

Impact of weak lensing on bright standard siren analyses

Charlie T. Mpetha^{1,*}, Giuseppe Congedo¹, Andy Taylor¹ and Martin A. Hendry²

¹*Institute for Astronomy, School of Physics and Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, United Kingdom*

²*SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom*



(Received 11 March 2024; accepted 6 June 2024; published 2 July 2024)

Gravitational waves from binary mergers at cosmological distances will experience weak lensing by large scale structure. This causes a (de)magnification, μ , of the wave amplitude, and a completely degenerate modification to the inferred luminosity distance d_L . The customary method to address this is to increase the uncertainty on d_L according to the dispersion of the magnification distribution at the source redshift, $\sigma_\mu(z)$. But this term is dependent on the cosmological parameters that are being constrained by gravitational wave “standard sirens,” such as the Hubble parameter H_0 , and the matter density fraction Ω_m . The dispersion $\sigma_\mu(z)$ is also sensitive to the resolution of the simulation used for its calculation. Tension in the measured value of H_0 from independent datasets, and the present use of weak-lensing fitting functions calibrated using outdated cosmological simulations, suggest $\sigma_\mu(z)$ could be underestimated. This motivates an investigation into the consequences of mischaracterizing $\sigma_\mu(z)$. We consider two classes of standard siren, supermassive black hole binary and binary neutron star mergers. Underestimating H_0 and Ω_m when calculating $\sigma_\mu(z)$ increases the probability of finding a residual lensing bias on these parameters greater than 1σ by 1.5–3 times. Underestimating $\sigma_\mu(z)$ by using low resolution/small sky-area simulations can also significantly increase the probability of biased results; the probability of a $1\sigma(2\sigma)$ bias in H_0 and Ω_m found from binary neutron star mergers is 54% (19%) in this case. For neutron star mergers, the mean bias on H_0 caused by magnification selection effects is $\Delta H_0 = -0.1 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The spread around this mean bias—determined by assumptions on $\sigma_\mu(z)$ —is $\Delta H_0 = \pm 0.25 \text{ km s}^{-1} \text{ Mpc}^{-1}$, comparable to the forecasted uncertainty. These effects do not impact merging neutron stars’ utility for addressing the H_0 tension, but left uncorrected they limit their use for precision cosmology. For supermassive black hole binaries, the spread of possible biases on H_0 is significant, $\Delta H_0 = \pm 5 \text{ km s}^{-1} \text{ Mpc}^{-1}$, but $\mathcal{O}(200)$ observations are needed to reduce the variance below the bias. To achieve accurate subpercent level precision on cosmological parameters using standard sirens, first much improved knowledge on the form of the magnification distribution and its dependence on cosmology is needed.

DOI: [10.1103/PhysRevD.110.023502](https://doi.org/10.1103/PhysRevD.110.023502)

I. INTRODUCTION

Waves propagating through the Universe are magnified by the gravitational potential of large scale structure. This *weak-lensing* effect on light, observed through its effect on galaxy shapes, has been intensely studied as a source of cosmological information [1–5]. Similarly to light, gravitational wave weak lensing (GW-WL) modifies the observed wave amplitude h , to $h' = h\sqrt{\mu}$, where a prime denotes a lensed quantity [6–10]. The magnification μ depends on the matter along the line of sight where $\mu = 1$

corresponds to no effect, $\mu < 1$ a net underdensity of matter, and $\mu > 1$ a net overdensity.

The magnification distribution over the sky, $p(\mu, z)$, broadens towards higher redshifts, and is peaked at values below 1 as there are more voids than clusters. The tail of $p(\mu, z)$ is weighted towards values greater than 1 since clustered matter can have a larger effect on the magnification than voids. Gravitational waves from low redshift merging binaries, such as those observed using the LIGO-Virgo-Kagra (LVK) network [11–14], will not experience impactful weak lensing, though they may still be strongly lensed [15,16]. The space-based GW detector LISA [17] expects to observe between a few and a few tens of supermassive black hole binary (SMBHB) mergers in the mid-2030s over its nominal five-year survey time with a duty cycle of 80% [18–20]. The effect of weak lensing on these observations will be significant, even for individual sources due to small uncertainties on their measured

*Contact author: c.mpetha@ed.ac.uk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

distances. A future network of third generation ground-based detectors (3G), such as the Einstein Telescope (ET) [21] and Cosmic Explorer (CE) [22] is also anticipated to begin operations in the mid-2030s. These detectors will observe binary neutron stars (BNSs) at $z \sim 2$ [23]. Weak lensing will not be as important for individual BNSs due to their larger distance uncertainties compared to those of SMBHBs. On the other hand, with the 3G network expected to detect many more BNSs than the numbers of SMBHBs observed by LISA, the effect of instrumental scatter on the cosmological parameters will be reduced. The bias due to lensing is likely to be even more significant.

Past works have studied how successfully a GW can be “delensed” using external data to estimate μ for each observed source [24–26]. The authors of Ref. [26] found that sufficient delensing using galaxy surveys for all SMBHB mergers observed with LISA will be extremely challenging. For both SMBHBs and BNSs, we cannot expect to remove or even meaningfully reduce the weak-lensing effect, unless further advances are made in delensing methods. An alternative delensing prospect which uses GW data alone is through measuring wave-optics features in a single waveform [27]. However, the expected probability for observing these features is low [28], and this approach is not able to directly probe the range of lens masses which contribute most to weak-lensing magnification.

Measuring the distance and redshift of a binary merger is the core of using these sources for cosmology, as “standard sirens” [29]. Bright standard sirens (BSS) are the ideal case where the GW is observed alongside an electromagnetic counterpart. Another method for using standard sirens for cosmology utilizes the weak lensing of a large number of GWs as information instead of noise, constraining structure as well as geometry [30–37]. But this is only viable with sufficient source numbers, likely not until at least the 2040s. In the meantime, many works have demonstrated how weak lensing can bias the distance redshift relation for the case of supernovae standard candles and GW standard sirens [38–50].

The luminosity distance d_L is an observable quantity with gravitational waves from binary mergers, without needing any external calibration from the cosmic distance ladder. At redshift z ,

$$d_L(z) = (1+z) \int_0^z \frac{cdz'}{H(z')}, \quad (1)$$

where the Hubble parameter is

$$\begin{aligned} \left(\frac{H(z)}{H_0}\right)^2 &= \Omega_m(1+z)^3 + \Omega_K(1+z)^2 \\ &+ \Omega_{DE}(1+z)^{3(1+w_0+w_a)} e^{\frac{-3w_a z}{1+z}}. \end{aligned} \quad (2)$$

Here $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ is the present expansion rate, Ω_m , Ω_K , and Ω_{DE} are the matter, curvature, and dark energy density fractions, respectively, and w_0 and w_a are equation of state parameters for dark energy [51,52].

The luminosity distance is inversely proportional to the observed wave amplitude. Therefore the observed, lensed distance is

$$d'_L = \frac{d_L}{\sqrt{\mu}}. \quad (3)$$

The total distance uncertainty is

$$\sigma_{d'_L}^2 = \sigma_{\text{GW}}^2 + \sigma_{\text{WL}}^2 + \left(\frac{\partial d_L}{\partial z}\right)^2 (\sigma_z^2 + \sigma_{v_{\text{pec}}}^2), \quad (4)$$

where $\partial d_L / \partial z$ converts redshift to distance uncertainties. σ_{GW} and $\sigma_{v_{\text{pec}}}$ are the uncertainty from the detector and source peculiar velocity, respectively. The weak-lensing term is

$$\frac{\sigma_{\text{WL}}(z)}{d_L(z)} = \frac{1}{2} \frac{\sigma_\mu(z)}{\langle \mu(z) \rangle} = \frac{1}{2} \sigma_\mu(z), \quad (5)$$

$\sigma_\mu(z)$ is the uncertainty in the magnification of the gravitational wave. It has been assumed that the mean of the magnification distribution $\langle \mu \rangle = 1$. For real observations this is in general not the case due to detector selection effects as explored in Ref. [53], henceforth CT21. The consequence is $\langle \mu \rangle \sim 1.02$ at $z = 2$.

Weak lensing could lead to a net overestimation or underestimation of luminosity distances at different (true) redshifts, so that the inferred cosmological parameters in Eq. (2) are prone to biases [50]. A mitigation might be to include a weak-lensing uncertainty according to Eq. (5), using an assumed functional form for $\sigma_\mu(z)$ [53–58]. The possible lensing bias is washed out by increasing distance errors. A caveat of this added weak-lensing uncertainty is that the form of the magnification function is not well known, is dependent on the uncertain values of cosmological parameters, and is affected by poorly understood physics influencing the growth of structure on small scales. There is tension in the value of H_0 [59] (recently exacerbated by JWST observations [60]), and present weak-lensing fitting functions have been created with cosmological simulations that could be missing large and small-scale power important in estimating $\sigma_\mu(z)$ [61] (henceforth T11). Both of these facts motivate an investigation into how underestimating or overestimating the weak-lensing uncertainty can impact a cosmological analysis using GWs.

We will evaluate the *residual lensing bias*, the lensing bias remaining even after including a weak-lensing uncertainty term. The authors of Ref. [50] found the residual lensing bias for one idealized source population,

demonstrating for the first time that the lensing bias can be greater than the statistical uncertainty for BNS mergers. Our goals differ from that work in that we are broadly assessing the impact of *mischaracterizing* the weak-lensing uncertainty term for realistic observations from future space and ground-based GW detectors.

We find that underestimating the weak-lensing uncertainty by using incorrect values for cosmological parameters leads to a factor 1.5 (3) increase in the probability of obtaining results biased by more than 1σ for BNSs (SMBHBs). We also demonstrate the importance of the resolution of simulations used to determine $\sigma_\mu(z)$, pointing out that missing power due to simulations not including very small and large scale magnification power could mean up to a factor of 10 greater probability of obtaining results biased by more than 1σ . For constraints from BNSs, we find the bias on H_0 caused by lensing is too small to affect their application to the H_0 tension. But the bias is at a similar level to the variance and so limits their use for precision cosmology. For SMBHBs the lensing bias is significant with a range $\sim 5 \text{ km s}^{-1} \text{ Mpc}^{-1}$, but will only become relevant with $\mathcal{O}(200)$ observations. If GWs from merging binaries are to be used to their full potential by the next generation of detectors, much improved knowledge of $\sigma_\mu(z)$ is needed.

Section II describes the method used to evaluate the residual lensing bias. In Sec. III we introduce the binary merger catalogs used in our analysis. Section IV details the generation of magnification distributions and weak-lensing uncertainties, and Sec. V gives our results. We conclude in Sec. VI.

II. METHOD

Our goal is to evaluate the probability of finding cosmological constraints that are biased due to lensing, in particular the *residual lensing bias* caused by underestimating the weak-lensing uncertainty. The following steps outline how we evaluate this residual lensing bias.

- (1) Set a true cosmology by choosing values for cosmological parameters in Eq. (2).
- (2) Choose a *population*: supermassive black hole binaries (SMBHBs) with a particular seed model observed with LISA [20] (see Sec. III A), or binary neutron stars (BNSs) observed with third generation ground based detectors (see Sec. III B). Steps 3–7 are repeated over many catalog *realizations*: independent sets of observations drawn from the population model.
- (3) Choose a catalog realization. For each merger in the catalog, randomly draw a magnification μ as described in Sec. IV from the redshift and cosmology dependent magnification distribution. Applying the magnification to the measured merger distance and its uncertainty σ_{GW} (a more magnified source has a higher SNR and smaller uncertainty, see

Appendix A) gives a lensed source, to be compared to the unlensed case. The unlensed case has $\mu = 1$ for all sources. Only sources with a SNR greater than a chosen threshold after applying the magnification go into the catalog.

- (4) For each merger generate an observed distance and redshift by drawing from a Gaussian distribution with a width equal to the instrumental uncertainty and mean equal to the true distance/redshift. Also add scatter to the redshift caused by peculiar velocities, motion of the binaries not due to Hubble flow. This is done for both the lensed and unlensed sources, with each having the same random draw to ensure we are only comparing the effects of lensing ($d_L^{\text{obs}} = d_L^{\text{obs}}/\sqrt{\mu}$).
- (5) In both the lensed and unlensed cases, include a weak-lensing uncertainty $\sigma_{\text{WL}}(z)$, as in Eq. (4), to the distance. The assumptions going into generating $\sigma_{\text{WL}}(z)$ are the focus of this investigation.
- (6) Given the observed values of d_L and z for each merger in the catalog, perform Monte Carlo Markov chain (MCMC) sampling of $d_L - z$ in Eq. (1) to reconstruct cosmological parameters in the lensed and unlensed cases. In this work we use the EMCEE package [62].
- (7) The absolute bias is the difference in the best fit recovered parameter value in the lensed and unlensed case. The bias in units of uncertainty is found from the parameter differences method of the TENSIO-METER package¹ [63,64]. This uses information from the full $h - \Omega_m$ 2D posterior to quantify the size of the bias, and does not assume Gaussianity in the parameter space.
- (8) Aggregate the biases for each catalog realization of a population. The resulting distribution of biases from all realizations characterizes the residual lensing bias that can be expected for that population, given the true cosmology and assumed weak-lensing uncertainty.
- (9) Repeat steps 3–8 a further ~ 300 times, and combine the distributions from each repetition. There is possible variation in observations due to the random drawing of magnifications and random scatter caused by distance and redshift uncertainties. By doing this, we have ensured we sample over both statistical variation in the catalogs, and statistical variation in the observations.

