Wu, Z.-F., Chan, L. W.L., Hendry, M. and Hannuksela, O. A. (2023) Reducing the impact of weak-lensing errors on gravitational-wave standard sirens. Monthly Notices of the Royal Astronomical Society, 522(3), pp. 4059-4077.

Copyright © 2023 The Authors. Reproduced under a Creative Commons Attribution 4.0 International License.

For the purpose of open access, the author(s) has applied a Creative Commons Attribution license to any Accepted Manuscript version arising.

https://eprints.gla.ac.uk/297833/

Deposited on: 4 May 2023

Enlighten – Research publications by members of the University of Glasgow
https://eprints.gla.ac.uk

Reducing the Impact of Weak-lensing Errors on Gravitational-wave Standard Sirens

Zhao-Feng Wu,[1]★ Lok W. L. Chan,[1]† Martin Hendry,[2]‡ Otto A. Hannuksela[1]§
[1]*Department of Physics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong*
[2]*SUPA, School of Physics and Astronomy, University of Glasgow, UK*

Accepted 2023 April 18. Received 2023 April 12; in original form 2023 January 11

ABSTRACT

The mergers of supermassive black hole binaries (SMBHBs) can serve as standard sirens: the gravitational wave (GW) analog of standard candles. The upcoming space-borne GW detectors will be able to discover such systems and estimate their luminosity distances precisely. Unfortunately, weak gravitational lensing can induce significant errors in the measured distance of these standard sirens at high redshift, severely limiting their usefulness as precise distance probes. The uncertainty due to weak lensing can be reduced if the lensing magnification of the siren can be estimated independently, a procedure called 'delensing'. With the help of up-to-date numerical simulations, here we investigate how much the weak-lensing errors can be reduced using convergence maps reconstructed from shear measurements. We also evaluate the impact of delensing on cosmological parameter estimation with bright standard sirens. We find that the weak-lensing errors for sirens at $z_s = 2.9$ can be reduced by about a factor of two on average, but to achieve this would require expensive ultra-deep field observations for every siren. Such an approach is likely to be practical in only limited cases, and the reduction in the weak-lensing error is therefore likely to be insufficient to significantly improve the cosmological parameter estimation. We conclude that performing delensing corrections is unlikely to be worthwhile, in contrast to the more positive expectations presented in previous studies. For delensing to become more practicable and useful in the future will require significant improvements in the resolution/depth of weak-lensing surveys and/or the methods to reconstruct convergence maps from these surveys.

Key words: gravitational lensing: weak – gravitational waves – distance scale – cosmological parameters

1 INTRODUCTION

A promising method to obtain accurate cosmological distance measurements in the future is to observe gravitational waves emitted by merging supermassive black hole binaries (SMBHBs). Schutz (1986) showed that measuring the GW signals of the inspiraling binary with a network of interferometers could estimate its luminosity distance, independently of the cosmic distance ladder. That is why these binaries are coined as 'standard sirens' – the gravitational wave analog of standard candles.

However, the redshift cannot be measured from the gravitational waves alone, so identifying an electromagnetic (EM) counterpart is also crucial for cosmological purposes (Dalal et al. 2006). Such an electromagnetic counterpart would exist if the merging source was a binary neutron star (BNS), in which case a potentially detectable gamma-ray burst would be associated with the merger (Nakar 2007), or may exist if the merging source was an SMBHB, in which case the accretion disk surrounding the black holes may give off an electromagnetic signal (Milosavljević & Phinney 2005; Tanaka et al. 2012; Yuan et al. 2021). If the EM counterpart of the coalescence can be observed, and thus the redshift of the host galaxy determined, the cosmological parameters can then be estimated by analyzing the relationship between the luminosity distance and the redshift of these standard sirens.

In this work, only GWs generated by the coalescence of supermassive (or massive) black hole binaries are considered. These systems are expected to be luminous in GWs such that these systems may be detectable even up to $z = 5$. However, the frequency of the expected GW signals from SMBHBs lies below the sensitive band of ground-based GW detectors and is thus better suited for space-based detection.

Scientists have proposed several space-borne GW detector projects to extend the GW spectrum to the millihertz band, such as *Laser Interferometer Space Antenna* (*LISA*) (Amaro-Seoane et al. 2017) and *TianQin* (*TQ*) (Luo et al. 2016). GW detectors in space benefit from their long baselines and the absence of seismic noise. A baseline of 2.5 million kilometers (Thorpe et al. 2019) between spacecraft allows the detection of GWs in the millihertz band.

The expected number and redshift distribution of SMBHBs that *LISA* and *TQ* will observe are very uncertain and strongly dependent on details of our model for the mergers of galaxy nuclei (Arun et al. 2009). However, observing even a handful of SMBHBs could constrain the standard cosmological model with impressive accuracy, as was first demonstrated by Holz & Hughes (2005). The expected measurement errors of the luminosity distance from the SMBHB merger can be dramatically small compared with the scatter present

★ E-mail: Foisonwzf@link.cuhk.edu.hk
† E-mail: lokwlc@link.cuhk.edu.hk
‡ E-mail: martin.hendry@glasgow.ac.uk
§ E-mail: hannuksela@phy.cuhk.edu.hk

© 2022 The Authors

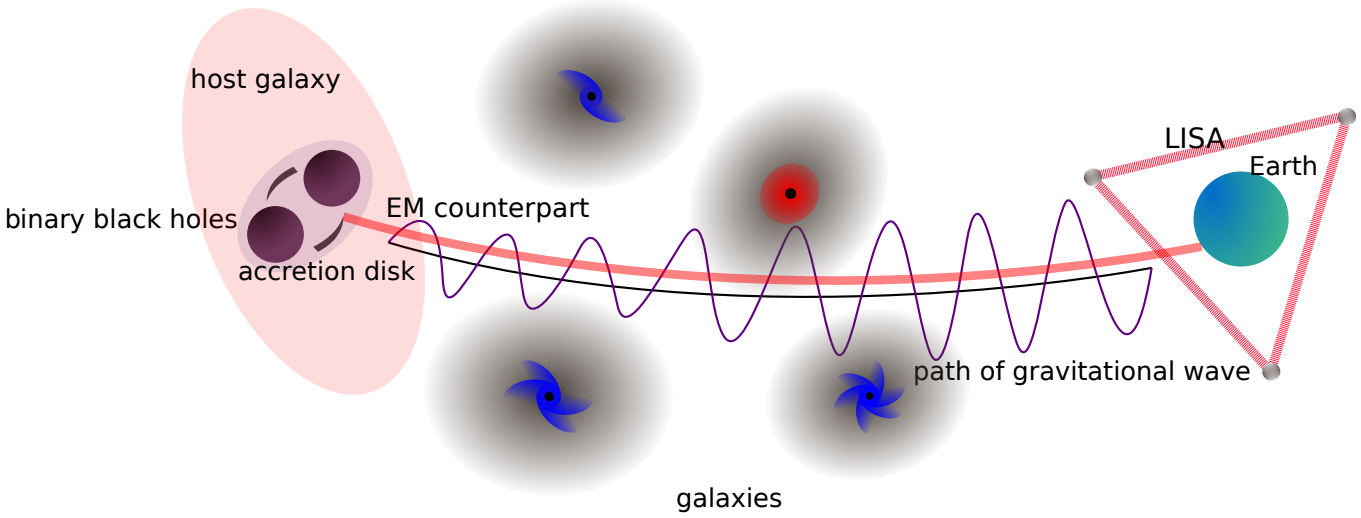

Figure 1. An illustration of the effect of weak gravitational lensing on the GWs emitted by binary black hole systems (SMBHBs in our case). We show here both the host galaxy of the binary and the circumbinary accretion disk[1] that may be responsible for the EM counterpart. The dark matter haloes around the intervening galaxies will (de)magnify the GW signals, adding uncertainties to the amplitude of the original waveforms. The GW signals will be measured by space-borne GW detectors (*LISA* in the illustration as an example) and the EM counterpart will be observed by electromagnetic telescopes.

for other cosmological probes. At the same time, many SMBHBs are anticipated to be detected at high redshift. These data will, therefore, potentially be of great value for the measurement of cosmological parameters such as H_0 and the dimensionless densities of dark matter and dark energy.

However, as was also pointed out in Holz & Hughes (2005), there is a huge caveat: GW signals from the SMBHBs would be affected by weak gravitational lensing, which is caused by fluctuations in the density of matter along the line of sight to the siren. When the sky position of the targeted siren coincides with a foreground galaxy or cluster, then strong lensing takes place, which can provide much more information than a standard siren alone (Pang et al. 2020; Hannuksela et al. 2020). While such strong lensing events will be very useful, their probability of occurring is extremely small. Besides, the uncertainties in the lens model also limit their usefulness. Consequently, only weak gravitational lensing will be considered in this work. Weak gravitational lensing is caused by extended dark matter halos that lie along the line of sight to the siren; the effect is weaker yet more prevalent for high-redshift standard sirens like SMBHBs (See Fig. 1 for illustration). The error induced on the luminosity distance by the weak-lensing effect is expected to be a few percent for sources at high redshift, which is comparable to the scatter of the SMBHB siren luminosity distance estimate itself (Holz & Linder 2005; Kocsis et al. 2006). Therefore, the weak-lensing effect substantially limits the power of the SMBHB sirens as precise cosmological distance probes.

To improve the performance of SMBHB sirens, corrections from EM observations may be applied to reduce the impact of weak-lensing errors. Shapiro et al. (2010) proposed a method based on estimating the weak-lensing signal in the direction of each siren using maps of the shear and flexion reconstructed from surveys of foreground galaxies. In this way we can infer the weak-lensing error affecting each siren from the reconstructed maps and hence correct for that error on the siren's luminosity distance estimate – a procedure called 'delensing'.

In previous studies of delensing (Shapiro et al. 2010; Hilbert et al. 2011), the authors assumed a model for the matter fluctuations across cosmic time either from analytical formulae or numerical simulations. In Shapiro et al. (2010), the analytical formulae used had been extended beyond their accepted range of accuracy to reach the required higher resolution. On the other hand, the numerical simulations used in Hilbert et al. (2011) had an adequate resolution, but the values of the cosmological parameters adopted by the simulations deviate from the more recently accepted values.[2] As pointed out in Hilbert et al. (2011), the optimal smoothing strategy and outcomes of delensing should be cosmologically dependent, so a revisit is necessary to investigate the reliability of the predictions made in that work.

Now with the availability of up-to-date numerical simulations, a more complete and consistent treatment can be applied to evaluate the potential of delensing by weak-lensing reconstruction. Compared to the previous studies, we focus on making the evaluation more realistic, and more in line with the expected performance of future galaxy surveys. Due to the uncertainties in the advances of technology, here we adopt optimistic settings for the measurement uncertainties to explore the optimal ability of delensing in the future. Moreover, in this paper we also use up-to-date SMBHB siren mock catalogs, allowing an explicit investigation of the impact of weak-lensing error reduction on cosmological parameter estimation. However, we do not perform a joint standard siren + weak gravitational lensing analysis, which has been discussed in other studies (Congedo & Taylor 2019; Mpetha et al. 2022; Balaudo et al. 2022). We also do not exploit the non-Gaussianity of the lensing distributions.

The goal of this paper is, therefore, to evaluate the potential of weak-lensing reconstruction for reducing weak-lensing errors on

[1] We neglect possible circum-single disks around each black hole and a probable binary cavity in the illustration, which may also help identify the EM counterparts.

[2] The cosmological parameters used in the backbone N-body simulations in Hilbert et al. (2011) are: $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $h = 0.73$, $n = 1$, $\sigma_8 = 0.9$.

SMBHB sirens, and to investigate the impact on cosmological parameter estimation. We attempt to strike a balance between optimistic and realistic assumptions on future surveys so that the predictions could be more useful and reliable in the future. In Sec. 2.1 and 2.3, we introduce the necessary background information for standard siren cosmology and weak-lensing reconstruction. A Bayesian analysis to investigate the impact of delensing on cosmological parameter estimation is then described in Sec. 2.2, and the construction of the best estimator of weak-lensing error is presented in Sec. 2.4. In Sec. 2.5, we introduce the numerical simulations adopted in our analysis. In Sec. 3, we compute the reduction in the weak-lensing error under different observational schemes. We discuss the outcomes of the previous delensing studies and compare them to ours in Sec. 4. Finally, we conclude in Sec. 5 along with a discussion about topics for further investigation.

2 METHODOLOGY

2.1 Standard siren cosmology

2.1.1 Standard sirens

For the inspiral of a compact binary system with component masses m_1 and m_2, the frequency domain GW waveform can be expressed as (Sathyaprakash & Schutz 2009),

$$\widetilde{h}(f) = \frac{1}{D_L}\sqrt{\frac{5}{24}}\frac{(G\bar{M}_z)^{5/6}}{\pi^{2/3}c^{3/2}} f^{-7/6}\exp(-i\Phi(f;\bar{M}_z,\eta)) \tag{1}$$

where G is the gravitational constant, $\bar{M} = \eta^{0.6}M$ is the chirp mass, $\eta = m_1m_2/M^2$ is the symmetric mass ratio, $M = m_1 + m_2$ is the total mass, and Φ is phase of the waveform. The source luminosity distance D_L directly appears in the waveform, which provides luminosity distance information free from the 'cosmic distance ladder'. Therefore, these compact binary systems are coined as 'standard sirens' and their measured luminosity distances may have small systematic and statistical errors compared to other approaches. One type of standard siren is the SMBHB system that generates GW signals detectable by space-borne GW detectors. We refer to SMBHB sirens simply as (standard) sirens hereafter.

However, according to Eq. (1), only the redshifted chirp mass $\bar{M}_z \equiv \bar{M}(1+z)$ is accessible in the waveform, and the redshift z is therefore degenerate with the chirp mass $\bar{M}$. Consequently, measurements of redshift must rely on extra information. In this work, the EM counterparts of the GW signals are assumed to be accessible so that the degeneracy between redshift and mass is broken – i.e. a scenario where the sirens are 'bright'.

Possible EM counterparts from SMBHBs include the broadband nonthermal EM emission from electrons accelerated at the external forward shock expected in post-merger relativistic jets from the coalescence (Yuan et al. 2021). If the binary's tidal torques are able to open a central cavity in the accreting gaseous disc, the SMBHB might have an unusually low soft X-ray luminosity and weak ultraviolet and broad optical emission lines, as compared to an AGN powered by a single supermassive black hole with the same total mass (Tanaka et al. 2012). This could also help identify the EM counterparts of the sirens.

2.1.2 Estimating cosmological parameters

In this paper, we focus on the estimation of dimensionless density parameters of dark matter and dark energy respectively, labelled Ω_M and Ω_Λ, without assuming a flat universe[3] under the ΛCDM model. This is because the SMBHBs are detectable even for redshift up to $z = 3$, where the difference in the expansion history becomes more distinguishable for different values of Ω_Λ and Ω_m compared to the low-redshift case.

Once the luminosity distance D_L and redshift z of each siren have been measured, we can use the theoretical relation between D_L and z, which depends on the expansion history of the universe, to estimate the cosmological parameters. The theoretical relationship is given by,

$$D_L = \frac{c(1+z)}{H_0}\begin{cases} \frac{1}{\sqrt{\Omega_K}}\sinh(\sqrt{\Omega_K}\int_0^z \frac{H_0}{H(z')}dz') & \text{for } \Omega_K > 0 \\ \int_0^z \frac{H_0}{H(z')}dz' & \text{for } \Omega_K = 0 \\ \frac{1}{\sqrt{|\Omega_K|}}\sin(\sqrt{|\Omega_K|}\int_0^z \frac{H_0}{H(z')}dz') & \text{for } \Omega_K < 0 \end{cases} \tag{2}$$

where $H_0 \equiv H(z=0)$ describes the current expansion rate of the Universe and c is the speed of light in vacuum. The Hubble constant for different redshifts $H(z)$ is governed by the normalized Friedmann equation,

$$H(z) = H_0\sqrt{\Omega_M(1+z)^3 + \Omega_K(1+z)^2 + \Omega_\Lambda}, \tag{3}$$

where Ω_M, Ω_K, and Ω_Λ are fractional densities of the total matter, curvature, and dark energy with respect to the critical density $\rho_c = 3H_0^2/8\pi G$ respectively.

The observed luminosity distance of each siren ${D_L}^{\rm obs}$ is subject to various sources of noise that include measurement uncertainties, the impact of host galaxy peculiar velocities and the effect of weak gravitational lensing. Measurement noise could be reduced by advances in the relevant GW detector technologies, while the SMBHBs are assumed to be distant enough so that the effect of peculiar velocity is negligible. In this work, we concentrate on the weak-lensing effect.

