

Future prospects on testing extensions to Λ CDM through the weak lensing of gravitational waves

Charlie T. Mpetha¹,* Giuseppe Congedo¹, and Andy Taylor

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

 (Received 18 August 2022; accepted 2 May 2023; published 16 May 2023)

With planned space-based and 3rd generation ground-based gravitational wave detectors (LISA, Einstein Telescope, Cosmic Explorer), and proposed DeciHz detectors (DECIGO, Big Bang Observer), it is timely to explore statistical cosmological tests that can be employed with the forthcoming plethora of data, 10^4 – 10^6 mergers a year. We forecast the combination of the standard siren measurement with the weak lensing of gravitational waves from binary mergers. For 10 years of 3rd generation detector runtime, this joint analysis will constrain the dark energy equation of state with marginalized 1σ uncertainties of $\sigma(w_0) \sim 0.005$ and $\sigma(w_a) \sim 0.04$. This is comparable to or better than forecasts for future galaxy/intensity mapping surveys, and better constraints are possible when combining these and other future probes with gravitational waves. We find that combining mergers with and without an electromagnetic counterpart helps break parameter degeneracies. Using DeciHz detectors in the post-LISA era, we demonstrate for the first time how merging binaries could achieve a precision on the sum of neutrino masses of $\sigma(\Sigma m_\nu) \sim 0.05$ eV using 3×10^6 sources up to $z = 3.5$ with a distance uncertainty of 1%, and \sim percent or subpercent precision also on curvature, dark energy, and other parameters, independently from other probes. Finally, we demonstrate how the cosmology dependence in the redshift distribution of mergers can be exploited to improve dark energy constraints if the cosmic merger rate is known, instead of relying on measured distributions as is standard in cosmology. In the coming decades gravitational waves will become a formidable probe of both geometry and large scale structure.

DOI: [10.1103/PhysRevD.107.103518](https://doi.org/10.1103/PhysRevD.107.103518)

I. INTRODUCTION

Gravitational waves (GWs) are becoming a powerful cosmological tool. In 2017, coincident GW [1] and electromagnetic (EM) [2,3] observations of a merging binary neutron star system heralded the beginning of multimessenger GW cosmology. This event, named GW170817, was used to measure the present expansion rate of the Universe H_0 . The associated gamma ray burst GRB170817A allowed localization of the binary to its host galaxy, providing a redshift. This, combined with the distance to the source found using properties of the GW waveform, gave a value of $H_0 = 70.0^{+12.0}_{-8.0}$ km s⁻¹ Mpc⁻¹ [4]. Precise, independent measurements of H_0 are pertinent due to the apparent tension between local Universe measurements of H_0 (such as type Ia supernovae [5] and lensed quasars [6]) and the value inferred from the cosmic

microwave background [7]. Just 50 multimessenger events similar to GW170817 would give us sufficient precision on H_0 to favor the early or late universe measurement [8]. The planned improvement to the current LIGO-Virgo-Kagra (LVK) network [9] makes this goal highly achievable within the next few years, motivating exploring the further possibilities of using GWs, beyond a standard siren measurement of H_0 . By the 2040's, several other GW detectors (Einstein Telescope [10], Cosmic Explorer [11], LISA [12]), both ground- and spaced-based, will be either taking data or in the later stages of their development. The combination of these detectors will allow an exploration of more GW frequency ranges (source types) than LVK, and with much improved sensitivity. This will dramatically increase the scope of GWs for astrophysics and cosmology. Large numbers of GWs from binary systems will be ideal for statistical cosmological tests. This study aims to evaluate the prospects of the weak lensing of gravitational waves, where correlations in the small perturbations of the GW propagation can be used to infer properties of the intervening matter distribution. Such observations open up the possibility of using GWs as a novel probe of not only the geometry of the Universe, but also of large scale structure.

*c.mpetha@ed.ac.uk

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

There are a few advantages of performing a weak lensing analysis using GWs instead of galaxies, besides it being a novel probe. It is a very clean measurement, avoiding systematics such as intrinsic alignments or blending which plague present and future weak lensing studies [13]. GW detectors will also observe mergers to much higher redshifts, up to $z \sim 100$. As in the standard siren case, the limiting factor in this analysis is the redshift determination. If there was a reliable GW only redshift inference method, for example sufficient source numbers for a “spectral siren” analysis [14], then properties of the large scale matter distribution could be probed up to very high redshifts. Although the lensing window function peaks at low redshift, direct observation of the “Dark Ages” would give us valuable observational information on structure formation, such that we do not need to extrapolate the properties of structure in this redshift regime from models. The key challenge of GW weak lensing compared to galaxy weak lensing is the difficulty of observing a GW and distinguishing it from detector noise.

The possibility for using the weak lensing of GWs (GW-WL) as a probe of cosmology was first discussed in Ref. [15] and developed in Ref. [16]. GW-WL forecasting was performed in Congedo and Taylor [17] (henceforth CT19), where it was demonstrated a joint analysis of GWs, utilizing both the standard siren distance measurement and the matter field information encapsulated in their weak lensing, can break key degeneracies giving better constraints than would be obtained individually. WL is largely insensitive to the Hubble parameter H_0 [18], while a joint standard siren + GW-WL analysis is sensitive to the expansion rate. The standard siren measurement provides a tight constraint on geometry parameters such as the expansion rate H_0 and matter density fraction Ω_m , without relying on an external dataset. In CT19 the authors assumed a number density of GW sources of 1 deg^{-2} (15000 sources), and that each was a well localized event similar to GW170817 (a binary neutron star merger with an EM counterpart). But the performance of future detectors is currently speculative, and highly dependent on the unknown distribution and number of binary mergers in the Universe. To fully explore the potential of cosmological constraints through a joint analysis of standard sirens and GW-WL, a range of assumptions of source properties is needed. The analysis should also be extended to varying models of dark energy and a varying neutrino mass—parameters of interest over the coming decades.

Here we build upon the work of CT19. Tomographic weak lensing is used to exploit all available cosmological information. A variety of assumptions on source and detector properties are made, and we extend to more cosmological models to explore how GWs can be used to answer present problems in cosmology. We also include sources without an associated EM counterpart, which will comprise the bulk of future GW detections. We find that

combining two merger populations (sources with and without an EM counterpart) helps break parameter degeneracies and improve constraints in the w_0 - w_a and Ω_m - S_8 planes, where the clustering parameter $S_8 = \sigma_8 \sqrt{\Omega_m}/0.3$. We show how a cosmologically varying source redshift distribution in the weak lensing analysis can further improve dark energy constraints. This is an advantage of gravitational wave studies where an astrophysics motivated cosmic merger rate density is possible. For a combination of GW detectors in the 2040’s with EM telescopes, we find a standard siren + weak lensing analysis will be capable of improving upon dark energy equation of state parameter constraints from galaxy surveys such as Euclid [19] and the Vera C. Rubin Observatory [20], and intensity mapping surveys such as HIRAX [21]. As we extend further into the future, if DeciHz GW detectors such as DECIGO [22] and the Big Bang Observer [23] come to fruition then the weak lensing of GWs will not only provide ultra precise constraints on geometry parameters, but could even give us valuable information on structure formation [24] and constrain the sum of neutrino masses Σm_ν . This is the first demonstration of GWs from merging binaries being an independent probe on neutrinos. We show that, by combining GWs with redshifts from galaxy surveys, GW-WL will be competitive with other single probes such as The Vera C. Rubin Observatory [25] and CMB-S4 [26] in constraining Σm_ν . Even in the most cosmology agnostic model used, $\nu k w$ CDM (curvature, dark energy and neutrinos all free parameters), we obtain percent or subpercent level errors on all parameters, apart from Σm_ν . The clean and well understood measurement from GWs could be a valuable source of extra information when combining with other future surveys.

Section II details future planned GW detectors, sources of GWs and their use for cosmology. In Sec. III the physics of the weak lensing of gravitational waves is introduced, before the modeling used in this analysis, including source assumptions and detector uncertainties, is described in Sec. IV. Finally the results are presented in Sec. V and we conclude in Sec. VI.

II. STANDARD SIRENS: SOURCES AND FUTURE DETECTORS

Observing gravitational waves from compact object mergers provides the luminosity distance d_L of the source, through determination of the amplitude and frequency evolution of the wave. This measurement is model-independent, giving it a clear advantage over, for example, type Ia supernovae which depend on calibration of the cosmic distance ladder. This direct measurement of the distance from the waveform makes GWs “standard sirens.” The observed gravitational wave amplitude is given by the linear combination of the + and \times polarization of the GW, combined with their corresponding antenna pattern functions F describing a detectors response to the wave:

$$h = F_+ h_+ + F_\times h_\times. \quad (1)$$

To leading order, the individual components of the gravitational wave are given by

$$h_+ = \frac{4}{d_L} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f}{c} \right)^{2/3} \frac{(1 + \cos^2 \iota)}{2} \cos[\Phi(f)], \quad (2)$$

$$h_\times = \frac{4}{d_L} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f}{c} \right)^{2/3} \cos \iota \sin[\Phi(f)]. \quad (3)$$

The speed of light is given by c . \mathcal{M}_c is the redshifted chirp mass, $\mathcal{M}_c = (1+z)(M_1 M_2)^{3/5} / (M_1 + M_2)^{1/5}$ (M_1 and M_2 are the component masses), f the frequency, ι is the inclination angle and Φ is the phase. Observable binary mergers include binary neutron stars (BNS), a black hole and a neutron star (BHNS) and binary black holes (BBH), and the inspiral of these sources is quite universal. They can, however, be distinguished during the binary merger and ringdown through higher order terms in the parametrized post-Newtonian formalism of the waveform. For more detail see Refs [27,28].

The expansion rate $H(z)$ at redshift z is related to d_L by

$$d_L(z) = (1+z) \int_0^z \frac{cdz'}{H(z')}, \quad (4)$$

where

$$\left(\frac{H(z)}{H_0} \right)^2 = \Omega_m (1+z)^3 + \Omega_K (1+z)^2 + \Omega_{\text{DE}} (1+z)^{3(1+w_0+w_a)} e^{\frac{-3w_a z}{1+z}}. \quad (5)$$

H_0 is the present Universe expansion rate ($h=H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is the dimensionless expansion rate), Ω_m , Ω_K and Ω_{DE} are the present matter, curvature and dark energy density parameters, while the widely used CPL parametrization [29,30] of a time-varying dark energy equation of state (EoS)

$$\frac{p_{\text{DE}}}{\rho_{\text{DE}}} \equiv w_{\text{DE}} = w_0 + w_a \frac{z}{1+z}, \quad (6)$$

is assumed. w_0 is the dark energy EoS today, and w_a is its growth rate with the scale factor a . This functional form can accurately recreate the EoS for alternative dark energy models, such as a scalar field dark energy. Encapsulated in Ω_m is the baryonic, cold dark matter and neutrino density components,

$$\Omega_m = \Omega_b + \Omega_{\text{CDM}} + \Omega_\nu. \quad (7)$$

In Λ CDM most of the matter in the Universe is cold dark matter (CDM), dark energy is a cosmological constant Λ

($\Omega_{\text{DE}} = \Omega_\Lambda$) with a constant EoS of $w_\Lambda = -1$ ($w_0 = -1$ and $w_a = 0$) and there is zero curvature ($\Omega_K = 0$).

If the redshift of the GW source can also be determined, then a $d_L - z$ relation can be used to probe the cosmological parameters in Eqs (4)–(7). This redshift determination is the limiting factor for cosmological applications of GWs. Due to the mass-redshift degeneracy in the GW waveform, some method of breaking this degeneracy, or external data, is needed to determine the redshift. GW sources can be broadly split into two categories, depending on whether they can or cannot be localized to their host galaxy. These two categories are bright standard sirens (BSS) and dark standard sirens (DSS). BSS represent an ideal case for cosmology. Either an associated EM counterpart, or very precise sky localization due to a high SNR, allows an accurate source redshift determination [31]. For multimessenger events, there is the possibility to probe deviations from GR through comparing the EM and GW propagation [32]. In the case of DSS, there is no associated electromagnetic counterpart and the sky localization is too poor to identify the host galaxy. These sources make up the bulk of gravitational wave detections, therefore statistical redshift inference methods for large numbers of DSS are receiving more attention in the ramp up toward 3rd generation (3G) GW detectors.

BNS mergers may produce an observable EM counterpart, depending on the merger distance and inclination angle. Supermassive black hole (SMBH) and BHNS mergers may produce an EM counterpart, but for SMBH mergers their rate is too small for statistical cosmology applications [33]. For BHNS mergers their rate is uncertain, as is the fraction of these that will produce a detectable counterpart, though the expectation is they will mostly be DSS [34]. Most DSS are stellar binary black hole (BBH) mergers, and for these several redshift inference methods have been proposed in the absence of a counterpart. These include: a statistical average of galaxies within the GW localization error box [31,35,36] which has already been performed on LVK data [37], correlating the clustering of GW sources with the clustering of galaxies [38,39] or another tracer such as HI intensity mapping [40], by breaking the mass-redshift degeneracy using information on the source frame mass distribution [14,41,42], tidal corrections in the late-inspiral signal of a BNS merger [43], or an expected cosmic merger rate density of sources [44,45].

The so-called 3G ground-based GW detectors expected during the 2030s–2040s include the triangular configuration Einstein Telescope (ET), and Cosmic Explorer (CE) which is similar to LVK with longer arm lengths. In the same time period we will see the first space-based GW detector, LISA. The combined ground-based detectors will observe in the frequency range 1–10³ Hz and be sensitive to most BBH and BNS mergers in the Universe [42,46], observing $\mathcal{O}(10^4\text{--}10^6)$ binary mergers every year [47,48].