III. BINARY MERGER CATALOGS

Our goal is to evaluate the residual lensing bias on cosmological parameters constrained with GWs. To this end, realistic catalogs of observations by the next

¹<https://github.com/mraveri/tensiometer>.

generation of gravitational wave detectors are needed. For the following statistical tests 100/500 catalog realizations are used for BSS observed with LISA/3G. Section III A details the subset of SMBHB catalogs from Ref. [20] used in this work, while the BNS catalogs we generated are described in Sec. III B. An example of one of the catalogs for each population, including instrumental distance/redshift uncertainties and magnification and instrumental scatter, can be seen in Fig. 1. The number of observations will depend on many factors; the performance of future GW and EM detectors, the fraction of the sky area covered by electromagnetic detectors for the multimessenger follow-up, the merger rate, and properties of the EM counterpart. For this reason we remain agnostic of rates and integration times, and instead use a broad range of source numbers in each catalog. We choose not to consider dark standard sirens in this analysis, for example 3G observed binary black holes, or extreme-mass-ratio inspirals observed with LISA. The large redshift uncertainty of these sources will dominate over the weak-lensing uncertainty for most cases. While Ref. [65] demonstrates that using several hundred 3G-observed SNR > 300 binary black holes could achieve comparable precision on H_0 to bright standard sirens, most of these will be too low redshift for important magnification effects.

All population catalogs have been generated assuming a Planck-like cosmology [66]. In this work we evaluate the residual lensing bias in different true cosmologies, and while the assumed cosmology would impact the source

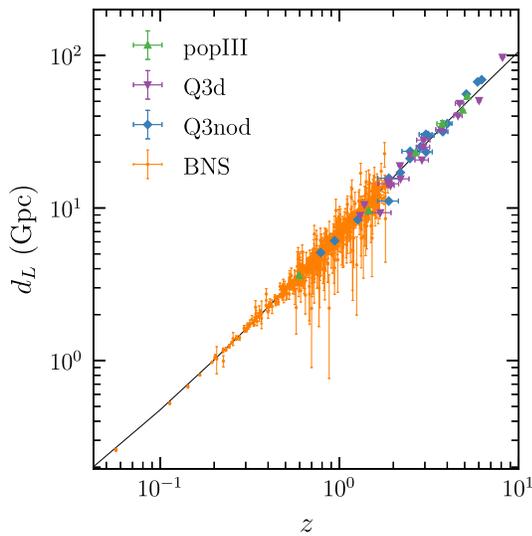


FIG. 1. Representative catalogs for each binary merger population used in this work. Every gravitational wave has been observed alongside an electromagnetic counterpart. Sources have instrumental uncertainty and scatter in distance and redshift, and scatter caused by weak lensing and peculiar velocity. Weak lensing causes large scatter towards high redshift. The popIII light-seed SMBHB model typically predicts fewer, lower redshift mergers than the Q3 heavy-seed models.

distributions and uncertainties, we note that this effect is subdominant to details on the seed model and host galaxy evolution.

A. LISA: Supermassive black hole binaries

We use a subset of the catalogs presented in Refs. [20,67], refer there for more detail on the catalog generation, and to Ref. [67] for cosmological forecasts using these catalogs. As is common in the literature, uncertainty on the seed model for supermassive black holes motivates including three possibilities; black holes are formed from high-redshift high-mass metal poor “popIII” stars [68], the “Q3” heavy seed model where massive black holes are formed from the collapse of protogalactic disks [69,70] with a time delay between the formation time of a binary and its merger (Q3d), and the Q3 model with no delay time between formation and merger (Q3nod). Only mergers with an electromagnetic counterpart are considered. Other methods of estimating the redshift of a single source often rely on galaxy catalogs which will not be possible for high-redshift SMBHBs. In these catalogs there are three possibilities for the redshift uncertainty depending on the magnitude and redshift of the host galaxy observed with the ELT.² If the host galaxy magnitude $m_{\text{gal,ELT}} < 27.2$ and $0.5 < z_{\text{gal}} \leq 5$, then $\sigma_z = 5 \times 10^{-4}$. If $27.2 < m_{\text{gal,ELT}} < 31.3$ and $0.5 < z_{\text{gal}} \leq 5$ ($z_{\text{gal}} > 5$), then $\sigma_z = 0.1$ ($\sigma_z = 0.25$). See Ref. [20] for more detail. We also include a peculiar velocity drawn from a Gaussian centered at zero with a width of 500 km s^{-1} . Sources (from all seed models) range from $z = 0.8$ to $z = 9.2$, with distance uncertainties from 0.02% to 30% having a median value of 0.8%/0.4% for the popIII/Q3 seed models. The average number of sources (median redshift) is 7 (2.4), 15 (3.2), and 20 (2.9) for popIII, Q3d, and Q3nod, respectively.

B. 3G: Binary neutron stars

For third generation ground-based detectors (3G), the most likely BSS are binary neutron star (BNS) mergers. It is not expected that many black hole–neutron star mergers will be observed alongside an EM counterpart [71], so this population is not included. BNSs are generated using the assumptions in Table 1 of Ref. [72], which involve uniform sampling over sky location and inclination, uniform sampling of spins aligned with the orbital angular momentum in the range $[-0.05, 0.05]$, and uniform sampling of mass in the range $[1, 2.5]M_{\odot}$. The adopted redshift distribution of BSS is described in Ref. [35]. The Fisher-matrix code GWFAST [73] is used to find measurement uncertainties of each source in the catalog. We assume a network of Einstein Telescope and two Cosmic Explorers (ET + 2CE)

²<https://elt.eso.org/>.

for the main analysis, considering other configurations in Appendix B.

$\mathcal{O}(10^5)$ BNSs will be observed every year by a 3G detector network [72,74]. Predictions on the corresponding number of BSS vary significantly, from 15 to 3500 yr⁻¹ [23,75–80] depending on assumptions on the GW and EM detectors, merger rates, and counterparts. To reflect this our 500 generated catalogs have source numbers sampled uniformly between 100 and 1000. The minimum is chosen to be 100 as this could be realized even with the lowest rate estimates after 10 years of run-time.

We impose a GW detector network SNR threshold of $\rho_{\text{lim}} = 12$ [81]. Because of the requirement of an EM counterpart, these observations come with selection effects beyond the requirements for a detection of the gravitational wave alone. We add a cut on sky localization of 20 deg² [82,83]. We also limit the distance uncertainty to $\sigma_{d_L}/d_L < 30\%$, around the upper limit of the $z - \sigma_{d_L}/d_L$ fitting function in Ref. [84] for BSS observations with ET. As pointed out in Ref. [10], knowledge of the sky location and inclination of a merger from its EM counterpart can improve constraints on GW parameters, quoting an order of magnitude improvement in σ_{d_L}/d_L . More recent studies use information on the inclination angle ι from EM observations to break the $d_L - \iota$ degeneracy and improve σ_{d_L}/d_L by several factors [85–87]. This was performed for GW170817 using the radio counterpart, improving the H_0 constraint by a factor of 2 [88]. In the context of LISA, Ref. [89] shows how from sky localization the wave amplitude and inclination uncertainties can both be improved by a factor of ~ 2 when $\iota \approx 90^\circ$. This factor increases to ~ 10 if there is also some prior knowledge from EM observations that constrains the inclination. Therefore we make the simplifying assumption that, for all sources with $\iota \leq 60^\circ$ and $\iota \geq 120^\circ$ remaining after making the SNR and sky localization cuts, the distance uncertainty is reduced by a factor of 2 using localization information from the EM counterpart. We note that there could also be a constraint on the inclination angle for a fraction of observations, but do not explicitly include this effect.

The final catalogs contain sources with redshifts from 0.018 to 2 and instrumental d_L uncertainties ranging from 0.008% to 30%, with a median value of 4%. Each merger has a peculiar velocity drawn from a Gaussian centered at zero with width 500 km s⁻¹. Finally, we assume spectroscopic redshift uncertainties of $\sigma_z = 5 \times 10^{-4}(1+z)$, based on electromagnetic follow-up from present and forthcoming spectroscopic redshift surveys, e.g., DESI [90], ELT, and MegaMapper [91]. The impact of this uncertainty is negligible.

IV. MAGNIFICATION DISTRIBUTIONS

The following section details the redshift and cosmology dependent magnification distribution, and how it is used to

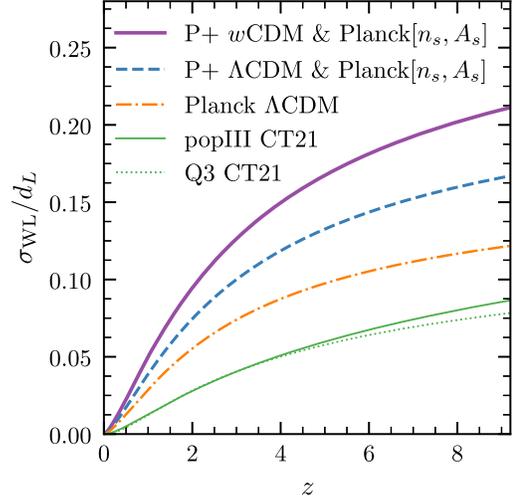


FIG. 2. The weak-lensing uncertainty fitting functions used in this work. Three have been created numerically in CAMB using different cosmologies: Planck 2018 [66] and the Pantheon+ (P+) Λ CDM and w CDM maximum posterior values [92]. Supernovae only constrain geometry, so Planck 2018 values for n_s and A_s are used. Also shown are results from CT21 using numerical simulations in T11, specifically for popIII/Q3 MBHBs observed with LISA, including selection effects.

generate a weak-lensing fitting function. We will show how the width of the magnification distribution depends strongly on the assumed cosmological parameters and modeling assumptions. This lack of knowledge of the true form motivates an investigation into the effect on a cosmological analysis of mischaracterizing the weak-lensing uncertainty. The various weak-lensing fitting functions used in this work are summarized in Fig. 2.

A. Background

The magnification is given by

$$\mu = \frac{1}{(1-\kappa)^2 - \gamma^2}. \quad (6)$$

The convergence κ is the integrated surface mass density along the line of sight, and γ is the shear field. We are interested in the variance of the magnification at a particular redshift $\sigma_\mu^2(z)$, as this provides $\sigma_{\text{WL}}(z)$ in Eq. (4). For the rest of this work, unless otherwise stated, we will drop the small shear corrections which have little impact on gravitational waves [93]. In that case, at a given redshift, the variance in the convergence field is directly related to the variance in the magnification [58],

$$\sigma_\mu^2(z|\boldsymbol{\theta}) \approx 4\sigma_\kappa^2(z|\boldsymbol{\theta}). \quad (7)$$

The convergence field is cosmology dependent. The cosmological parameters influencing the convergence field are $\boldsymbol{\theta} = \{h, \Omega_m, \Omega_b, \Sigma m_\nu, n_s, A_s, \Omega_K, w_0, w_a\}$. Ω_b is the

baryon density fraction, Σm_ν the sum of neutrino masses, A_s is the amplitude of scalar perturbations, and n_s the spectral tilt. Now, using Eq. (5), the variance in the convergence field can be related to the weak-lensing uncertainty we assume for a gravitational wave source,

$$\frac{\sigma_{\text{WL}}(z|\boldsymbol{\theta})}{d_L(z|\boldsymbol{\theta})} \approx \sigma_\kappa(z|\boldsymbol{\theta}). \quad (8)$$

Knowing the dispersion, a magnification probability distribution, $p(\mu, z|\boldsymbol{\theta})$, can be constructed. Poorly understood small-scale physics significantly impacts the high-magnification tail of $p(\mu, z|\boldsymbol{\theta})$. To avoid introducing modeling uncertainty, we restrict our analysis to the weak-lensing regime, with $\mu \leq 1.75$.³ Therefore a simple shifted log-normal model as presented in Ref. [58] is sufficient, but for comparison we also use $p(\mu, z)$ from the $N_{\text{side}} = 4096$ and $N_{\text{side}} = 16384$ simulations described in Ref. [94]⁴ (henceforth T17).