2.1.3 Weak gravitational lensing & delensing

Unlike the impact of peculiar velocities, which will reduce with distance, the uncertainties due to weak gravitational lensing become more dominant for high-redshift sirens (Hirata et al. 2010). GWs generated from those sources have a larger probability to be significantly lensed by the intervening dark matter halos as they propagate through the universe. Therefore, reducing the errors from weak gravitational lensing for the high-redshift sirens should be important. A brief introduction about the weak-lensing effect and the idea of 'delensing' is as follows. A more detailed and technical introduction is given in Section 2.3.

In the weak-lensing and geometrical optics limit, the observed luminosity distance to a siren is related to its true luminosity distance by

$${D_L}^{\rm obs} = D_L/\sqrt{\mu} \approx D_L(1 - \delta\mu/2), \tag{4}$$

where μ is the lensing magnification and $\delta\mu = \mu - 1$ is the deviation from the unlensed case (Holz & Hughes 2005; Takahashi 2006). Note that $\delta\mu = 0$ if the siren is not lensed. If the lensing magnification experienced by each siren can be estimated from other measurements,

[3] However, the simulations used in this paper all assumed *true* values of the cosmological parameters consistent with a flat universe, i.e. $\Omega_K = 0$, when generating the data.

then we can compensate for the weak-lensing errors and obtain corrected observed luminosity distances $D_L^{\rm obs}$. This is the fundamental motivation of this work and the origin of the name 'delensing'.

2.2 Bayesian analysis on impact of delensing

In this section, we introduce a Bayesian analysis to quantify the impact of delensing on cosmological parameter estimation by standard sirens. The posterior on the cosmological parameters is given by[4]

$$p(\vec{\Omega}|\,\mathbf{D}_{\rm GW}\,,\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens}) \propto p(\vec{\Omega}|\,\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens}) p(\mathbf{D}_{\rm GW}\,|\,\vec{\Omega},\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens})\,, \quad (5)$$

where $\vec{\Omega}$ is e.g. a two-dimensional vector consisting of Ω_M and Ω_Λ, $\mathbf{D}_{\rm GW}$ is the vector of GW data, $\mathbf{z}_{\rm s}$ is the vector of siren's redshifts determined by EM counterparts, $\mathbf{d}_{\rm lens}$ is the lensing data used for delensing. In this paper, the lensing data $\mathbf{d}_{\rm lens}$ include the sky positions, observed ellipticities, and redshifts of galaxies.

Note that here we are assuming that H_0 is already known and fixed in value; we could also consider the joint inference of H_0 and the other cosmological parameters, but we expect that the H_0 tension will be resolved and H_0 can be approximated as a fixed parameter by the time *LISA* operates[5]. Even if the tension between far- and near-field observations remains, the precision of the near-field measurements will be sufficient to narrow down the prior for H_0 to the degree that the approximation is reasonable and *LISA* can only marginally improve the inference of H_0.

For parallel channel checking, the ground-based GW detectors should already have detected many BNS events to constrain the value of H_0 by that time (Califano et al. 2023), so *LISA* could only contribute weakly and H_0 could be approximated as a fixed parameter again. In principle, BNS and SMBHBs serve in the same way as standard sirens and thus experience similar systematics. However, SMBHBs are much more luminous than BNS and thus detectable up to higher redshift with stronger power in constraining Ω_m and Ω_Λ. Therefore, we focus on inferring Ω_m and Ω_Λ in this paper.

In this paper we also do not perform joint cosmological inference. Therefore, we may write $p(\vec{\Omega}|\,\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens}) = p(\vec{\Omega})$ and we assume for simplicity flat priors on the cosmological parameters. The posterior on cosmological parameters then becomes

$$p(\vec{\Omega}|\,\mathbf{D}_{\rm GW}\,,\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens}) \propto p(\mathbf{D}_{\rm GW}\,|\,\vec{\Omega},\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens})\,. \quad (6)$$

The $\mathbf{D}_{\rm GW}$ term includes the observed luminosity distance $D_L^{\rm obs}$ and their measurement uncertainties σ_{D_L}, where the latter are obtained from the Fisher matrix formalism with first-order approximation (Zhu et al. 2022; Chassande-Mottin et al. 2019). Therefore, the likelihood $p(\mathbf{D}_{\rm GW}\,|\,\vec{\Omega},\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens})$ can be derived as[6]

$$p(\mathbf{D}_{\rm GW}|\,\vec{\Omega},\mathbf{z}_{\rm s},\mathbf{d}_{\rm lens}) \propto \prod_i \int \exp\left[D_L^{\rm cor}(\vec{\Omega},z_{\rm s},\mu-\mu_{\rm est}) - D_L^{\rm obs,cor})^2/2\sigma_{D_L}^2\right] \times p(\mu-\mu_{\rm est}|\,\mathbf{d}_{\rm lens},z_{\rm s})\,d\mu\,, \quad (7)$$

where $\mu_{\rm est}$ is the estimated magnification, μ is the true magnification, and $z_{\rm s}$ is the redshift of each siren. The corrected luminosity distances $D_L^{\rm cor}(\vec{\Omega},z_{\rm s},\mu-\mu_{\rm est})$ and $D_L^{\rm obs,cor}$ are determined by,

$$\begin{aligned} D_L^{\rm cor}(\vec{\Omega},z_{\rm s},\mu-\mu_{\rm est}) &= D_L(\vec{\Omega},z_{\rm s}) \times (1+(\mu_{\rm est}-\mu)/2)\,, \\ D_L^{\rm obs,cor} &= D_L^{\rm obs}\sqrt{\mu_{\rm est}} \end{aligned} \quad (8)$$

where $D_L(\vec{\Omega},z_{\rm s})$ is the luminosity distance calculated by Eq. (2).

The $p(\mu-\mu_{\rm est}|\,\mathbf{d}_{\rm lens},z_{\rm s})$ term is the conditional distribution of the error in magnification estimation $\mu-\mu_{\rm est}$ given the lensing data $\mathbf{d}_{\rm lens}$ at redshift $z_{\rm s}$. The construction of $\mu_{\rm est}$ and the estimation of $p(\mu-\mu_{\rm est}|\,\mathbf{d}_{\rm lens},z_{\rm s})$ is given in Sec. 2.4.

2.3 Weak-lensing reconstruction simulation

In Section 2.1.3, it was noted that the weak-lensing error is connected with the observed luminosity distance $D_L^{\rm obs}$ by means of the magnification μ. The magnification is further related to the lensing convergence κ by (Takahashi 2006; Shapiro et al. 2010),

$$\mu \approx 1+2\kappa. \quad (9)$$

Therefore, if we can resolve the convergence by constructing an accurate convergence map around the siren's location, then we can estimate the siren's weak-lensing error. The convergence map can be constructed from other weak-lensing fields, which is a method called weak-lensing reconstruction. Here we only focus on the use of weak-lensing shear fields, leaving a discussion about using other lensing fields like flexion until Sec. 4.

In this paper, we make use of the simulated weak-lensing maps from the extended Scinet Light Cone Simulations (SLICS) (Harnois-Deraps et al. 2018). The SLICS contain flat sky weak-lensing maps constructed by ray-tracing with the Multiple-Lens-Plane technique (Vale & White 2003) under the Born approximation (Schneider et al. 1998; White & Vale 2004). It has been shown that these approximations are in good agreement with the full treatment and only deviate slightly at the smallest scale (Harnois-Déraps & van Waerbeke 2015; Hilbert et al. 2020). The lensing maps neglect baryonic effects and consider lensing by dark matter only. A brief introduction to the Born approximation and Multiple-Lens-Plane technique, together with the reconstruction algorithm, is as follows.

2.3.1 Ray-tracing

The trajectories of photons are deflected gravitationally by intervening matter inhomogeneities before they reach the observer. This deflection is called gravitational lensing, which causes position shifts in the observed image (Schneider et al. 2006). Therefore, the weak-lensing simulations are generally constructed by integrating over null geodesics to calculate the total deflection along the past light cone (Harnois-Déraps & van Waerbeke 2015). Under the Born approximation, the integrations are simplified to be calculated on straight lines (rather than photons' trajectories). Then the weak-lensing convergence $\kappa(\boldsymbol{\theta})$ can be obtained by integrating over the density contrast $\delta(\boldsymbol{\theta},\chi)$ along the line of sight:

$$\kappa(\boldsymbol{\theta},\chi_s) = \frac{3H_0^2\Omega_m}{2c^2}\int_0^{\chi_s}\frac{\chi(\chi_s-\chi)}{\chi_s}\delta(\boldsymbol{\theta},\chi)(1+z)\,d\chi, \quad (10)$$

where χ is the comoving distance and χ_s is the comoving distance to the source plane in which the targeted siren is embedded. The density contrast $\delta(\boldsymbol{\theta},\chi)$ is defined by

$$\delta(\boldsymbol{\theta},\chi) \equiv \frac{\rho(\boldsymbol{\theta},\chi)}{\bar{\rho}(\chi)} - 1\,, \quad (11)$$

[4] See Appendix A for a detailed derivation.

[5] As pointed out by Lahav & Liddle (2022), while the tension remains highly significant, its severity has somewhat lessened in the past two years. Many recent near-field observations yield similar results as the CMB measurements for H_0, including one using Type Ia supernovae but with a different calibration. The hope that the Hubble tension will be resolved in the coming years is reasonable.

[6] See Appendix A for a detailed derivation.

where ρ is the matter density and $\bar{\rho} = 3H_0^2\Omega_m/8\pi G a^3$ is the mean density of the universe at different time. H_0 and ω_m are the present-day values of the Hubble constant and density parameter.

Furthermore, the matter distribution in the light cone can be approximated as a set of discrete lens planes. Then the 3D matter density distribution can be collapsed into planes of two-dimensional density fluctuations, given by,

$$\delta_{2D}(\boldsymbol{\theta}, \chi_{\rm lens}) = \frac{1}{\Delta\chi}\int_{\chi_{\rm lens}-\frac{1}{2}\Delta\chi}^{\chi_{\rm lens}+\frac{1}{2}\Delta\chi} \delta(\boldsymbol{\theta}, \chi)\, d\chi, \tag{12}$$

where $\Delta\chi$ denotes the comoving length of the collapsed region. This effectively turns the integration along the light ray into a discrete sum at the lens locations and manifests the name of the technique. The convergence κ is then computed by,

$$\kappa(\boldsymbol{\theta}, \chi_s) = \frac{3H_0^2\Omega_m}{2c^2}\sum_{\chi=\chi_1}^{\chi=\chi_n}\frac{\chi(\chi_s-\chi)}{\chi_s}\delta_{2D}(\boldsymbol{\theta}, \chi)(1+z(\chi))\Delta\chi, \tag{13}$$

where χ_1 is the comoving distance to the first lens plane and χ_n is the comoving distance to the last lens plane before the source.

2.3.2 KS inversion method

The weak-lensing shear and convergence are both combinations of derivatives of a scalar lensing potential field $\psi(\boldsymbol{\theta})$. The two components of the shear can be written in terms of $\psi(\boldsymbol{\theta})$ by,

$$\gamma_1 = \frac{1}{2}(\partial_1^2 - \partial_2^2)\psi, \quad \gamma_2 = \partial_1\partial_2\psi\,, \tag{14}$$

where the partial derivatives ∂_i with $i = 1, 2$ correspond to the angular coordinates θ_i. The convergence $\kappa(\boldsymbol{\theta})$ can also be expressed in terms of $\psi(\boldsymbol{\theta})$ by

$$\kappa = \frac{1}{2}(\partial_1^2 + \partial_2^2)\psi\,. \tag{15}$$

Then the shear maps $\gamma_{1,2}(\boldsymbol{\theta})$ can be computed via Fourier transformation by,

$$\hat{\gamma}_{1,2} = \hat{\kappa}\,\hat{P}_{1,2}\,, \tag{16}$$

where the hat symbol denotes Fourier transform and $\hat{P}_1(\boldsymbol{\ell}), \hat{P}_2(\boldsymbol{\ell})$ are given by,

$$\hat{P}_1(\boldsymbol{\ell}) = \frac{\ell_1^2-\ell_2^2}{\ell^2}, \quad \hat{P}_2(\boldsymbol{\ell}) = \frac{2\ell_1\ell_2}{\ell^2}, \tag{17}$$

with $\ell^2 \equiv \ell_1^2 + \ell_2^2$ and ℓ_i being the wave numbers with respect to the angular coordinates θ_i.

The order is reversed in practice as the convergence map is not directly observable in a galaxy survey. By contrast, the shear maps $\gamma_{1,2}(\boldsymbol{\theta}, z)$ can be derived from the weighted average ellipticities of the galaxies around the position $\boldsymbol{\theta}$ in the image[7]. The weighting takes factors like the redshift distribution of galaxies into consideration and more details will be provided in the second half of Sec. 3.1. Then the KS inversion method (Kaiser & Squires 1993; Bartelmann & Schneider 2001) can be used to reconstruct the convergence map[8]. The KS inversion method is simply the inverse of Eq. (16),

$$\hat{\kappa} = \hat{P}_1\hat{\gamma}_1 + \hat{P}_2\hat{\gamma}_2\,, \tag{18}$$

where the hat symbol denotes Fourier transforms and $\hat{P}_1(\boldsymbol{\ell}), \hat{P}_2(\boldsymbol{\ell})$ are the same as above. According to Eq. (18), the shear maps $\gamma_{1,2}(\boldsymbol{\theta})$ are related to the convergence map $\kappa(\boldsymbol{\theta})$ by simple algebraic equations in the Fourier domain. Then the reconstructed convergence map can be obtained by inverse Fourier transforms.

2.3.3 Mass-sheet degeneracy problem

There is a degeneracy when reconstructing κ from $\gamma_{1,2}$ when $\ell_1 = \ell_2 = 0$. Consequently, the mean value of the reconstructed convergence field $\bar{\kappa}$ cannot be determined only from shear information, which is the so-called mass-sheet degeneracy (Bartelmann 1995). In practice, the observational area is finite, resulting in a lower bound $\ell_{\rm min}$ for the wave numbers of the reconstructed Fourier modes. Basically, the reconstructed convergence would differ from the true convergence field by a constant determined by fluctuations larger than the field area.

For wide-field surveys, we often set the reconstructed modes with $\ell \le \ell_{\rm min}$ to zero, as the fluctuations larger than the field area are negligible. This is a reasonable assumption for wide-field reconstruction (Massey et al. 2007). For deep-field surveys, the observation area is much smaller and thus the error from mass-sheet degeneracy must be treated seriously.

In principle, the mass-sheet degeneracy problem for deep-field surveys may be alleviated if additional wide-field surveys around that region are available (Shapiro et al. 2010). The basic idea behind this is that the deep images will be used to measure small-scale convergence fluctuations, while the wide images will pick up modes larger than the size of the deep images. In practice, the effectiveness of hybridizing wide and deep survey maps is subject to weak-lensing shape noise, which highly depends on the size of the deep-field survey as well as the galaxy density $n_{\rm gal}(\boldsymbol{\theta}, z_{\rm gal})$ of the wide-field survey. It is possible that the hybridization has no improvement because the galaxy density of the wide-field survey is not enough to obtain a satisfying signal-to-noise ratio (SNR) even in probing large-scale fluctuations.

2.3.4 Weak-lensing shape noise

The gravitational shears can be derived from the ellipticities $\epsilon_{1,2}$ of the background galaxies. However, in reality, the projected shapes of galaxies are not intrinsically circular. The measured ellipticity is a combination of their intrinsic ellipticity and the gravitational lensing shear. The shear is also subject to measurement noise and uncertainties in the PSF (point spread function) correction. All these effects can be modelled as additive noises[9] to both components of the shear field (Pires et al. 2020),

$$\epsilon_i = \gamma_i + N_i, \tag{19}$$

where $i = 1, 2$. The noises N_1 and N_2 are assumed to be Gaussian and independent of each other, with zero mean and standard deviation given by,

$$\sigma_s^i(\boldsymbol{\theta}) = \frac{\sigma_\gamma}{\sqrt{n_g(\boldsymbol{\theta})}}, \tag{20}$$

[7] The ellipticity of a galaxy is a point estimate for the reduced shear at that sky position and redshift, which is the only real observable in galaxy surveys. We ignore the reduced factor as it is close to one in the weak-lensing regime.

[8] If there is no available information about some line-of-sights (due to sparsity of galaxies or masks), then advanced methods, such as the one in Pires et al. (2020), should be applied to overcome the missing-data problem.