The large uncertainty is due to the uncertain cosmic merger rates for different binary populations (see e.g [49] for a summary). LISA, observing in the 10^{-4} –1 Hz frequency range, will observe SMBH inspirals, stellar binaries before their inspiral and merger is detected by ground-based observatories [50], as well as many extreme mass-ratio inspirals (EMRI), which can also be used for cosmology [51]. Space-based GW detectors observing in the DeciHz regime, such as DECIGO [22] and Big Bang Observer (BBO) [23], are possible successors to LISA and will bridge the gap between these two frequency ranges. Their design is far more ambitious, with four LISA-like constellations (with smaller arm lengths) in heliocentric orbits, two of which at the same location and the other two distributed around the Sun. They could observe the inspiral of binaries over periods of months before they enter the ground-based frequency range during their merger. This, coupled with their changing position and orientation, allows observations of similar numbers of mergers to 3G detectors but with much improved sky localization, sufficient to localize a merger to its host galaxy even in the absence of a transient EM counterpart. For example BBO could contribute $\mathcal{O}(10^5)$ highly localized (\sim few arcsec) binary mergers per year [15]. But while 3G ground-based detectors such as ET and CE are in the later stages of their development, the difficulty of space missions makes the future of detectors such as DECIGO and BBO less certain, and strongly dependent on the success of LISA.

To summarize, from 3G detectors we can expect $\mathcal{O}(10^5\text{--}10^7)$ sources observed by ground-based observatories by the end of the 2040s, and comparable numbers with much improved sky localization in the second half of the century should space-based DeciHz detectors be launched.

III. WEAK LENSING OF GRAVITATIONAL WAVES

The gravitational potential that is created by large-scale structure in the Universe induces small perturbations in the path of propagating radiation. The radiation is weakly lensed, and the resulting magnification of both electromagnetic [52,53] and gravitational [54] radiation in the geometric optics regime is given by

$$\mu \approx 1 + 2\kappa, \quad (8)$$

where κ is the lensing convergence, a weighted projection of density perturbations along the line-of-sight.

For GWs the situation is seen in Fig. 1. The amplitude of the wave is increased as the wave is lensed by an overdensity, leading to a smaller inferred luminosity distance. By relating the lensed and unlensed fluxes we obtain an expression for the observed, lensed luminosity distance d_L^{obs} in terms of the true d_L and the convergence:

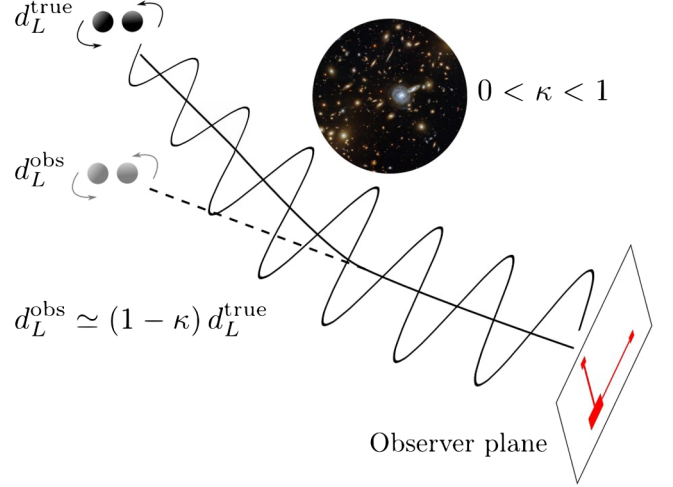


FIG. 1. Gravitational waves being lensed by a large matter overdensity. The magnification results in a larger wave amplitude, and a smaller inferred distance to the source. The size of the deflection and change in amplitude has been exaggerated for illustration purposes.

$$d_L^{\text{obs}} = \frac{1}{\sqrt{\mu}} d_L^{\text{true}} \simeq (1 - \kappa) d_L^{\text{true}}. \quad (9)$$

There is also a small modification to the gravitational wave phase. In the geometric optics limit any phase fluctuations will just correspond to an arrival time shift making them observationally irrelevant in the case of weak lensing [55]. The strong lensing of gravitational waves, in the wave optics regime, is another science application of lensed GWs. In this case, it is useful for constraining properties of the lens and source populations [56], or as a further test of geometry [57]. To use GWs to explore the large scale matter distribution we turn to their weak lensing.

From a set of many mergers, a best fit $d_L - z$ relation can be constructed, sensitive to cosmological parameters. This is the standard siren measurement. The weak lensing of each source will create scatter around this best fit curve of order 1%, and due to the cosmological principle an average from large numbers of sources will recover the true relationship. Hence from this scatter, each source provides a point estimate of the convergence field. Though the lensing error dominates toward higher redshifts, uncertainties introduced by the source redshift, distance measurement and peculiar velocities of the sources make this a noisy estimate of the convergence field.

Sources, in our case binary mergers, are binned in redshift intervals. The variance of κ at different multipole modes ℓ (the Fourier conjugate of the angular separation) across tomographic redshift bins i and j is found through a two-point correlation function of the Fourier transformed convergence field in each bin,

$$\langle \tilde{\kappa}_{(i)}(\ell) \tilde{\kappa}_{(j)}^*(\ell') \rangle = (2\pi)^2 \delta_D(\ell - \ell') C_{(ij)}^{\kappa\kappa}(\ell), \quad (10)$$

where $C_{(ij)}^{\kappa\kappa}(\ell)$ is the convergence power spectrum. This power spectrum is sensitive to cosmological parameters, and provides a complementary probe of information to the standard sirens, through a different filtering of the same data. The auto-correlation for each redshift bin corresponds to $i = j$. Cross-correlating between bins adds information on the time variation of cosmological parameters. The convergence power spectrum between the i th and j th redshift bin is given by

$$C_{(ij)}^{\kappa\kappa}(\ell) = \int_0^{z_{\max}} \frac{c(1+z)^2}{H(z)} \mathcal{W}_{(i)}(z) \mathcal{W}_{(j)}(z) \times P_{\delta}\left(k = \frac{\ell + 1/2}{\mathcal{K}(r)}, z\right) dz, \quad (11)$$

where

$$\mathcal{W}_{(i)}(z) = \frac{3}{2} \Omega_{m,0} \left(\frac{H_0}{c}\right)^2 \int_{z_i}^{z_{i+1}} p(z') \frac{\mathcal{K}(r' - r)}{\mathcal{K}(r')} dz' \quad (12)$$

is the Window function, containing the normalized observed source distribution $p(z)$ and the effect of the lens on this source due to their separation. The comoving distance r is related to the transverse comoving distance $\mathcal{K}(r)$ by

$$\mathcal{K}(r) = \begin{cases} \frac{1}{\sqrt{|K|}} \sinh(\sqrt{|K|}r), & K < 0, \\ r, & K = 0, \\ \frac{1}{\sqrt{|K|}} \sin(\sqrt{|K|}r), & K > 0, \end{cases} \quad (13)$$

where K is the curvature.

The Limber approximation for k -modes, $k = (\ell + 1/2)/\mathcal{K}(r)$, is shown in the matter power spectrum $P_{\delta}(k, z)$, and is used for large values of ℓ . A common parameter to normalize the matter power spectrum is the linear normalization of its amplitude in spheres of $8 h^{-1}$ Mpc, σ_8 [58],

$$\sigma_8^2 = \sigma_{R=8 h^{-1} \text{Mpc}}^2 = \int \frac{dk k^3 P_{\delta}(k, z)}{2\pi^2} W^2(k, R), \quad (14)$$

where $W(k, R)$ is a spherical top hat window function.

IV. MODELING BINARY MERGER OBSERVATIONS

A. Properties of binary merger populations

In this study we use certain assumptions on the properties of the BSS and DSS, summarized by the following.

- (i) For 3G detectors, all BSS events are assumed to be from merging binary neutron stars (BNS), while DSS are from both binary black hole (BBH) and black hole neutron star (BHNS) mergers. Though many observed BNS will also be DSS, BBH and

BHNS will be the dominant contributors. For DeciHz detectors we assume well-localized BBH and BHNS mergers also contribute to the BSS sample. This is because for these detectors the constellations of satellites distributed around the Sun provide excellent source sky localization [23], allowing many GWs to be localized to a host galaxy from which a redshift can be determined.

- (ii) The redshift uncertainty for BSS,

$$\sigma_{z,\text{BSS}} = \sigma_{0,\text{BSS}}(1+z), \quad (15)$$

is spectroscopic, with $\sigma_{0,\text{BSS}} = 0.001$ [4].

- (iii) For DSS, a localization uncertainty of 1 deg^2 is adopted [59,60]. The redshift is determined by a statistical average of galaxies in the localization error box. We set

$$\sigma_{z,\text{DSS}} = \frac{\sigma_{0,\text{DSS}}}{\sqrt{f(z)g(z)}}(1+z). \quad (16)$$

Here, $\sigma_{0,\text{DSS}} = 0.03$ representing expected photometric uncertainties from the Vera C. Rubin Observatory [61]. Future high redshift spectroscopic galaxy surveys, such as MegaMapper [62], could allow a statistical DSS method with spectroscopic redshifts (spec- z). The two functions $f(z)$ and $g(z)$ represent the GW detector and galaxy survey selection functions respectively, and are discussed further below.

- (iv) BSS events are observed up to $z_{\max} = 2$ based on 3G GW detector horizons [11,63] and future spectroscopic galaxy surveys [19,64].
- (v) For DSS, $z_{\max} = 3.5$, from the limit of future photometric galaxy surveys [65]. These galaxies are needed to estimate the source redshift.
- (vi) There are typically expected to be a factor of 10 more BBH than BNS detections [10,66–68]. Only a fraction of BNS mergers will have EM counterparts. From this we assume $N_{\text{DSS}}/N_{\text{BSS}} = 100$ for 3G detectors, consistent with current LVK detections [69].

The observed redshift distribution of binary mergers $p(z)$ is a combination of different factors,

$$p(z) = p^{(\text{th})}(z)g(z)f(z). \quad (17)$$

For a given population of binaries, $p^{(\text{th})}(z)$ is a theoretical probability distribution derived from a theoretical cosmic merger rate density $\mathcal{R}(z)$, which can be modeled from the star formation rate and binary synthesis properties. The source number distribution is found from

$$n^{(\text{th})}(z) = 4\pi \frac{\mathcal{R}(z)}{(1+z)} \frac{dV_c}{dz} = 4\pi c \frac{\mathcal{R}(z)}{(1+z)} \frac{r^2(z)}{H(z)}, \quad (18)$$

where V_c is the comoving volume. We use the BPASS predictions for transients [70], which provides $\mathcal{R}(z)$ for different binary populations separately, including BNS, BBH and BHNS. This number distribution is then normalized to find the probability distribution,

$$p^{(\text{th})}(z) = \frac{n^{(\text{th})}(z)}{\int_0^z n^{(\text{th})}(z') dz'}. \quad (19)$$

Depending on GW source and detector parameters such as the sensitivity curve, antenna patterns, inclination angle, compact object masses etc., the GW may or may not be detected. This is modeled by the selection function $g(z)$. In Leandro *et al.* [71], using a Fisher matrix analysis and SNR limit of 8 for merging BBHs, the authors find for a choice of 3G detector,

$$g(z) \propto e^{-(r(z)/r_{\text{cut}})^3}, \quad (20)$$

where the value r_{cut} for BBH mergers is given as 7.9 Gpc, which we adopt for the DSS case. This form is found by fitting a distribution of realistic sources that are above the SNR threshold as a function of comoving distance. In the BSS case a value of $r_{\text{cut}} = 4$ Gpc is found by assuming the same functional form, but with a detector redshift limit of $z = 2$, as will be the case for BNS observations.

The inclusion of the galaxy survey selection function $f(z)$ in Eq. (17) is important as we are assuming the redshift determination of merging compact objects is through galaxy observations, either from host galaxy determination, or from a statistical average of galaxies in a photometric galaxy survey. The inclusion of a source in the observed distribution depends on whether or not it has an associated redshift, hence depends on a relevant galaxy survey selection function. A simplistic choice has been made in this case, of a tophat function with a probability of 0.9 (completeness), then a smooth decay before reaching the maximum redshift. The shape is determined by a pivot redshift ($z_{\text{pivot}} = 1.8$ for BSS [72] and 2.8 for DSS [65]),

$$f(z) = \frac{0.9(1 - \tanh(z - z_{\text{pivot}}))}{2wz_{\text{pivot}}}, \quad (21)$$

$$w = \frac{z_{\text{max}} - z_{\text{pivot}}}{2z_{\text{max}}}. \quad (22)$$

These functions are also used to estimate the redshift uncertainty in the DSS case by acting as a shot noise scaling as shown in Eq. (16), recovering the expected behavior of near-photometric uncertainties at low redshift [51,68], and a rapid dilution to higher redshifts as the number of GW and galaxy observations decreases. The resulting $p^{(\text{obs})}(z)$ is shown for both BSS and DSS in Fig. 2. The sensitivity of the results to the forms of $g(z)$ and $f(z)$ is explored in Appendix A 1.