To include the impact of selection effects, we must modify the magnification probability distribution. Following Refs. [53,95],

$$\mathcal{P}(\mu, z|\boldsymbol{\theta}) = \mathcal{C} p(\mu, z|\boldsymbol{\theta}) \frac{d\mathcal{N}(\mu, z)}{dz}, \quad (9)$$

where

$$\frac{d\mathcal{N}(\mu, z)}{dz} = \int_{\rho_{\text{lim}}/\sqrt{\mu}}^{\infty} N(z, \rho) d\rho. \quad (10)$$

The SNR is given by ρ , $N(z, \rho)$ is the number density of sources as a function of redshift and SNR, ρ_{lim} is the SNR threshold, and \mathcal{C} is a normalization constant. This modified magnification distribution is drawn from randomly to find the magnification of sources. Figure 16 of Ref. [20] shows that BSS SMBHBs have SNRs far above any possible SNR threshold, hence have no selection effects. This is not the case for BSS BNSs.

The dispersion of the magnification distribution as a function of redshift can be found using observations, simulations, or numerical calculation using a halo model. These possibilities are explored in the following sections.

B. Numerical calculation

Reference [58] provides a tool to generate the dispersion in a redshift and cosmology dependent convergence field, $\sigma_\kappa(z|\boldsymbol{\theta})$, for varying cosmological parameters using the Boltzmann code CAMB.⁵ We made minor modifications to

their notebook; using the HMcode-2020 nonlinear model [96], setting $k_{\text{max}} = 100h \text{ Mpc}^{-1}$ as this is where HMcode switches from calling the linear power spectrum to using a linear extrapolation, and including neutrinos and dark energy to the fitting. The main uncertainty in this approach is the modeling of nonlinear scales. HMcode is known to not be reliable for $z > 2$ [96], and calculating the convergence power spectrum requires a power-law extrapolation of the Weyl potential at high k modes which is not well motivated. This calculation involves integrating the convergence power spectrum over the multipole ℓ to find the total variance. This integral converges for $\ell_{\text{max}} = 10^7$ [58], and while we cannot expect to accurately predict the power spectrum at such high multipoles, omitting small scales could lead to underpredicting the variance. Thus, knowledge of much smaller scales than presently available is needed for accurate modeling of $\sigma_\kappa(z|\boldsymbol{\theta})$.

Using this modified notebook, $\sigma_\kappa(z|\boldsymbol{\theta})$ is found for different choices of cosmological parameters: Planck 2018 base Λ CDM constraints from TT, TE, EE + lowE + lensing + BAO [66] (henceforth labeled “Planck”), and the best fit parameters from the Pantheon + analysis assuming either Λ CDM or w CDM [92]. These choices highlight the current tension in constraints of H_0 . Supernovae only constrain geometry so n_s and A_s are fixed to their Planck values when generating these fitting functions. The weak-lensing uncertainty for various cosmologies is plotted in Fig. 2. The value of h , Ω_m , w_0 , and w_a used for each cosmology is given in Table I.

C. Simulations

The present industry standard weak-lensing fitting function [53–57,97] was created in 2010 [97] using the magnification distribution of the 1998 work Ref. [98]. It has since been modified in CT21 using the ray-tracing simulations presented in T11. The use of these simulations gives cause for concern; they have limited resolution over a very small patch of the sky and therefore could underpredict the width of the magnification distribution, they are dark matter only simulations neglecting baryonic physics, and they were created using outdated values of cosmological parameters ($h = 0.705$, $\Omega_m = 0.274$). The authors of Ref. [99] demonstrate how the inclusion of baryons in hydrodynamic simulations modifies the high magnification tail ($\mu > 3$), past the weak-lensing regime. This suggests

TABLE I. We will evaluate the residual lensing bias when one of these three cosmologies is used to generate a weak-lensing fitting function, while the truth is given by another.

Cosmology	h	Ω_m	w_0	w_a
Planck	0.677	0.311	−1	0
Pantheon+ Λ CDM	0.736	0.344	−1	0
Pantheon+ w CDM	0.733	0.403	−0.93	−0.1

³Imposing this limit causes $\langle \mu \rangle < 1$, but this effect is small (0.01% at $z = 2$ and 0.5% at $z = 8$) and we confirmed it does not impact results for the size of the lensing bias.

⁴http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky_raytracing/.

⁵<https://camb.info/>.

that dark matter only simulations are sufficient for evaluating the effects of weak lensing. But more recent works using the IllustrisTNG [100] and MilleniumTNG [101] simulation suites suggest that baryons, and in particular feedback effects, have a larger impact, causing a narrower $p(\mu, z)$ as matter is redistributed from high to low density regions. Furthermore, the magnification distribution is highly dependent on cosmological parameters. Lower values of h and Ω_m predict a smaller $\sigma_\kappa(z|\theta)$. More recent constraints on Ω_m favor larger values [66,92] than used in T11, and there is the possibility h favors the near-Universe measurement [92,102]. Therefore current weak-lensing fitting functions in the literature could be underestimating the dispersion of $p(\mu, z|\theta)$.

The function calibrated using T11, as presented in CT21, where the authors consider lensing selection effects for the popIII and Q3 SMBHB seed models separately, is plotted as a green solid and green dotted line in Fig. 2. The fitting function, whose form was originally found in Ref. [97], is given by

$$\frac{\sigma_{\text{WL}}}{d_L} = \frac{C}{2} \left(\frac{(1 - (1+z)^{-\beta})}{\beta} \right)^\alpha, \quad (11)$$

where $C = 0.061$, $\beta = 0.264$, and $\alpha = 1.89$ for the popIII seed model and $C = 0.096$, $\beta = 0.62$, and $\alpha = 2.36$ for the Q3 model. Note the factor of $1/2$, which is missing in Ref. [55]. This is because the original expression is for the uncertainty in $\sigma(\ln d_L^2) = 2\sigma(d_L)/d_L$ [53,65].

The difference between these curves and those generated using CAMB is significant. This is due to the simulation's resolution. As previously mentioned, for the numerical calculation there is an integral over the convergence power spectrum from $\ell_{\min} = 1$ to $\ell_{\max} = 10^7$. The ray-tracing simulations of T11 use a $1 \times 1 \text{ deg}^2$ box with a grid resolution of $3 \text{ kpc}/h$. The small box will fail to include superclusters or huge voids, both of which will broaden the magnification distribution. This sets a high ℓ_{\min} , while ℓ_{\max} increases with redshift but is smaller than 10^7 . Hence, the variance found from these simulations will be smaller. Setting ℓ_{\min} and ℓ_{\max} in the numerical calculation based on the simulation parameters leads to comparable predictions for $\sigma_{\text{WL}}(z)$ to Refs. [97] and CT21.

One could also consider using the ray-tracing simulations of T17 to find the dispersion in the magnification distribution over finer redshift slices. They are full sky, providing access to lower multipoles, but have significantly lower resolution in the high ℓ regime, and are available up to a maximum redshift of 5.3, compared to $z_{\max} = 20$ in T11 (though there are only seven redshift bins from $z = 1$ to $z = 20$). They use a similar set of cosmological parameters to T11, with $h = 0.7$, $\Omega_m = 0.279$.

⁶Incidentally, including this factor of $1/2$ would improve results in numerous forecasting papers, e.g., Refs. [32,35,55,56,76,103].

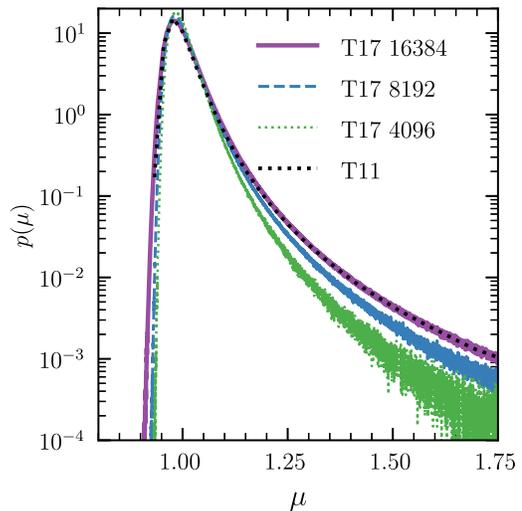


FIG. 3. The magnification probability distribution at $z = 1.0334$ found from the full-sky ray-tracing simulations of T17. Increasing the N_{side} of the simulation from 4096 to 16384 leads to a broader distribution, and therefore a larger assumed weak-lensing uncertainty. Also overlaid is the $z = 1$ probability density function (PDF) from the T11 $1 \times 1 \text{ deg}^2$ ray-tracing simulations.

To demonstrate the importance of simulation resolution, the magnification distribution from simulations of T17 for the same snapshot at $z = 1.0334$, but with a different N_{side} , can be seen in Fig. 3. Both the convergence and shear fields are used to find $p(\mu)$ from these simulations. There is a marked difference in the width of $p(\mu)$, highlighting the importance of resolution (or equivalently, ℓ_{\max}) in estimating the weak-lensing uncertainty. Also shown is $p(\mu)$ from T11 at $z = 1$, which shows close similarity to the $N_{\text{side}} = 16384$ case. This suggests that even finer resolution full-sky simulations could give a broader distribution than found in T11. Corresponding weak-lensing uncertainties found using the method of Ref. [97] can be seen in Fig. 4. The weak-lensing uncertainty found from the full-sky T17 $N_{\text{side}} = 16384$ simulations follows the fitting function of Eq. (11) with $C = 0.076$, $\beta = 0.344$, and $\alpha = 2.25$ with no selection effects, and $C = 0.092$, $\beta = 0.164$, and $\alpha = 2.09$ considering selection effects from an ET + 2CE network. The finer redshift slices and full-sky nature make this fitting function more appropriate for the redshift regime where it is applicable, $z \leq 5.3$. We suggest this fitting function should be used for the lensing uncertainty on BNS mergers. However, higher resolution full-sky simulations will likely lead to a wider predicted magnification distribution.

Weak-lensing data products from other cosmological simulations exist [104–109] and it would be interesting to see how choice of simulation affects the derived weak-lensing uncertainty. To our knowledge, the only other example of a weak-lensing uncertainty predicted from simulations is shown in Ref. [54]. The authors use the ELUCID simulations [109] to estimate $\sigma_{\text{WL}}(z)$ for

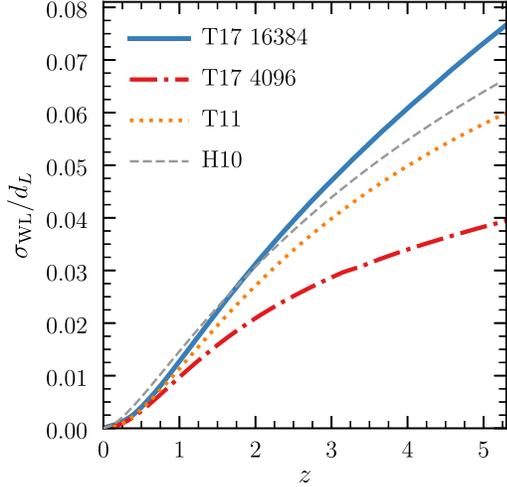


FIG. 4. A comparison of weak-lensing uncertainties from full-sky ray tracing simulations with different resolutions (T17), high resolution ray-tracing simulations over a $1 \times 1 \text{ deg}^2$ patch of sky (T11), and from the analytic PDF in Ref. [98] used in Ref. [97] (H10). No selection effects have been included.

observations of binary black holes at $z \leq 1$. For $z > 1$ they also use results from T11, hence find general agreement with the curves in CT21.

D. Observations

Ideally, $p(\mu, z)$ would be inferred from observations. Through the weak lensing of galaxies we measure reduced shear $g = \gamma/(1 - \kappa)$, which is either approximated as shear, or γ is recovered through an iterative scheme [110]. The convergence is then reconstructed from the shear. This is a very noisy process, and measuring enough galaxies for accurate reconstruction requires deep observations. Measuring $p(\mu, z)$ directly over the full sky in fine redshift slices is not achievable in the near future. The best we can do is calibrate simulations to lensing statistics from forthcoming large scale galaxy surveys, such as the Vera C. Rubin Observatory [111] or Euclid [112].

V. RESULTS

A. Evaluating the residual lensing bias

For each combination of true and assumed cosmology, the steps in Sec. II are performed to gain a large distribution of possible lensing biases for each population. Using this distribution, we can characterize the impact of weak lensing through the mean bias, and the dispersion of biases. Positively magnified sources are more likely to be observed, this is a weak-lensing selection effect. When there are no selection effects $\langle \mu \rangle = 1$ and the mean bias ~ 0 . But with low SNR sources that can be shifted inside the SNR limit through a weak-lensing magnification, $\langle \mu \rangle \neq 1$, and there is a bias in the recovered cosmological parameters. This depends primarily on the source redshift

TABLE II. The magnification selection effect on gravitational waves causes a bias \bar{b} on recovered cosmological parameters. This depends on the source redshift distribution: results for the BNS catalogs used in this work are given here. The bias has been averaged over 300 statistical realizations of 500 catalogs. The bias also depends on the width of magnification distribution, which is affected by cosmological parameters and simulation resolution.