[9] In reality, the measurement noise and uncertainties in the PSF correction are much more complicated and highly detector dependent. The effects should be a combination of additive or multiplicative errors/biases. Here, we are evaluating the potential of delensing under very ideal conditions.

where $n_{\rm gal}(\boldsymbol{\theta})$ is the number of galaxies within the pixel at $\boldsymbol{\theta}$. The shear dispersion per galaxy, σ_γ, arises both from the measurement uncertainties and the intrinsic scatter of galaxy shapes. The Gaussian assumption is reasonable (Hirata et al. 2010) in the weak-lensing regime and here we further assume that the shear dispersion for each galaxy is independent of each other. In other words, we ignore the intrinsic alignment of neighbour galaxies which can bias the shear estimate. In light of Eq. (19), we derive in Fourier space,

$$\hat{\kappa}^{\rm noisy} = \hat{\kappa} + \hat{P}_1\hat{N}_1 + \hat{P}_2\hat{N}_2\,. \tag{21}$$

Then the reconstructed convergence map is also subject to an additive noise N_κ by,

$$N_\kappa = P_1 * N_1 + P_2 * N_2\,, \tag{22}$$

where the asterisk denotes convolution, P_1 and P_2 are the inverse Fourier transforms of $\hat{P}_1$ and $\hat{P}_2$.

When the shear noises N_1 and N_2 are Gaussian and independent across the field with a constant standard deviation, then the induced noise in convergence maps is also Gaussian and independent at each position, with a standard deviation $\sigma_\kappa = \sigma_s^1 = \sigma_s^2$ according to Eq. (21). In reality, the number of galaxies varies slightly across the field. Even at the same position, the variances of N_1 and N_2 might also be slightly different. These effects introduce noise correlations and variations in the reconstructed convergence maps, but they were found to remain negligible compared to other effects studied in this paper.

The pixelation of the observed ellipticity fields already smooths the fields with a smoothing scale equal to the pixel size. However, it may be insufficient to obtain a satisfactory signal-to-noise ratio for the estimated convergence. Then the raw estimated convergence should be further smoothed by a Gaussian filter with filter scale θ_s,

$$\kappa_{\rm smooth}(\boldsymbol{\theta}) \propto \int exp(-\frac{(\boldsymbol{\theta}-\boldsymbol{\theta}')^2}{2\theta_s^2})\kappa_{\rm raw}(\boldsymbol{\theta}')\,d^2\boldsymbol{\theta}' \tag{23}$$

where the integration is conducted on the whole field area and the proportionality constant is determined by normalization. The smoothing can also be done by using Wiener filters, but it turns out that the improvements are marginal (Hilbert et al. 2011). There should exist an optimal smoothing scale determined by a trade-off between reconstructing small-scale features of the convergence and reducing shape noise by averaging over galaxy images (Dalal et al. 2003).

2.3.5 *Redshift distribution of galaxies*

Finally, the galaxies do not all lie at the redshift $z_{\rm s}$ of the standard candle/siren but have a certain redshift distribution $n_{\rm gal}(\boldsymbol{\theta}, z_{\rm gal})$ that depends on their intrinsic redshift distribution and on the depth of the survey. A smoothed version of the effective convergence will be reconstructed if the individual galaxy redshifts are not utilized:

$$\kappa_{\rm eff}(\boldsymbol{\theta}) = \int n_{\rm gal}(\boldsymbol{\theta}, z_{\rm gal})\,\kappa(\boldsymbol{\theta}, z_{\rm gal})\,dz_{\rm gal}. \tag{24}$$

If the redshift distribution of the galaxy number density $n_{\rm gal}(\boldsymbol{\theta}, z_{\rm gal})$ is not sharply peaked around the redshift $z_{\rm s}$ of the standard siren, the effective convergence $\kappa_{\rm eff}(\boldsymbol{\theta})$ may deviate substantially from the true convergence $\kappa(\boldsymbol{\theta}, z_{\rm s})$. Variations in the redshifts of galaxies contribute to an additional noise source.

2.4 Estimator of magnification from noisy convergence maps

The reconstructed convergence maps from shear measurements can be converted directly into the estimated magnification maps simply by Eq. (9).[10]

However, as pointed out by Hilbert et al. (2011), this simple estimate might fail if the estimated convergence $\kappa_{\rm est}(\boldsymbol{\theta})$ deviates substantially from the true convergence $\kappa(\boldsymbol{\theta}, z_{\rm s})$. We quantify this deviation by the residual magnification,

$$\mu_{\rm res} = \mu - \mu_{\rm est}. \tag{25}$$

The difference may originate from the weak-lensing shape noise, mass-sheet degeneracy, or realistic redshift distribution of galaxies. These effects make the simple estimate perform even worse than just using the lensing prior in certain cases (i.e., assuming $\mu_{\rm est}(\boldsymbol{\theta}) = 0$).

To improve this, one should construct an unbiased magnification estimator which minimizes the residual dispersion $\sigma_{\mu_{\rm res}}$ (Hilbert et al. 2011). If the conditional distribution $p(\mu|\kappa_{\rm est})$ of the true magnification μ for a given estimated convergence $\kappa_{\rm est}$ is known, then the estimator satisfying the above requirements can be derived as,

$$\mu_{\rm est}(\boldsymbol{\theta}) = \langle\mu\rangle_{\mu|\kappa_{\rm est}}(\kappa_{\rm est}(\boldsymbol{\theta})), \tag{26}$$

where $\langle\mu\rangle_{\mu|\kappa_{\rm est}}(\kappa_{\rm est})$ is the expectation value of the true magnification μ for a given estimated convergence $\kappa_{\rm est}$:

$$\langle\mu\rangle_{\mu|\kappa_{\rm est}}(\kappa_{\rm est}) = \int_0^{\mu_{\rm max}} \mu\, p(\mu|\kappa_{\rm est})\,d\mu, \tag{27}$$

where $\mu_{\rm max}$ is the maximal allowed value of the magnification which remains in the weak-lensing regime. In the subsequent discussions and analyses, $\mu_{\rm max} = 1.5$ is chosen so that all the data above that value are abandoned.

The conditional distribution of the true magnification given the estimated convergence can be inferred from numerical simulations, thus allowing the optimal magnification estimator to be constructed. However, the magnification estimator constructed in this way will be dependent on the values of the cosmological parameters (and other settings, e.g. the galaxy biasing scheme) adopted by the numerical simulations that are used. Strictly, therefore, this magnification estimator ought to account for uncertainties in these model parameters and settings through appropriate marginalisation – with the result that, in practice, the optimal magnification estimator would likely perform somewhat more poorly than considered here. In what follows, however, we will not consider this issue further since our purpose in this work is to investigate the limitations of delensing methods in a realistic setting but under more favourable conditions.

In Appendix A of Hilbert et al. (2011) it is shown that this estimator is optimal in the sense that no other magnification estimator based on the estimated convergence $\kappa_{\rm est}$ yields a smaller dispersion in the residual magnification. For instance, the residual magnification dispersion $\sigma_{\mu_{\rm res}}$ for the best estimator is never larger than the dispersion in the uncorrected case.

The conditional distributions $p(\mu_{\rm res}|\kappa_{\rm est})$[11] characterizes the accuracy of the weak-lensing reconstruction and is equivalent to $p(\mu - \mu_{\rm est}|\,\mathbf{d}_{\rm lens}, z_{\rm s})$ in Sec. 2.2. The redshift dependence does not show explicitly in $p(\mu_{\rm res}|\kappa_{\rm est})$ for simplicity of notation. To evaluate

[10] Here we assume the weak-lensing limit, where the lensing convergence, shear and rotation are small.

[11] $p(\mu_{\rm res}|\kappa_{\rm est})$ and $p(\mu|\kappa_{\rm est})$ only deviate in their mean and have the same standard deviation $\sigma_{\mu_{\rm res}|\kappa_{\rm est}} = \sigma_{\mu|\kappa_{\rm est}}$ by construction.

the general delensing performance, $p(\mu_{\rm res}|\kappa_{\rm est})$ could be marginalized to obtain the distribution $p(\mu_{\rm res})$ for the specific observation scheme at a particular redshift.

Although the lensing distributions like $p(\mu_{\rm res})$ and $p(\mu)$ are in principle non-Gaussian (Hirata et al. 2010; Shang & Haiman 2011), we approximate them as Gaussian and only focus on the standard deviations $\sigma_{\mu_{\rm res}}$ when discussing the weak-lensing error and delensing performance.

2.5 Simulations

2.5.1 SLICS Catalogs

The SLICS (Scinet LIght Cone Simulations) (Harnois-Déraps & van Waerbeke 2015) were designed as a massive upgrade of the CLONE simulations (Harnois-Deraps et al. 2012). In our work, we used the expanded version of the SLICS suite (Harnois-Deraps et al. 2018), including hybrid mock catalogs that represent future lensing data at the level of LSST (*Large Synoptic Survey Telescope*, now designated the Vera Rubin Observatory) (Ivezić et al. 2019). The most significant advantage of the LSST-like hybrid mock catalogs is that they simultaneously contain lensing and galactic properties.

The SLICS are based on 1025 N-body simulations produced by the high-performance gravity solver CUBEP3M (Harnois-Deraps et al. 2013). The series of N-body simulations were used to construct dark matter halo catalogs, which later served as the galaxy catalogs' skeleton. Then mock galaxy catalogs are produced from Halo Occupation Distribution (HOD) models with only mass dependence. The luminosity of the whole galaxy in the r-band is also given for each galaxy in the catalogs, consistent with the HOD used.

Throughout this work, we neglect the higher-order correction caused by wave optics lensing[12]. The reason is that the mass resolution of SLICS is not adequate to resolve the matter fluctuations with scales sensitive to the wave nature of GWs at *LISA* frequencies (Takahashi 2006; Oguri & Takahashi 2020; Harnois-Déraps & van Waerbeke 2015). In reality, the wave optics effect would slightly suppress the lensing magnification experienced by the GWs detected by *LISA*.

The same set of N-body simulations of dark matter was used to construct shear and convergence maps via the ray-tracing algorithm with the multiple-plane tiling technique. The maps were all flat-sky, 100 deg^2 maps with 7745^2 pixels, computed on specified lens planes. The redshifts of the lens planes $z_{\rm lens}$ and source planes $z_{\rm source}$ are listed in Table 1. Once the galaxy catalogs are given, the lensing information can be linearly interpolated at the galaxy coordinates and redshifts from the lens planes. The interpolation is only done along the redshift direction since the resolution of the lens maps already approaches the limitation in the mass resolution of the N-body simulations. In addition to the shear, the observed ellipticity is also included in the LSST-like hybrid catalogs. The observed ellipticity $\epsilon_{1,2}$ deviates from the true shear $\gamma_{1,2}$ by a Gaussian error with width $\sigma_\gamma = 0.29$ per galaxy, where the error takes both the intrinsic shape noise and measurement errors into account.

All the information about the LSST-like hybrid catalogs (LSST-like catalogs hereafter) related to our work is summarized in Table 2. For details of the SLICS catalogs, please refer to the extended SLICS paper (Harnois-Deraps et al. 2018).

[12] Wave optics lensing refers to the suppression of magnification when the gravitational-wave wavelength matches approximately the Schwarzschild radius of the gravitational lens.

2.5.2 Siren Catalogs

Catalogs of standard sirens are necessary to emulate the realistic delensing outcomes. The construction of these catalogs depends on our understanding of the universe, specifically on the SMBHB merger models. Apart from modelling the sources, the observed GW signals vary with the detector configurations. In this paper, we used the siren catalogs generated by Zhu et al. (2022), which consider all the factors mentioned above and include all the ingredients needed for our purpose.

The siren catalogs in Zhu et al. (2022) began with catalogs of SMBHB mergers. These preliminary catalogs depend on possible formation models of supermassive black holes (SMBHs) including 'popIII', 'Q3d' and 'Q3nod'. Basically, the redshift and mass distributions as well as the average number of mergers depend on the formation models. An introduction to the formation models can be found in Zhu et al. (2022) and they are not our main focus in this paper. Note that for both 'popIII', 'Q3d' and 'Q3nod', the evolution of the SMBHs should be deeply correlated with the evolution of their host galaxies (Ferrarese & Merritt 2000; Ding et al. 2020).

Then the GW waveforms were generated from the IMRPhenomPv2 (Hannam et al. 2014) according to the merger catalogs. After the waveform generation, the siren catalogs were constructed based on the waveforms and multiple configurations for the space-borne GW detectors. The detector configurations included individuals and combinations of potential space-borne GW detectors. Here we concentrate on the detector configuration *TQ+LISA*. Only sirens with redshift $z_s < 3$ and producing GW signals with SNR $\rho > 8$ were presented in the siren catalogs[13].

Each realization of the siren catalogs includes the following information about each siren: the true luminosity distance D_L, the measurement error of the luminosity distance σ_{D_L}, the source redshift z_s, and the total mass of the SMBHB $M_{\rm tot}$. For more details about the siren catalogs, please refer to Zhu et al. (2022).

2.5.3 Combining two simulations

In Section 2.5.2, we pointed out that the evolution of the SMBHs is deeply coupled with the evolution of their host galaxies. Therefore, the siren catalogs should be correlated with the LSST-like galaxy catalogs in a complicated way. However, the catalogs of standard sirens that we are using are based on a statistical description of the formation models and the detector configurations without referring to any specific galaxy catalogs. This freedom enables us to tailor the siren catalogs to the LSST-like hybrid catalogs – i.e. we can choose host galaxies that are appropriate to the sirens as long as their redshift and mass distributions are consistent.

Here we adopt an empirical relationship reported by Ding et al. (2020) to match sirens with their host galaxies. The empirical relationship connects the mass of the central SMBH with the luminosity of its host galaxy in the r-band, shown in Eq. (28):

$$\log\left(\frac{M_{BH}}{10^7 M_\odot}\right) = 0.49 + 0.90\log\left(\frac{L_r}{10^{10} L_\odot}\right) \tag{28}$$

[13] For signals with SNR just below the threshold, lensing might help the signals to be detected by slightly magnifying their waveforms. However, the probability of such events occurring should be small and the errors in the distance estimation should be large even after being magnified. Therefore, we ignore this issue in our analysis. Notice that this is not the case in the strong-lensing regime as the magnification can be very large so that lensing can significantly increase the SNR of the signals.

$z_{\rm lens}$	0.042	0.130	0.221	0.317	0.418	0.525	0.640	0.764	0.897	1.041	1.199	1.373	1.562	1.772	2.007	2.269	2.565	2.899
$z_{\rm source}$	0.086	0.175	0.268	0.366	0.471	0.582	0.701	0.829	0.968	1.118	1.283	1.464	1.664	1.886	2.134	2.412	2.727	3.084

Table 1. The redshift of lens and source planes used for ray-tracing with the Multiple-Lens-Plane technique to construct the weak-lensing maps.

LSST-like hybrid catalogs contain information about each galaxy:				
sky position x, y ; redshift z ; Luminosity in r-band L_r ; convergence κ ; shear $\gamma_{1,2}$; observed ellipticity $\epsilon_{1,2}$				
Total area of the field	The redshift range	Galaxy number density	Resolution of lensing maps	Shear dispersion (single galaxy)
$10 \times 10\,{\rm deg}^2$	$0 - 3$	$28.3\,{\rm arcmin}^{-2}$	$\approx$ 4.6 arcsec	$\sigma_\gamma = 0.29$

Table 2. Relevant information about the LSST-like hybrid catalogs.

In general, there should be a considerable time lag between the merger of galaxies and the merger of their embedded SMBHs (Volonteri et al. 2007). Therefore, we assume that the galaxy has reached its new equilibrium before the inspiraling phase of the SMBHB. Then the properties of the siren are highly correlated to the properties of its host galaxy. Therefore, the above empirical relationship holds with the mass of central SMBH equal to the total mass of the binary.

We interpret the empirical relationship as a loose constraint with respect to the siren catalogs. We start by binning the galaxies into redshift intervals with bin size $\Delta z = 0.1$. Then we implement the empirical relationship to find the most suitable host galaxy for each siren within the redshift bin.

However, many expected host galaxies of the detected sirens would lie below the luminosity threshold for observation associated with the SLICS catalogues. The sirens that produce GWs detectable by *LISA* and *TQ* have a total mass that is approximately independent of redshift, even in the high redshift region. Thus, the expected luminosity of those host galaxies does not change too much. This presents a problem as we go to higher redshifts, where those galaxies would be too dim to be observed. Furthermore, the empirical relationship might not be valid at high redshift in any case. These issues require future study, and for the present work we simply assign the galaxy with the closest redshift as the host for siren where there is no more suitable choice. We will discuss this problem later in Sec. 5.

The more technical problem is that the two suites of simulations (i.e. those underpinning the SLICS catalogs and our siren catalogs) have slightly different sets of cosmological parameters. From Eq. (13) and Eq. (19), it is clear that the weak-lensing maps depend on the cosmological parameter configurations. Since we focus more on the impact of weak gravitational lensing in this paper, for our sirens we adopt a flat ΛCDM model with cosmological parameters taken to be the same as in SLICS. As mentioned above, the redshift and mass distributions are the core assumptions for the statistical behaviors of standard sirens. Hence we recompute the luminosity distance of each siren with its redshift z unchanged but with the cosmological parameters re-tuned to match those of the SLICS configurations.