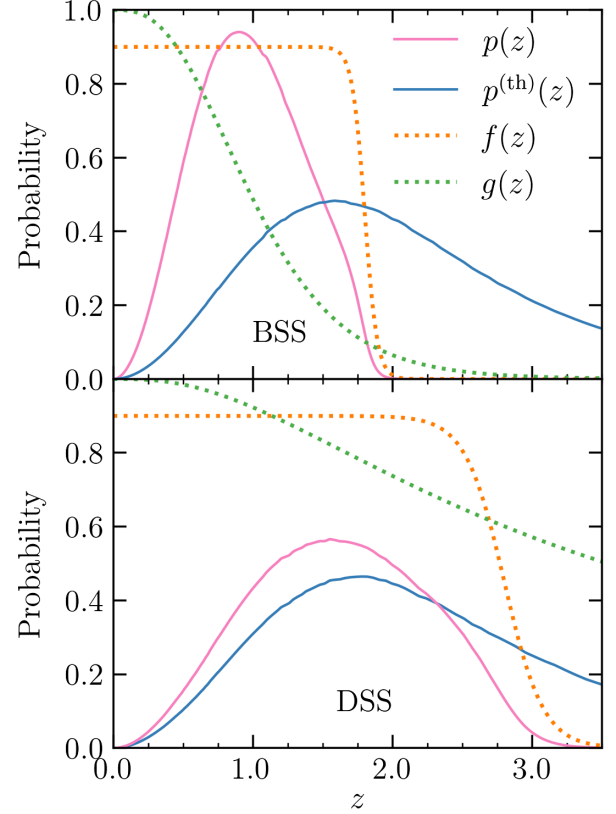


FIG. 2. $p(z)$ is the observed redshift distribution of gravitational waves from binary mergers for 3G ground-based detectors. This is split into two populations; binary neutron stars with an EM counterpart (bright standard sirens, top), and a combined population of binary black holes and black hole-neutron stars with a statistically obtained redshift (dark standard sirens, bottom). $p(z)$ is obtained by combining the theoretical normalized redshift distribution $p^{(\text{th})}(z)$ (found from BPASS simulated merger rate densities [70]) with approximate forms of the relevant galaxy survey selection function $f(z)$ and 3rd generation gravitational wave detector selection function $g(z)$.

B. Source and detector uncertainties

When using GWs as standard sirens, there are several sources of uncertainty. The GW detector instrumental uncertainty (σ_{GW}) is assumed to have a range of values from $0.004d_L$ to $0.1d_L$. The source redshift uncertainty (σ_z) is defined above. Uncertainty in the redshift owing to the source peculiar velocity ($\sigma_{v_{\text{pec}}}$) is given the form

$$\sigma_{v_{\text{pec}}} = \frac{\sqrt{\langle v_{\text{pec}}^2 \rangle}}{c} (1 + z). \quad (23)$$

An rms value of $\sqrt{\langle v_{\text{pec}}^2 \rangle} \sim 500 \text{ km s}^{-1}$ is adopted [73]. Finally is the uncertainty in the true value of d_L due to the weak lensing of the GWs (σ_{WL}). Different predictions for the size of σ_{WL} exist in the literature. A (pessimistic) fit used for Einstein Telescope forecasts [74,75] assumes 5%

uncertainty at $z = 1$ and linearly extrapolates: $\sigma_{\text{WL}}/d_L = 0.05z$. A more appropriate expression derived in Ref. [76] using properties of the magnification probability distribution up to $z = 3$ is given by

$$\frac{\sigma_{\text{WL}}}{d_L} = 0.066 \left(\frac{1 - (1+z)^{-0.25}}{0.25} \right)^{1.8}. \quad (24)$$

Other forms for σ_{WL} exist in the literature [77,78], Eq. (24) is used here as a middle estimate. There has been development in “delensing” the GW, by observing the matter distribution along the line of sight of the GW source [79]. This can reduce the lensing uncertainty—however it is not useful for our purposes, where we wish to extract information from the lensing of GWs. The combined uncertainty is

$$\sigma_{d_L}^2 = \sigma_{\text{GW}}^2 + \sigma_{\text{WL}}^2 + \left(\frac{\partial d_L}{\partial z} \right)^2 \left(\sigma_z^2 + \sigma_{v_{\text{pec}}}^2 \right). \quad (25)$$

The observed power spectrum is the underlying convergence power spectrum, modified by a shot noise term due to our finite number of sources.

$$\mathcal{C}_{(ij)}^{\text{obs}} = \mathcal{C}_{(ij)}^{\kappa\kappa}(\ell) + \delta_{ij} \frac{\sigma_{\kappa}^2}{\bar{n}} e^{(\ell/\ell_{\text{max}})^2}, \quad (26)$$

$$\sigma_{\kappa}^2 = \sigma_{\text{GW}}^2 + \left(\frac{\partial d_L}{\partial z} \right)^2 \left(\sigma_z^2 + \sigma_{v_{\text{pec}}}^2 \right), \quad (27)$$

where \bar{n} is the number density of GW sources, and δ_{ij} is the Kronecker delta ensuring shot noise only contributes for autocorrelations. Possible high-order correlations between the WL and GW signal are not included. To include the effects of sky localization on the convergence power spectrum, we assume that a Gaussian kernel damps the signal. The total observed power spectrum is then deconvolved with the same kernel, leaving the inverse-Gaussian term in the shot-noise error seen in Eq. (26) [39]. The maximum ℓ mode that can be probed is directly related to the localization area, the uncertainty blows up past this ℓ mode. If the localization area is given by $A_{\text{loc}} = \Omega \text{ deg}^2$, then $\ell_{\text{max}} \sim 180/\sqrt{\pi\Omega}$.

Finally, the covariance is given by

$$\begin{aligned} \text{Cov}[\mathcal{C}_{(ij)}^{\text{obs}}(\ell), \mathcal{C}_{(mn)}^{\text{obs}}(\ell')] \\ = \frac{\delta_{\ell\ell'}}{f_{\text{sky}}(2\ell+1)\Delta\ell} \left(\mathcal{C}_{(im)}^{\text{obs}}(\ell)\mathcal{C}_{(jn)}^{\text{obs}}(\ell) + \mathcal{C}_{(in)}^{\text{obs}}(\ell)\mathcal{C}_{(jm)}^{\text{obs}}(\ell) \right), \end{aligned} \quad (28)$$

where trispectrum terms have been neglected. In this work we assume an observable sky area of 15000 deg^2 giving $f_{\text{sky}} = 0.36$ based on Euclid. This value considers the sky blocked from view by the galactic disk and zodiacal plane,

and while GW observations are not limited by this, host galaxy redshifts are.

V. FORECASTING A JOINT STANDARD SIREN + WEAK LENSING ANALYSIS

The real strength of this approach comes from the combination of the standard siren and weak lensing analyses, where the standard sirens essentially provide a strong constraint on geometry in the weak lensing analysis, without needing to rely on external datasets. The Bayesian framework is outlined in CT19. Analytic derivatives can be found in the standard siren case, and numerical derivatives are needed when finding the Jacobian of the convergence power spectrum. Care is needed for these derivatives as the response of the power spectrum to a cosmological parameter can vary greatly over ℓ -modes and redshift bins. Our solution is to, for each cosmological parameter, use the median value from the optimal steps of all modes and redshift bin correlations, then test it for optimality and stability. More detail can be found in Appendix B. A Fisher information matrix is constructed and used to find the Cramér-Rao lower bound on physical parameters, which the model depends on. These are then mapped to observed parameter uncertainties using Monte Carlo methods, as the mapping is often nonlinear. We sum Fisher matrices to combine the results from the standard siren and weak lensing analyses, and all quoted constraints are 1σ marginalized uncertainties.

CLASS [80] with HMCODE [81] is used to compute σ_8 values and the auto- and cross- convergence power spectra between six redshift bins, where bin edges are defined such that there is an approximately equal number of sources per bin. A maximum ℓ mode of $\ell = 3000$ is used to stay clear of the highly nonlinear regime, where the statistical properties of the convergence field are close to Gaussian [82]. Otherwise we could produce overly optimistic forecasts by assuming perfect knowledge of a regime with uncertain modeling. This is due to the importance of baryonic physics on such small scales [83]. Nevertheless, there is still an assumption of much improved understanding of nonlinearity in the matter power spectrum by the 2040s–2050s. We investigate the effect of decreasing the maximum ℓ mode on forecasts in Appendix C. Fiducial parameter choices and the range of their flat priors are given in Table I, as well as a summary of the cosmological models used in this analysis, demonstrating which parameters are allowed to vary in which models. In this work we set $h^2\Omega_b = 0.0224$ as a fixed parameter, motivated by its strong CMB constraint from the Planck satellite [7].

A. 2040s: 3G GW detectors

In the 3rd generation of GW detectors, consisting of a network of ground-based interferometers such as the Einstein Telescope and Cosmic Explorer, there is expected

TABLE I. Parameters used in the Fisher forecasting analysis, their flat prior ranges, and the cosmological models in which they are treated as free (✓) or set to constant (✗). $k\Lambda$ CDM, w CDM, and ν CDM are Λ CDM with curvature, dark energy or neutrinos respectively as a free parameter. The most cosmology agnostic model in this analysis, νkw CDM, treats all these as free.

Parameter	Fiducial	Prior	Λ CDM	$k\Lambda$ CDM	Model w CDM	ν CDM	νkw CDM
Dimensionless Hubble parameter h	0.673	[0.5, 0.9]	✓	✓	✓	✓	✓
Total matter density parameter Ω_m	0.316	[0.1, 0.9]	✓	✓	✓	✓	✓
Scalar spectral index n_s	0.965	[0.5, 1.5]	✓	✓	✓	✓	✓
Scalar amplitude $\ln(10^{10}A_s)$	3.05	[2, 4]	✓	✓	✓	✓	✓
Curvature density parameter Ω_K	0	[-0.3, 0.3]	✗	✓	✗	✗	✓
Dark energy EoS parameter w_0	-1	[-2, 0]	✗	✗	✓	✗	✓
Dark energy EoS parameter w_a	0	[-2, 2]	✗	✗	✓	✗	✓
Sum of neutrino masses Σm_ν [eV]	0.06	[0.005, 1]	✗	✗	✗	✓	✓

to be up to $\sim 10^5$ yr $^{-1}$ observations of BNS mergers [84] and $\sim 10^6$ yr $^{-1}$ for BBH mergers [10,11,48]. In Ref. [85] the authors forecast the expected number of BNS mergers with an associated gamma ray burst counterpart (making them BSS), using different assumptions on the GW and gamma ray detectors. An upper limit from CE + GECAM of $N \sim 3 \times 10^4$ (translating to $\bar{n} \sim 2$ deg $^{-2}$ using our assumed fractional sky area of $f_{\text{sky}} = 0.36$) over a 10 year runtime is found, and the number could be higher when considering other GW and EM detectors. The inclusion of the ET would give comparable or slightly higher numbers—there would likely be numerous coincident detections. Predictions for joint GW + EM observations (gamma ray bursts or x-rays) are also found in Refs. [86,87], and are generally comparable or lower, agreeing with the range given in Ref. [85]. These numbers are fairly small for statistical cosmological tests. In this case of relatively small numbers of BSS, then it is vital the abundant DSS detections are also used for cosmology. With a redshift determined from a statistical average of galaxies in the localization error box of the GW, the redshift uncertainties of DSS are expected to be much larger than BSS. Despite this, the deeper redshift range and larger numbers means they can still be informative.

By the late 2030s, there may be other spectroscopic redshift surveys, for example the proposed MegaMapper [62] which would observe spec- z 's of $\sim 10^8$ galaxies from $2 < z < 5$. While not a benefit to spectroscopic follow up of BNS mergers, whose GW detector selection function drops off around $z = 2$, MegaMapper could be used to perform statistical dark standard siren redshift estimation using spec- z 's. To test this possibility, we reran the analysis replacing the redshift uncertainty in Eq. (16) with $\sigma_{0,\text{DSS}} = 0.001$. This decreases the combined BSS + DSS parameter constraints by a factor of 0.5–0.8.

Figure 3 gives marginalized 1σ and 2σ contours for constraining the dark energy equation of state parameters w_0 and w_a , and matter density and clustering parameters

Ω_m and $S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$ in the w CDM cosmological model (see Table I), using expected numbers after 10 years of 3G detector runtime. BSS are assumed to have an instrumental d_L error $\sigma_{\text{GW}}/d_L = 2\%$, while DSS have $\sigma_{\text{GW}}/d_L = 10\%$ [47]. This is motivated by the smaller number of well measured BSS, and the large number of DSS which will have a large range of instrumental uncertainties. It can be seen how, due to the different degeneracy lines between BSS and DSS, their combination leads to improved constraints. This demonstrates the importance of including the large population of DSS to the sample. Despite their noisy redshift estimate, they probe higher redshifts leading to an improved Ω_m and greater sensitivity to the time-variation of dark energy parameterized by w_a . But the larger sky localization uncertainty means the C_ℓ^{KK} result is degraded, hence the much weaker constraint on the clustering parameter S_8 , which can only be probed through the weak lensing measurement of GWs. The dark energy equation of state parameters can be constrained through the standard siren measurement alone, and this provides the most information. Combining the standard siren measurement with weak lensing tomography, as is done for all results, improves constraints on geometry parameters by $\sim 10\%$. Using 2D lensing (no tomographic bins) gives marginal added information for geometry parameters.