$p(\mu)$ cosmology	$\bar{b}(h)$	$\bar{b}(\Omega_m)$
Planck log norm	-0.0011	0.0074
Pantheon+ Λ CDM log norm	-0.0015	0.0098
Pantheon+ w CDM log norm	-0.0020	0.0153
T17 ($N_{\text{side}} = 4096$)	-0.0007	0.0044
T17 ($N_{\text{side}} = 16384$)	-0.0013	0.0083

distribution and the magnification distribution. For a chosen population, the mean residual lensing bias found by averaging over all statistical samples of a single catalog realization shows little variation from realization to realization. Even though the realizations have a range of source numbers, this will only impact the parameter uncertainties and not the mean lensing bias. For the BNS catalogs considered in this work, we report the mean absolute bias on h and Ω_m due to selection effects in Table II. Results are shown for the three cosmologies used to generate the numerical estimate of $\sigma_{\text{WL}}(z)$, and when magnifications are drawn directly from the $N_{\text{side}} = 4096$ and $N_{\text{side}} = 16384$ simulations of T17. In the $N_{\text{side}} = 4096$ case the mean bias is smaller as the limited resolution leads to a smaller ℓ_{max} than in the $N_{\text{side}} = 16384$ and numerical cases. BSS SMBHBs need a high SNR to accurately localize and associate with an EM counterpart so there are no magnification selection effects on these sources, and their mean bias fluctuates around zero across each catalog.

Next, we go on to consider the impact of the assumption on $\sigma_{\text{WL}}(z)$. This will primarily affect the width of the bias distribution. From Table II it can already be seen how mischaracterizing the true cosmology, or using low resolution simulations, when correcting for magnification effects could lead to a residual lensing bias. For the SMBHB catalogs, the high redshift sources lead to large magnifications and large absolute biases. But the small source numbers mean large parameter uncertainties as the bias is below the level of the variance. For BNSs the opposite is true: they have both smaller magnifications and uncertainties. The absolute bias is small, but its impact is larger. We quote results in two ways: first the probability of the bias on h (Ω_m) exceeding a value of ± 0.0025 (± 0.015) for BNSs and ± 0.02 (± 0.05) for SMBHBs (this indicates the spread of biases across catalogs); then the probability of there being a discordance in the parameter space greater than a certain number of standard deviations (this takes into account the parameter uncertainties to show the impact of the residual lensing bias on parameter constraints).

TABLE III. Mean probability of finding a lensing bias when a lensing uncertainty term on the distance is not included. Both the probability of the bias on h and Ω_m exceeding a set value, and the probability of the bias exceeding a number of standard deviations in the 2D $h - \Omega_m$ posterior, are shown. The true cosmology used is Planck 2018 [66]. For BNSs $x_h = 0.0025$, $x_{\Omega_m} = 0.015$. For SMBHBs $x_h = 0.02$, $x_{\Omega_m} = 0.05$.

Population	$P(b(h) \geq \pm x_h)$	$P(b(\Omega_m) \geq \pm x_{\Omega_m})$	$P(b \geq 1\sigma[3\sigma])$
popIII	0.65	0.72	0.90 [0.69]
Q3	0.78	0.77	0.98 [0.79]
Q3nod	0.56	0.62	0.98 [0.81]
BNS	0.62	0.55	0.79 [0.44]

1. No added weak-lensing uncertainty

First we will evaluate the impact of not including a weak-lensing uncertainty term in Eq. (4). If this term is not added, the probability of a lensing bias is given in Table III, where the true cosmology is Planck. It can be seen that the probability of obtaining highly significant biases from LISA sources is very large, due to the small uncertainties on the high redshift, high SNR SMBHBs, and the lack of a weak-lensing uncertainty term. Even for the ground-based observed BNSs the probability is high, demonstrating that weak lensing is an important consideration for these sources.

This scenario highlights the necessity of including $\sigma_{\text{WL}}(z)$. Any cosmological analysis using well measured binary mergers beyond $z \sim 0.5$ that does not could be significantly biased.

2. Correct uncertainty implemented

Now we test the situation where $\sigma_{\text{WL}}(z)$ is generated using the same distribution from which magnifications are drawn. There is perfect knowledge of the true $p(\mu, z|\theta)$. This does not mean the magnification can be removed from the GW, for this an external measurement of μ is needed for each source. Magnifications are drawn from $p(\mu, z|\theta)$ generated using Planck parameters and the added uncertainty is the Planck curve in Fig. 2. Results are presented in Table IV.

In all populations, the probability of a large parameter shift remains high. The absolute bias due to lensing is primarily dependent on the shape of the magnification distribution, which is the same as it was in Table III.

In the case of SMBHBs, for a bias $b \geq 1\sigma$ the probability is ≤ 0.05 , and for $b \geq 2\sigma$ the probability is ≤ 0.01 . The lensing bias has been effectively removed by increasing uncertainties according to the dispersion of $p(\mu, z|\theta)$.

For BNSs we are more likely to observe sources with a positive magnification. There is an average negative/positive shift on h/Ω_m , also shown in Ref. [50]. Because of this selection effect the probability of a residual bias $b \geq 1\sigma$ ($b \geq 2\sigma$) is 0.18(0.01). The lensing selection

TABLE IV. Mean probability of finding a lensing bias. Both the probability of the bias on h and Ω_m exceeding a set value, and the probability of the bias exceeding a number of standard deviations in the 2D $h - \Omega_m$ posterior, are shown. The weak-lensing uncertainty used is based on perfect knowledge of the shape of the magnification distribution, created using Planck 2018 parameters. For BNSs $x_h = 0.0025$, $x_{\Omega_m} = 0.015$. For SMBHBs $x_h = 0.02$, $x_{\Omega_m} = 0.05$.

Population	$P(b(h) \geq \pm x_h)$	$P(b(\Omega_m) \geq \pm x_{\Omega_m})$	$P(b \geq 1\sigma[2\sigma])$
popIII	0.45	0.47	0.03 [0.00]
Q3	0.54	0.53	0.03 [0.00]
Q3nod	0.32	0.40	0.05 [0.01]
BNS	0.21	0.14	0.18 [0.01]

effect could in principle be corrected for using the method of CT21, outlined in their Eq. (10). Following this we confirmed that, if source numbers are large enough, binning them by redshift, finding $\langle d_L^2(z) \rangle$ in each bin and multiplying by $\langle \mu(z) \rangle$ for that bin removed the lensing bias caused by selection effects. The key caveats of this method are needing sufficient source numbers to obtain a reliable $\langle d_L(z) \rangle$, and crucially, accurate knowledge of $p(\mu, z|\theta)$. As will be demonstrated in the next section, mischaracterizing the magnification distribution could exacerbate the lensing bias instead of mitigating it.

3. Imperfect correction

In reality, we do not know the correct form of $p(\mu, z|\theta)$. First we will consider the case when the uncertainty is underestimated, and then when it is overestimated. If the weak-lensing uncertainty is underestimated, then the lensing bias demonstrated in Table III will not be mitigated, as it has mostly been in Sec. VA 2. There are two main reasons the weak-lensing uncertainty could be underestimated:

- (A) using too small values of H_0 or Ω_m to predict $\sigma_{\text{WL}}(z)$; and
- (B) incorrect modeling of the shape of the magnification distribution.

We investigate case A by assuming the true cosmology used to generate $p(\mu, z|\theta)$ is either the Pantheon+ Λ CDM or Pantheon+ w CDM best fit parameters, and the weak-lensing uncertainty is given by the Planck curve in Fig. 2. Figure 5 shows an example of the residual lensing bias from ET + 2CE observations of BNSs with an electromagnetic counterpart. Probabilities of obtaining a lensing bias are presented in Table V. A key result is the importance of accurate modeling of $p(\mu, z|\theta)$ even for large numbers of BSS BNSs, where the lensing bias is greater than that caused by selection effects alone. We also tested the bias when BNSs are used to probe extensions to the base Λ CDM, including $k\Lambda$ CDM (curvature) and w_0 CDM (dark energy) in Appendix C. For LISA SMBHBs, despite the

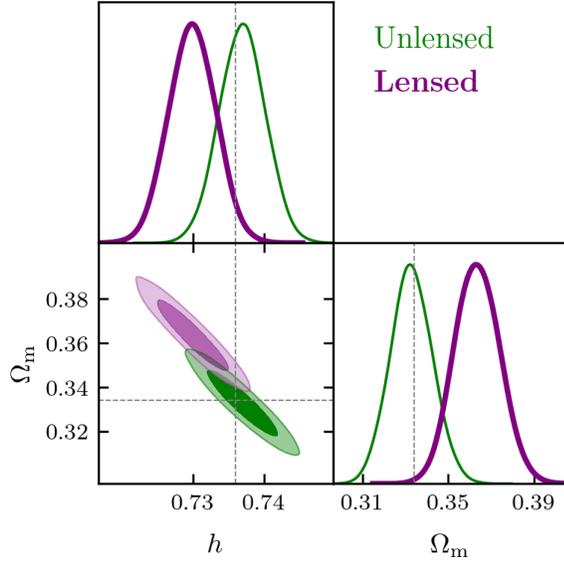


FIG. 5. The lensed and unlensed contours are constraints from the same binary neutron star catalog with 641 sources. The *true* cosmology is the best fit Pantheon+ Λ CDM parameters. In both cases, the distance uncertainty is increased according to a weak-lensing uncertainty fitting function generated using Planck 2018 parameters: this is the assumed cosmology. The green “Unlensed” contour is the constraint on h and Ω_m when all sources experience magnification $\mu = 1$. The purple “Lensed” contour is when each source experiences a magnification drawn from a redshift dependent magnification distribution created using the *true* cosmology. The parameters of this model give a broader magnification distribution than the weak-lensing uncertainty fitting function predicts, hence the lensing bias is not completely mitigated; there is a total bias of 2.0σ , or an absolute bias of $b(h) = -0.007$ and $b(\Omega_m) = 0.031$.

large probability of a significant parameter shift, the broad parameter uncertainties make the probability of a large discordance caused by lensing small.

To assess the impact of underestimating cosmological parameters, probabilities in Table V are compared with those in Table IV, which describe when the weak-lensing uncertainty is correctly estimated. First consider if the true cosmology is Pantheon+ Λ CDM, and assumed is Planck. For all populations, the probability of the absolute bias on h

and Ω_m of exceeding a certain value has increased. This is due to the wider magnification distribution leading to more significant magnification effects. For SMBHBs, the probability of obtaining results biased by more than 1σ is increased by a factor ~ 2.5 – 3 , but still remains low due to the large parameter uncertainties caused by few observations. For BNSs, the probability of results biased by more than 1σ has increased by a factor of 1.5 , and is close to $1/3$. If the true cosmology is in fact Pantheon+ w CDM, probabilities increase by an extra factor of ~ 2 .

To test case B, we compare the fitting functions of CT21 with curves created using CAMB. $\sigma_{\text{WL}}(z|\theta)$ generated using the numerical method of Sec. IV B and the cosmological parameters in T11 and T17 are very similar to the Planck curve in Fig. 2 (orange dash-dotted line). This is because the increase in Ω_m and decrease in h between the parameters used in the simulations and Planck are competing effects and there is a partial cancellation of biases. The difference in σ_{WL} is $\lesssim 5\%$, so using Planck as the true cosmology to draw magnifications, and the fitting functions of CT21 (green solid and dotted lines) as the lensing uncertainty acts as a model comparison. The probability of obtaining a residual lensing bias in this case is presented in Table VI for each population. Once again, as the true cosmology is the same as in Sec. VA 2, we can compare bias probabilities with Table IV. In this case the uncertainty is too small to wash out the bias. For SMBHBs the probability of results biased by $\geq 1\sigma$ increases by a factor of ~ 10 , and there is now a significant probability of results biased by $\geq 2\sigma$. For BNSs, there is now greater than 50% probability of finding results biased by more than 1σ and for $b \geq 2\sigma$ the probability has jumped from 1% to 19%. Because of the impact of selection effects on BNSs and their stronger constraining power, combining the SMBHB and BNS catalogs has a small impact on the BNS results.

If the assumption on $\sigma_{\text{WL}}(z|\theta)$ in fact overestimates the dispersion in $p(\mu, z|\theta)$, then the uncertainty on the cosmological parameters is excessively diluted. When the true cosmology is Planck and the assumed cosmology is Pantheon+ Λ CDM (w CDM), the increased uncertainty at low redshift causes a degradation of parameter constraints of 10%(15%)/8%(12%) for SMBHBs/BNSs.

TABLE V. Mean probability of finding a lensing bias. Both the probability of the bias on h and Ω_m exceeding a set value, and the probability of the bias exceeding a number of standard deviations in the 2D $h - \Omega_m$ posterior, are shown. The weak-lensing uncertainty is a fitting function based on the Planck 2018 cosmological parameters. The column headers give the true cosmology. For BNSs $x_h = 0.0025$, $x_{\Omega_m} = 0.015$. For SMBHBs $x_h = 0.02$, $x_{\Omega_m} = 0.05$.