The cosmological parameters adopted in SLICS relied on the best fit WMAP9 + BAO + SN parameters (Hinshaw et al. 2013): $\Omega_m = 0.2905$, $\Omega_\Lambda = 0.7095$, $\Omega_b = 0.0473$, $h = 0.6898$, $\sigma_8 = 0.826$ and $n_s = 0.969$. This choice lies close to the mid-point between the cosmic shear and the Planck best-fit values in the $[\sigma_8 - \Omega_m]$ plane.

3 RESULTS

3.1 Perfect delensing

We first consider the case of perfect reconstruction, which can be regarded as a hard limit for the more realistic delensing scenarios that we consider later. In this case, the noise-free shear information around the siren at the same redshift z_s is directly accessible, and the information for each line-of-sight is viewed as a sample from the underlying distribution $p(\mu_{\rm res}|\kappa_{\rm est})$.

Fig. 2 shows the outcomes of perfect delensing. The uncorrected case ($\mu_{\rm est} = 1$) has a standard deviation $\sigma_\mu = 0.12$ for standard sirens at redshift $z_s = 2.9$. In contrast, the perfectly corrected case has five times smaller standard deviation for the residual $\sigma_{\mu_{\rm res}} = 0.024$. Note that the dispersion of residual magnification in the perfect case is small but non-vanishing. This discrepancy originates from the interpolation setup in SLICS, where the shear maps are constructed by Fast Fourier Transform from the convergence maps, but the interpolation on the pixels is done at the very end, on the shear and convergence maps all at once (Harnois-Deraps et al. 2012). Therefore, this small error may be interpreted as the contribution from small-scale fluctuations beneath the resolution of the simulation.

Although the residual error inherited from the numerical simulation, $\sigma_{\mu_{\rm sim}}$, is annoying, Fig 3 shows that $\sigma_{\mu_{\rm sim}}$ decreases rapidly with smoothing. In realistic cases, smoothing is inevitable to suppress the noise, and hence the interpolation errors are negligible. Therefore, this error will be ignored hereafter unless otherwise specified.

Even with perfect measurements for individual galaxies, a magnification estimate involves a certain amount of pixelation and smoothing of the shear observation maps, since the inversion algorithm only applies to smooth fields. Also, in realistic cases, the noises in the shape measurements can be reduced by smoothing and averaging. According to Eq. (18), the smoothing on the shear measurements eventually propagates into the constructed convergence fields and thus leads to errors in estimating the magnification. The impact of smoothing on the dispersion of the residual magnification $\sigma_{\mu_{\rm res}}$ at $z = 2.9$ is shown in Fig. 4, where the interpolation error has already been neglected. As shown in Fig. 4, the magnification correction is already degraded substantially (reaching 50% of the uncorrected dispersion) if the convergence map is smoothed with a filter scale[14] $\theta_s \approx 10$ arcsec. This illustrates that it is necessary to reconstruct the convergence on scales of a few arcseconds in order to reduce the

[14] The definition of smoothing filter scale θ_s is given in Eq. (23).

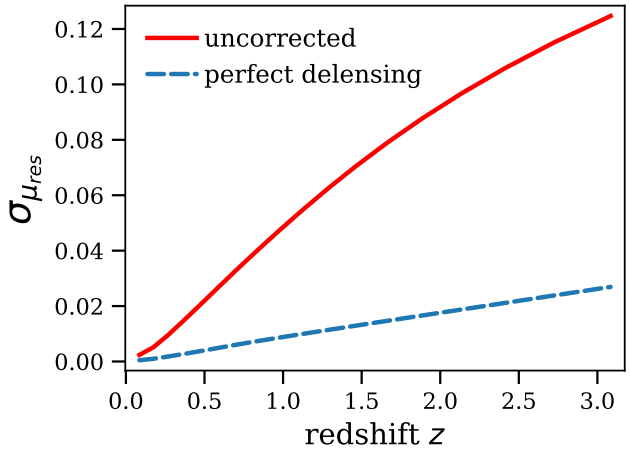

Figure 2. The dispersion of the residual weak-lensing magnification $\sigma_{\mu_{\rm res}}$ for the uncorrected case ($\mu_{\rm est} = 1$, solid line) and the perfectly corrected case as a function of redshift z. The perfectly corrected case has a much smaller but non-vanishing standard deviation for the residual $\sigma_{\mu_{\rm res}}$. The small dispersion is due to interpolation error $\sigma_{\mu_{\rm sim}}$ in SLICS.

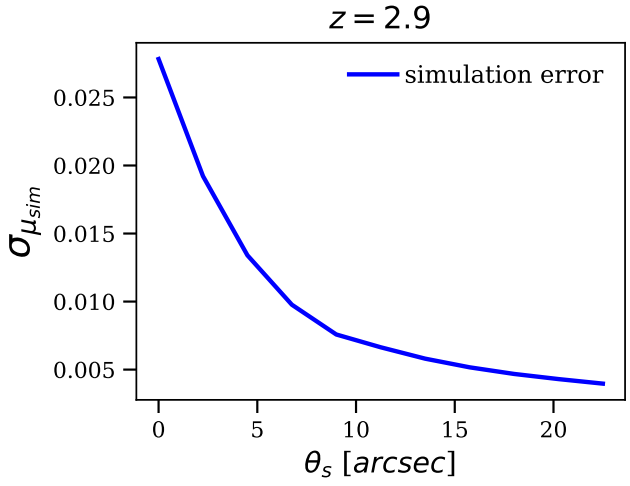

Figure 3. The dispersion due to interpolation uncertainties $\sigma_{\mu_{\rm sim}}$ as a function of the smoothing scale θ_s, for a siren at redshift $z = 2.9$. The $\sigma_{\mu_{\rm sim}}$ decreases rapidly with smoothing and thus can be ignored in realistic cases.

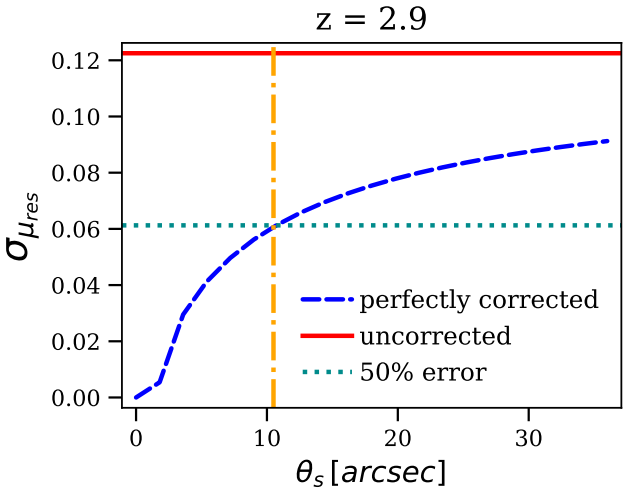

Figure 4. The dispersion of residual magnification $\sigma_{\mu_{\rm res}}$ (dashed line) for sirens at redshift $z_s = 2.9$ as a function of the smoothing filter scale θ_s from a perfect reconstruction. The smoothing wipes away the small-scale fluctuations and hence the non-vanishing $\sigma_{\mu_{\rm res}}$ is contributed from the modes with scale below θ_s. The solid horizontal line marks the magnification dispersion without correction and the smoothing scale corresponding to 50% of the uncorrected dispersion is denoted in the figure by the dotted horizontal line. The interpolation errors $\sigma_{\mu_{\rm sim}}$ have already been eliminated.

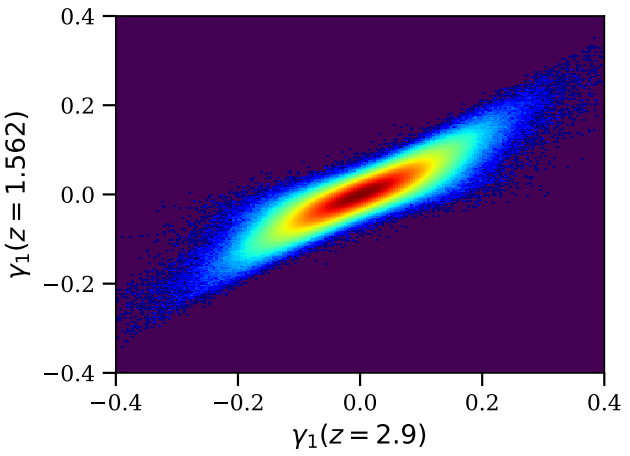

Figure 5. The joint distribution of the shear component γ_1 at redshift $z = 1.562$ and redshift $z_s = 2.9$, at the same sky position. Lighter areas correspond to higher probability densities on a logarithmic scale. The joint distribution of γ at different redshifts along the same line of sight is highly correlated and well approximated by a linear relation. Note that the joint distribution for the second shear components $\gamma_2(z)$ and $\gamma_2(z_s)$ is the same as for the first shear components.

lensing error by a factor of two or better, compared to the uncorrected case.

The perfect case also helps to reduce the error caused by the realistic redshift distribution of galaxies. As pointed out in Eq. (24), if the redshift information of the galaxies is not available, only an effective convergence $\kappa_{\rm eff}(\boldsymbol{\theta})$ can be deduced from the realistic redshift distribution of galaxies, which may deviate substantially from the actual convergence at the sirens $\kappa(\boldsymbol{\theta}, z_{\rm s})$. This error further propagates into the estimation of magnification. If the redshift information, either photometric or spectroscopic, about the galaxies is accessible, the variations of the lensing properties as a function of redshift can be exploited to perform a three-dimensional reconstruction of the estimated magnification maps. However, the correlation across lens planes has been explicitly broken due to the way of construction in SLICS, hence any 3D reconstruction should be performed only inside individual lens sub-volumes. Here, we take a simpler approach that is similar to Hilbert et al. (2011), which does not attempt to construct the 3D matter structures causing the lensing along the line of sight but only exploits the statistical relation between lensing quantities at different discrete redshifts.

As Fig. 5 illustrates, the shear values at two different redshifts along the same line of sight are strongly correlated and the shear at

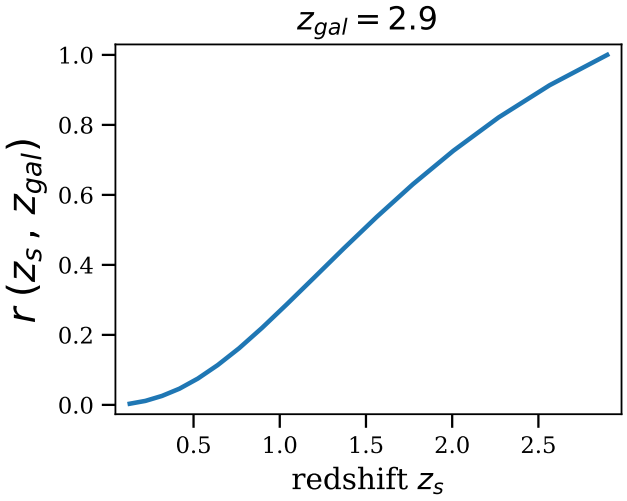

Figure 6. The ratio $r_\gamma(z_s; z_s)$ between the mean shear at redshift z_s of the siren and the shear at redshift $z_{\rm gal} = 2.9$ of the source galaxies. The redshift of the source galaxies is fixed and the $r(z_s, z_{\rm gal})$ is calculated by $\langle\gamma(z_s)\rangle/\langle\gamma(z_{\rm gal})\rangle$ to the shear maps in SLICS.

lower redshift is smaller on average, so there should be some non-vanishing systematic error introduced by including galaxies with redshift far from that of the siren. Also, Fig. 5 shows that the joint distribution of γ at different redshifts along the same line of sight is highly correlated and well approximated by a linear relation. This motivates the expression for the unbiased estimates of the shear at siren's redshift z_s given the shear of the source galaxies with redshift $z_{\rm gal}$,

$$\gamma(z_s)_{|\gamma(z_{\rm gal})} = r(z_s, z_{\rm gal})\gamma(z_{\rm gal}) \tag{29}$$

where $r(z_s, z_{\rm gal}) = \langle\gamma(z_s)\rangle/\langle\gamma(z_{\rm gal})\rangle$ is a factor that depends on both the redshift of the siren and the source galaxies. Then the shape noise of the galaxies at different redshifts is scaled by the ratio $r(z_s, z_{\rm gal})$. Fig. 6 shows the redshift dependence of the ratio factor $r(z_s, z_{\rm gal})$ with a fixed redshift of the source galaxies at $z_{\rm gal} = 2.9$.

Galaxies from different redshift bins could provide separate estimations on the shear at the siren's position with errors that are also different from each other. Prior to smoothing and finite truncation of the observation field, the errors in estimating the shear at the siren's position are composed of two parts: the equivalent shape noise σ_{μ_s} and the scatter around the linear relation σ_{μ_z} given by Eq. (29). The equivalent shape noise σ_{μ_s} is given by,

$$\sigma_{\mu_s} = r(z_s, z_{\rm gal})\frac{\sigma_\gamma}{\sqrt{n_{\rm gal}(z_{\rm gal})}} \tag{30}$$

where σ_γ is the effective shape noise for one single galaxy and $n_{\rm gal}$ is the average number of galaxies within one pixel[15]. According to Eq. (29), σ_{μ_s} is sensitive to both the galaxy density $n_{\rm gal}(z)$ and ratio factor $r(z_s, z_{\rm gal})$, leading to different values for different redshift intervals. The scatter around the linear relation σ_{μ_z} also has a distinct value for different redshifts, and it has spatial correlation prior to smoothing originated from the intrinsic spatial correlation of the shear at a single redshift.

[15] Although the galaxy number varies among pixels at the same redshift, the difference is marginal for the purpose of this paper, especially after smoothing.

Because estimations from different redshift bins have different errors, proper weighting must be applied to optimize the integrated results. In principle, the optimal weighting should depend both on σ_{μ_s} and σ_{μ_z}, but the latter error is affected uniquely by smoothing at different redshifts since the intrinsic spatial correlation varies for different redshifts. Fortunately, it turns out that the latter is much smaller than the equivalent shape noise for the smoothing scales and galaxy number densities considered in this paper. Thus we approximate the redshift weighting by,

$$w_{z,\gamma}(z_s, z_{\rm gal}) = \frac{1}{\sigma^2_{\mu_s}(z_s, z_{\rm gal})} \tag{31}$$

where $\sigma_{\mu_s}(z_s, z_{\rm gal})$ is given by Eq. (30). Since the shape noise does not have spatial correlation prior to smoothing and the magnitude of the shape noise is the same for pixels at the same redshift, the smoothing only adds a fixed proportionality constant in front of the equivalent shape noise $\sigma_{\mu_s}(z_s, z_{\rm gal})$ at every redshift $z_{\rm gal}$, leading to unchanged weighting after normalization. Therefore, the optimal weighting is independent of smoothing if σ_{μ_z} is ignored, which is a valid assumption and simplifies the problem significantly. We will follow this choice of the weighting $w_{z,\gamma}(z_s, z_{\rm gal})$ and ratio factor $r(z_s, z_{\rm gal})$ for the subsequent analyses.

In practice, the weightings should also reflect the image quality of individual galaxies. However, the SLICS catalogs do not emulate the observational details of weak-lensing measurements. Therefore, we neglect technical observation caveats such as the neighbour-exclusion bias (Harnois-Deraps et al. 2018) as well as the magnification and size bias (Liu et al. 2014), leaving them to future studies.

3.2 Wide-field delensing

As introduced in Sec. 2.5.1, the LSST-like galaxy catalogs can serve well as a test of the delensing procedure using wide-field surveys in the near future. The conditional distribution $p(\mu_{\rm res}|\kappa_{\rm est})$ used to estimate the siren's magnification can be sampled by the simulated galaxies in the LSST-like galaxy catalogs because it is assumed that each siren has a host galaxy (Sec. 2.5.2). In principle, galaxies inside the catalogs should be weighted by factors during the construction of $p(\mu_{\rm res}|\kappa_{\rm est})$ depending on the properties of the targeted siren since the properties of the siren are expected to be highly correlated to the properties of its host galaxy (Sec. 2.5.2 and Sec. 2.5.3). These weightings can in principle be used to study the selection effects on sirens with different properties. However, here we only adopt an equal weighting for all galaxies and sirens within the same redshift bin, leaving the discussion on selection effects until the end of Sec. 4 and middle of Sec. 5.

The key features related to the reconstruction behaviours are listed in Table 2. Additionally, the redshift dependence of the galaxy number per arcmin^2 is shown in Fig. 7. To emulate what happens in practice, the galaxies in the LSST-like catalogs are divided into 17 redshift bins according to their true redshift. The choice of the bin number and bin edges are consistent with the SLICS lensing simulation setup (Harnois-Deraps et al. 2018), where the bin edges are placed at the redshifts of the source planes $z_{\rm source}$ in Table 1.

