymbol{\nabla}_\perp\cdot\boldsymbol{\delta D}$, $\boldsymbol{\delta D}(\boldsymbol{x}_\perp) = \boldsymbol{D}(\boldsymbol{x}_\perp) - \boldsymbol{D}_0$ and $\boldsymbol{D}_0 = \boldsymbol{D}(0)$. $\boldsymbol{\delta D} \ll \boldsymbol{x}_\perp$ and can be treated as small perturbations since $\kappa \ll 1$.

Fourier transforming equation (29),

$$E_\kappa(\boldsymbol{k}_\perp) = \int \mathrm{d}^2x_\perp E_\kappa(\boldsymbol{x}_\perp)\mathrm{e}^{-\mathrm{i}\boldsymbol{k}_\perp\cdot\boldsymbol{x}_\perp} = \frac{1}{L}\int \mathrm{d}^3x\,\tilde{T}^\kappa_{\mathrm{w}_1}(\boldsymbol{x})\tilde{T}^\kappa_{\mathrm{w}_2}(\boldsymbol{x})\mathrm{e}^{-\mathrm{i}\boldsymbol{k}_\perp\cdot\boldsymbol{x}_\perp} - (2\pi)^2\delta^{2\mathrm{D}}(\boldsymbol{k}_\perp)V_\kappa. \tag{A3}$$

$\tilde{T} = \tilde{T}_\mathrm{b} + n$, and noise is uncorrelated with the signal. The product in real space can be represented as a convolution in Fourier space:

$$\int \mathrm{d}^3x\mathrm{e}^{-\mathrm{i}\boldsymbol{k}_\perp\cdot\boldsymbol{x}_\perp}\tilde{T}^\kappa_{\mathrm{w}_1}(\boldsymbol{x})\tilde{T}^\kappa_{\mathrm{w}_2}(\boldsymbol{x}) = \int \frac{\mathrm{d}^3k'}{(2\pi)^3}\tilde{T}^\kappa_{\mathrm{w}_1}(\boldsymbol{k}'_\perp, k'_\parallel)\tilde{T}^\kappa_{\mathrm{w}_2}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel), \tag{A4}$$

$$\tilde{T}_\mathrm{b}(\boldsymbol{k}) = \int \mathrm{d}^3x\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}T_\mathrm{b}(\boldsymbol{x}_\perp - \boldsymbol{D}(\boldsymbol{x}_\perp), x_\parallel) = \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_\perp\cdot\boldsymbol{D}_0}[T_\mathrm{b}(\boldsymbol{k}) - \int \mathrm{d}^3x\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\boldsymbol{\nabla}_\perp T_\mathrm{b}(\boldsymbol{x}_\perp, x_\parallel)\cdot\boldsymbol{\delta D}(\boldsymbol{x}_\perp + \boldsymbol{D}_0)], \tag{A5}$$

and the lensing introduced term can be further simplified as

$$\begin{aligned}\int \mathrm{d}^3x\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\boldsymbol{\nabla}_\perp T_\mathrm{b}(\boldsymbol{x}_\perp, x_\parallel)\cdot\boldsymbol{\delta D}(\boldsymbol{x}_\perp + \boldsymbol{D}_0) &= \int \mathrm{d}^3x\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}T_\mathrm{b}(\boldsymbol{x}_\perp, x_\parallel)(\mathrm{i}\boldsymbol{k}_\perp - \boldsymbol{\nabla}_\perp)\cdot\boldsymbol{\delta D}(\boldsymbol{x}_\perp + \boldsymbol{D}_0)\\ &= \int \frac{\mathrm{d}^2k'_\perp}{(2\pi)^2}T_\mathrm{b}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, k_\parallel)\mathrm{i}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp)\cdot\boldsymbol{\delta D}(\boldsymbol{k}'_\perp)\mathrm{e}^{\mathrm{i}\boldsymbol{k}'_\perp\cdot\boldsymbol{D}_0}.\end{aligned} \tag{A6}$$

The quadratic term in equation (A3) can be written as

$$\begin{aligned}\int \mathrm{d}^3x\mathrm{e}^{-\mathrm{i}\boldsymbol{k}_\perp\cdot\boldsymbol{x}_\perp}\tilde{T}^\kappa_{\mathrm{w}_1}(\boldsymbol{x}_\perp, x_\parallel)\tilde{T}^\kappa_{\mathrm{w}_2}(\boldsymbol{x}_\perp, x_\parallel) = \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_\perp\cdot\boldsymbol{D}_0}\int \frac{\mathrm{d}^3k'}{(2\pi)^3}W^\kappa_1(\boldsymbol{k}'_\perp, k'_\parallel)W^\kappa_2(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel)[T_\mathrm{b}(\boldsymbol{k}'_\perp, k'_\parallel)T_\mathrm{b}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel)\\ - T_\mathrm{b}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel)\int \frac{\mathrm{d}^2k''_\perp}{(2\pi)^2}T_\mathrm{b}(\boldsymbol{k}'_\perp - \boldsymbol{k}''_\perp, k'_\parallel)\mathrm{i}(\boldsymbol{k}'_\perp - \boldsymbol{k}''_\perp)\cdot\boldsymbol{\delta D}(\boldsymbol{k}''_\perp)\mathrm{e}^{\mathrm{i}\boldsymbol{k}''_\perp\cdot\boldsymbol{D}_0}\\ - T_\mathrm{b}(\boldsymbol{k}'_\perp, k'_\parallel)\int \frac{\mathrm{d}^2k'''_\perp}{(2\pi)^2}T_\mathrm{b}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp - \boldsymbol{k}'''_\perp, -k'_\parallel)\mathrm{i}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp - \boldsymbol{k}'''_\perp)\cdot\boldsymbol{\delta D}(\boldsymbol{k}'''_\perp)\mathrm{e}^{\mathrm{i}\boldsymbol{k}'''_\perp\cdot\boldsymbol{D}_0}]\\ + \text{noise}.\end{aligned} \tag{A7}$$