Population	Pantheon+ Λ CDM			Pantheon+ w CDM		
	$P(b(h) \geq \pm x_h)$	$P(b(\Omega_m) \geq \pm x_{\Omega_m})$	$P(b \geq 1\sigma[2\sigma])$	$P(b(h) \geq \pm x_h)$	$P(b(\Omega_m) \geq \pm x_{\Omega_m})$	$P(b \geq 1\sigma[2\sigma])$
popIII	0.55	0.56	0.07 [0.02]	0.62	0.65	0.15 [0.04]
Q3	0.66	0.64	0.10 [0.02]	0.72	0.73	0.18 [0.05]
Q3nod	0.44	0.53	0.13 [0.02]	0.49	0.64	0.21 [0.05]
BNS	0.32	0.24	0.29 [0.02]	0.50	0.53	0.49 [0.10]

TABLE VI. Mean probability of finding a lensing bias. Both the probability of the bias on h and Ω_m exceeding a set value, and the probability of the bias exceeding a number of standard deviations in the 2D $h - \Omega_m$ posterior, are shown. The true cosmology is given by the Planck parameters and the weak-lensing uncertainty is given by CT21. For BNSs $x_h = 0.0025$, $x_{\Omega_m} = 0.015$. For SMBHBs $x_h = 0.02$, $x_{\Omega_m} = 0.05$.

Population	Planck and CT21		
	$P(b(h) \geq \pm x_h)$	$P(b(\Omega_m) \geq \pm x_{\Omega_m})$	$P(b \geq 1\sigma[2\sigma])$
popIII	0.53	0.55	0.32 [0.09]
Q3	0.64	0.64	0.38 [0.11]
Q3nod	0.37	0.44	0.46 [0.15]
BNS	0.32	0.25	0.54 [0.19]

B. Reducing the probability of a lensing bias

1. Setting requirements

In Sec. VA we calculated the residual lensing bias in several test cases, demonstrating the bias can exceed the variance significantly if either cosmological parameters are underestimated, or too-low resolution simulations are underestimating the weak-lensing uncertainty. We now set requirements on the required precisions to which the input cosmology should be known, and discuss prospects for using state-of-the-art simulations to prevent the residual lensing bias becoming a limiting factor for bright standard siren analyses.

The dependence of the residual lensing bias on H_0 and Ω_m to the biased input parameters can be easily tested in our framework. For a fixed true cosmology, the average lensing bias is constant as it only depends on the magnification and source distributions. When we vary the input cosmology used to calculate the weak-lensing uncertainty, we are just scaling the average uncertainty on recovered

cosmological parameters [smaller assumed $\sigma_{\text{WL}}(z|\theta)$ leads to smaller parameter uncertainties]. So the scaling of the average bias in units of σ will simply depend on the scaling of σ , the parameter uncertainties. The reason we have cast the change of bias in this way is to remain agnostic of the true source numbers/distribution and the true cosmology.

In Fig. 6 we show the change in the lensing bias (the color of the heatmap) as a function of the *input bias* on cosmological parameters (the x and y axes). The bias change is computed as a fractional change relative to the lensing bias when the input bias is zero; this is the $(0, 0)$ point on these plots. When the input bias is zero, this can be considered a best-case scenario. If the input bias is negative, the lensing bias is increased and the bias change is greater than 1. If the input bias is positive, then the bias change is less than 1, but the uncertainties on the recovered H_0 and Ω_m have been increased, weakening their constraints. We confirmed that the results in Fig. 6 are general for different source numbers and input cosmologies. The left plot shows that if the input Ω_m is known to a precision of 0.02 or better, the increase in bias is kept at or below the 5% level. The H_0 offset is less impactful; weak lensing is less sensitive to the Hubble parameter [113], but H_0 should be known in combination with Ω_m to $\sim 2 \text{ km s}^{-1} \text{ Mpc}^{-1}$. In the right plot we also consider the amplitude of the linear matter power spectrum in spheres of radius $8 \text{ Mpc}^{-1} h^{-1}$, σ_8 . We find that, if Ω_m is to be known to 0.02 or better, then σ_8 should be known to a precision of 0.07 to keep the increase in bias below the 5% level. The required precisions for σ_8 and Ω_m are well within parameter uncertainties from, e.g., Planck, and are consistent with differences in these parameters observed from different probes (e.g., Refs. [5,66,92]). While the required precision on H_0 is also well within current constraints, it does not span the H_0 tension.

It is reassuring that the precision to which we need to know H_0 and Ω_m is much larger than the expected

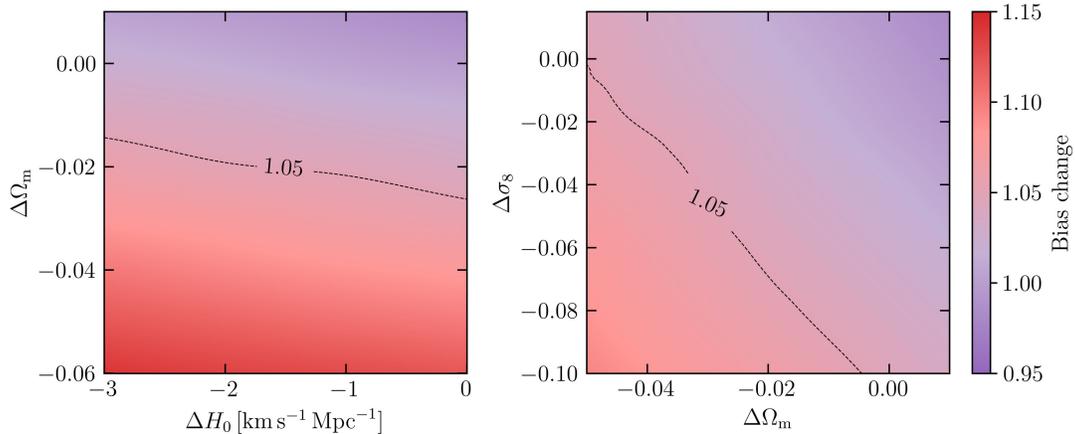


FIG. 6. Heatmap displaying the lensing bias change due to incorrect cosmological parameters when generating the weak-lensing uncertainty. The x and y axes show the difference between the parameter used to generate the weak-lensing uncertainty and the true value of that parameter. The color shows the resulting change in the average residual lensing bias, relative to the bias when the input cosmology is correct, $(0, 0)$ on these plots. Also overlaid is a contour showing the boundary where the change in the bias is at the 5% level.

constraints from standard sirens on these parameters, but we only have one Universe and one set of observations. As demonstrated in the previous section, the probability for large lensing biases can still be high, even when we are close to being correct to the input cosmology.

Testing the required simulation resolution is more challenging, as we cannot know what the true magnification distribution should be in this case. We use the full-sky simulations of Ref. [94], using the expected weak-lensing uncertainty from each N_{side} . We also downsample the $N_{\text{side}} = 4096$ simulations to $N_{\text{side}} = 2048$ and $N_{\text{side}} = 1024$. We find a linear scaling between the average bias and $\log_2(N_{\text{side}})$ (with and without the downsampled simulations), which is largely the same, regardless of the true cosmology. This is due to a constant factor difference between these weak-lensing fitting functions at $z \geq 1$. For every power of 2 increase in N_{side} , the average bias drops by $\sim 0.1\sigma$, where σ is the recovered parameter uncertainty.

To fully investigate the effect of the simulation resolution on the magnification distribution, and relate this to expected magnifications of real sources, baryonic effects at small scales are required. The AbacusSummit full simulation suite includes simulations with baryons, and weak-lensing maps of these are forthcoming [104]. Weak-lensing maps for full hydrodynamical simulations with baryons have been calculated in Ref. [101], but for a single resolution of $N_{\text{side}} = 12288$. In that work, the authors find that baryons reduce the width of the magnification distribution. If this effect could be probed down to smaller scales/greater resolution, this may indicate a convergence of the width of the magnification distribution with resolution, which would provide a best-guess weak-lensing uncertainty.

A similar result could also be obtained using baryonic feedback effects in the halo model used to numerically calculate $\sigma_{\text{WL}}(z|\theta)$. A feedback model, HMcode2020_feedback, is already included in CAMB [96], but this model was not used in this work due to a divergence in the variance of the power spectrum at small scales.

2. Other approaches

We consider two methods to address imperfect knowledge on the weak-lensing uncertainty. The first assumes we are confident in the modeling of $\sigma_{\text{WL}}(z)$ and know how it varies with cosmology, but are ignorant of the true values of cosmological parameters. This could be achieved either through improved modeling of small scales in numerical codes such as CAMB, or a thorough investigation on how the width of the magnification distribution varies across high resolution full-sky simulations using different assumed cosmologies, such as is available with AbacusSummit. The generation of $\sigma_{\text{WL}}(z|\theta)$ is added as an ingredient to the MCMC analysis with each new sampling of h , Ω_m and any other parameters being constrained. This does not guarantee results are unbiased, but removes the sensitivity to an assumed cosmology.

Alternatively, we are not confident in our modeling of $\sigma_{\text{WL}}(z)$, and we do not know how it varies with cosmology. Here, we consider introducing an extra parameter into the fitting which gives an amplitude modification to the variance, similar to that given in Ref. [114]:

$$\sigma'_{\text{WL}}(z) = A\sigma_{\text{WL}}(z). \quad (12)$$

A realistic prior on A is needed. The fitting will try to increase the variance as much as it can, as this will minimize $(\text{data} - \text{model})^2/\text{variance}$. The prior on A could be based on the difference in predictions for the weak-lensing uncertainty across simulations or cosmological parameters. This is a simple choice, and neglects that we are also uncertain on the variation of uncertainty with redshift. A second shape changing parameter could be introduced, but this would also require a well justified prior. This method attempts to find an acceptable bias-variance trade-off, increasing uncertainties within a defined range as you fit the data to a model. It also allows you to jointly constrain cosmology with the form of the weak-lensing uncertainty.

If the goal is reliability of parameter constraints, the simplest approach may be to use a large enough weak-lensing uncertainty to encompass all possible curves, increasing parameter uncertainties by $\sim 15\%$ but strongly reducing the probability of obtaining biased results.

VI. CONCLUSIONS

Gravitational waves from binary mergers will be weakly lensed. To address the unknown magnification of each source, a weak-lensing uncertainty term must be added to the observations. We aim to investigate the impact of mischaracterizing this term on a cosmological analysis.

Our key conclusions are:

- (i) Any cosmological analysis using binary mergers at $z \gtrsim 0.5$ not including a weak-lensing term has a high probability of being biased by more than 1σ .
- (ii) Present weak-lensing uncertainty fitting functions could be drastically underestimating the width of the redshift dependent magnification distribution. This is caused by the use of $1 \times 1 \text{ deg}^2$ simulations using too small cosmological parameters. We provide a new form for the weak-lensing fitting formula in Eq. (11) using the full-sky ray-tracing simulations of T17 in Sec. IV C, noting this reduces the probability of a residual lensing bias for $z \leq 5.3$ bright standard sirens, such as merging binary neutron stars.
- (iii) If cosmological parameters are underestimated when generating $\sigma_{\text{WL}}(z)$, the probability of parameter constraints biased by more than 1σ increases by a factor 1.5–3 compared to the case when these parameters are not underestimated.
- (iv) If the magnification distribution is in fact much broader than predicted by small sky-area ray-tracing simulations, but they are used to estimate the

weak-lensing uncertainty, bias probabilities increase by a factor 3 (10) for BNSs (SMBHBs) compared to the case when the correct $p(\mu)$ is used. The probability of a $1\sigma(2\sigma)$ bias in H_0 and Ω_m found from BNS mergers is 54%(19%).

- (v) Considering the small absolute value of the bias for BNSs, with a spread of $\Delta b(H_0) \sim 0.25 \text{ km s}^{-1} \text{ Mpc}^{-1}$, magnification selection effects or incorrect assumptions on the true form of $p(\mu|z, \theta)$ are unlikely to impact the efficacy of BNSs on addressing the H_0 tension. However should more BSS SMBHBs be detected than anticipated, the absolute bias having a spread $\Delta b(H_0) \sim 5 \text{ km s}^{-1} \text{ Mpc}^{-1}$ would be significant. An order of magnitude more SMBHBs than considered in these catalogs (from ~ 20 to ~ 200) would be needed for the bias to become a serious consideration.
- (vi) Utilizing the full potential of bright standard sirens requires accurate modeling of $p(\mu|z, \theta)$ using H_0 , Ω_m , and σ_8 known to a precision better than $2 \text{ km s}^{-1} \text{ Mpc}^{-1}$, 0.02, and 0.07, respectively.
- (vii) The most promising approach to estimate $\sigma_{\text{WL}}(z)$ is using high-resolution weak-lensing maps, including baryons, over a range of resolutions. A convergence of the width of $p(\mu, z)$ with resolution caused by baryonic feedback effects would provide a reliable estimate for the weak-lensing uncertainty. This investigation will be possible in the near future [104].