In more realistic situations, only the photometric redshifts of the galaxies are accessible and thus it would be possible to have the galaxies categorized into the wrong redshift bins. Here we ignore this effect and do not explore the optimal choice of the redshift bins. The exploration of the optimal choice of redshift bins should rely on some weak-lensing simulations that do not have the same set of special redshifts for all realizations (such as different sets of $z_{\rm source}$ and $z_{\rm lens}$) and treat the weak-lensing properties along redshift more seriously, i.e. not just linear interpolation.

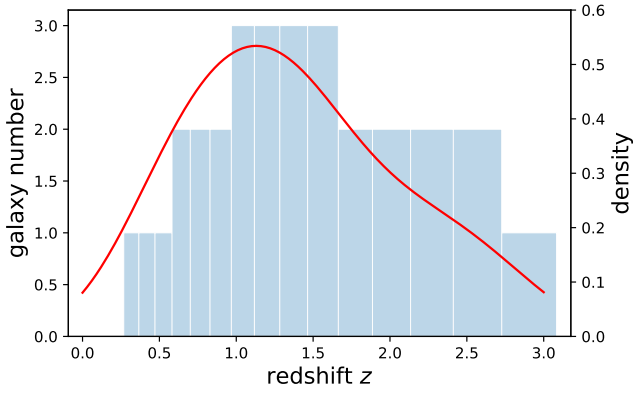

Figure 7. The galaxy distribution $n_{\rm gal}$ along redshift z for the LSST-like catalogs. The galaxy number denotes the total number of galaxies within one square arcminute and redshift bin. The red curve shows the corresponding density distribution function. The peak of the distribution is between $z = 1$ and $z = 1.5$.

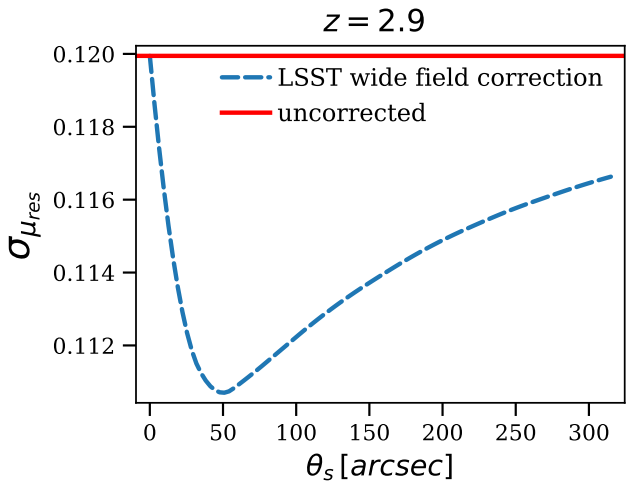

Figure 8. The dispersion of residual magnification $\sigma_{\mu_{\rm res}}$ (dashed line) for sirens at redshift $z_s = 2.9$ as a function of the smoothing filter scale θ_s for an LSST-like wide-field survey. The solid horizontal line marks the magnification dispersion without correction. The optimal smoothing filter scale is around 50 arcsecond with a reduction of the residual dispersion being less than 10%.

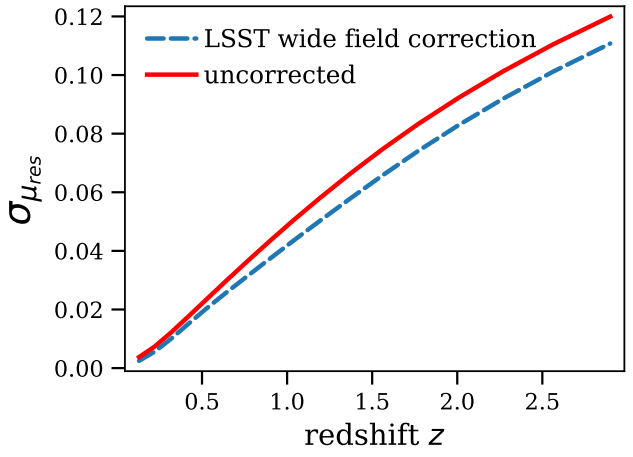

Figure 9. The dispersion of residual magnification $\sigma_{\mu_{\rm res}}$ (dashed line) as a function of the siren's redshift z. The dispersion without correction ($\mu_{\rm est} = 1$, solid line) is compared to the residual dispersion with convergence reconstructed from an LSST-like wide-field survey under the optimal filter scale (dashed line). The weak-lensing errors are only reduced by about 10% at all redshifts.

In another more realistic scenario, the observed ellipticities of the galaxies in the mock catalogs are noisy due to measurement error and shape noise. Thus, certain smoothing must be applied to reduce the noise level. Fig. 8 quantifies how much the dispersion of the residual magnification $\sigma_{\mu_{\rm res}}$ for a standard siren at $z_s = 2.9$ can be reduced with an LSST-like futuristic wide-field survey for particular choices of the filter scale θ_s. With the optimal filter scale $\theta_s \approx 50$ arcsec, Fig. 9 then shows how $\sigma_{\mu_{\rm res}}$ changes as a function of the siren's redshift z_s. As Fig. 9 illustrates, the weak-lensing errors are only reduced by about 10% compared to the uncorrected case at all redshifts. There is little benefit from delensing using an LSST-like wide-field survey.

Since the total area of the field is large enough ($10 \times 10\,{\rm deg}^2$), the mass-sheet degeneracy problem is negligible. The poor performance of the LSST-like wide-field surveys mainly originates from the inadequate number density of the galaxies. Since the equivalent shape noise σ_{μ_s} within one pixel depends on the galaxy density and can only be suppressed by smoothing, an insufficiently high galaxy density will lead to a large filter scale θ_s to obtain an acceptable signal-to-noise ratio for our magnification estimate. The large filter scale smooths away the small-scale contributions to the magnification, resulting in very poor performance of delensing. To significantly improve the delensing result, we therefore must have a deep and dedicated survey that points to the targeted siren.

3.3 Golden siren selection

The end of Sec. 3.2 demonstrates that an LSST-like futuristic wide-field survey is not sufficient to reduce the weak-lensing error significantly. Therefore, we must turn to a deep-field survey with sufficient resolution and depth to obtain images with adequate galaxy number density. Because such a deep-field survey is very expensive in practice, it may only be feasible to conduct the deep-field survey around one specific siren.

We refer to the best siren inside a catalog for carrying out the delensing process as a 'golden siren'[16]. The siren is chosen based on how well the delensing process can improve the accuracy of parameter estimation, after reducing the dispersion of the residual magnification for the siren. The performance of the siren is affected by the measurement error on its estimated luminosity distance and the redshift of the siren. With a smaller measurement error, the siren can better constrain the cosmological parameters. On the other hand, given the same measurement errors, a siren can generally better constrain the cosmological parameters when the siren is further away, according to Eq. (2). Therefore, we only consider sirens with $z_s > 1$

[16] This terminology has already been introduced in the literature (Nissanke et al. 2013) to refer to a particularly well-localized, nearby bright siren, and we adapt the term to the current context.

as candidates for being the 'golden siren'. We note, however, that the siren with the highest redshift may not have the smallest measurement errors. The trade-off between smaller measurement errors and higher redshift is complicated. We attempt to reweight the parameter estimation results using the likelihood of individual sirens with reduced $\sigma_{\mu_{\rm res}}$ to select the golden siren from each of our siren catalogs, and investigate the improvements in cosmological parameter estimation for different $\sigma_{\mu_{\rm res}}$.

The likelihood of parameter estimation after correction can be expanded as

$$p(d_1 \cdots d_i^{\text{reduced } \sigma_{\mu_{\rm res}}} \cdots d_n|\vec{\Omega}) \propto p(d_1 \cdots d_n|\vec{\Omega}) \times \frac{p(d_i^{\text{reduced } \sigma_{\mu_{\rm res}}}|\vec{\Omega})}{p(d_i|\vec{\Omega})}, \quad (32)$$

where $p(d_i^{\text{reduced } \sigma_{\mu_{\rm res}}}|\vec{\Omega})/p(d_i|\vec{\Omega})$ is the factor used to reweight the estimation results. Here d_i is the GW data of the chosen siren, and $d_{j\neq i}$ is the GW data of other sirens. $\vec{\Omega}$ is the vector of cosmological parameters and other information such as $\mathbf{z}_{\rm S}$ is not shown explicitly.

By comparing the contour area in the posterior of the cosmological parameters, one can evaluate quantitatively how the reduction of $\sigma_{\mu_{\rm res}}$ for the chosen siren improves the cosmological parameter estimation.

Fig. 10 shows examples of the credible regions obtained for the cosmological parameters[17] before (labelled 'Uncorrected', shown in the right panels) and after (labelled 'Golden siren', shown in the left panels) delensing of the golden siren identified in each mock siren catalog. Example results are shown for each of the three formation models considered in our analyses: popIII (top), Q3d (mid), Q3nod (bottom). In each case the residual dispersion is fixed to $\sigma_{\mu_{\rm res}} = 5\%$, i.e. a factor of two lower than the uncorrected $\sigma_\mu \approx 10\%$. These plots show that reducing $\sigma_{\mu_{\rm res}}$ for the golden siren can not reduce the contour area of the posterior of the cosmological parameters appreciably, as the change is marginal and insignificant. Moreover we note, again, that these results are for the case where the Hubble constant is fixed to equal its true value – i.e. this parameter is considered to be already known (or very tightly constrained) from other observations. Even in this idealised case, the delensing of the golden siren does not result in any significant reduction in the area of the credible region for the other cosmological parameters. In practice, the uncertainty in H_0 can be considered as an extra error on the estimated luminosity distance and the effect of delensing is more insignificant.

Fig. 11 shows how the area of the credible region contour inferred for the cosmological parameters changes with increasing dispersion of the residual weak-lensing magnification. The left, middle and right plots are for the popIII, Q3d and Q3nod models respectively. The overall increasing trend of the contour area as the dispersion increases is clear, although there are some random Poissonian errors from the sampling processes for high residual dispersions. However, even after delensing the golden siren, the contour area is comparable in size to the uncorrected case when $\sigma_{\mu_{\rm res}} \approx 5\%$, and shows a significant (factor of two or greater) reduction in size only if $\sigma_{\mu_{\rm res}} \leq 2\%$. The feasibility of reducing $\sigma_{\mu_{\rm res}}$ to such a level using convergence maps reconstructed from deep-field surveys is evaluated quantitatively in the following sections.

[17] We show here contours for the dimensionless density parameters only, in the idealised case where the value of the Hubble constant is considered to be known already.

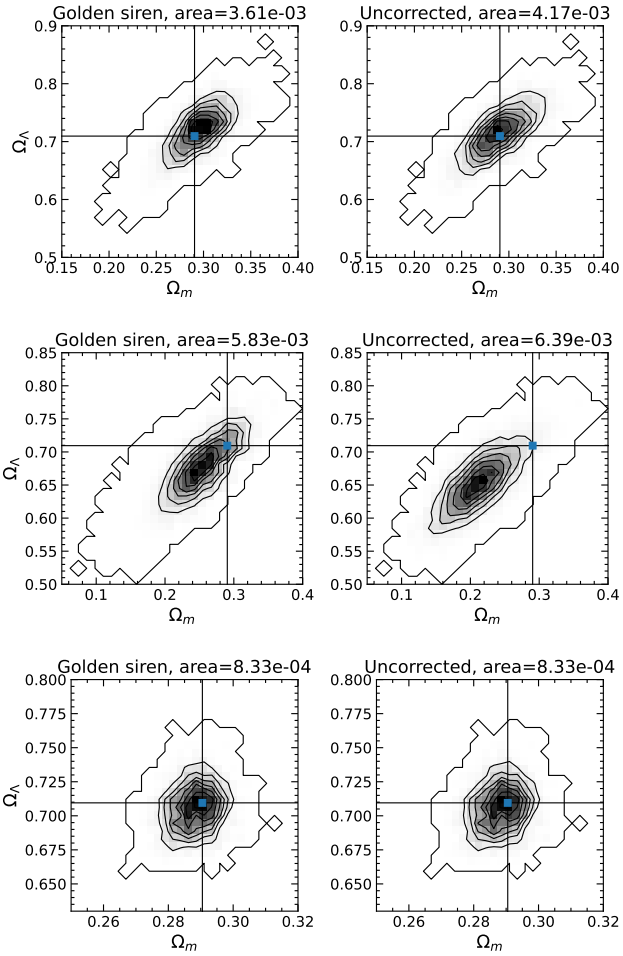

Figure 10. Examples of credible region contours for the inferred cosmological parameters before (labelled 'Uncorrected', right panels) and after (labelled 'Golden siren', left panels) delensing of the golden siren identified in each mock siren catalog. Examples are shown for each of the three formation models considered: popIII (top), Q3d (mid), Q3nod (bottom). The residual dispersion is fixed to $\sigma_{\mu_{\rm res}} = 5\%$, i.e. a factor of two lower than the uncorrected $\sigma_\mu \approx 10\%$. The blue square shows the true values of the cosmological parameters. The change in the contour areas of the credible regions is marginal and insignificant.

3.4 Deep-field delensing

3.4.1 Construction of deep-field survey

The SLICS simulations do not contain deep-field mock galaxy catalogs, so the construction of such catalogs proved necessary. To emulate state-of-the-art technologies, here we adopt a configuration similar to the *James Webb Space Telescope* (JWST) ultra-deep field (UDF) for the construction of the deep-field survey.

The observational setup and the redshift distribution of galaxies that we adopt follow the simulated JWST UDF catalogs (Yung et al. 2022). Since the satellites might be too dim to perform accurate measurements of their shapes, only the field galaxies within the catalogs are counted towards the galaxy number density. Fig. 12 shows how the galaxy density (i.e. galaxy number within one arcmin^2 and redshift bin) changes with redshift in our simulated deep-field survey. Note that the measurement errors in galaxy shape for the JWST UDF are complicated and small compared to the shape noise.

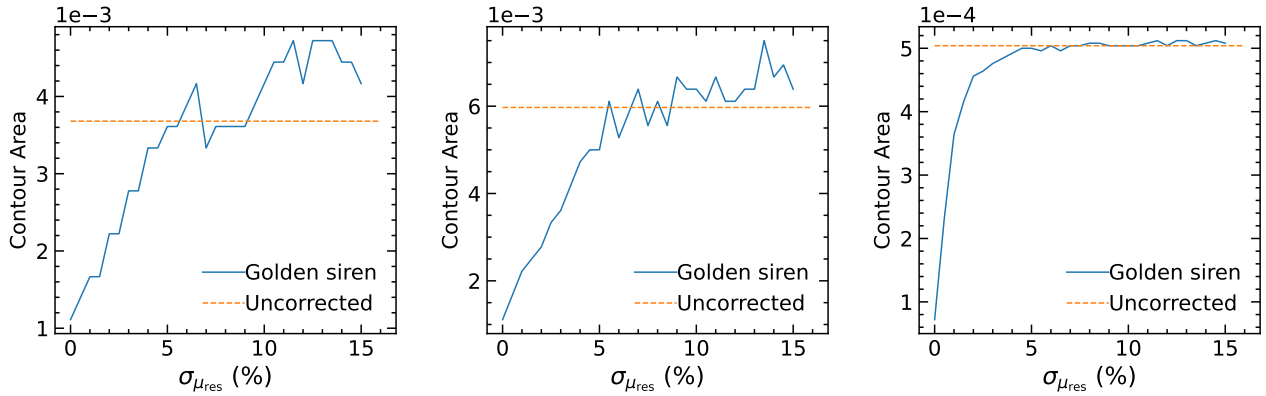

Figure 11. The area of the credible region contour as a function of the dispersion of residual magnification $\sigma_{\mu_{\rm res}}$ for the delensed golden siren under formation models popIII (left), Q3d (mid) and Q3nod (right). The horizontal dotted line denotes the area in the uncorrected case. The plots show the increasing trend of the contour area with growing $\sigma_{\mu_{\rm res}}$. Random Poissonian errors from the sampling processes are responsible for the fluctuations at high $\sigma_{\mu_{\rm res}}$. To improve the cosmological parameter estimation significantly (factor of two or greater reduction), the residual dispersion $\sigma_{\mu_{\rm res}}$ for the golden siren after delensing should be less than 2%.

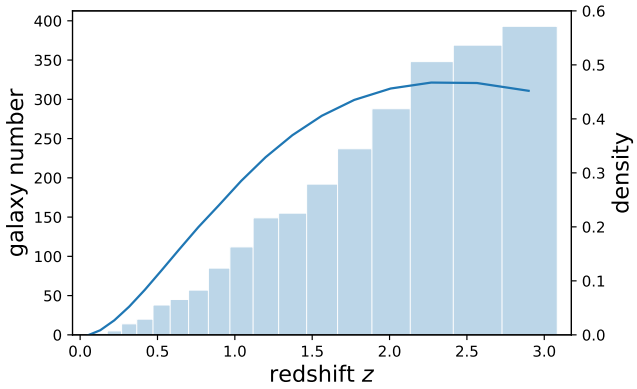

Figure 12. The galaxy distribution $n_{\rm gal}$ as a function of redshift z for the JWST-like galaxy catalogs used in our study. The galaxy number denotes the total number of galaxies within one square arcminute and redshift bin. The blue curve shows the corresponding density distribution function. There is a sudden truncation at $z = 3$, aligning with the maximal redshift for lensing maps in SLICS.