Using the relationship that

$$\langle T_{\rm b}(\boldsymbol{k}'_\perp, k'_\parallel) T_{\rm b}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel)\rangle = (2\pi)^3 \delta^{\rm 3D}(\boldsymbol{k}_\perp, 0) P_{\rm 3D}(\boldsymbol{k}'_\perp, k'_\parallel), \tag{A8}$$

we found that the expectation value of the first terms and the noise term in equation (A7) can cancel the last term in equation (A3). Note $\delta^{\rm D}(0) = \lim_{\Delta k\to 0} (\Delta k)^{-1} \sim (L/2\pi)$, and $W_2^\kappa(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel) \sim W_2^\kappa(\boldsymbol{k}'_\perp, k'_\parallel)$ since $\delta^{\rm 2D}(\boldsymbol{k}_\perp)$ is non-zero only when $\boldsymbol{k}_\perp = 0$. Similarly, the last two terms can be simplified. Both $e^{i\boldsymbol{k}''_\perp\cdot\boldsymbol{D}_0}$ and $e^{i\boldsymbol{k}'''_\perp\cdot\boldsymbol{D}_0}$ terms cancel $e^{-i\boldsymbol{k}_\perp\cdot\boldsymbol{D}_0}$ since $\boldsymbol{k}''_\perp = \boldsymbol{k}_\perp$ and $\boldsymbol{k}'''_\perp = \boldsymbol{k}_\perp$, respectively. The normalization factor

$$b_\kappa(\boldsymbol{k}_\perp) = \frac{2}{k_\perp^2}\int \frac{{\rm d}^3k'}{(2\pi)^3} W_1^\kappa(\boldsymbol{k}'_\perp, k'_\parallel) W_2^\kappa(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel)[(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp)\cdot\boldsymbol{k}_\perp P_{\rm 3D}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel) + \boldsymbol{k}'_\perp\cdot\boldsymbol{k}_\perp P_{\rm 3D}(\boldsymbol{k}'_\perp, k'_\parallel)]. \tag{A9}$$

Similarly, replacing W_1^κ, W_2^κ by $W_1^{\gamma_1}$, $W_2^{\gamma_1}(W_1^{\gamma_2}$, $W_2^{\gamma_2})$, and $k_\perp^2$ by $k_\perp^2\cos 2\theta_{\boldsymbol{k}_\perp}(k_\perp^2\sin 2\theta_{\boldsymbol{k}_\perp})$, we find the normalization factor for $\gamma_1(\gamma_2)$.

The noise of the estimator can be calculated in the absence of lensing: $\langle|\hat\kappa(\boldsymbol{k}_\perp)|^2\rangle = \langle\hat\kappa(\boldsymbol{k}_\perp)\hat\kappa^\star(\boldsymbol{k}_\perp)\rangle$. Since $\langle|\hat\kappa(\boldsymbol{k}_\perp)|^2\rangle = (2\pi)^2\delta^{\rm 2D}(0)N_\kappa(\boldsymbol{k}_\perp)$ and $\delta^{\rm 2D}(0) = \lim_{\Delta k\to 0}(\Delta k)^{-2} \sim (L/2\pi)^2$, Wick's theorem gives

$$\begin{aligned} N_\kappa(\boldsymbol{k}_\perp) = \frac{1}{b(\boldsymbol{k}_\perp)^2 L}\int\frac{{\rm d}^2k'_\perp}{(2\pi)^2}\int\frac{{\rm d}k'_\parallel}{(2\pi)}\{&P_{\rm 3D}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel)P_{\rm 3D}(\boldsymbol{k}'_\perp, k'_\parallel)[W_1^\kappa(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel)W_2^\kappa(\boldsymbol{k}'_\perp, k'_\parallel)]^2 \\ &+ P_{\rm 3D}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel)P_{\rm 3D}(\boldsymbol{k}'_\perp, k'_\parallel)\mathcal{F}^\kappa(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp, -k'_\parallel)\mathcal{F}^\kappa(\boldsymbol{k}'_\perp, k'_\parallel)\}. \end{aligned} \tag{A10}$$

The first term is the convolution of $P_{\rm 3D}(\boldsymbol{k})W_1^\kappa(\boldsymbol{k})^2$ and $P_{\rm 3D}(\boldsymbol{k})W_2^\kappa(\boldsymbol{k})^2$, and the second term is the convolution of $P_{\rm 3D}(\boldsymbol{k})\mathcal{F}^\kappa(\boldsymbol{k})$ with itself. The dimensionless quantity $k_\perp^2\, N_\kappa(\boldsymbol{k}_\perp)/(2\pi)$ is equivalent to $l^2C_l/(2\pi)$ in the literature.

This paper has been typeset from a TEX/LATEX file prepared by the author.