A similar, independent analysis was performed in Ref. [50]. There the true cosmology is that used in the AbacusSummit simulations [104] with $h = 0.6736$, $\Omega_m = 0.3138$, similar to the Planck values used in this work, and the weak-lensing uncertainty is the fitting function of Ref. [55]. The magnification distributions from AbacusSummit, as seen in Fig. 1 of Ref. [50], are broader than those from the simulations of T11 from which the WL uncertainty is derived, contributing to their derived bias. The authors of Ref. [50] include larger lensing magnifications while we limit to $\mu \leq 1.75$ to avoid modeling uncertainties. Therefore we could be under-representing the probability of a bias.

With enough data, there is the possibility of constraining the parameters of the weak-lensing fitting function in Eq. (11) jointly with cosmology. We briefly explored this in Eq. (12), where the amplitude of $\sigma_{\text{WL}}(z)$ is treated as a free parameter. This is a hybrid approach, using the information from observations to avoid fixing $\sigma_{\text{WL}}(z)$ based only on simulations, and was originally envisioned in Ref. [32]. Also, as pointed out in Ref. [30], the relative abundance of magnified and demagnified sources reveals the non-Gaussianity of $p(\mu|z)$. We leave as future work a detailed investigation into the prospects of jointly constraining cosmology and the variance of the magnification distribution.

An interesting consideration is how these results impact delensing prospects. Present delensing studies assume a form for the magnification distribution, which if incorrect

could lead to a biased delensing procedure. Future delensing prospects, and the impact of the magnification distribution on delensing, will be explored in a future work. Other future work includes assessing the impact of weak lensing on numerous high redshift dark standard sirens, and investigating the effect of the residual lensing bias for analyses focusing on high redshift SMBHBs to study, for example, modified gravity [76,103,115].

ACKNOWLEDGMENTS

We thank Alberto Mangliagli for use of their SMBHB catalogs. C. T. M. thanks Ryuichi Takahashi for helpful discussion and feedback, and Jack Elvin-Poole and Pierre Burger for useful discourse. C. T. M. is supported by a Science and Technology Facilities Council (STFC) Studentship Grant. A. T. is supported by a STFC Consolidate Grant. M. H. acknowledges the support of the Science and Technology Facilities Council (Grant Ref. ST/V005634/1). The Python packages NumPy, SciPy, Matplotlib, GetDist, EMCEE, GWFAST and TENSIMETER have been used in this work.

APPENDIX A: LENSED DISTANCE UNCERTAINTY

Consider two cases, a source at d_L with no magnification ($\mu = 1$), and the same source detected by the same detector but with a magnification μ .

From Fisher formalism,

$$\sigma_{d_L} = \frac{1}{\sqrt{\Gamma_{d_L}}}, \quad (\text{A1})$$

$$\Gamma_{d_L} = \left(\frac{\partial \tilde{h}(f)}{\partial d_L} \middle| \frac{\partial \tilde{h}(f)}{\partial d_L} \right), \quad (\text{A2})$$

$$\Gamma_{d_L} = 4 \int_{f_{\min}}^{f_{\max}} \frac{1}{S_h(f)} \frac{\partial \tilde{h}^*(f)}{\partial d_L} \frac{\partial \tilde{h}(f)}{\partial d_L} df. \quad (\text{A3})$$

For simplicity assume

$$\tilde{h}(f) = \frac{A}{d_L} M_z^{5/6} f^{-7/6} e^{i\Phi(f)}. \quad (\text{A4})$$

Then

$$\Gamma_{d_L} = \int_{f_{\min}}^{f_{\max}} \frac{1}{S_h(f)} \frac{A^2}{d_L^4} M_z^{5/3} f^{-7/6} df, \quad (\text{A5})$$

$$\sigma_{d_L} = \frac{d_L^2}{A} \left(M_z^{5/3} \int_{f_{\min}}^{f_{\max}} \frac{1}{S_h(f)} f^{-7/6} df \right)^{-1/2}. \quad (\text{A6})$$

Hence, the uncertainty on d_L

$$\sigma_{d_L} \propto d_L^2. \quad (\text{A7})$$

Therefore, using

$$d'_L = \frac{d_L}{\sqrt{\mu}}, \quad (\text{A8})$$

where $'$ corresponds to lensed quantities,

$$\frac{\sigma'_{d_L}}{(d'_L)^2} = \frac{\sigma_{d_L}}{d_L^2}, \quad (\text{A9})$$

$$\sigma'_{d_L} = \frac{\sigma_{d_L} (d_L / \sqrt{\mu})^2}{d_L^2}, \quad (\text{A10})$$

$$\sigma'_{d_L} = \frac{\sigma_{d_L}}{\mu}. \quad (\text{A11})$$

This relation has been verified by passing a catalog, both before and after lensing, to the GW Fisher parameter uncertainty code GWFISH [116].

APPENDIX B: OTHER GROUND-BASED DETECTOR NETWORKS

1. ET and CE

If there is only one CE-like detector in the network, the sky localization is degraded. We use a single catalog of 500 sources to test this case, finding a factor 2 increase in the probability of exceeding 1σ when the Planck curve in Fig. 2 is used for $\sigma_{\text{WL}}(z)$ and the true cosmology is given by the Pantheon+ Λ CDM constraints. There is a factor 5 increase in bias probabilities when CT21 is used for $\sigma_{\text{WL}}(z)$ and Planck is the true cosmology. The residual lensing bias is less impactful due to the degraded measurement uncertainties, but still significant.

2. ET and Voyager

Here we consider a combination of Einstein Telescope and a next generation LVK detector [12], to aid sky localization. Because of the EM counterpart selection effects, such as a sky localization $< 20 \text{ deg}^2$ and a distance uncertainty $< 30\%$, the redshift of BSS BNS from this detector combination are ≤ 0.5 . Selection effects now cause a small bias of $b(h) = -0.0001$ and $b(\Omega_m) = 0.0013$.

If we relax the assumption on the sky localization to $< 200 \text{ deg}^2$, the bias due to selection effects increases by a factor of 5 due to the inclusion of higher redshift sources. Because the uncertainty on h and Ω_m when being jointly constrained by BSS observed in this GW detector network is large, the probability of the residual lensing bias exceeding 1σ is 0.04 even when no weak-lensing uncertainty term is included. The measurement uncertainty dominates.

We can conclude that the impact of residual lensing biases is most important for high-redshift binaries observed

by LISA, or BNS mergers observed through a ground-based network of two or more 3G detectors.

APPENDIX C: RESIDUAL LENSING BIAS IN EXTENSIONS TO Λ CDM

1. Curvature

We consider the scenario where the true cosmology is the Pantheon+ Λ CDM best fit values, and the weak-lensing uncertainty is derived using Planck 2018 parameters. The source population is BNSs observed by the 3G detector network. We find the mean parameter biases caused by magnification selection effects are $\bar{b}(h) = 0.0004$, $\bar{b}(\Omega_m) = 0.0161$, $\bar{b}(\Omega_k) = -0.03$. We find the probabilities $P(b(h) \geq \pm 0.0025) = 0.86$, $P(b(\Omega_m) \geq \pm 0.015) = 0.60$, and $P(b(\Omega_k) \geq \pm 0.1) = 0.52$. It can be seen that, for Ω_k particularly, the probability of a residual lensing bias that is both large in terms of the parameter, and large in terms of the parameter uncertainties, is high. The probability of a discordance in the dataset $\geq 1\sigma$ ($\geq 2\sigma$) is 0.19 (0.02). An example of the lensing bias in this cosmological model can be seen in Fig. 7.

2. Dark energy

We use the same scenario as in the case of curvature—Pantheon+ Λ CDM for the true cosmology and Planck to generate the assumed weak-lensing uncertainty. We find the mean parameter biases caused by magnification selection effects are $\bar{b}(h) = 0.0007$, $\bar{b}(\Omega_m) = 0.0236$, and

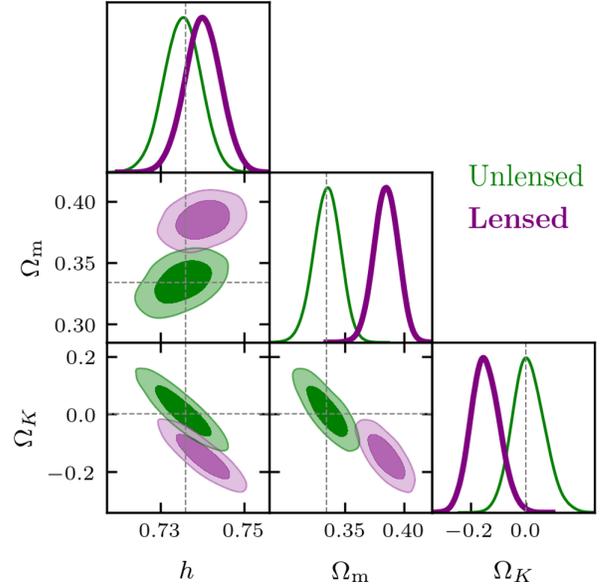


FIG. 7. Similar to Fig. 5 but for the $k\Lambda$ CDM cosmological model, where we are also constraining curvature. In this case lensing leads to a 2.7σ bias in this parameter space, and a $> 2\sigma$ preference for negative curvature.

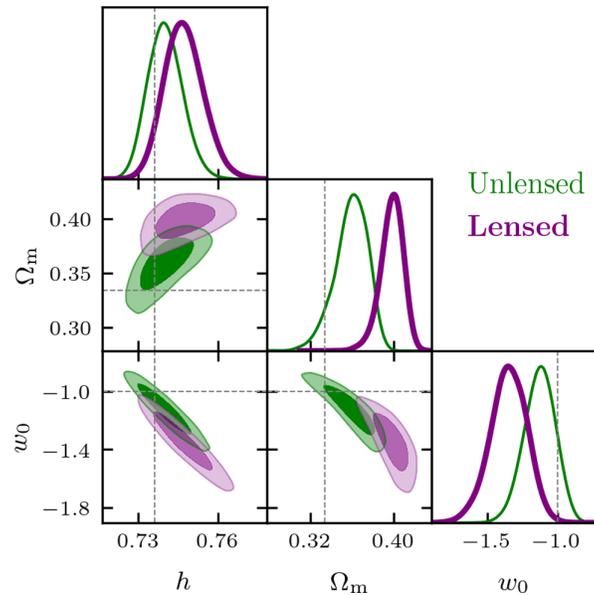


FIG. 8. Similar to Fig. 5 but for the w_0 CDM cosmological model, where we are also constraining a constant dark energy equation of state. In this case lensing leads to a 1.2σ bias in this parameter space, and a $>2\sigma$ preference for a phantom dark energy.

$\bar{b}(w_0) = -0.0574$. We find the probabilities $P(b(h) \geq \pm 0.0025) = 0.72$, $P(b(\Omega_m) \geq \pm 0.015) = 0.77$, and $P(b(w_0) \geq \pm 0.1) = 0.54$. Once again, the parameter extension to Λ CDM causes significant biases. The probability of a discordance in the dataset $\geq 1\sigma$ ($\geq 2\sigma$) is 0.54(0.01). An example of the lensing bias in this cosmological model can be seen in Fig. 8.