For simplicity, therefore, only the shape noise will be counted in determining the shear dispersion per galaxy, i.e. $\sigma_\gamma = \sigma_{\rm shape} = 0.26$ (Schaan et al. 2017). Other information about the simulated deep-field survey is summarized in Table 3. Note that the underlying weak-lensing maps are the same as in the LSST-like wide-field survey.

To make a consistent construction of the deep-field survey with respect to the lensing maps, we produced deep-field mock galaxy catalogs so that the positions of the galaxies trace the underlying dark matter distribution with a fixed bias. We do this by utilizing the projected 2D density contrast $\delta_{2D}(\boldsymbol{\theta}, z)$ defined by Eq. (12) such that the density distribution of galaxies in each redshift bin is proportional to the mass distribution projected within the bin (Harnois-Deraps et al. 2018). The proportionality constant is the linear galaxy bias which is assumed to be $b = 1.18$, the same as the bias in the KiDS-HOD mocks included in SLICS.[18] The redshifts of the galaxies within a bin follow a random uniform distribution for each line-of-sight.

This construction has the advantage that both the lensing maps and deep-field galaxy catalog arise from the same set of density contrasts $\delta_{2D}(\boldsymbol{\theta}, z)$. However, the bias here is a fixed parameter instead of being redshift, scale and mass dependent. The linear bias model might also not be sufficient for explaining the small-scale galaxy/matter distribution. Future studies are necessary to investigate these effects.

The augmented galaxies added to our deep-field catalog could be interpreted as galaxies too faint to be observed in a wide-field survey. However, since the approach to simulate the augmented galaxies is too simplistic, these galaxies are not considered as the potential host galaxies for SMBHBs and hence not viewed as samples towards the distributions $p(\mu_{\rm res}|\kappa_{\rm est})$. This issue will be further discussed in Sec. 5.

3.4.2 *Deep-field delensing results*

The high number density of galaxies strongly suppresses the shape noise and only a small smoothing scale is required to obtain an acceptable signal-to-noise ratio. This is very beneficial for extracting small-scale signals in reconstruction and thus allows one to substantially reduce the residual dispersion $\sigma_{\mu_{\rm res}}$ for standard sirens. In Sec. 3.3, the golden siren is assumed to have redshift $z_s > 1$ and hence the distribution $p(\mu_{\rm res}|\kappa_{\rm est})$ will be investigated at $z > 1$ only.

As Fig. 13 shows, the residual dispersion $\sigma_{\mu_{\rm res}}$ can be reduced by delensing to around 60% of the uncorrected dispersion at $z_s = 2.9$ with the optimal filter scale $\theta_s \approx 8$ arcsec. Under the best filter scale $\theta_s \approx 8$ arcsec, the redshift dependence of the residual dispersion $\sigma_{\mu_{\rm res}}$ is shown in Fig. 14. Obviously, the delensing outcomes are significantly better than the LSST-like wide-field surveys.

Apart from smoothing, which wipes away the small-scale fluctuations, the mass-sheet degeneracy originating from the limited size of the deep survey also contributes to the residual dispersion $\sigma_{\mu_{\rm res}}$. The size of the JWST UDF is much larger than the Hubble Space Telescope UDF, but it is still not sufficient for the mass-sheet degeneracy

[18] Both the LSST-like mock catalogs and the KiDS-HOD mocks are straightforward extensions of a single HOD model.

Total area of the field	The redshift range	Galaxy number density	Resolution of lensing maps	Shear dispersion (single galaxy)
11×11 arcmin^2	$0 - 3$	$\approx$ 2500 arcmin^{-2}	$\approx$ 4.6 arcsec	$\sigma_\gamma = 0.26$

Table 3. Relevant information about the JWST-like deep-field catalog used in this study.

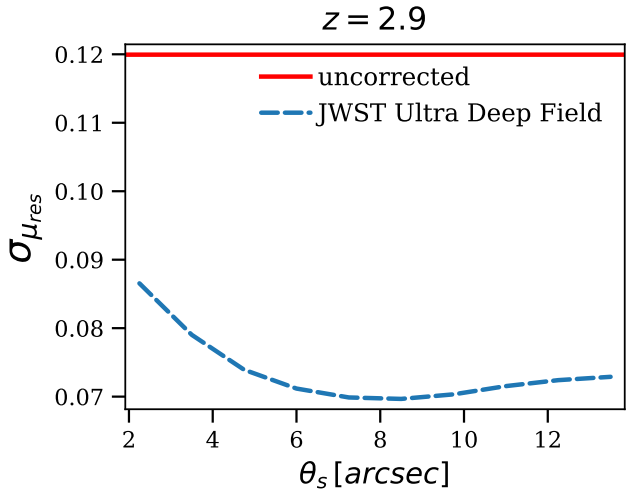

Figure 13. The dispersion of residual magnification $\sigma_{\mu_{\rm res}}$ (dashed line) for sirens at redshift $z_s = 2.9$ as a function of the smoothing filter scale θ_s from a JWST-like deep-field survey. The solid horizontal line marks the magnification dispersion without correction. The optimal smoothing filter scale is around 8 arcsecond, resulting in a 42% reduction of the residual dispersion.

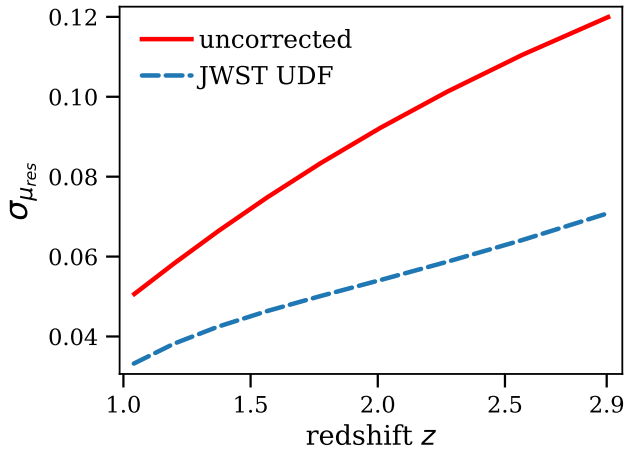

Figure 14. The dispersion of residual magnification $\sigma_{\mu_{\rm res}}$ (dashed line) as a function of the siren's redshift z. The dispersion without correction ($\mu_{\rm est} = 1$, solid line) is compared to the residual dispersion with convergence reconstructed from a JWST-like deep-field survey under the optimal filter scale (dashed line). The weak-lensing errors are reduced by around 30 $\sim$ 40% at all redshifts, significantly better than the outcomes from LSST-like wide-field surveys.

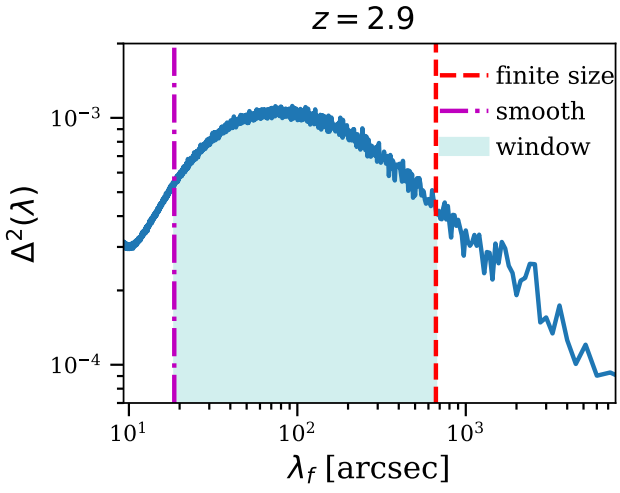

Figure 15. The logarithmic band power of the convergence field at $z = 2.9$. The area under the curve at different intervals of the wavenumbers λ_f for the fluctuation modes quantifies their contribution to the total variance. The fluctuation modes with wavenumbers $\lambda_f \approx 80$ arcsec contribute the most. The detection window for the fluctuation modes limited by smoothing (lower limit) and finite size of the field (upper limit) is also denoted. Note that the log scale in the x axis has been compensated so that the proportion of the area under the curve is the same as in linear scale. However, no manipulations have been carried out on the y axis.

problem to be negligible. Basically, the smoothing and finite size of the observation field put lower and upper bounds on the detectable scale in convergence fluctuations. Fig. 15 illustrates the logarithmic band power of the convergence field at $z = 2.9$, which illustrates the contribution at different fluctuation scales to the total convergence variance by the area below the curve. The best window of detection is also denoted in Fig. 15.

Although the peak in the logarithmic band power of convergence has been included in the window, signals out of the detection window are not inconsiderable, especially the part from the large-scale fluctuations. Fortunately, as mentioned in Sec. 2.3.3, fluctuations larger than the deep survey area can be partially eliminated by an accompanying wide-field survey. The large-scale fluctuations beyond the size of the deep images can be picked up by a wide-field survey that has an adequate field size. We will present the method and results of this combination in the following section.

3.5 Hybrid observation delensing

Sec. 3.2 demonstrated that LSST-like wide-field surveys are blind to sub-arcminute convergence fluctuations that would require significantly high galaxy densities to probe them. However, they have the potential to detect fluctuations larger than a few tens of arcminutes since wide-field surveys of this type have enough width and the shape noise problem is less serious for these scales of consideration. To take full advantage of the potential of delensing, we can therefore estimate the siren's magnification from a hybrid observation com-

posed of both deep- and wide-survey data, similar to the composite approach considered in Shapiro et al. (2010). The wide survey is dedicated to probing fluctuation modes larger than the size of the deep image and the deep survey is able to recover fine features.

Although the underlying matter fluctuations are non-Gaussian in principle, we assumed that the fluctuation modes larger than the size of the deep survey do not couple to the modes within the deep image. This can be partially justified by the relatively large area of the JWST-like UDF ($11\times11\ arcmin^2$) where the fluctuations beyond the size of the UDF are within an almost linear regime and well approximated by a Gaussian random field. Future studies should investigate this caveat and how it depends on the specific choice of the deep field survey.

In this scenario, we are unconcerned about small-scale fluctuations in the LSST-like wide-field survey. Therefore, the pixel size used in reconstruction for the dedicated wide-field survey should be the same as the field size of the deep survey. Then the number of galaxies within one pixel increases dramatically, making the shape noise less severe.

However, the galaxy density may not be sufficient to obtain an acceptable signal-to-noise ratio even under this chosen pixelation and certain smoothing can be helpful to reduce the noise level. To quantify the ability of a wide-field survey for removing the mass-sheet degeneracy of the deep survey, we defined the residual mass-sheet degeneracy as,

$$\mu_{\rm res(ms)} = \mu_{\rm ms} - \mu_{\rm est(ms)} \tag{33}$$

where $\mu_{\rm ms}$ is the part of magnification contributed from the fluctuation modes larger than the deep survey size and $\mu_{\rm est(ms)}$ is the estimated value for $\mu_{\rm ms}$ from the wide-field survey.

Fig. 16 quantifies how the dispersion of the residual mass-sheet degeneracy $\sigma_{\mu_{\rm res(ms)}}$ changes as a function of the filter scale θ_s for a siren located at $z = 2.9$. It illustrates that the mass-sheet degeneracy uncertainty in a JWST-like deep-field survey can be reduced to $\sim 60\%$ with the help of an LSST-like wide-field survey.

With the optimal filter scale θ_s, Fig. 17 illustrates the dispersion of the residual magnification $\sigma_{\mu_{\rm res}}$ as a function of the siren's redshift z_s under different observation schemes. The weak-lensing errors can be reduced by half for sirens at $z_s = 2.9$ on average under a futuristic hybrid observation (JWST-like + LSST-like surveys). A similar but smaller reduction is achievable for sirens in the range $1 < z < 3$. It is clear from this Figure that the mass-sheet degeneracy problem in a JWST-like deep-field survey is not dominant, since the total elimination of mass-sheet degeneracy by an (artificial) perfect wide-field survey does not improve the results significantly, and an LSST-like wide-field survey already produces results close to the perfect case and hence is good enough for breaking the mass-sheet degeneracy in a JWST-like deep-field survey. This conclusion may change if the total area of the adopted deep-field survey deviates from the JWST UDF configuration, and the galaxy density of the chosen wide-field survey is also important to this conclusion.

In a Bayesian approach, one might prefer to use the individual probability distributions $p(\mu_{\rm res}|\kappa_{\rm est})$ of the residuals $\mu_{\rm res}$ given the individual estimate of convergence $\kappa_{\rm est}$ rather than using a common residual distribution $p(\mu_{\rm res})$. The probability distribution of estimated convergence $p(\kappa_{\rm est})$ at $z = 2.9$ is represented in Fig. 18. It is clear that the median is less than zero, i.e. it is more probable to obtain an estimated convergence less than zero. Fig. 19 quantifies the value of $\sigma_{\mu_{\rm res}|\kappa_{\rm est}}$ as a function of the estimated convergence $\kappa_{\rm est}$ from different observation schemes at $z = 2.9$. As Fig. 19 depicts, low estimated convergence corresponds to small residual dispersion, which appears more encouraging than the marginalized residual dispersion. The vertical dashed line indicates the median of $\kappa_{\rm est}$, which

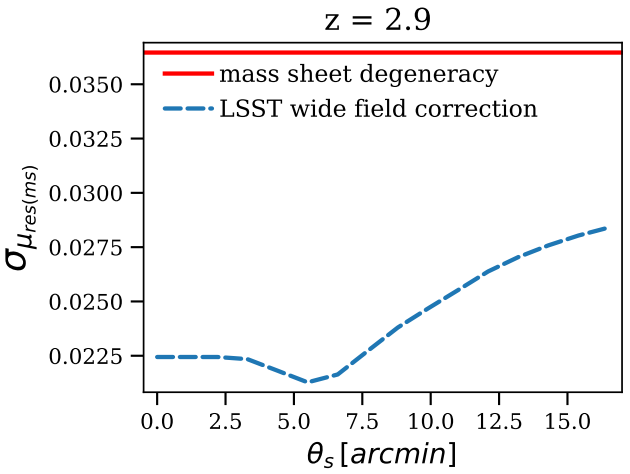

Figure 16. The dispersion of the residual mass-sheet degeneracy $\sigma_{\mu_{\rm res(ms)}}$ (dashed line) for sirens at redshift $z_s = 2.9$ as a function of the smoothing filter scale θ_s from an LSST-like wide-field survey. The solid horizontal line marks the residual mass-sheet degeneracy without correction. The optimal smoothing filter scale is ≈ 5.5 arcminute with a $\approx 40\%$ reduction.

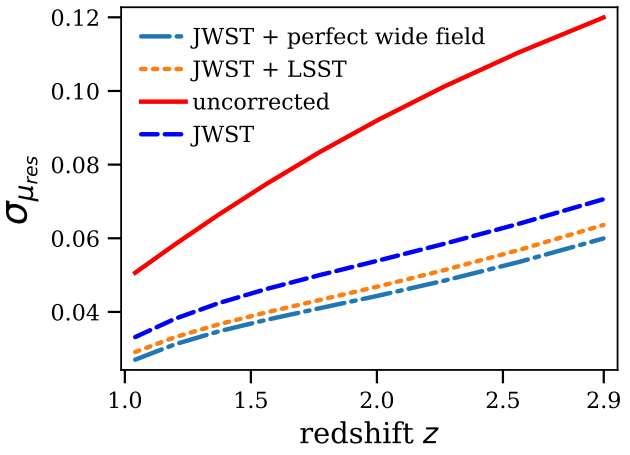

Figure 17. The dispersion of residual magnification $\sigma_{\mu_{\rm res}}$ as a function of the siren's redshift z under different observational schemes. Solid: the dispersion without correction ($\mu_{\rm est} = 1$). Dashed: the residual dispersion with convergence reconstructed from a JWST-like deep-field survey. Dotted: the residual dispersion with convergence reconstructed from a hybrid observation consisting of a JWST-like deep-field survey and an LSST-like wide-field survey. Dot-dashed: same as the dashed case but free from the mass-sheet degeneracy problem. The filter scale in all cases is tuned to be optimal.

is the same as the dashed line in Fig. 18. Therefore, under a futuristic hybrid observation scenario, there is a 50% probability that the weak-lensing error for a siren at $z = 2.9$ can at least be reduced by half and the reduction can reach $\sim 65\%$ in the most fortuitous case.