Figure 4 shows constraints for w_0 and w_a in w CDM against source numbers, which is a proxy for detector runtime. Using the uncertainty on the LVK inferred local Universe BNS merger rate $\mathcal{R}(0)$ [69], and assuming the same BPASS form for $\mathcal{R}(z)$, an estimate for the range of total mergers occurring over all cosmic history can be obtained, as an indication of future detection limits. This is done by normalizing the BPASS $\mathcal{R}(z)$ to each upper and lower bound of $\mathcal{R}(0)$ in turn, substituting into Eq. (18) and integrating over the redshift range. This range over a 10 year runtime is shown by the red vertical lines. The forecasts for the combined standard siren and GW weak lensing constraints

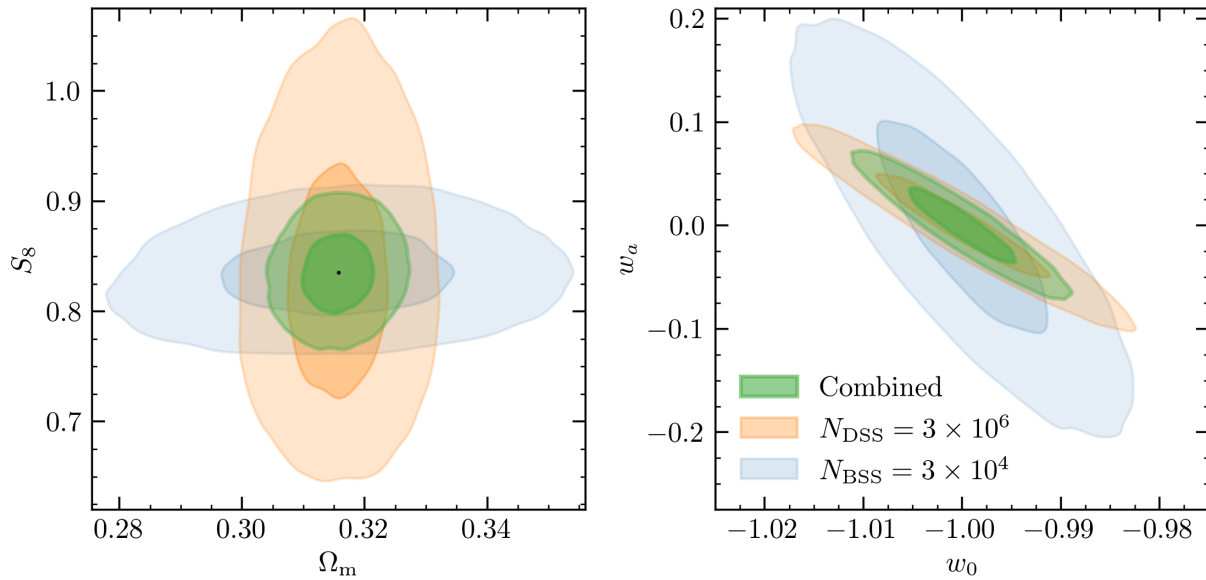


FIG. 3. Forecasts for combined gravitational wave weak lensing + standard siren constraints on cosmological parameters in the w CDM model (see Table I) using realistic numbers for 3G detectors. Bright standard sirens (BSS) have a smaller redshift range but more accurate redshift determination through localization to a host galaxy, while dark standard sirens (DSS) can be observed up to larger redshifts, but their redshift is estimated using statistical methods and so is much noisier. We set the GW detector distance uncertainty to 2% for BSS and 10% for DSS.

using a population of BSS only and a range of values for the instrumental d_L uncertainty from 1%–10% (green lines) are in broad agreement with those presented in Ref. [85]. The

combination of these populations of BSS with a population of DSS with $N_{\text{DSS}} = 100 N_{\text{BSS}}$ and an instrumental uncertainty of 10% (blue lines) demonstrates the importance of

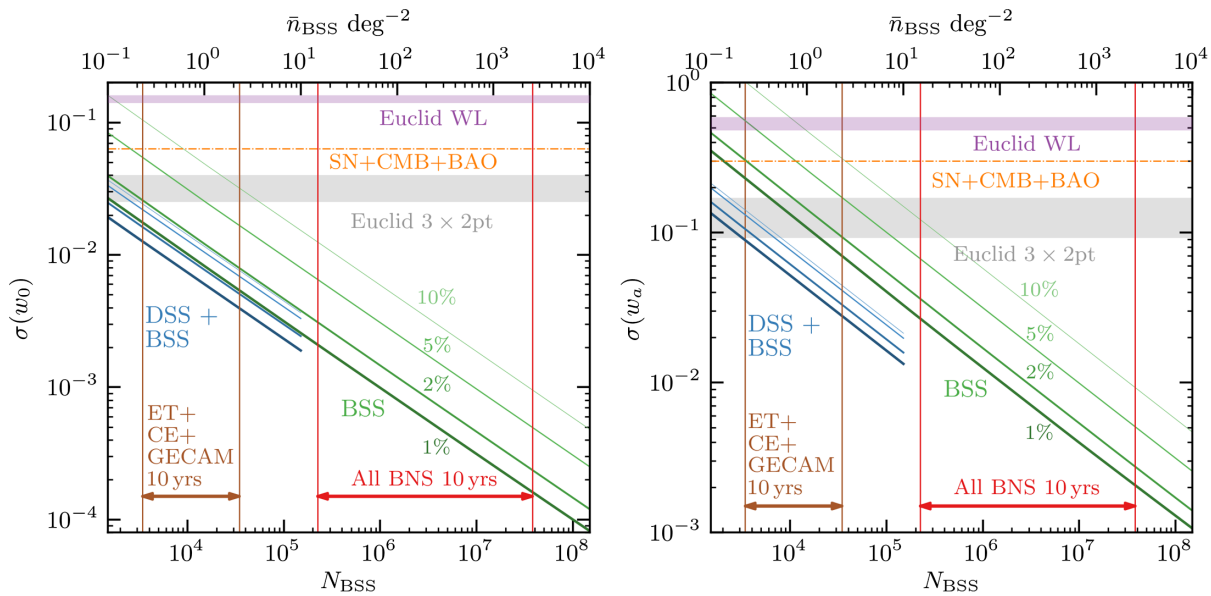


FIG. 4. Forecasted errors in the w CDM model (see Table I) on the dark energy equation of state parameters w_0 (left) and w_a (right) found from a joint gravitational wave weak lensing + standard siren analysis, as a function of the number (density) of bright standard siren (BSS) observations. The green lines use a population of BSS with spectroscopic redshifts and a range of GW detector d_L uncertainties, from 1%–10%. The blue lines combine these populations with a set of dark standard sirens (DSS) from binary black holes and black hole-neutron stars, with a statistical redshift determination and detector d_L uncertainty of 10%. It is assumed that $N_{\text{DSS}} = 100 N_{\text{BSS}}$. Horizontal bands give Euclid forecasts from either weak lensing (WL) alone, or the combined 3×2 pt analysis. The orange dash-dot line is the most precise present constraint from Supernovae + CMB + BAO. The regions bound by vertical lines give expected numbers of BSS from the combination of 3G detectors Einstein Telescope (ET) and Cosmic Explorer (CE) with the gamma ray detector GECAM in brown, and a theoretical range for all BNS mergers in the Universe in red, both over 10 years of detector runtime.

including the large number of noisier, but higher redshift, sources. The blue lines stop at $\bar{n}_{\text{BSS}} = 10 \text{ deg}^{-2}$ as DSS source numbers greater than 10^3 deg^{-2} are unlikely, indicated by the LVK inferred limits of all BNS mergers. For expected numbers of sources from 3G GW detectors, by combining BSS and DSS observations, constraints on w_0 and w_a will improve upon those by the Euclid 3×2 pt analysis [88] by a factor of 3–10. The Vera C. Rubin Observatory predicts similar constraints [20,89], therefore a joint standard siren + GW-WL analysis using 3G GW detectors will improve further dark energy constraints from future galaxy surveys. Expected forecasts from 3G detectors for all parameters can be found in Appendix D.

Fig. 4 shows how increasing the detector uncertainty of the BSS sources from 1% to 10% in the combined BSS + DSS constraints (blue lines) increases these combined constraints by a factor of 2, indicating that the forecasts are sensitive to the assumed instrumental error. However, there is still a significant gain in information when combining the two merger populations, demonstrated by the difference between the green and blue lines. These forecasts are also sensitive to the observed distribution of mergers and hence on the choice of population synthesis code used, in this case BPASS [70]. It has been shown how the choice of the star formation rate and metallicity evolution used by the population synthesis code can have a large impact on the distribution of BBH mergers, less so on the BNS merger distribution [90]. Forecasts were also produced using different $p(z)$ assumptions to test the sensitivity of our results to this choice, see Appendix A 2. As expected different redshift distributions can alter the parameter degeneracies and resulting constraints, but this effect is marginal, and does not affect the main conclusions of the paper.

B. 2050s and beyond: DeciHz GW detectors

DeciHz detectors will observe the inspiral of orbiting binaries for months to years before merger as the satellites orbit the Sun. The resulting precise d_L determination and sky localization will allow identification to a host galaxy for many sources, without the need for an EM transient counterpart [15,91–93]. This greatly increases the expected number of BSS, comparable to the number of DSS.

In Fig. 4, when we extend to source numbers only possible with DeciHz detectors the precision will be significantly improved (for example see Table III). This huge population of sources with accurate redshifts allows weakly lensed GWs to be used for more challenging statistical cosmological tests. The value of the sum of neutrino masses Σm_ν is presently only constrained by upper bounds. The most stringent constraint is $\Sigma m_\nu < 0.12 \text{ eV}$ [95% confidence limit (C.L.)] coming from Planck 2018 including lensing + BAO [7]. The value of Σm_ν impacts structure formation, and precise determinations could shed light on whether neutrinos follow a normal or inverted

hierarchy [94], adding valuable information to the question of whether neutrinos are Dirac or Majorana fermions [95]. The normal hierarchy predicts $\Sigma m_\nu > 0.06 \text{ eV}$, while in the inverted hierarchy scenario $\Sigma m_\nu > 0.10 \text{ eV}$, so they can be distinguished in the case where neutrinos follow a normal hierarchy and $\Sigma m_\nu < 0.1 \text{ eV}$.

It is interesting to investigate how informative the weak lensing of GWs can be on the sum of neutrino masses. Refs [96–98] explored how combining a standard siren measurement of GWs with other probes (CMB, Supernovae and BAO) could be used to improve constraints on the sum of neutrino masses, but as far as we are aware our work demonstrates the first use of GWs from merging binaries as an independent probe of the sum of neutrino masses. We include a more optimistic source distribution where BSS are comprised of highly localized BNS, BBH, and BHNS mergers and have an associated spectroscopic redshift up to $z_{\text{max}} = 3.5$, similar to that used in [15] where they assumed $z_{\text{max}} = 5$. Figure 5 shows constraints in $\nu\Lambda\text{CDM}$ in the Σm_ν vs σ_8 plane, which can only be investigated with

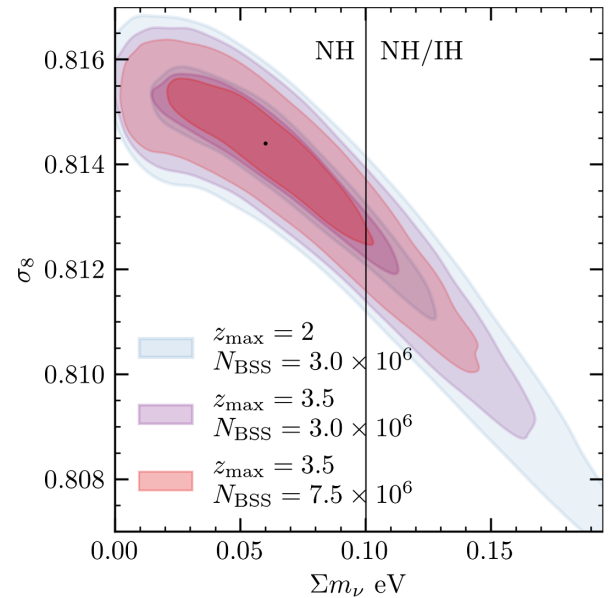


FIG. 5. Constraints in the sum of neutrino masses vs σ_8 plane in the $\nu\Lambda\text{CDM}$ cosmological model (see Table I) from a joint gravitational wave weak lensing + standard siren analysis. $z_{\text{max}} = 2$ corresponds to merging binary neutron stars with an observed electromagnetic (EM) transient counterpart (spec- z), observed up to a maximum redshift of 2. $z_{\text{max}} = 3.5$ is a more optimistic population with spectroscopic redshifts up to 3.5, and the inclusion of sources without an EM counterpart (merging black holes or black hole–neutron stars) whose host galaxy has been identified through exquisite sky localization of the gravitational wave (GW) signal. This is only possible with DeciHz GW detectors. Here it is assumed the instrumental uncertainty $\sigma_{d_L}/d_L = 0.4\%$ [15]. The region left of the black vertical can only correspond to the normal neutrino mass hierarchy (NH), while that to the right can correspond to either the normal or inverted hierarchy (NH/IH).

the added lensing information and so represents an entirely novel use of GWs. We see the degeneracy slope between these two parameters, where the free-streaming of higher mass neutrinos impedes clustering. Because of the improved d_L measurement in DeciHz detectors we set $\sigma_{\text{GW}}/d_L = 0.4\%$ based on the median BBH distance uncertainty at $z \sim 1.5$ of Ref. [15]. More conservative results can be seen in Appendix E. In Fig. 5 the results for $z_{\text{max}} = 2$ correspond to the same BSS population as in Fig. 3. With the higher

redshift population and larger source numbers, a $\sim 1\sigma$ preference for NH can be determined with $\sigma(\Sigma m_\nu) = 0.04$ eV. $\sigma(\Sigma m_\nu) = 0.05$ eV is still achievable with the same population and a much smaller source number, even when increasing the distance uncertainty to 1%. Assuming the smaller number and lower redshift scenario with $\sigma_{\text{GW}}/d_L = 1\%$ gives $\sigma(\Sigma m_\nu) = 0.07$ eV. Due to cosmic variance and the assumed f_{sky} , improvement with number density diminishes past $\bar{n} = 500$ deg $^{-2}$.