-
- [1] M. Bartelmann and P. Schneider, Weak gravitational lensing, *Phys. Rep.* **340**, 291 (2001).
 - [2] R. Mandelbaum, Weak lensing for precision cosmology, *Annu. Rev. Astron. Astrophys.* **56**, 393 (2018).
 - [3] C. Heymans, T. Tröster, M. Asgari, C. Blake, H. Hildebrandt, B. Joachimi, K. Kuijken, C.-A. Lin *et al.*, KiDS-1000 cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints, *Astron. Astrophys.* **646**, A140 (2021).
 - [4] R. Dalal, X. Li, A. Nicola, J. Zuntz, M. A. Strauss, S. Sugiyama, T. Zhang, M. M. Rau *et al.*, Hyper Suprime-Cam Year 3 results: Cosmology from cosmic shear power spectra, *Phys. Rev. D* **108**, 123519 (2023).
 - [5] T. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis *et al.*, Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing, *Phys. Rev. D* **105**, 023520 (2022).
 - [6] D. Marković, Possibility of determining cosmological parameters from measurements of gravitational waves emitted by coalescing, compact binaries, *Phys. Rev. D* **48**, 4738 (1993).
 - [7] Y. Wang, A. Stebbins, and E. L. Turner, Gravitational lensing of gravitational waves from merging neutron star binaries, *Phys. Rev. Lett.* **77**, 2875 (1996).
 - [8] R. Takahashi and T. Nakamura, Wave effects in the gravitational lensing of gravitational waves from chirping binaries, *Astrophys. J.* **595**, 1039 (2003).
 - [9] R. Takahashi, T. Suyama, and S. Michikoshi, Scattering of gravitational waves by the weak gravitational fields of lens objects, *Astron. Astrophys.* **438**, L5 (2005).
 - [10] D. E. Holz and S. A. Hughes, Using gravitational-wave standard sirens, *Astrophys. J.* **629**, 15 (2005).
 - [11] B. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, C. Adams *et al.* (The LIGO Scientific Collaboration), Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, 061102 (2016).
 - [12] The LIGO Scientific Collaboration, Gravitational wave astronomy with LIGO and similar detectors in the next decade, [arXiv:1904.03187](https://arxiv.org/abs/1904.03187).
 - [13] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto, Interferometer design of the KAGRA gravitational wave detector, *Phys. Rev. D* **88**, 043007 (2013).
 - [14] C. S. Unnikrishnan, IndIGO and LIGO-India: Scope and plans for gravitational wave research and precision metrology in India, *Int. J. Mod. Phys. A* **22**, 1341010 (2013).
 - [15] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari *et al.*,

- Search for lensing signatures in the gravitational-wave observations from the first half of LIGO–Virgo’s third observing run, *Astrophys. J.* **923**, 14 (2021).
- [16] K. Liao, X.-L. Fan, X. Ding, M. Biesiada, and Z.-H. Zhu, Precision cosmology from future lensed gravitational wave and electromagnetic signals, *Nat. Commun.* **8**, 1148 (2017).
- [17] P. A. Seoane, S. Aoudia, H. Audley, G. Auger, S. Babak, J. Baker, E. Barausse, S. Barke *et al.* (The eLISA Consortium), The gravitational universe, [arXiv:1305.5720](https://arxiv.org/abs/1305.5720).
- [18] A. Klein, E. Barausse, A. Sesana, A. Petiteau, E. Berti, S. Babak, J. Gair, S. Aoudia *et al.*, Science with the space-based interferometer eLISA: Supermassive black hole binaries, *Phys. Rev. D* **93**, 024003 (2016).
- [19] M. Bonetti, A. Sesana, F. Haardt, E. Barausse, and M. Colpi, Post-Newtonian evolution of massive black hole triplets in galactic nuclei—IV. Implications for LISA, *Mon. Not. R. Astron. Soc.* **486**, 4044 (2019).
- [20] A. Mangiagli, C. Caprini, M. Volonteri, S. Marsat, S. Vergani, N. Tamanini, and H. Inchauspé, Massive black hole binaries in LISA: Multimessenger prospects and electromagnetic counterparts, *Phys. Rev. D* **106**, 103017 (2022).
- [21] M. Maggiore, C. V. D. Broeck, N. Bartolo, E. Belgacem, D. Bertacca, M. A. Bizouard, M. Branchesi *et al.*, Science case for the Einstein Telescope, *J. Cosmol. Astropart. Phys.* **03** (2020) 050.
- [22] D. Reitze, R. X. Adhikari, S. Ballmer, B. Barish, L. Barsotti, G. Billingsley, D. A. Brown, Y. Chen *et al.*, Cosmic Explorer: The U.S. contribution to gravitational-wave astronomy beyond LIGO, *Bull. Am. Astron. Soc.* **51**, 035 (2019), [arXiv:1907.04833](https://arxiv.org/abs/1907.04833).
- [23] E. Belgacem, Y. Dirian, S. Foffa, E. J. Howell, M. Maggiore, and T. Regimbau, Cosmology and dark energy from joint gravitational wave-GRB observations, *J. Cosmol. Astropart. Phys.* **08** (2019) 015.
- [24] S. Hilbert, J. R. Gair, and L. J. King, Reducing distance errors for standard candles and standard sirens with weak-lensing shear and flexion maps, *Mon. Not. R. Astron. Soc.* **412**, 1023 (2011).
- [25] C. Shapiro, D. J. Bacon, M. Hendry, and B. Hoyle, Delensing gravitational wave standard sirens with shear and flexion maps, *Mon. Not. R. Astron. Soc.* **404**, 858 (2010).
- [26] Z.-F. Wu, L. W. L. Chan, M. Hendry, and O. A. Hannuksela, Reducing the impact of weak-lensing errors on gravitational-wave standard sirens, *Mon. Not. R. Astron. Soc.* **522**, 4059 (2023).
- [27] S. Savastano, G. Tambalo, H. Villarrubia-Rojo, and M. Zumalacárregui, Weakly lensed gravitational waves: Probing cosmic structures with wave-optics features, *Phys. Rev. D* **108**, 103532 (2023).
- [28] M. Çalıřkan, N. Anil Kumar, L. Ji, J. M. Ezquiaga, R. Cotesta, E. Berti, and M. Kamionkowski, Probing wave-optics effects and low-mass dark matter halos with lensing of gravitational waves from massive black holes, *Phys. Rev. D* **108**, 123543 (2023).
- [29] B. F. Schutz, Determining the Hubble constant from gravitational wave observations, *Nature (London)* **323**, 310 (1986).
- [30] C. Cutler and D. E. Holz, Ultrahigh precision cosmology from gravitational waves, *Phys. Rev. D* **80**, 104009 (2009).
- [31] S. Camera and A. Nishizawa, Beyond concordance cosmology with magnification of gravitational-wave standard sirens, *Phys. Rev. Lett.* **110**, 151103 (2013).
- [32] G. Congedo and A. Taylor, Joint cosmological inference of standard sirens and gravitational wave weak lensing, *Phys. Rev. D* **99**, 083526 (2019).
- [33] S. Mukherjee, B. D. Wandelt, and J. Silk, Multimessenger tests of gravity with weakly lensed gravitational waves, *Phys. Rev. D* **101**, 103509 (2020).
- [34] S. Mukherjee, B. D. Wandelt, and J. Silk, Probing the theory of gravity with gravitational lensing of gravitational waves and galaxy surveys, *Mon. Not. R. Astron. Soc.* **494**, 1956 (2020).
- [35] C. T. Mpetha, G. Congedo, and A. Taylor, Future prospects on testing extensions to Λ CDM through the weak lensing of gravitational waves, *Phys. Rev. D* **107**, 103518 (2023).
- [36] A. Balardo, A. Garoffolo, M. Martinelli, S. Mukherjee, and A. Silvestri, Prospects of testing late-time cosmology with weak lensing of gravitational waves and galaxy surveys, *J. Cosmol. Astropart. Phys.* **06** (2023) 050.
- [37] G. Scelfo, M. Berti, A. Silvestri, and M. Viel, Testing gravity with gravitational waves \times electromagnetic probes cross-correlations, *J. Cosmol. Astropart. Phys.* **02** (2023) 010.
- [38] I. Ben-Dayan, M. Gasperini, G. Marozzi, F. Nugier, and G. Veneziano, Do stochastic inhomogeneities affect dark-energy precision measurements?, *Phys. Rev. Lett.* **110**, 021301 (2013).
- [39] I. Ben-Dayan, M. Gasperini, G. Marozzi, F. Nugier, and G. Veneziano, Average and dispersion of the luminosity-redshift relation in the concordance model, *J. Cosmol. Astropart. Phys.* **06** (2013) 002.
- [40] N. Kaiser and J. A. Peacock, On the bias of the distance-redshift relation from gravitational lensing, *Mon. Not. R. Astron. Soc.* **455**, 4518 (2015).
- [41] D. Sarkar, A. Amblard, D. E. Holz, and A. Cooray, Lensing and supernovae: Quantifying the bias on the dark energy equation of state, *Astrophys. J.* **678**, 1 (2008).
- [42] M. Smith, D. J. Bacon, R. C. Nichol, H. Campbell, C. Clarkson, R. Maartens, and C. B. D’Andrea, B. A. Bassett *et al.*, The effect of weak lensing on distance estimates from supernovae, *Astrophys. J.* **780**, 24 (2013).
- [43] C. Shang and Z. Haiman, Cosmology with standard sirens: The importance of the shape of the lensing magnification distribution, *Mon. Not. R. Astron. Soc.* **411**, 9 (2010).
- [44] P. Fleury, C. Clarkson, and R. Maartens, How does the cosmic large-scale structure bias the Hubble diagram?, *J. Cosmol. Astropart. Phys.* **03** (2017) 062.
- [45] J. Adamek, C. Clarkson, L. Coates, R. Durrer, and M. Kunz, Bias and scatter in the Hubble diagram from cosmological large-scale structure, *Phys. Rev. D* **100**, 021301 (2019).
- [46] C.-M. Yoo, K.-I. Nakao, H. Kozaki, and R. Takahashi, Lensing effects on gravitational waves in a clumpy universe: Effects of inhomogeneity on the distance-redshift relation, *Astrophys. J.* **655**, 691 (2007).
- [47] G. Fanizza, M. Gasperini, G. Marozzi, and G. Veneziano, Generalized covariant prescriptions for averaging

- cosmological observables, *J. Cosmol. Astropart. Phys.* **02** (2020) 017.
- [48] G. Fanizza, B. Fiorini, and G. Marozzi, Cosmic variance of H_0 in light of forthcoming high-redshift surveys, *Phys. Rev. D* **104**, 083506 (2021).
- [49] T. Schiavone, E. Di Dio, and G. Fanizza, The skewness of the distance-redshift relation in Λ CDM, *J. Cosmol. Astropart. Phys.* **02** (2024) 050.
- [50] S. Canevarolo and N. E. Chisari, Lensing bias on cosmological parameters from bright standard sirens, [arXiv:2310.12764](https://arxiv.org/abs/2310.12764).
- [51] M. Chevallier and D. Polarski, Accelerating Universes with scaling dark matter, *Int. J. Mod. Phys. D* **10**, 213 (2001).
- [52] E. V. Linder, Exploring the expansion history of the Universe, *Phys. Rev. Lett.* **90**, 091301 (2003).
- [53] G. Cusin and N. Tamanini, Characterization of lensing selection effects for LISA massive black hole binary mergers, *Mon. Not. R. Astron. Soc.* **504**, 3610 (2021).
- [54] X. Shan, C. Wei, and B. Hu, Lensing magnification: Gravitational waves from coalescing stellar-mass binary black holes, *Mon. Not. R. Astron. Soc.* **508**, 1253 (2021).
- [55] N. Tamanini, C. Caprini, E. Barausse, A. Sesana, A. Klein, and A. Petiteau, Science with the space-based interferometer eLISA. III: Probing the expansion of the universe using gravitational wave standard sirens, *J. Cosmol. Astropart. Phys.* **04** (2016) 002.
- [56] L. Speri, N. Tamanini, R. R. Caldwell, J. R. Gair, and B. Wang, Testing the quasar Hubble diagram with LISA standard sirens, *Phys. Rev. D* **103**, 083526 (2021).
- [57] L. Dai, T. Venumadhav, and K. Sigurdson, Effect of lensing magnification on the apparent distribution of black hole mergers, *Phys. Rev. D* **95**, 044011 (2017).
- [58] M. Martinelli, F. Scarcella, N. B. Hogg, B. J. Kavanagh, D. Gaggero, and P. Fleury, Dancing in the dark: Detecting a population of distant primordial black holes, *J. Cosmol. Astropart. Phys.* **08** (2022) 006.
- [59] L. Verde, T. Treu, and A. G. Riess, Tensions between the early and late universe, *Nat. Astron.* **3**, 891 (2019).
- [60] A. G. Riess, G. S. Anand, W. Yuan, S. Casertano, A. Dolphin, L. M. Macri, L. Breuval, D. Scolnic, L. M. Macri, L. Breuval, D. Scolnic, M. Perrin, and R. I. Anderson, Crowded no more: The accuracy of the Hubble constant tested with high-resolution observations of cepheids by JWST, *Astrophys. J. Lett.* **956**, L18 (2023).
- [61] R. Takahashi, M. Oguri, M. Sato, and T. Hamana, Probability distribution functions of cosmological lensing: Convergence, shear and magnification, *Astrophys. J.* **742**, 15 (2011).
- [62] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, EMCEE: The MCMC hammer, *Publ. Astron. Soc. Pac.* **125**, 306 (2013).
- [63] M. Raveri, G. Zacharegkas, and W. Hu, Quantifying concordance of correlated cosmological datasets, *Phys. Rev. D* **101**, 103527 (2020).
- [64] M. Raveri and C. Doux, Non-Gaussian estimates of tensions in cosmological parameters, *Phys. Rev. D* **104**, 043504 (2021).
- [65] N. Muttoni, D. Laghi, N. Tamanini, S. Marsat, and D. Izquierdo-Villalba, Dark siren cosmology with binary black holes in the era of third-generation gravitational wave detectors, *Phys. Rev. D* **108**, 043543 (2023).
- [66] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro *et al.* (The Planck Collaboration), Planck 2018 results, *Astron. Astrophys.* **641**, A6 (2020).
- [67] A. Mangiagli, C. Caprini, S. Marsat, L. Speri, R. R. Caldwell, and N. Tamanini, Massive black hole binaries in LISA: Constraining cosmological parameters at high redshifts, [arXiv:2312.04632](https://arxiv.org/abs/2312.04632).
- [68] P. Madau and M. J. Rees, Massive black holes as population III remnants, *Astrophys. J.* **551**, L27 (2001).
- [69] M. C. Begelman, M. Volonteri, and M. J. Rees, Formation of supermassive black holes by direct collapse in pregalactic haloes, *Mon. Not. R. Astron. Soc.* **370**, 289 (2006).
- [70] M. Volonteri, G. Lodato, and P. Natarajan, The evolution of massive black hole seeds, *Mon. Not. R. Astron. Soc.* **383**, 1079 (2007).
- [71] G. Fragione, Black-hole–neutron-star mergers are unlikely multimessenger sources, *Astrophys. J. Lett.* **923**, L2 (2021).
- [72] F. Iacovelli, M. Mancarella, S. Foffa, and M. Maggiore, Forecasting the detection capabilities of third-generation gravitational-wave detectors using GWFIRST, *Astrophys. J.* **941**, 208 (2022).
- [73] F. Iacovelli, M. Mancarella, S. Foffa, and M. Maggiore, GWFIRST: A Fisher information matrix Python code for third-generation gravitational-wave detectors, *Astrophys. J. Suppl. Ser.* **263**, 2 (2022).
- [74] S. Borhanian and B. S. Sathyaprakash, Listening to the universe with next generation ground-based gravitational-wave detectors, [arXiv:2202.11048](https://arxiv.org/abs/2202.11048).
- [75] G. Stratta, R. Ciolfi, L. Amati, E. Bozzo, G. Ghirlanda, E. Maiorano, L. Nicastro, A. Rossi *et al.*, THESEUS: A key space mission concept for multi-messenger astrophysics, *Adv. Space Res.* **62**, 662 (2018).
- [76] T. Yang, Gravitational-wave detector networks: Standard sirens on cosmology and modified gravity theory, *J. Cosmol. Astropart. Phys.* **05** (2021) 044.
- [77] S. Ronchini, M. Branchesi, G. Oganessian, B. Banerjee, U. Dupletsa, G. Ghirlanda, J. Harms, M. Mapelli, and F. Santoliquido, Perspectives for multimessenger astronomy with the next generation of gravitational-wave detectors and high-energy satellites, *Astron. Astrophys.* **665**, A97 (2022).
- [78] V. Alfradique, M. Quartin, L. Amendola, T. Castro, and A. Toubiana, The lure of sirens: Joint distance and velocity measurements with third-generation detectors, *Mon. Not. R. Astron. Soc.* **517**, 5449 (2022).
- [79] H.-Y. Chen, P. S. Cowperthwaite, B. D. Metzger, and E. Berger, A program for multimessenger standard siren cosmology in the era of LIGO A+, Rubin Observatory, and beyond, *Astrophys. J.* **908**, L4 (2021).
- [80] J. W. Gardner, L. Sun, S. Borhanian, P. D. Lasky, E. Thrane, D. E. McClelland, and B. J. J. Slagmolen, Multimessenger astronomy with a Southern-hemisphere gravitational-wave observatory, *Phys. Rev. D* **108**, 123026 (2023).