4 DISCUSSION

In Sec. 3.5, the main conclusion presented is that, even for sirens at $z_s = 2.9$, the weak-lensing errors can be reduced by about a factor of two on average using convergence maps reconstructed from galaxy

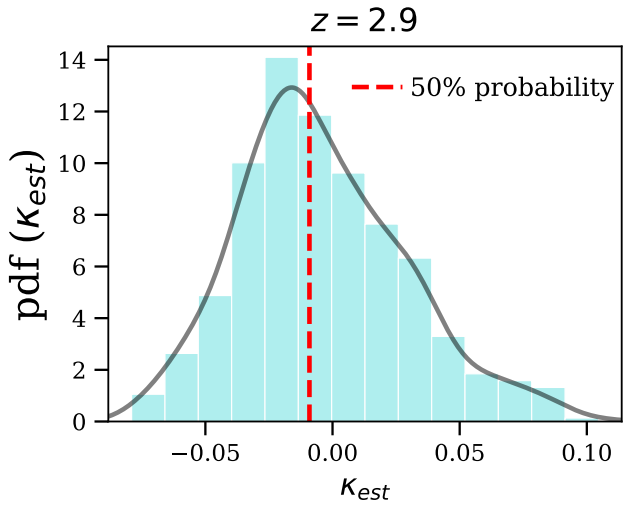

Figure 18. The probability distribution of estimated convergence $\kappa_{\rm est}$ at $z = 2.9$. The black solid curve shows the corresponding probability density function. The red dashed line denotes the median of the estimated convergence which is clearly less than zero.

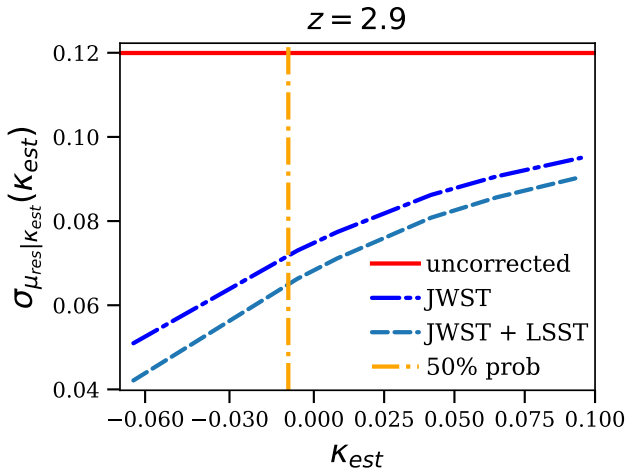

Figure 19. The dispersion $\sigma_{\mu_{\rm res}|\kappa_{\rm est}}(\kappa_{\rm est})$ of the conditional distribution $p(\mu_{\rm res}|\kappa_{\rm est})$ of the residual magnification $\mu_{\rm res} = \mu - \mu_{\rm est}$ as a function of the estimated convergence $\kappa_{\rm est}$ at redshift $z = 2.9$. Different observation schemes are considered and the vertical dashed line is the same as in Fig. 18. The solid red line indicates the dispersion of the uncorrected magnification. There is a 50% probability that the weak-lensing error for a siren at $z_s = 2.9$ can at least be reduced by half and the reduction can reach ~ 65% in the most favourable case.

shape information, provided we have access to future hybrid observations that combine wide- and deep-field surveys. Similar results were obtained in Shapiro et al. (2010) but with a much lower expectation on the depth of the deep-field survey. The deep-field survey adopted in Shapiro et al. (2010) has galaxy density $n_{\rm gal} = 1000\,{\rm arcmin}^{-2}$ while the JWST-like UDF adopted in our work assumes a galaxy density $n_{\rm gal} \approx 2500\,{\rm arcmin}^{-2}$. Moreover, Shapiro et al. (2010) only assumed 2D galaxy shape maps while we make use of the redshift information for individual galaxies. The reason why Shapiro et al.

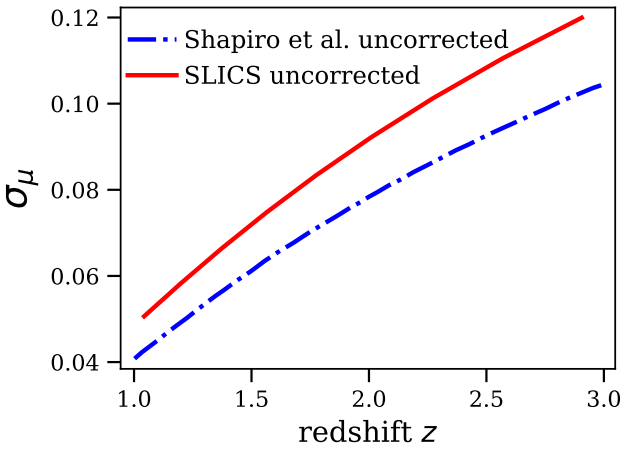

Figure 20. Comparison between the uncorrected magnification in Shapiro et al. (2010) and in SLICS (Harnois-Deraps et al. 2018). The uncorrected magnification is larger in SLICS at all redshifts, indicating that Shapiro et al. (2010) underestimated the small-scale fluctuations according to their assumptions.

(2010) could achieve essentially an equivalent result but with simpler requirements is that some assumptions made in that work turn out to be invalid.

First, Shapiro et al. (2010) did not make use of any numerical N-body simulations to obtain the matter fluctuation distribution, instead extending the Smith et al. (2003) fitting formula for the matter power spectrum beyond its accepted range of validity and assuming that matter fluctuations remain Gaussian even at the smallest scales. This approximation should be tested by numerical N-body simulations like SLICS, which compute the non-linear evolution of dark matter under gravity and thus resolve the structure formation deep in the non-linear regime.

As Fig. 20 shows, the uncorrected magnification in Shapiro et al. (2010) is smaller at all redshifts than the uncorrected magnification in SLICS. This indicates that the approximation and extension made in Shapiro et al. (2010) appears to underestimate the matter power spectrum, especially at the small scales where the fitting formula is no longer valid. This also appears to render the shape noise problem less serious than it is in reality. Thus, the galaxy density required to achieve a factor of two reduction of the weak-lensing error appeared to be lower than it is found to be in our case.

Shapiro et al. (2010) also underestimated the shear dispersion for a single galaxy σ_γ. They assumed $\sigma_\gamma = 0.2$, which incorporates intrinsic shape dispersion, background noise from the sky and the camera, and noise due to an imperfect deconvolution of the PSF. As the paradigm has shifted over the past decade, now the consensus is that the intrinsic shape dispersion for each galaxy is up to $\sigma_{\rm shape} = 0.26$, and higher for σ_γ since other noises are included (Schaan et al. 2017).

Finally, Shapiro et al. (2010) made use of the first weak gravitational flexion $\mathcal{F}$ and second weak gravitational flexion $\mathcal{G}$ maps to reconstruct the convergence maps. The weak gravitational flexions are derived from the third derivative of the lensing potential $\psi(\boldsymbol{\theta})$ and hence they are more sensitive to small-scale fluctuations than is shear. Including flexion allows one to supplement the shear measurements when reconstructing the small-scale modes.

However, in practice, only the reduced flexion fields F and G are directly observable and measurements of the second reduced

flexion G are extremely delicate and generally dominated by noise (Rowe et al. 2013). Furthermore, the reduced flexion dispersion per galaxy is underestimated in Shapiro et al. (2010). The first flexion dispersion per galaxy is assumed to be $\sigma_{\mathcal{F}} = 0.5/arcmin$ while a later paper demonstrated that $\sigma_F = 1.56/arcmin$ [19] inferred from the outcomes of *Hubble's Advanced Camera for Surveys* with realistic galaxy morphology (Rowe et al. 2013). Other works based on the *Hubble Space Telescope* (HST) including Cain et al. (2011) and Lanusse et al. (2016) reported or adopted the same value for σ_F.

In principle, σ_F should be different in distinct deep-field surveys due to the unique PSF and PSF correction for each survey, but the values are expected to be around the same level. Further studies are needed to confirm this hypothesis. Nonetheless, given that the recent estimate of σ_F for the *Hubble* UDF is already three times larger than the estimate in Shapiro et al. (2010), we expect that the inclusion of flexion maps would not improve the result significantly and hence we have discarded the use of flexion in this analysis. Therefore, Shapiro et al. (2010) overestimated the contribution from flexions $\mathcal{F}$ and $\mathcal{G}$ in weak-lensing reconstruction. All of the reasons above account for why Shapiro et al. (2010) obtained similar delensing outcomes but with a much lower expectation value for the depth and number density of the deep-field survey.

Hilbert et al. (2011) also investigated the accuracy of weak-lensing reconstruction to infer the siren's magnification. They made use of galaxy shear and flexion maps to reconstruct the convergence maps, following the same assumptions for the single galaxy shape dispersion as in Shapiro et al. (2010). Therefore, Hilbert et al. (2011) also underestimated the values for σ_γ, $\sigma_{\mathcal{F}}$ and $\sigma_{\mathcal{G}}$. Their analysis indicated that the weak-lensing error for a siren with $z_s = 3$ can be reduced to 72% if only the shear data is available.

Although Hilbert et al. (2011) underestimated the single galaxy shape dispersion, they obtained worse error reduction results compared to our outcomes mainly because their adopted weak-lensing survey had a much lower galaxy number density, i.e. $n_{\rm gal} = 500\,{\rm arcmin}^{-2}$.

Another possible reason for their worse performance is because of the adopted N-body simulation. The backbone dark matter N-body simulation used to generate the weak-lensing maps in Hilbert et al. (2011) is the Millennium Simulation (MS) by Springel et al. (2005). While the comoving side length of a grid cube in the MS, $L_{\rm box}$, is comparable to SLICS, the MS has a smaller dark matter particle mass m_p and a higher particle number n_p within one comoving cube. However, the difference in m_p and n_p is not significant, so the resolution of MS is only slightly higher. The higher mass resolution in the backbone N-body simulation enables Hilbert et al. (2011) to build weak-lensing maps with higher resolution. Therefore, the resolution of weak-lensing maps in Hilbert et al. (2011) is 3.5 arcsec, slightly higher than the resolution of 4.6 arcsec of the weak-lensing maps in SLICS.

However, the adopted cosmological parameters in the MS are somewhat outdated: $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $h = 0.73$, $n_s = 1$ and $\sigma_8 = 0.9$, which differ from the recent best-fit values[20]. Hilbert et al. (2011) used these parameters and simulated weak-lensing maps via the Multiple-Lens-Plane ray-tracing algorithm. Since the techniques and resolution are similar, the deviations in the weak-lensing maps are expected to be the consequence of differences in cosmology. These will lead to different initial amplitudes and evolution of the matter density fluctuations. Moreover, the comoving distance as a function of redshift will also be affected by cosmology, which plays a vital role in calculating the lensing properties at specific redshifts. Therefore, the cosmology dependence of the weak-lensing maps is rather complicated.

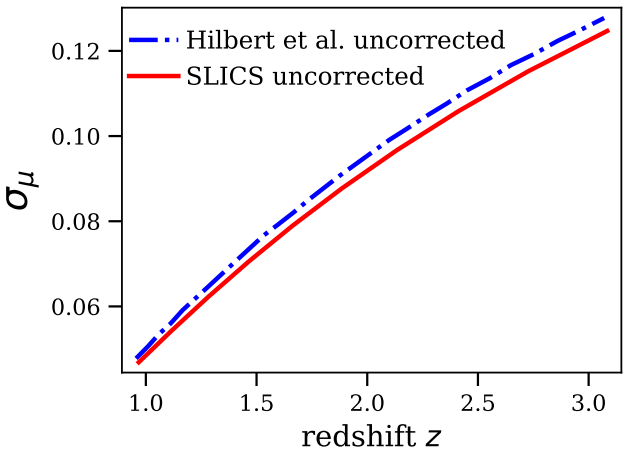

Figure 21. Comparison between the uncorrected magnification in Hilbert et al. (2011) and in SLICS (Harnois-Deraps et al. 2018). The uncorrected magnification in Hilbert et al. (2011) is slightly larger at all redshifts. The small deviation is expected to originate from differences in the cosmological model, as the adopted σ_8 is somewhat larger in Hilbert et al. (2011).

As shown in Fig 21, the uncorrected magnification in Hilbert et al. (2011) is slightly larger than the uncorrected magnification in SLICS at all redshifts, which is expected as the σ_8 is larger in Hilbert et al. (2011). Fig. 22 further demonstrates that the excess of the uncorrected magnification in Hilbert et al. (2011) compared to SLICS is specifically contributed from small-scale fluctuations. Note that the shear measurements are less sensitive to small-scale fluctuations due to shape noise. Therefore, even if Hilbert et al. (2011) could have adopted the same deep-field survey used in our work, we would anticipate that the outcomes would still be worse compared to ours.

Surprisingly, we find that the statistical relations between lensing properties at different redshifts have similar trends regardless of the different cosmologies adopted in the two studies. This can be demonstrated by comparing Fig. 5 and Fig. 6 to the corresponding figures in Hilbert et al. (2011) (Fig. 16 and Fig. 17). Therefore, the correlations between lensing quantities at different redshifts may be insensitive to cosmology, which would be interesting to investigate further in the future.

One more issue in Hilbert et al. (2011) is the lack of simulations of wide-field surveys to break the mass-sheet degeneracy for the simulated deep-field survey. Hilbert et al. (2011) assumed perfect removal of mass-sheet degeneracy by a wide-field survey, and hence the reduction of the weak-lensing errors should be even smaller in practice.

Finally, Shapiro et al. (2010) and Hilbert et al. (2011) studied only the homogeneous galaxy distribution at one specific redshift slice. By contrast, the galaxy catalogs used in our work consider both configurations of the observational program and the underlying matter distribution. Thus, the distribution is generally inhomogeneous at each redshift slice. Therefore, we believe that the shape noise, which is related to the number of galaxies within one pixel, is better modelled in our work.

Furthermore, the selection effects in estimating lensing distribu-

[19] The reduced factor is negligible in the weak-lensing regime.
[20] See the end of Sec. 2.5.3 for the best fit values from current observations.

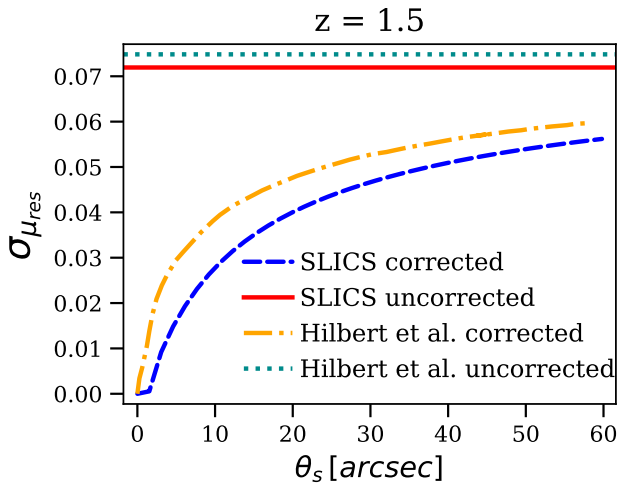

Figure 22. The dispersion of residual magnification $\sigma_{\mu_{\rm res}}$ at redshift $z = 1.5$ as a function of the smoothing filter scale θ_s from a perfect reconstruction in Hilbert et al. (2011) (dot-dashed) and SLICS (Harnois-Deraps et al. 2018) (dashed). Same as in Fig 4, $\sigma_{\mu_{\rm res}}$ is contributed from the modes with scale below θ_s. The horizontal lines mark the magnification dispersion without correction in Hilbert et al. (2011) (dotted) and SLICS (Harnois-Deraps et al. 2018) (solid). The dot-dashed line is higher than the dashed line, especially at small θ_s, indicating that the convergence maps in Hilbert et al. (2011) have larger small-scale fluctuations compared to SLICS.

tions for standard sirens was not investigated carefully in both Shapiro et al. (2010) and Hilbert et al. (2011). The selection effects are induced by the correlation between the siren's total mass and the underlying mass density field since a high-mass siren will tend to be located in the denser part of the density field. Hence the lensing distributions for sirens with different total masses are distinct. However, Shapiro et al. (2010) assumed that every line-of-sight in the observation field can serve as an equal sample towards the distributions of weak-lensing properties for the sirens. Hilbert et al. (2011) made a similar assumption but weighted each line-of-sight by its inverse magnification to account for this selection effect.

To model the selection effect more carefully, we made use of the empirical relation given in Eq. (28) to correlate the total mass of the SMBHB with the luminosity of the host galaxy in the r-band. The luminosities of galaxies in the LSST-like catalogs are simulated by a HOD method and hence trace the underlying matter distribution and lensing fields. By selecting galaxies with a specific range of luminosities, we can model the selection effects and evaluate the dependence of lensing distributions on the siren's total mass. Nonetheless, a complete treatment using this approach will require future studies, which will be discussed further in the following section.