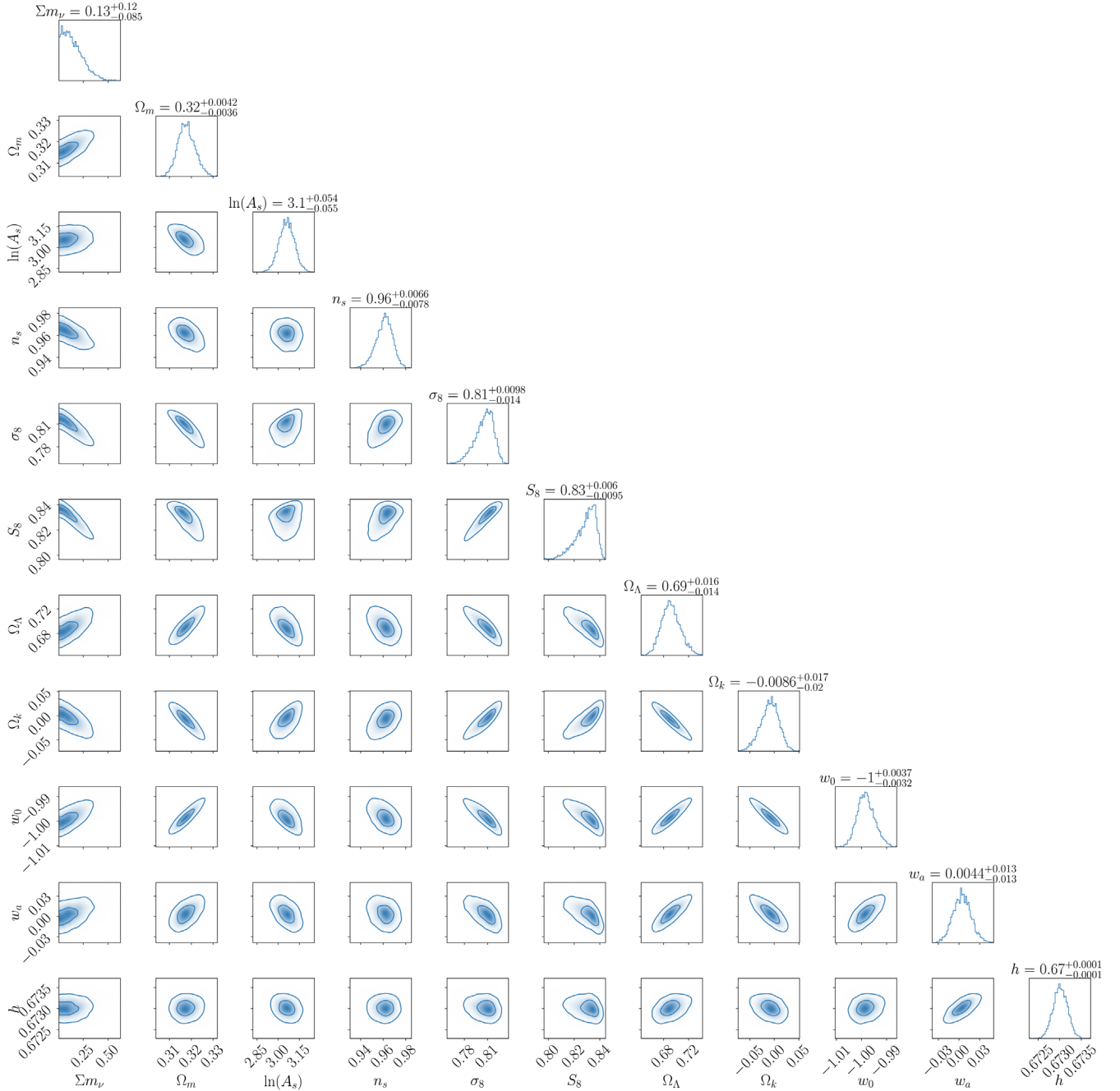


FIG. 6. Constraints on all cosmological parameters in the νkwCDM cosmological model (see Table I) from the joint inference analysis of gravitational waves (GWs). This involves combining a standard siren distance measurement with a weak lensing analysis of the GWs. Here we use 3×10^6 sources up to $z_{\text{max}} = 2$ with $\sigma_{\text{GW}}/d_L = 1\%$. Created using the CORNER package [100].

The most cosmology agnostic model used in our framework is the ν_{kw} CDM model. Allowing both curvature and dark energy to vary changes some parameter degeneracies [99]. Figure 6 shows contours for all parameters, in this case there is $\bar{n}_{\text{BSS}} = 200 \text{ deg}^{-2}$ ($N = 3 \times 10^6$) with $\sigma_{\text{GW}}/d_L = 1\%$. Even in this case we obtain percent or subpercent accuracy on all parameters, besides those relating to neutrinos. This demonstrates the power of this joint standard siren+GW-WL method. Forecasts for constraints on all parameters in each cosmological model from DeciHz detectors can be found in Appendix D.

C. Cosmologically varying $p(z)$

For the weak lensing of gravitational waves, in this framework we generate the observed redshift distribution using a fixed $\mathcal{R}(z)$ in Eq. (18). If the form of $\mathcal{R}(z)$ is well-known, then the cosmology dependence of $p(z)$, through the comoving volume term, can be exploited. Expected astrophysical parameters such as spin and mass distributions contributing to the form of $\mathcal{R}(z)$ can be found from many observations, and through this added knowledge we can perform a first order decoupling of astrophysics and cosmology. This will be possible with 3G detectors as we start to constrain the formation channels of binary mergers, leading to an observationally motivated expression for $\mathcal{R}(z)$ [101–103]. Knowing $\mathcal{R}(z)$ means the cosmology dependence of $p(z)$ can be exploited, and also reduces our

sensitivity to selection effects usually present in an observed $p(z)$. Again we use BPASS for $\mathcal{R}(z)$'s. We choose this nonparametric model due to its good agreement with observations. A parametric model, while possible to marginalize over, could produce wildly incorrect rates and therefore not be a useful model. Figure 7 shows a comparison between forecasts found when $p(z)$ is kept fixed with the fiducial cosmology, or allowed to vary in the derivative calculation. These contours show constraints given by the weak lensing of GWs only (not combined with the standard siren measurement). It can be seen how the w_0 - w_a degeneracy is shifted, and information is gained on w_a . For other parameters, allowing $p(z)$ to vary leads to some cancellation of parameter dependence and therefore weaker constraints. The combined standard siren and weak lensing constraints are only marginally affected. Tests were performed to check whether uncertainties in $\mathcal{R}(z)$ could affect parameter uncertainties, and we found that even large errors in $\mathcal{R}(z)$ do not have an effect on quoted results.

VI. DISCUSSION AND CONCLUSIONS

Future 3rd generation gravitational wave detectors will observe huge numbers of binary mergers, sufficient for challenging statistical cosmological tests such as weak lensing. Two confirmed detector proposals, the Einstein Telescope and Cosmic Explorer plan to come online in the late 2030s–early 2040s. Looking further, there will be an incredible boom for GW cosmology with the advent of DeciHz observatories, but due to the technical difficulty their status will not be clear until after the pioneering space-based GW observatory LISA has been launched, also in the 2030s. Therefore, as we consider 10 years of data taking, these forecasts are for far in the future, when GW cosmology has become a tried and tested method. Here we explore those cosmological parameters that are presently of interest, such as the dark energy equation of state parameters and the sum of neutrino masses, to illustrate the power of the method and compare with other surveys which provide predictions for the same parameters.

We find GW weak lensing tomography is set to be a valuable cosmological tool for 3G detectors. Bright standard sirens found from multimessenger gravitational wave and electromagnetic observations could constrain the dark energy equation of state parameters w_0 and w_a with comparable precision to Euclid, depending on the number of sources and their uncertainty. It is likely that true numbers of BSS will be similar to the upper limits quoted in Ref. [85] due to the existence of other GW (ET, LISA) and electromagnetic (Fermi-GBM, Swift-BAT etc.) observatories. In the case where we combine a population of BSS with realistic numbers of dark standard sirens coming from binary black hole mergers, constraints on w_0 and w_a improve upon Euclid, and give comparable or better constraints to forecasts for a CMB-S4-like experiment with 10^9 spectrometer hours [104], and a 1024 dish

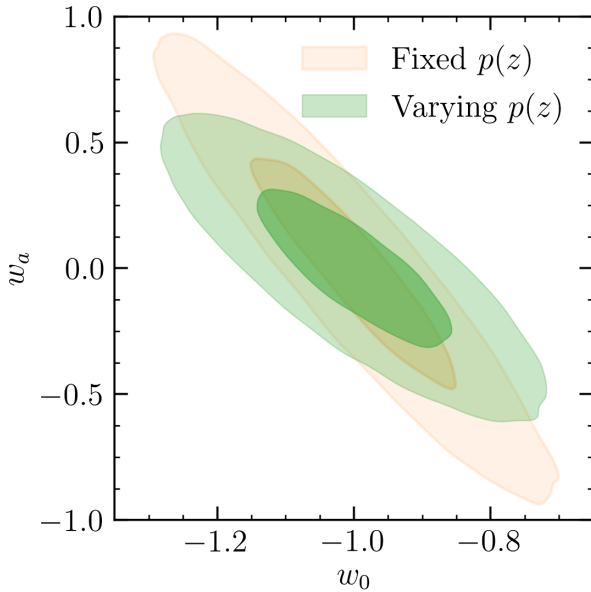


FIG. 7. Constraints on dark energy equation of state parameters w_0 and w_a in the w CDM cosmological model (see Table I) from the weak lensing of gravitational waves only. The orange contour is the situation of a fixed, observed $p(z)$ in the weak lensing analysis. The green contour is when we use prior knowledge about the expected cosmic merger rate density of sources, and then exploit the dependence of $p(z)$ on cosmological parameters through the comoving volume.

HIRAX experiment which expects a precision \sim few percent [105]. Therefore it is vital the tools necessary for statistical inference of DSS redshifts are developed. The situation could be improved even further by high redshift spectroscopic galaxy surveys such as MegaMapper [62], which could allow a spec- z statistical redshift determination for DSS mergers. The advantage of GW-WL over intensity mapping is we have a much better understanding of the physics underlying how GWs are lensed (essentially just general relativity), compared to the complex relationship between the distribution of, for example, neutral Hydrogen and dark matter. The advantage over a CMB experiment is the lack of foregrounds confusing the GW signal. A final benefit of GW-WL is the ability to use an astrophysically motivated cosmic merger rate density of sources. Large numbers of future observations will help to constrain the formation channels of binary mergers, allowing us to exploit the cosmology dependence of $p(z)$. We find allowing $p(z)$ to vary with cosmology leads to improved constraints on the time-variation dark energy, but degrades other geometry parameters such as h and Ω_m . The main difficulty to contend with for GWs is observing a signal over the background with a high enough SNR to accurately measure its parameters, such as distance and chirp mass.

For space-based DeciHz GW detectors, many more BSS observations are expected due to the incredible sky localization of these detectors, including DECIGO and the Big Bang Observatory. There will be exquisite precision on geometry parameters due to very large number of $d_L - z$ data points with small d_L errors. These detectors could achieve a 1σ preference for the neutrino normal hierarchy using a fractional sky coverage of $f_{\text{sky}} = 0.36$. While not significant enough as a single probe to determine the neutrino mass hierarchy, these results demonstrate for the first time how the weak lensing of GWs could provide valuable extra information on the sum of neutrino masses when combined with future CMB experiments [26] and HI intensity mapping surveys [106]. For these results to be possible, the substantial challenge to overcome is building and deploying the space-based DeciHz GW detectors into their heliocentric orbits. Another recently explored possibility is a Lunar-based DeciHz GW detector, which would be easier to maintain and have comparable sensitivity to the proposed space-based DeciHz detectors [107].

A limitation of this study is using single values for uncertainties for the whole population of GW sources. A natural extension to this work is to produce a realistic set of GW measurements through a Fisher forecast of their parameter uncertainties, then perform the $d_L - z$ and power spectrum fitting on this realistic set of measurements.

ACKNOWLEDGMENTS

C. T. M. thanks Danny Laghi for helpful discussion on statistical redshift inference, and Suvodip Mukherjee

for useful exchange on the shot noise modification due to the gravitational wave localization uncertainty. C. T. M. is supported by a Science and Technology Facilities Council (STFC) Studentship Grant. A. T. is supported by an STFC Consolidated Grant. The PYTHON packages NumPy, SciPy, NUMDIFFTOOLS, MATPLOTLIB and CORNER have been used in this work.

APPENDIX A: CHOICE OF GW SOURCE DISTRIBUTIONS

1. Selection functions

Here assumptions on our survey selection functions are varied to test the sensitivity of results to these choices, which are based on predictions for forthcoming electromagnetic and gravitational wave detectors. The galaxy survey selection function $f(z)$, seen in Eq. (21), depends only on the pivot redshift z_{pivot} . Instead of changing the form, we investigate the impact of the choice of this pivot redshift in the BSS case. It is found that, because the bulk of the sources are at lower redshifts that are accessible to spectroscopic surveys, varying the pivot redshift and hence the number of sources in the high redshift tail has a small impact on results.

For the GW selection function, we modify the power in the exponential in Eq. (20) and the value of r_{cut} . Varying the third power to a quadratic or quartic term has a very small effect as the change in the shape of the distribution leads to less sources at lower redshifts but more at higher redshifts. Changing the value of r_{cut} has the largest impact on the presented results, as we are essentially changing the sensitivity of our GW detector. Decreasing r_{cut} by 1 Gpc degrades forecasts by $\sim 10\%$ without affecting the parameter degeneracies. The general conclusions of the paper including the benefits of combining BSS and DSS, and a standard siren+GW-WL analysis producing competitive forecasts with other future probes, are not affected.

2. Observed $p(z)$

Different choices for the observed $p(z)$ can be made. We can use a shifted galaxy survey $p(z)$,

$$p(z) \propto z^2 e^{-(z/z_0)^2}, \quad (\text{A1})$$

with $z_0 = 2/\sqrt{2}$ which is higher than the Euclid value of $z_0 = 0.9/\sqrt{2}$ to reflect the greater redshift range of 3G GW detectors. This is the distribution used in CT19, and will be referred to as galaxy survey-like (GS-like).