- [81] B. Sathyaprakash, M. Abernathy, F. Acernese, P. Ajith, B. Allen, P. Amaro-Seoane, N. Andersson, S. Aoudia *et al.*, Scientific objectives of Einstein Telescope, *Classical Quantum Gravity* **29**, 124013 (2012).
- [82] H.-Y. Chen, P. S. Cowperthwaite, B. D. Metzger, and E. Berger, A program for multimessenger standard siren cosmology in the era of LIGO A+, Rubin Observatory, and beyond, *Astrophys. J.* **908**, L4 (2021).
- [83] I. Andreoni, R. Margutti, O. S. Salafia, B. Parazin, V. A. Villar, M. W. Coughlin, P. Yoachim, K. Mortensen *et al.*, Target-of-opportunity observations of gravitational-wave events with Vera C. Rubin Observatory, *Astrophys. J. Suppl. Ser.* **260**, 18 (2022).
- [84] W. Zhao, C. V. D. Broeck, D. Baskaran, and T. G. F. Li, Determination of dark energy by the Einstein Telescope: Comparing with CMB, BAO, and SNIa observations, *Phys. Rev. D* **83**, 023005 (2011).
- [85] H.-Y. Chen, S. Vitale, and R. Narayan, Viewing angle of binary neutron star mergers, *Phys. Rev. X* **9**, 031028 (2019).
- [86] E. Nakar and T. Piran, Afterglow constraints on the viewing angle of binary neutron star mergers and determination of the Hubble constant, *Astrophys. J.* **909**, 114 (2021).
- [87] J. M. S. de Souza and R. Sturani, Luminosity distance uncertainties from gravitational wave detections of binary neutron stars by third generation observatories, *Phys. Rev. D* **108**, 043027 (2023).
- [88] K. Hotokezaka, E. Nakar, O. Gottlieb, S. Nissanke, K. Masuda, G. Hallinan, K. P. Mooley, and A. T. Deller, A Hubble constant measurement from superluminal motion of the jet in GW170817, *Nat. Astron.* **3**, 940 (2019).
- [89] S. Shah, G. Nelemans, and M. van der Sluys, Using electromagnetic observations to aid gravitational-wave parameter estimation of compact binaries observed with LISA, *Astron. Astrophys.* **553**, A82 (2013).
- [90] C. Hahn, M. J. Wilson, O. Ruiz-Macias, S. Cole, D. H. Weinberg, J. Moustakas, A. Kremin, J. L. Tinker *et al.*, The DESI bright galaxy survey: Final target selection, design, and validation, *Astron. J.* **165**, 253 (2023).
- [91] D. J. Schlegel, J. A. Kollmeier, G. Aldering, S. Bailey, C. Baltay, C. Bebek, S. BenZvi, R. Besuner *et al.*, The MegaMapper: A stage-5 spectroscopic instrument concept for the study of inflation and dark energy, [arXiv:2209.04322](https://arxiv.org/abs/2209.04322).
- [92] D. Brout, D. Scolnic, B. Popovic, A. G. Riess, A. Carr, J. Zuntz, R. Kessler, T. M. Davis *et al.*, The Pantheon+ analysis: Cosmological constraints, *Astrophys. J.* **938**, 110 (2022).
- [93] R. Takahashi, Amplitude and phase fluctuations for gravitational waves propagating through inhomogeneous mass distribution in the universe, *Astrophys. J.* **644**, 80 (2006).
- [94] R. Takahashi, T. Hamana, M. Shirasaki, T. Namikawa, T. Nishimichi, K. Osato, and K. Shiroyama, Full-sky gravitational lensing simulation for large-area galaxy surveys and cosmic microwave background experiments, *Astrophys. J.* **850**, 24 (2017).
- [95] G. Cusin, R. Durrer, and I. Dvorkin, Strong and weak lensing of gravitational waves: A semi-analytical approach, [arXiv:1912.11916](https://arxiv.org/abs/1912.11916).
- [96] A. J. Mead, S. Brieden, T. Tröster, and C. Heymans, HMcode-2020: Improved modelling of non-linear cosmological power spectra with baryonic feedback, *Mon. Not. R. Astron. Soc.* **502**, 1401 (2021).
- [97] C. M. Hirata, D. E. Holz, and C. Cutler, Reducing the weak lensing noise for the gravitational wave Hubble diagram using the non-Gaussianity of the magnification distribution, *Phys. Rev. D* **81**, 124046 (2010).
- [98] D. E. Holz and R. M. Wald, New method for determining cumulative gravitational lensing effects in inhomogeneous universes, *Phys. Rev. D* **58**, 063501 (1998).
- [99] T. Castro, M. Quartin, C. Giocoli, S. Borgani, and K. Dolag, The effect of baryons in the cosmological lensing PDFs, *Mon. Not. R. Astron. Soc.* **478**, 1305 (2018).
- [100] K. Osato, J. Liu, and Z. Haiman, κ TNG: Effect of baryonic processes on weak lensing with IllustrisTNG simulations, *Mon. Not. R. Astron. Soc.* **502**, 5593 (2021).
- [101] F. Ferlito, V. Springel, C. T. Davies, C. Hernández-Aguayo, R. Pakmor, M. Barrera, S. D. M. White, A. M. Delgado *et al.*, The MillenniumTNG project: The impact of baryons and massive neutrinos on high-resolution weak gravitational lensing convergence maps, *Mon. Not. R. Astron. Soc.* **524**, 5591 (2023).
- [102] K. C. Wong, S. H. Suyu, G. C.-F. Chen, C. E. Rusu, M. Millon, D. Sluse, V. Bonvin, C. D. Fassnacht *et al.*, HOLICOW—XIII. A 2.4 per cent measurement of H_0 from lensed quasars: 5.3 σ tension between early- and late-Universe probes, *Mon. Not. R. Astron. Soc.* **498**, 1420 (2019).
- [103] E. Belgacem, G. Calcagni, M. Crisostomi, C. Dalang, Y. Dirian, J. M. Ezquiaga, M. Fasiello, S. Foffa *et al.*, Testing modified gravity at cosmological distances with LISA standard sirens, *J. Cosmol. Astropart. Phys.* **07** (2019) 024.
- [104] B. Hadzhiyska, S. Yuan, C. Blake, D. J. Eisenstein, J. Aguilar, S. Ahlen, D. Brooks, T. Claybaugh *et al.*, Synthetic light-cone catalogues of modern redshift and weak lensing surveys with AbacusSummit, *Mon. Not. R. Astron. Soc.* **525**, 4367 (2023).
- [105] J. Harnois-Déraps, A. Amon, A. Choi, V. Demchenko, C. Heymans, A. Kannawadi, R. Nakajima, E. Sirks *et al.*, Cosmological simulations for combined-probe analyses: Covariance and neighbour-exclusion bias, *Mon. Not. R. Astron. Soc.* **481**, 1337 (2018).
- [106] J. DeRose, R. H. Wechsler, M. R. Becker, M. T. Busha, E. S. Rykoff, N. MacCrann, B. Erickson, A. E. Evrard *et al.*, The buzzard flock: Dark energy survey synthetic sky catalogs, [arXiv:1901.02401](https://arxiv.org/abs/1901.02401).
- [107] D. Korytov, A. Hearin, E. Kovacs, P. Larsen, E. Rangel, J. Hollowed, A. J. Benson, K. Heitmann *et al.*, CosmoDC2: A synthetic sky catalog for dark energy science with LSST, *Astrophys. J. Suppl. Ser.* **245**, 26 (2019).
- [108] T. Kacprzak, J. Fluri, A. Schneider, A. Refregier, and J. Stadel, CosmoGridV1: A simulated Λ CDM theory prediction for map-level cosmological inference, *J. Cosmol. Astropart. Phys.* **02** (2023) 050.
- [109] C. Wei, G. Li, X. Kang, Y. Luo, Q. Xia, P. Wang, X. Yang, H. Wang *et al.*, Full-sky ray-tracing simulation of weak lensing using ELUCID simulations: Exploring galaxy intrinsic alignment and cosmic shear correlations, *Astrophys. J.* **853**, 25 (2018).

- [110] S. Pires, V. Vandenbussche, V. Kansal, R. Bender, L. Blot, D. Bonino, A. Boucaud, J. Brinchmann *et al.*, Euclid: Reconstruction of weak-lensing mass maps for non-Gaussianity studies, *Astron. Astrophys.* **638**, A141 (2020).
- [111] LSST Dark Energy Science Collaboration, Large Synoptic Survey Telescope: Dark Energy Science Collaboration, [arXiv:1211.0310](https://arxiv.org/abs/1211.0310).
- [112] R. Laureijs, J. Amiaux, S. Arduini, J.L. Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin *et al.*, Euclid definition study report, [arXiv:1110.3193](https://arxiv.org/abs/1110.3193).
- [113] A. Hall, Cosmology from weak lensing alone and implications for the Hubble tension, *Mon. Not. R. Astron. Soc.* **505**, 4935 (2021).
- [114] E. L. Wagoner, E. Rozo, X. Fang, M. Crocce, J. Elvin-Poole, and N. W. and, Linear systematics mitigation in galaxy clustering in the dark energy survey year 1 data, *Mon. Not. R. Astron. Soc.* **503**, 4349 (2021).
- [115] M. Corman, A. Ghosh, C. Escamilla-Rivera, M. A. Hendry, S. Marsat, and N. Tamanini, Constraining cosmological extra dimensions with gravitational wave standard sirens: From theory to current and future multimessenger observations, *Phys. Rev. D* **105**, 064061 (2022).
- [116] U. Dupletsa, J. Harms, B. Banerjee, M. Branchesi, B. Goncharov, A. Maselli, A. Oliveira, S. Ronchini, and J. Tassinio, GWFISH: A simulation software to evaluate parameter-estimation capabilities of gravitational-wave detector networks, *Astron. Comput.* **42**, 100671 (2023).