5 CONCLUSIONS AND OUTLOOK

SMBHB systems are standard sirens that can provide accurate distance measurements based on their well-understood GW signals. However, gravitational lensing induces significant errors in the measured luminosity distances of high-redshift SMBHBs and the lensing errors are in general comparable to their measurement errors. Moreover, the lensing errors can not be averaged out due to the paucity of SMBHBs. Therefore, gravitational lensing of SMBHBs severely limits their usefulness as standard sirens and their power to constrain cosmological parameters.

In this work, we investigated how much the weak-lensing errors can be reduced by weak-lensing reconstruction, making use of the latest numerical simulations. We then considered the potential impact of 'delensing' strategies on cosmological parameter estimation.

The main conclusion of our work is that the weak-lensing errors for sirens at $z_s = 2.9$ can be reduced by about a factor of two on average, under a futuristic hybrid observation involving wide- and deep-field galaxy surveys. However, such a hybrid observation requires an expensive ultra-deep field and hence might only be feasible for one particular siren in practice. Consequently, performing such a correction is unlikely to be worthwhile since the reduction of the error on the estimated luminosity distance for just one siren is likely insufficient to significantly improve the inference on the cosmological parameters.

Our analysis suggests that galaxy surveys can not sufficiently recover lensing structures at small scales due to the shape noise and thus the delensing performance is not satisfactory. Nonetheless, our conclusions might change if for example CMB lensing maps were incorporated. Other EM observations that can infer the density field along the trajectories of GW, such as tomographically-reconstructed galaxy density fields, might also be helpful for delensing. Future developments in the application of high-precision flexion measurements may also improve the outcomes significantly.

However, as pointed out at the end of Sec. 2.4, we do not fully consider the non-Gaussian nature of the weak-lensing fields as we ignore the discussions about higher moments of the lensing distributions, which should be addressed in future work. Additionally, we have ignored higher-order weak-lensing effects such as reduced shear and flexion. The contaminating effects such as intrinsic alignments are also neglected in our work.

Another limitation of this work is related to the modelling of the selection effects. As demonstrated in Sec. 2.5.3, many high-redshift sirens do not have appropriate host galaxy candidates within the LSST-like catalogs since the suitable galaxies are below the luminosity threshold at the siren's redshift. This can be solved by constructing HOD-based deep-field catalogs from the same set of dark matter N-body simulations. Note that the mock deep-field catalogs in our work only simulate the galaxies in a simple way and do not contain information about the mass or luminosity of the galaxies. This is the reason why we do not consider the galaxies in the deep-field survey as SMBHB candidates and do not investigate the dependence of lensing distributions on the siren's total mass. Therefore, a complete treatment of the selection effects should rely on future deep-field catalogs that simultaneously include weak-lensing maps and dim galaxies with the necessary properties.

Furthermore, the lensing convergence power spectra at different redshifts are related to cosmology. Therefore, the construction of an optimal magnification estimator should depend on the cosmology and the misalignment of the assumed cosmology to the real cosmology might also introduce a bias in the analysis. Additionally, the accuracy of the estimated magnification might also be cosmology-dependent. In this work, we only compare outcomes between two sets of cosmological parameters. The cosmology dependence of error reduction should be further explored in future studies.

Besides, future studies might incorporate weak-lensing data directly into the estimation of cosmological parameters by making use of their cosmology dependence. The weak-lensing data can be obtained solely from the lensing of the GW signals (Congedo & Taylor 2019) or derived from the synergy of galaxy surveys and GW experiments (Balaudo et al. 2022). The cross-correlation of the GW lensing signals with the cosmic density field probed by EM observations has the potential to validate the standard model of cosmology and also

the general theory of relativity (Mukherjee et al. 2020b; Mukherjee et al. 2020a).

Moreover, the effect of masks in the observed shear maps should also be accounted for properly. Since the KS inversion method is non-local, then realistic weak-lensing reconstructions suffer from the missing-data problem even if the targeted siren lying outside the mask. To solve the problem, methods equipped with inpainting techniques like the KS+ method can be helpful (Pires et al. 2020).

Finally, the weak-lensing reconstruction can be improved by new approaches for galaxy shear measurements such as the kinematic lensing method (S. et al. 2022). The kinematic lensing method combines photometric shape measurements with resolved spectroscopic observations to infer the intrinsic galaxy shape and directly estimate the lensing shear. The kinematic lensing method has an order of magnitude improvement over the traditional method, which is very useful in the context of weak-lensing reconstruction. Methods for 3D weak-lensing reconstructions might also help to reduce the weak-lensing errors for standard sirens (Leonard et al. 2014).

ACKNOWLEDGMENTS

The authors give our special thanks to Joachim Harnois-Deraps for providing detailed information on the SLICS catalogs and offering useful suggestions. We also thank Catherine Heymans, Ben Giblin, and David Bacon for useful discussion and insight. The authors are grateful for the siren catalogs provided by Liang-Gui Zhu and his team. M. H. is supported by the Science and Technology Facilities Council (Ref. ST/L000946/1). Finally, the authors acknowledge the many useful comments of the referee which have improved the clarity of this work.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding authors.

REFERENCES

Amaro-Seoane P., et al., 2017, arXiv e-prints, p. arXiv:1702.00786
Arun K. G., et al., 2009, Classical and Quantum Gravity, 26, 094027
Balaudo A., Garoffolo A., Martinelli M., Mukherjee S., Silvestri A., 2022, arXiv e-prints, p. arXiv:2210.06398
Bartelmann M., 1995, A&A, 303, 643
Bartelmann M., Schneider P., 2001, Physics Reports, 340, 291
Cain B., Schechter P. L., Bautz M., 2011, ApJ, 736, 43
Califano M., de Martino I., Vernieri D., Capozziello S., 2023, Physical Sciences Forum, 7
Chassande-Mottin E., Leyde K., Mastrogiovanni S., Steer D., 2019, Phys. Rev. D, 100, 083514
Congedo G., Taylor A., 2019, Phys. Rev. D, 99, 083526
Dalal N., Holz D. E., Chen X., Frieman J. A., 2003, The Astrophysical Journal, 585, L11
Dalal N., Holz D. E., Hughes S. A., Jain B., 2006, Phys. Rev. D, 74, 063006
Ding X., et al., 2020, ApJ, 888, 37
Ferrarese L., Merritt D., 2000, The Astrophysical Journal, 539, L9
Hannam M., Schmidt P., Bohé A., Haegel L., Husa S., Ohme F., Pratten G., Pürrer M., 2014, Phys. Rev. Lett., 113, 151101
Hannuksela O. A., Collett T. E., Çalışkan M., Li T. G. F., 2020, Monthly Notices of the Royal Astronomical Society, 498, 3395
Harnois-Deraps J., Vafaei S., Van Waerbeke L., 2012, Monthly Notices of the Royal Astronomical Society, 426, 1262
Harnois-Deraps J., Pen U.-L., Iliev I. T., Merz H., Emberson J. D., Desjacques V., 2013, Monthly Notices of the Royal Astronomical Society, 436, 540
Harnois-Deraps J., et al., 2018, Monthly Notices of the Royal Astronomical Society, 481, 1337
Harnois-Déraps J., van Waerbeke L., 2015, Monthly Notices of the Royal Astronomical Society, 450, 2857
Hilbert S., Gair J. R., King L. J., 2011, Monthly Notices of the Royal Astronomical Society, 412, 1023
Hilbert S., et al., 2020, Monthly Notices of the Royal Astronomical Society, 493, 305
Hinshaw G., et al., 2013, ApJS, 208, 19
Hirata C. M., Holz D. E., Cutler C., 2010, Phys. Rev. D, 81, 124046
Holz D. E., Hughes S. A., 2005, ApJ, 629, 15
Holz D. E., Linder E. V., 2005, The Astrophysical Journal, 631, 678
Ivezić Ž., et al., 2019, ApJ, 873, 111
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kocsis B., Frei Z., Haiman Z., Menou K., 2006, The Astrophysical Journal, 637, 27
Lahav O., Liddle A. R., 2022, arXiv e-prints, p. arXiv:2201.08666
Lanusse F., Starck J.-L., Leonard A., Pires S., 2016, A&A, 591, A2
Leonard A., Lanusse F., Starck J.-L., 2014, Monthly Notices of the Royal Astronomical Society, 440, 1281
Liu J., Haiman Z., Hui L., Kratochvil J. M., May M., 2014, Phys. Rev. D, 89, 023515
Luo J., et al., 2016, Class. Quantum Grav., 33, 035010
Massey R., et al., 2007, Nature, 445, 286
Milosavljević M., Phinney E. S., 2005, ApJ, 622, L93
Mpetha C. T., Congedo G., Taylor A., 2022, arXiv e-prints, p. arXiv:2208.05959
Mukherjee S., Wandelt B. D., Silk J., 2020a, Phys. Rev. D, 101, 103509
Mukherjee S., Wandelt B. D., Silk J., 2020b, MNRAS, 494, 1956
Nakar E., 2007, Phys. Rep., 442, 166
Nissanke S., Kasliwal M., Georgieva A., 2013, ApJ, 767, 124
Oguri M., Takahashi R., 2020, Astrophys. J., 901, 58
Pang P. T. H., Hannuksela O. A., Dietrich T., Pagano G., Harry I. W., 2020, Monthly Notices of the Royal Astronomical Society, 495, 3740
Pires S., et al., 2020, A&A, 638, A141
Rowe B., Bacon D., Massey R., Heymans C., Häußler B., Taylor A., Rhodes J., Mellier Y., 2013, Monthly Notices of the Royal Astronomical Society, 435, 822
S. P. R., Krause E., Huang H.-J., Huff E., Xu J., Eifler T., Everett S., 2022, arXiv e-prints, p. arXiv:2209.11811
Sathyaprakash B. S., Schutz B. F., 2009, Living Rev. Relativ., 12, 2
Schaan E., Krause E., Eifler T., Doré O., Miyatake H., Rhodes J., Spergel D. N., 2017, Phys. Rev. D, 95, 123512
Schneider P., Van Waerbeke L., Jain B., Kruse G., 1998, Monthly Notices of the Royal Astronomical Society, 296, 873
Schneider P., Kochanek C. S., Wambsganss J., 2006, Gravitational lensing: strong, weak, and micro. No. 33 in Saas-Fee advanced course / Swiss Society for Astrophysics and Astronomy, Springer, Berlin ; New York
Schutz B. F., 1986, Nature, 323, 310
Shang C., Haiman Z., 2011, MNRAS, 411, 9
Shapiro C., Bacon D. J., Hendry M., Hoyle B., 2010, Monthly Notices of the Royal Astronomical Society, 404, 858
Smith R. E., et al., 2003, Monthly Notices of the Royal Astronomical Society, 341, 1311
Springel V., et al., 2005, Nature, 435, 629
Takahashi R., 2006, The Astrophysical Journal, 644, 80
Tanaka T., Menou K., Haiman Z., 2012, MNRAS, 420, 705
Thorpe J. I., et al., 2019, in Bulletin of the American Astronomical Society. p. 77 (arXiv:1907.06482), doi:10.48550/arXiv.1907.06482
Vale C., White M., 2003, ApJ, 592, 699
Volonteri M., Lodato G., Natarajan P., 2007, Monthly Notices of the Royal Astronomical Society, 383, 1079
White M., Vale C., 2004, Astroparticle Physics, 22, 19
Yuan C., Murase K., Zhang B. T., Kimura S. S., Mészáros P., 2021, ApJL, 911, L15
Yung L. Y. A., et al., 2022, Monthly Notices of the Royal Astronomical Society, 515, 5416
Zhu L.-G., Hu Y.-M., Wang H.-T., Zhang J.-d., Li X.-D., Hendry M., Mei J., 2022, Phys. Rev. Research, 4, 013247

APPENDIX A: DERIVATION OF THE POSTERIOR AND LIKELIHOOD

The posterior on the cosmological parameters should be derived as,

$$p(\vec{\Omega}|\,\mathbf{D}_{\rm GW}\,,\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens}) = \frac{p(\vec{\Omega}|\,\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens})p(\mathbf{D}_{\rm GW}\,|\,\vec{\Omega},\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens})}{p(\,\mathbf{D}_{\rm GW}|\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens})}.$$

where the Bayesian evidence $p(\,\mathbf{D}_{\rm GW}|\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens})$ is ignored as it is irrelevant to the derivation. $\mathbf{z}_{\rm s}\,,\mathbf{d}_{\rm lens}$ can be regarded as background information.

The $\mathbf{D}_{\rm GW}$ includes the observed luminosity distance $D_L^{\rm obs}$ and their measurement uncertainties σ_{D_L}. In this paper, we only consider bright sirens, so we ignore the errors in redshifts since they are expected to be much lower than the errors in luminosity distance measurements. We also ignore the dependence of σ_{D_L} on other parameters for simplicity. Then the likelihood becomes

$$p(\mathbf{D}_{\rm GW}|\,\vec{\Omega},\mathbf{z}_{\rm s},\mathbf{d}_{\rm lens}) \propto \prod_i p(D_L^{\rm obs}|\,\sigma_{D_L},\vec{\Omega},z_{\rm s},\mathbf{d}_{\rm lens}) \tag{A1}$$

We ignore the cosmology dependence of the weak-lensing data as assumed in Sec. 2.2 and assume that weak gravitational lensing does not affect the measurement uncertainties σ_{D_L}[21]. The observed luminosity distance is related to the lensing magnification by 4. Then the likelihood is

$$\propto \int \prod_i p(D_L^{\rm obs}|\,\sigma_{D_L},\vec{\Omega},z_{\rm s},\mathbf{d}_{\rm lens},\mu)\; p(\mu|\,\mathbf{d}_{\rm lens},z_{\rm s})\; d\mu\,, \tag{A2}$$

where μ is the true lensing magnification for each siren. The lensing data $\mathbf{d}_{\rm lens}$ can yield an estimate of the magnification $\mu_{\rm est}$ to correct the luminosity distance. Therefore,

$$\propto \int \prod_i p(D_L^{\rm obs}|\,\sigma_{D_L},\vec{\Omega},z_{\rm s},\mu_{\rm est},\mu)\; p(\mu|\,\mathbf{d}_{\rm lens},z_{\rm s})\; d\mu\,. \tag{A3}$$

Since the measurement uncertainties σ_{D_L} are obtained from the Fisher matrix formalism with first-order approximation, the measurement error distribution should be Gaussian. The predicted luminosity distance D_L (Eq. (2)) and observed luminosity distance $D_L^{\rm obs}$ for each siren should be corrected by the estimated magnification $\mu_{\rm est}$, then

$$\propto \prod_i \int \exp\left[(D_L(\vec{\Omega},z_{\rm s})/\sqrt{\mu}\times\sqrt{\mu_{\rm est}} - D_L^{\rm obs}\times\sqrt{\mu_{\rm est}})^2/2\sigma_{D_L}^2\right]\times p(\mu|\,\mathbf{d}_{\rm lens},z_{\rm s})\; d\mu\,, \tag{A4}$$

Since both μ and $\mu_{\rm est}$ are close to one with high probability (within 2σ), then $\sqrt{\mu_{\rm est}}/\sqrt{\mu} \approx (1+(\mu_{\rm est}-\mu)/2)$. Let $D_L^{\rm cor}(\vec{\Omega},z_{\rm s},\mu-\mu_{\rm est}) = D_L(\vec{\Omega},z_{\rm s})\times(1+(\mu_{\rm est}-\mu)/2)$. Finally, the likelihood is given by,

$$\propto \prod_i \int \exp\left[(D_L^{\rm cor}(\vec{\Omega},z_{\rm s},\mu-\mu_{\rm est}) - D_L^{\rm obs,cor})^2/2\sigma_{D_L}^2\right]\times p(\mu-\mu_{\rm est}|\,\mathbf{d}_{\rm lens},z_{\rm s})\; d\mu\,, \tag{A5}$$

where $D_L^{\rm obs,cor} = D_L^{\rm obs}\times\sqrt{\mu_{\rm est}}$ and $p(\mu|\,\mathbf{d}_{\rm lens},z_{\rm s})$ is equivalent to $p(\mu-\mu_{\rm est}|\,\mathbf{d}_{\rm lens},z_{\rm s})$ since $\mu_{\rm est}$ depends only on $\mathbf{d}_{\rm lens}$ and $z_{\rm s}$.

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

[21] For high SNR signals, the ratio of the measurement errors σ_{D_L} for the unlensed and lensed signals is equal to the ratio of their SNRs, by the Fisher Information Matrix formalism. The SNR of the original signal and the (de)magnified signal only differ by a factor of $\sqrt{\mu}$. Therefore, weak lensing will only induce a small change (probably within a few percent) in the measurement error σ_{D_L} of the luminosity distance, which we regard as negligible.