Another route is changing the assumption on the cosmic merger rate density $\mathcal{R}(z)$. One way is by using a specific delay time distribution (DTD) between the formation of binaries, which follows the star formation rate (SFR), and the eventual inspiral and merger. This is similar to the BPASS results [70], without any modeling of the complicated physics of binary inspirals.

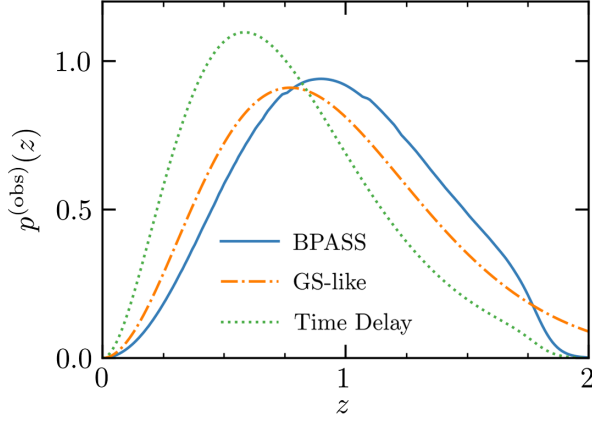


FIG. 8. Alternative observed redshift distributions to those used in the main analysis (BPASS, a population synthesis code). “GS-like” is a galaxy-survey like redshift distribution. The last model assumes binary populations follow the star formation rate, with some time delay between the formation and merger.

$$\mathcal{R}(z) \propto \int_0^{t_{\max}} \text{SFR}(\tau) P(\tau) d\tau. \quad (\text{A2})$$

$P(\tau)$ is the DTD, and t_{\max} is the age of the Universe minus the lookback time to the galaxy. To test the sensitivity of our results we use the more extreme “slow” model in Ref. [108]. These distributions are seen in Fig. 8.

Although different assumptions on the source distributions do affect the forecasts, the effect is marginal as demonstrated in Fig. 9.

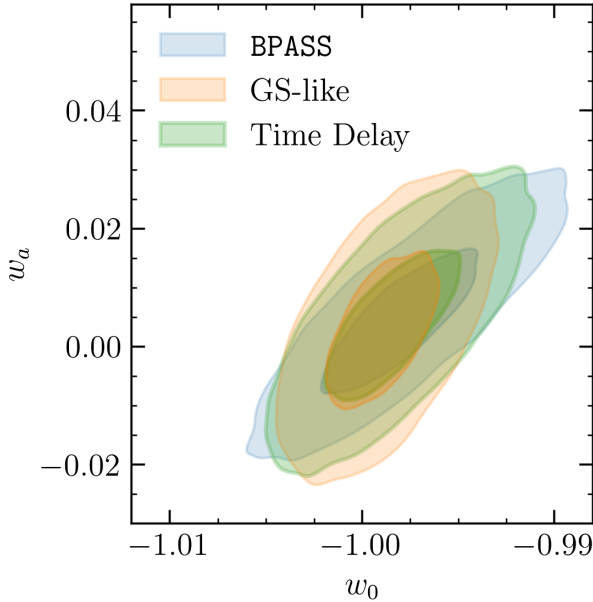


FIG. 9. Contours comparing forecasts in the w_0 - w_a plane using three different assumptions on the observed redshift distribution of binary mergers. Either a BPASS $\mathcal{R}(z)$, a redshift distribution using a delay time distribution between the SFR and the merger, or a galaxy-survey like (GS-like) distribution.

APPENDIX B: NUMERICAL DERIVATIVES

As discussed in Ref. [88] Sec. 4.5.2, great care is needed when finding numerical derivatives of the power spectrum—especially in the case of a tomographic analysis. This is because the response of the power spectrum to a changing cosmological parameter can vary greatly over different ℓ modes and redshift bin correlations. The choice of the step size for the derivative becomes nontrivial, and tests are needed to ensure results are not overly sensitive to this choice.

The method adopted here uses the NUMDIFFTOOLS package. The Derivative function is an adaptive central difference approximation scheme which can calculate the derivative of C_ℓ^{kk} for a range of step sizes, returning the optimal step size for each ℓ mode independently. This optimal step can vary greatly over ℓ modes and tomographic bin combinations. But using different values of the step size for different ℓ modes causes discontinuities in the derivative and an unrealistic forecast. Therefore for each parameter, we use the median of the optimal step sizes of each ℓ mode and correlation returned by the Derivative function. This may introduce inaccuracies as too large/small a step is being used in some cases. To test whether this median step is indeed optimal for the derivative of the whole C_ℓ^{kk} , we define a merit function. We find the absolute difference between the derivative found using a step size and the previous step. This difference between derivatives is found over all ℓ modes and bin correlations. Then we find the mean of all of these differences (similar results are obtained if the maximum or median difference is used). For an optimal step size, we expect this merit function to be minimized. This implies successive steps are narrowing in on the true value for the derivative. At too small step sizes the difference between successive steps can vary a lot due to numerical error in the calculation. At too large step sizes, we are not picking up the true shape of the function, and the derivative can vary a lot depending on the smaller scale features.

Once we have checked for optimality using this merit function, stability of results around this step size needs to be tested. Stability was tested for by recalculating forecasts for a large range of step sizes around this median step. We find that the parameter uncertainties vary by at most $\sim \pm 3\%$ of the uncertainty found using the optimal step size when the step size is varied in a range of at least $-\log(2) \leq \log(\text{Step}/\text{Optimal Step}) \leq \log(2)$ for each parameter. This demonstrates that the results are optimal and robust. These results can be seen in Fig. 10. Each parameter’s merit function has a minimum plateau region, and the optimal step size (vertical dotted line), along with the region where the results are robust (gray shaded region), lie within this minimum plateau. A method of improving the accuracy could be using a different step size for each bin correlation, though from Fig. 10 it is not expected this will have a significant impact on the results presented.

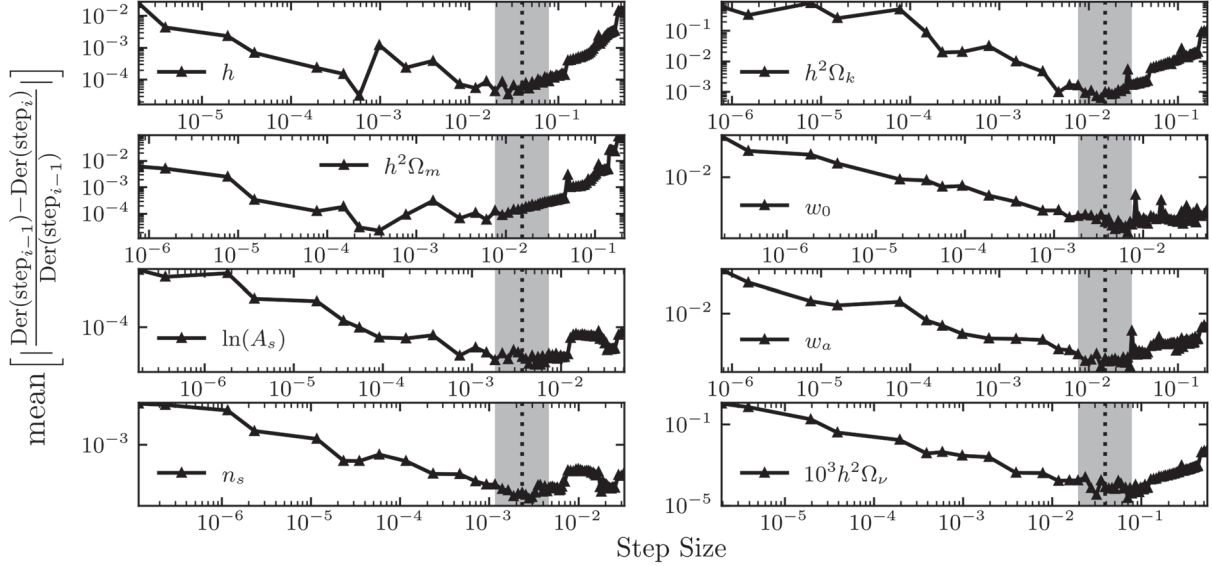


FIG. 10. For each parameter a chosen merit function for the derivative of the convergence power spectrum C_ℓ^{kk} as a function of the derivative step size is displayed. This is the mean difference between the derivative at a step i and $i - 1$. We expect this to be minimized for an optimal step size. Also shown by the vertical black dotted line is the step size found by taking the median of all the “optimal” step sizes returned by the `numdifftools.Derivative` function for each ℓ mode and redshift bin correlation. The gray shaded region shows the step sizes which produce parameter uncertainties within $\sim \pm 3\%$ of the uncertainty found using the step size at the dotted line.

APPENDIX C: REDUCING THE VALUE OF ℓ_{\max}

The results in Figs. 3, 5 were recreated using different assumptions on the maximum ℓ mode. Figure 11 shows the fractional difference between parameter constraints found using $\ell_{\max} = 3000$ (as in the main results) and other

choices of ℓ_{\max} . It can be seen that for the dark energy equation of state constrained using 3G GW detectors, the largest difference is only a few tenths of a percent of the original value. The difference is more significant for those parameters constrained only through GW-WL,

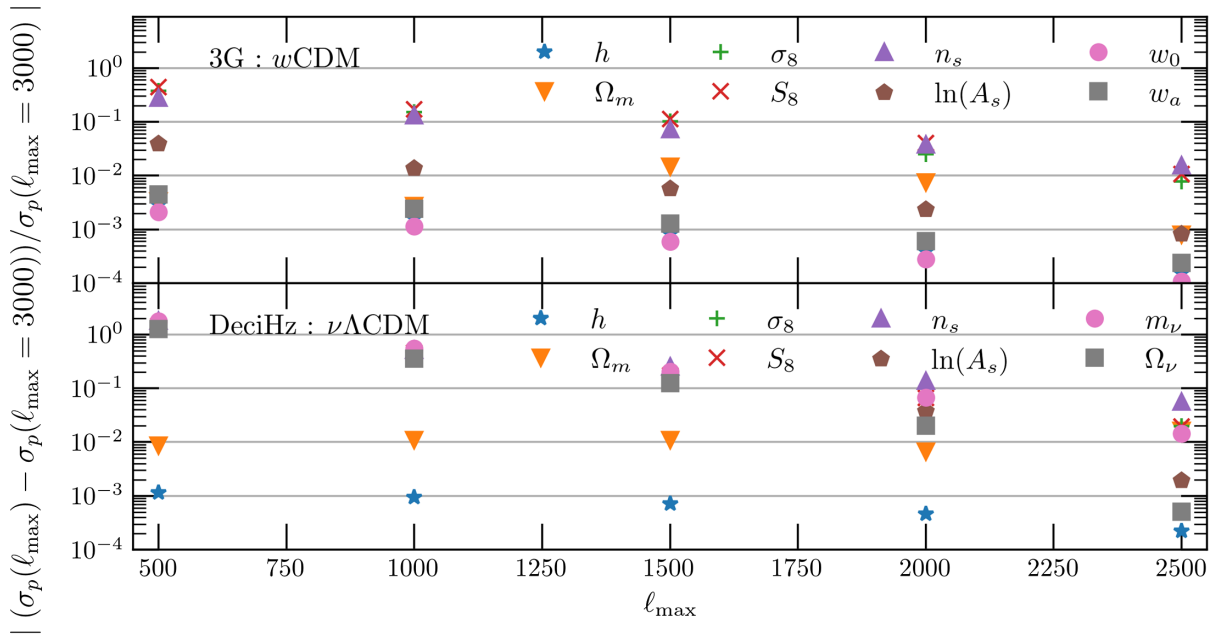


FIG. 11. The fractional difference between parameter constraints using a maximum ℓ mode of $\ell_{\max} = 3000$ and other choices of ℓ_{\max} . The comparison is made for the results in Fig. 3 (top) and Fig. 5 (bottom).

TABLE II. 1σ marginalized uncertainties for parameters in the 3G case. Case A uses 30000 BSS and 3×10^6 DSS. Case B uses 3000 BSS and 3×10^5 DSS. In both cases $\sigma_{\text{GW}}/d_L = 2\%$ for BSS and $\sigma_{\text{GW}}/d_L = 10\%$ for DSS.

Parameter	ΛCDM		$k\Lambda\text{CDM}$		$w\text{CDM}$		νCDM		$\nu kw\text{CDM}$	
	A	B	A	B	A	B	A	B	A	B
h [$\times 10^{-3}$]	0.22	0.69	0.49	1.5	1.4	4.4	0.22	0.69	2.8	7.0
Ω_m [$\times 10^{-3}$]	0.70	2.2	2.1	6.6	5.7	18	0.70	2.2	28	31
$\ln(10^{10}A_s)$	0.057	0.50	0.068	0.51	0.064	0.51	0.66	5.3	1.2	5.5
n_s	0.069	0.55	0.070	0.25	0.070	0.55	0.12	1.1	0.14	1.1
σ_8	0.034	0.22	0.035	0.21	0.035	0.21	0.036	0.20	0.069	0.23
S_8	0.035	0.22	0.061	0.12	0.035	0.22	0.037	0.20	0.23	0.25
Ω_Λ	\times	\times	0.014	0.045	\times	\times	\times	\times	0.29	0.31
Ω_K	\times	\times	0.0072	0.023	\times	\times	\times	\times	0.15	0.16
w_0	\times	\times	\times	\times	0.0054	0.017	\times	\times	0.087	0.15
w_a	\times	\times	\times	\times	0.036	0.11	\times	\times	0.32	0.61
Σm_ν [eV]	\times	\times	\times	\times	\times	\times	1.4	11	2.8	11

TABLE III. 1σ marginalized uncertainties for parameters in the DeciHz case assuming a number density of bright standard sirens of 200 deg^{-2} , 3×10^6 sources using $f_{\text{sky}} = 0.36$. The maximum redshift these sources are observed to is $z_{\text{max}} = 2$. Case A uses $\sigma_{\text{GW}}/d_L = 0.4\%$, case B uses $\sigma_{\text{GW}}/d_L = 1\%$.

Parameter	ΛCDM		$k\Lambda\text{CDM}$		$w\text{CDM}$		νCDM		$\nu kw\text{CDM}$	
	A	B	A	B	A	B	A	B	A	B
h [$\times 10^{-3}$]	0.022	0.028	0.045	0.061	0.11	0.16	0.022	0.028	0.12	0.17
Ω_m [$\times 10^{-3}$]	0.098	0.12	0.59	0.68	1.1	1.4	0.096	0.12	2.7	3.9
$\ln(10^{10}A_s)$ [$\times 10^{-3}$]	4.9	6.3	9.3	12	9.0	12	41	51	42	56
n_s [$\times 10^{-3}$]	4.1	5.4	4.1	5.5	4.2	5.5	5.1	6.7	7.5	8.9
σ_8 [$\times 10^{-3}$]	0.78	0.98	1.3	1.6	1.4	1.7	2.8	3.5	8.4	12
S_8 [$\times 10^{-3}$]	0.80	1.0	0.86	1.1	0.91	1.1	2.9	3.6	5.7	7.9
Ω_Λ [$\times 10^{-3}$]	\times	\times	0.85	1.0	\times	\times	\times	\times	10	15
Ω_K [$\times 10^{-3}$]	\times	\times	1.4	1.7	\times	\times	\times	\times	13	19
w_0 [$\times 10^{-3}$]	\times	\times	\times	\times	0.43	0.57	\times	\times	2.8	4.1
w_a [$\times 10^{-3}$]	\times	\times	\times	\times	5.7	7.3	\times	\times	10	14
Σm_ν [meV]	\times	\times	\times	\times	\times	\times	57	67	81	100

and here the maximum difference is around $\times 1.4$. The situation is worse in the DeciHz case constraining the $\nu\Lambda$ CDM model, this is due to the importance of nonlinear scales in constraining the neutrino mass. Here using $\ell_{\max} = 500$ increases the errors by a factor of 2 for the power spectrum parameters.

APPENDIX D: FULL PARAMETER CONSTRAINTS

Here we show constraints on all free and derived parameters for each cosmological model, for 3G GW detectors in Table II, and DeciHz detectors in Table III.

APPENDIX E: DECIHZ DETECTOR ASSUMPTIONS

In Fig. 12 we show probability density functions for different assumptions on the instrumental uncertainty and source numbers in DeciHz detectors. In the main analysis we assumed DeciHz detector's very accurate sky localization would allow well measured luminosity distances (the two measurements are degenerate).

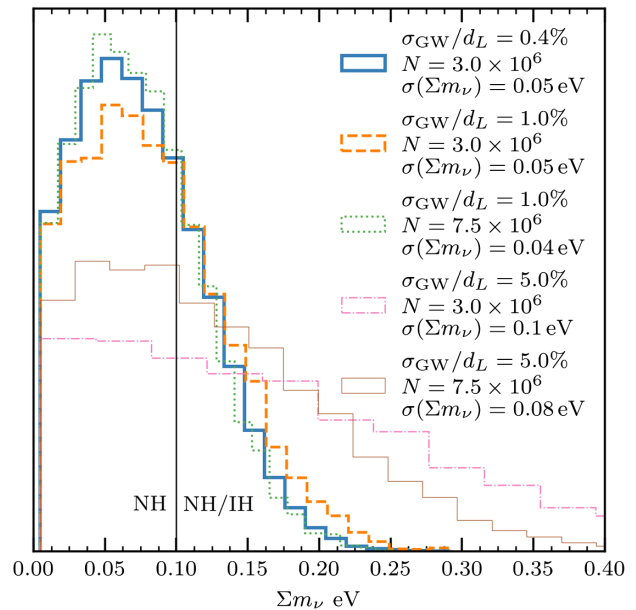


FIG. 12. Probability distributions comparing the high redshift population in Fig. 5 (see Table I) with different source numbers and distance uncertainties for DeciHz detectors.

- [1] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso *et al.* (The LIGO Scientific and Virgo Collaborations), GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett.* **119**, 161101 (2017).
- [2] A. Goldstein, P. Veres, E. Burns, M. S. Briggs, R. Hamburg, D. Kocevski, C. A. Wilson-Hodge, R. D. Preece, S. Poolakkil, O. J. Roberts *et al.*, An ordinary short gamma-ray burst with extraordinary implications: Fermi-GBM detection of GRB 170817A, *Astrophys. J.* **848**, L14 (2017).
- [3] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, Multi-messenger observations of a binary neutron star merger, *Astrophys. J.* **848**, L12 (2017).
- [4] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, K. Ackley, C. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt *et al.* (The LIGO Scientific and the Virgo Collaborations), A gravitational-wave standard siren measurement of the Hubble constant, *Nature (London)* **551**, 8588 (2017).
- [5] D. Brout, D. Scolnic, B. Popovic, A. G. Riess, J. Zuntz, R. Kessler, A. Carr, T. M. Davis, S. Hinton, D. Jones *et al.*, The Pantheon + Analysis: Cosmological constraints, *Astrophys. J.* **938**, 110 (2022).
- [6] K. C. Wong, S. H. Suyu, G. C.-F. Chen, C. E. Rusu, M. Millon, D. Sluse, V. Bonvin, C. D. Fassnacht, S. Taubenberger, M. W. Auger *et al.*, H0LiCOW XIII. A 2.4 per cent measurement of H0 from lensed quasars: 5.3 σ tension between early- and late-Universe probes, *Mon. Not. R. Astron. Soc.* **498**, 14201439 (2019).
- [7] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak *et al.*, Planck 2018 results, *Astron. Astrophys.* **641**, A6 (2020).
- [8] H.-Y. Chen, M. Fishbach, and D. E. Holz, A two per cent Hubble constant measurement from standard sirens within five years, *Nature (London)* **562**, 545547 (2018).
- [9] The LIGO Scientific Collaboration, Gravitational wave astronomy with LIGO and similar detectors in the next decade, arXiv:1904.03187.
- [10] M. Maggiore, C. V. D. Broeck, N. Bartolo, E. Belgacem, D. Bertacca, M. A. Bizouard, M. Branchesi, S. Clesse, S. Foffa, J. Garca-Bellido *et al.*, Science case for the Einstein telescope, *J. Cosmol. Astropart. Phys.* **03** (2020) 050.
- [11] D. Reitze, R. X. Adhikari, S. Ballmer, B. Barish, L. Barsotti, G. Billingsley, D. A. Brown, Y. Chen, D. Coyne, R. Eisenstein *et al.*, Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO, *Bull. Am. Astron. Soc.* **51**, 035 (2019).
- [12] The eLISA Consortium, P. A. Seoane, S. Aoudia, H. Audley, G. Auger, S. Babak, J. Baker, E. Barausse, S. Barke, M. Bassan, V. Beckmann *et al.*, The Gravitational Universe, arXiv:1305.5720.
- [13] R. Mandelbaum, Weak lensing for precision cosmology, *Annu. Rev. Astron. Astrophys.* **56**, 393433 (2018).
- [14] J. M. Ezquiaga and D. E. Holz, Spectral Sirens: Cosmology from the Full Mass Distribution of Compact Binaries, *Phys. Rev. Lett.* **129**, 061102 (2022).

- [15] C. Cutler and D. E. Holz, Ultrahigh precision cosmology from gravitational waves, *Phys. Rev. D* **80**, 104009 (2009).
- [16] S. Camera and A. Nishizawa, Beyond Concordance Cosmology with Magnification of Gravitational-Wave Standard Sirens, *Phys. Rev. Lett.* **110**, 151103 (2013).
- [17] G. Congedo and A. Taylor, Joint cosmological inference of standard sirens and gravitational wave weak lensing, *Phys. Rev. D* **99**, 083526 (2019).
- [18] A. Hall, Cosmology from weak lensing alone and implications for the Hubble tension, *Mon. Not. R. Astron. Soc.* **505**, 4935 (2021).
- [19] R. Laureijs, J. Amiaux, S. Arduini, J. L. Augures, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet *et al.*, Euclid definition study report, [arXiv:1110.3193](https://arxiv.org/abs/1110.3193).
- [20] LSST Dark Energy Science Collaboration, Large synoptic survey telescope: Dark energy science collaboration, [arXiv:1211.0310](https://arxiv.org/abs/1211.0310).
- [21] D. Crichton, M. Aich, A. Amara, K. Bandura, B. A. Bassett, C. Bengaly, P. Berner, S. Bhatporia, M. Bucher, T.-C. Chang *et al.*, Hydrogen intensity and real-time analysis experiment: 256-element array status and overview, *J. Astron. Telesc. Instrum. Syst.* **8**, 011019 (2022).
- [22] S. Kawamura, M. Ando, N. Seto, S. Sato, M. Musha, I. Kawano, J. Yokoyama, T. Tanaka, K. Ioka, T. Akutsu *et al.*, Current status of space gravitational wave antenna DECIGO and B-DECIGO, *Prog. Theor. Exp. Phys.* **2021**, 05A105 (2021).
- [23] G. M. Harry, P. Fritschel, D. A. Shaddock, W. Folkner, and E. S. Phinney, Laser interferometry for the big bang observer, *Classical Quantum Gravity* **23**, 4887 (2006); **23**, 7361(E) (2006).
- [24] J. Baker, T. Baker, C. Carbone, G. Congedo, C. Contaldi, I. Dvorkin, J. Gair, Z. Haiman, D. F. Mota, A. Renzini *et al.*, High angular resolution gravitational wave astronomy, *Exp. Astron.* **51**, 1441 (2021).
- [25] S. Mishra-Sharma, D. Alonso, and J. Dunkley, Neutrino masses and beyond- Λ CDM cosmology with LSST and future CMB experiments, *Phys. Rev. D* **97**, 123544 (2018).
- [26] S. Aiola, Y. Akrami, K. Basu, M. Boylan-Kolchin, T. Brinckmann, S. Bryan, C. M. Casey, J. Chluba, S. Clesse, F.-Y. Cyr-Racine *et al.* (HD Collaboration), Snowmass2021 CMB-HD White Paper, [arXiv:2203.05728](https://arxiv.org/abs/2203.05728).
- [27] M. Maggiore, *Gravitational Waves. Volume 1: Theory and Experiments* (Oxford University Press, New York, 2008).
- [28] M. Maggiore, *Gravitational Waves. Vol. 2: Astrophysics and Cosmology* (Oxford University Press, New York, 2018).
- [29] M. Chevallier and D. Polarski, Accelerating universes with scaling dark matter, *Int. J. Mod. Phys. D* **10**, 213 (2001).
- [30] E. V. Linder, Exploring the Expansion History of the Universe, *Phys. Rev. Lett.* **90**, 091301 (2003).
- [31] B. F. Schutz, Determining the Hubble constant from gravitational wave observations, *Nature (London)* **323**, 310 (1986).
- [32] N. Yunes and X. Siemens, Gravitational-wave tests of general relativity with ground-based detectors and pulsar-timing arrays, *Living Rev. Relativity* **16**, 9 (2013).
- [33] A. Klein, E. Barausse, A. Sesana, A. Petiteau, E. Berti, S. Babak, J. Gair, S. Aoudia, I. Hinder, F. Ohme *et al.*, Science with the space-based interferometer eLISA: Supermassive black hole binaries, *Phys. Rev. D* **93**, 024003 (2016).
- [34] G. Fragione, Black-hole-Neutron-star mergers are unlikely multimessenger sources, *Astrophys. J. Lett.* **923**, L2 (2021).
- [35] C. L. MacLeod and C. J. Hogan, Precision of Hubble constant derived using black hole binary absolute distances and statistical redshift information, *Phys. Rev. D* **77**, 043512 (2008).
- [36] W. Del Pozzo, Inference of cosmological parameters from gravitational waves: Applications to second generation interferometers, *Phys. Rev. D* **86**, 043011 (2012).
- [37] M. Soares-Santos, A. Palmese, W. Hartley, J. Annis, J. Garcia-Bellido, O. Lahav, Z. Doctor, M. Fishbach, D. E. Holz, H. Lin *et al.*, First measurement of the Hubble constant from a dark standard siren using the dark energy survey galaxies and the LIGO/Virgo binary black-hole merger GW170814, *Astrophys. J.* **876**, L7 (2019).
- [38] S. Mukherjee and B. D. Wandelt, Beyond the classical distance-redshift test: Cross-correlating redshift-free standard candles and sirens with redshift surveys, [arXiv:1808.06615](https://arxiv.org/abs/1808.06615).
- [39] 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**, 19561970 (2020).
- [40] G. Scelfo, M. Spinelli, A. Raccanelli, L. Boco, A. Lapi, and M. Viel, Gravitational waves \times HI intensity mapping: Cosmological and astrophysical applications, *J. Cosmol. Astropart. Phys.* **01** (2022) 004.
- [41] W. M. Farr, M. Fishbach, J. Ye, and D. E. Holz, A future percent-level measurement of the Hubble expansion at redshift 0.8 with Advanced LIGO, *Astrophys. J.* **883**, L42 (2019).
- [42] Z.-Q. You, X.-J. Zhu, G. Ashton, E. Thrane, and Z.-H. Zhu, Standard-siren cosmology using gravitational waves from binary black holes, *Astrophys. J.* **908**, 215 (2021).
- [43] C. Messenger, K. Takami, S. Gossan, L. Rezzolla, and B. S. Sathyaprakash, Source Redshifts from Gravitational-Wave Observations of Binary Neutron Star Mergers, *Phys. Rev. X* **4**, 041004 (2014).
- [44] C. Ye and M. Fishbach, Cosmology with standard sirens at cosmic noon, *Phys. Rev. D* **104**, 043507 (2021).
- [45] S. Mastroianni, K. Leyde, C. Karathanasis, E. Chassande-Mottin, D. A. Steer, J. Gair, A. Ghosh, R. Gray, S. Mukherjee, and S. Rinaldi, On the importance of source population models for gravitational-wave cosmology, *Phys. Rev. D* **104**, 062009 (2021).
- [46] M. Evans, R. X. Adhikari, C. Afle, S. W. Ballmer, S. Biscoveanu, S. Borhanian, D. A. Brown, Y. Chen, R. Eisenstein, A. Gruson *et al.*, A horizon study for cosmic explorer: Science, observatories, and community, [arXiv:2109.09882](https://arxiv.org/abs/2109.09882).
- [47] 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).
- [48] F. Iacovelli, M. Mancarella, S. Foffa, and M. Maggiore, Forecasting the detection capabilities of third-generation gravitational-wave detectors using GWFFAST, *Astrophys. J.* **941**, 208 (2022).

- [49] I. Mandel and F. S. Broekgaarden, Rates of compact object coalescences, *Living Rev. Relativity* **25**, 1 (2022).
- [50] A. Sesana, Prospects for Multiband Gravitational-Wave Astronomy after GW150914, *Phys. Rev. Lett.* **116**, 231102 (2016).
- [51] D. Laghi, N. Tamanini, W. Del Pozzo, A. Sesana, J. Gair, S. Babak, and D. Izquierdo-Villalba, Gravitational-wave cosmology with extreme mass-ratio inspirals, *Mon. Not. R. Astron. Soc.* **508**, 45124531 (2021).
- [52] P. Schneider, Gravitational lensing as a probe of structure, [arXiv:astro-ph/0306465](https://arxiv.org/abs/astro-ph/0306465).
- [53] M. Bartelmann and P. Schneider, Weak gravitational lensing, *Phys. Rep.* **340**, 291 (2001).
- [54] R. Takahashi, T. Suyama, and S. Michikoshi, Scattering of gravitational waves by the weak gravitational fields of lens objects, *Astron. Astrophys.* **438**, L5L8 (2005).
- [55] R. Takahashi, Amplitude and phase fluctuations for gravitational waves propagating through inhomogeneous mass distribution in the universe, *Astrophys. J.* **644**, 8085 (2006).
- [56] F. Xu, J. M. Ezquiaga, and D. E. Holz, Please repeat: Strong lensing of gravitational waves as a probe of compact binary and galaxy populations, *Astrophys. J.* **929**, 9 (2022).
- [57] 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).
- [58] P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, Princeton, NJ, 1980).
- [59] W. Zhao and L. Wen, Localization accuracy of compact binary coalescences detected by the third-generation gravitational-wave detectors and implication for cosmology, *Phys. Rev. D* **97**, 064031 (2018).
- [60] S. Vitale and M. Evans, Parameter estimation for binary black holes with networks of third-generation gravitational-wave detectors, *Phys. Rev. D* **95**, 064052 (2017).
- [61] Z. J. Zhang, C. Chang, P. Larsen, L. F. Secco, and J. Zuntz (The LSST Dark Energy Science Collaboration), Transitioning from stage-III to stage-IV: Cosmology from galaxy \times CMB lensing and shear \times CMB lensing, *Mon. Not. R. Astron. Soc.* **514**, 2181 (2022).
- [62] D. J. Schlegel, J. A. Kollmeier, G. Aldering, S. Bailey, C. Baltay, C. Bebek, S. BenZvi, R. Besuner, G. Blanc, A. S. Bolton *et al.*, Astro2020 APC White Paper: The MegaMapper: A $z > 2$ spectroscopic instrument for the study of inflation and dark energy, *Bull. Am. Astron. Soc.* **51**, 229 (2019).
- [63] 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.
- [64] T. Eifler, H. Miyatake, E. Krause, C. Heinrich, V. Miranda, C. Hirata, J. Xu *et al.*, Cosmology with the roman space telescope multiprobe strategies, *Mon. Not. R. Astron. Soc.* **507**, 17461761 (2021).
- [65] M. L. Graham, A. J. Connolly, W. Wang, S. J. Schmidt, C. B. Morrison, v. Ivezić, S. Fabbro, P. Côté, S. F. Daniel, R. L. Jones *et al.*, Photometric redshifts with the LSST. II. The impact of near-infrared and near-ultraviolet photometry, *Astron. J.* **159**, 258 (2020).
- [66] F. Calore, A. Cuoco, T. Regimbau, S. Sachdev, and P. D. Serpico, Cross-correlating galaxy catalogs and gravitational waves: A tomographic approach, *Phys. Rev. Res.* **2**, 023314 (2020).
- [67] M. Y. M. Lau, I. Mandel, A. Vigna-Gomez, C. J. Neijssel, S. Stevenson, and A. Sesana, Detecting double neutron stars with LISA, *Mon. Not. R. Astron. Soc.* **492**, 30613072 (2020).
- [68] W. Del Pozzo, A. Sesana, and A. Klein, Stellar binary black holes in the LISA band: A new class of standard sirens, *Mon. Not. R. Astron. Soc.* **475**, 34853492 (2018).
- [69] The LIGO Scientific and the Virgo Collaborations, GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, [arXiv:2108.01045](https://arxiv.org/abs/2108.01045).
- [70] J. J. Eldridge, E. R. Stanway, and P. N. Tang, A consistent estimate for gravitational wave and electromagnetic transient rates, *Mon. Not. R. Astron. Soc.* **482**, 870880 (2018).
- [71] H. Leandro, V. Marra, and R. Sturani, Measuring the Hubble constant with black sirens, *Phys. Rev. D* **105**, 023523 (2022).
- [72] G. D. Racca, R. Laureijs, L. Stagnaro, J.-C. Salvignol, J. Lorenzo Alvarez, G. Saavedra Criado, L. Gaspar Venancio, A. Short, P. Strada, T. Bnke *et al.*, The Euclid mission design, Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave, [10.1117/12.2230762](https://doi.org/10.1117/12.2230762) (2016).
- [73] P. J. E. Peebles, Origin of the large-scale galaxy peculiar velocity field: A minimal isocurvature model, *Nature (London)* **327**, 210 (1987).
- [74] B. S. Sathyaprakash, B. F. Schutz, and C. Van Den Broeck, Cosmography with the Einstein Telescope, *Classical Quantum Gravity* **27**, 215006 (2010).
- [75] 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).
- [76] 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).
- [77] X. Shan, C. Wei, and B. Hu, Lensing magnification: Gravitational waves from coalescing stellar-mass binary black holes, *Mon. Not. R. Astron. Soc.* **508**, 12531261 (2021).
- [78] 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.
- [79] 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).
- [80] D. Blas, J. Lesgourgues, and T. Tram, The cosmic linear anisotropy solving system (CLASS). Part II: Approximation schemes, *J. Cosmol. Astropart. Phys.* **07** (2011) 034.
- [81] A. J. Mead, S. Brieden, T. Trster, and C. Heymans, HMcode-2020: Improved modelling of non-linear cosmological power spectra with baryonic feedback, *Mon. Not. R. Astron. Soc.* **502**, 1401 (2021).

- [82] M. Takada and B. Jain, Cosmological parameters from lensing power spectrum and bispectrum tomography, *Mon. Not. R. Astron. Soc.* **348**, 897915 (2004).
- [83] H.-J. Huang, T. Eifler, R. Mandelbaum, and S. Dodelson, Modelling baryonic physics in future weak lensing surveys, *Mon. Not. R. Astron. Soc.* **488**, 1652 (2019).
- [84] S. R. Taylor and J. R. Gair, Cosmology with the lights off: Standard sirens in the Einstein Telescope era, *Phys. Rev. D* **86**, 023502 (2012).
- [85] J. Yu, H. Song, S. Ai, H. Gao, F. Wang, Y. Wang, Y. Lu, W. Fang, and W. Zhao, Multimessenger detection rates and distributions of binary neutron star mergers and their cosmological implications, *Astrophys. J.* **916**, 54 (2021).
- [86] S. Ronchini, M. Branchesi, G. Oganesyan, B. Banerjee, U. Dupletsa, G. Ghirlanda, J. Harms, M. Mapelli, and F. Santoliquido, Perspectives for multi-messenger astronomy with the next generation of gravitational-wave detectors and high-energy satellites, *Astron. Astrophys.* **665**, A97 (2022).
- [87] 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).
- [88] A. Blanchard, S. Camera, C. Carbone, V. F. Cardone, S. Casas, S. Clesse, S. Ili, M. Kilbinger, T. Kitching *et al.* (Euclid Collaboration), Euclid preparation, *Astron. Astrophys.* **642**, A191 (2020).
- [89] R. Mandelbaum, T. Eifler, R. Hložek, T. Collett, E. Gawiser, D. Scolnic, D. Alonso, H. Awan, R. Biswas, J. Blazek *et al.* (The LSST Dark Energy Science Collaboration), The LSST dark energy science collaboration (desc) science requirements document, [arXiv:1809.01669](https://arxiv.org/abs/1809.01669).
- [90] F. Santoliquido, M. Mapelli, M. C. Artale, and L. Boco, Modelling the host galaxies of binary compact object mergers with observational scaling relations, *Mon. Not. R. Astron. Soc.* **516**, 3297 (2022).
- [91] J. Crowder and N. J. Cornish, Beyond LISA: Exploring future gravitational wave missions, *Phys. Rev. D* **72**, 083005 (2005).
- [92] M. A. Sedda, C. P. L. Berry, K. Jani, P. Amaro-Seoane, P. Auclair, J. Baird, T. Baker, E. Berti, K. Breivik, A. Burrows *et al.*, The missing link in gravitational-wave astronomy: Discoveries waiting in the decihertz range, *Classical Quantum Gravity* **37**, 215011 (2020).
- [93] B. C. Seymour, H. Yu, and Y. Chen, Multiband gravitational wave cosmography with dark sirens, [arXiv:2208.01668](https://arxiv.org/abs/2208.01668).
- [94] S. R. Choudhury and S. Hannestad, Updated results on neutrino mass and mass hierarchy from cosmology with Planck 2018 Likelihoods, *J. Cosmol. Astropart. Phys.* **07** (2020) 037.
- [95] R. Jimenez, T. Kitching, C. Pea-Garay, and L. Verde, Can we measure the neutrino mass hierarchy in the sky?, *J. Cosmol. Astropart. Phys.* **05** (2010) 035.
- [96] L.-F. Wang, X.-N. Zhang, J.-F. Zhang, and X. Zhang, Impacts of gravitational-wave standard siren observation of the Einstein Telescope on weighing neutrinos in cosmology, *Phys. Lett. B* **782**, 87 (2018).
- [97] A. Agrawal, T. Okumura, and T. Futamase, Standard candles and sirens rescue H_0 , *Astrophys. J.* **904**, 169 (2020).
- [98] S.-J. Jin, R.-Q. Zhu, L.-F. Wang, H.-L. Li, J.-F. Zhang, and X. Zhang, Impacts of gravitational-wave standard siren observations from Einstein Telescope and Cosmic Explorer on weighing neutrinos in interacting dark energy models, *Commun. Theor. Phys.* **74**, 105404 (2022).
- [99] R. Hložek, M. Cortês, C. Clarkson, and B. Bassett, Dark energy degeneracies in the background dynamics, *Gen. Relativ. Gravit.* **40**, 285 (2008).
- [100] D. Foreman-Mackey, corner.py: Scatterplot matrices in python, *J. Open Source Software* **1**, 24 (2016).
- [101] N. Singh, T. Bulik, K. Belczynski, and A. Askar, Exploring compact binary populations with the Einstein Telescope, *Astron. Astrophys.* **667**, A2 (2022).
- [102] S.-X. Yi, F. Stoppa, G. Nelemans, and E. Cator, The GW-Universe Toolbox II: Constraining the binary black hole population with second and third generation detectors, *Astron. Astrophys.* **663**, A156 (2022).
- [103] S. Vitale, W. M. Farr, K. K. Y. Ng, and C. L. Rodriguez, Measuring the star formation rate with gravitational waves from binary black holes, *Astrophys. J.* **886**, L1 (2019).
- [104] K. S. Karkare, A. M. Dizgah, G. K. Keating, P. Breyse, and D. T. Chung, Snowmass 2021 cosmic frontier white paper: Cosmology with millimeter-wave line intensity mapping, [arXiv:2203.07258](https://arxiv.org/abs/2203.07258).
- [105] D. Karagiannis, R. Maartens, and L. Randrianjanahary, Cosmological constraints from the power spectrum and bispectrum of 21 cm intensity maps, *J. Cosmol. Astropart. Phys.* **11** (2022) 003.
- [106] S. Ferraro, N. Sailer, A. Slosar, and M. White, Snowmass2021 cosmic frontier white paper: Cosmology and fundamental physics from the three-dimensional large scale structure, [arXiv:2203.07506](https://arxiv.org/abs/2203.07506).
- [107] K. Jani and A. Loeb, Gravitational-wave lunar observatory for cosmology, *J. Cosmol. Astropart. Phys.* **06** (2021) 044.
- [108] K. S. McCarthy, Z. Zheng, and E. Ramirez-Ruiz, Constraining delay time distribution of binary neutron star mergers from host galaxy properties, *Mon. Not. R. Astron. Soc.* **499**, 52205229 (2020).