Testing the Quasar Hubble Diagram with LISA Standard Sirens

Lorenzo Speri[1,a], Nicola Tamanini[1], Robert R. Caldwell[2], Jonathan R. Gair[1], and Benjamin Wang[3]

[1] *Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut,*
Am Mühlenberg 1, 14476 Potsdam-Golm, Germany
[2] *Department of Physics & Astronomy, Dartmouth College,*
6127 Wilder Laboratory, Hanover, New Hampshire 03755 USA
[3] *University of California - Los Angeles, Los Angeles, California 90095 USA*
(Dated: October 20, 2020)

Quasars have recently been used as an absolute distance indicator, extending the Hubble diagram to high redshift to reveal a deviation from the expansion history predicted for the standard, ΛCDM cosmology. Here we show that the Laser Interferometer Space Antenna (LISA) will efficiently test this claim with standard sirens at high redshift, defined by the coincident gravitational wave (GW) and electromagnetic (EM) observations of the merger of massive black hole binaries (MBHBs). Assuming a fiducial ΛCDM cosmology for generating mock standard siren datasets, the evidence for the ΛCDM model with respect to an alternative model inferred from quasar data [*Nat. Astron.* **3**, 272 (2019)] is investigated. By simulating many realizations of possible future LISA observations, we find that for 50% of these realizations (median result) 4 MBHB standard siren measurements will suffice to strongly differentiate between the two models, while 14 standard sirens will yield a similar result in 95% of the realizations. In addition, we investigate the measurement precision of cosmological parameters as a function of the number of observed LISA MBHB standard sirens, finding that 15 events will on average achieve a relative precision of 5% for H_0, reducing to 3% and 2% with 25 and 40 events, respectively. Our investigation clearly highlights the potential of LISA as a cosmological probe able to accurately map the expansion of the universe at $z \gtrsim 2$, and as a tool to cross-check and cross-validate cosmological EM measurements with complementary GW observations.

I. INTRODUCTION

A key goal for cosmology is to characterize the accelerated cosmic expansion [1, 2] and understand the underlying physics [3, 4]. Major efforts are underway to measure the expansion history using a variety of techniques, across a wide span of redshift, in hopes of revealing a clue to the nature of dark energy. To date, the Hubble diagram has been constructed out to redshift $z \sim 1.4$ using the local distance ladder [5] and type Ia supernovae as standard candles [6]. In the near future, dedicated observatories and experiments such as DESI [7], Euclid [8], the Rubin Observatory [9], and the Roman Space Telescope [10] will refine and extend the Hubble diagram out to $z \sim 3$. Other methods, using for example the cosmic microwave background (CMB) [11], can retrieve information on the cosmic expansion at higher redshift, but they assume a cosmological model to infer the expansion rate and can thus yield only indirect measurements.

A new, direct method to measure the expansion history out to significantly higher redshift has recently been implemented, using quasars as absolute distance indicators [12, 13]. Determination of the quasar luminosity distance relies on a phenomenological relationship between the ultraviolet and X-ray emissions of the accreting plasma [14]. Risaliti & Lusso [13] analyzed a large catalog of X-ray and UV observations of quasars to calibrate this relationship and build a Hubble diagram that extends out to $z \simeq 5.5$. They found that their quasar sample yields distances that are in agreement with the standard, ΛCDM cosmological model at $z \lesssim 1.4$. However, at larger redshifts there appears to be significant disagreement between the quasar distances and the ΛCDM model predictions [13, 15]. This innovation necessarily raises new questions. Are quasars truly reliable distance indicators?

There are no clear answers at present, and secondary analyses of the quasar data have come to varying conclusions about the state of the standard cosmology [16–19]. Is there a crisis? After all, the local Hubble diagram has come under close scrutiny, recently, due to the apparent tension with parameter values inferred from the CMB [5, 11, 20, 21]. For certain, there is a need for new data and new methods to probe the expansion history.

The nascent field of gravitational wave (GW) astronomy offers a way to address these and other questions, providing a new, independent probe of the expansion using standard sirens [22–24]. An absolute distance is obtained from the inspiral and merger gravitational waveform, which, coupled with an electromagnetic (EM) measurement of the redshift, yields a new, independent determination of the cosmological parameters, for example of the Hubble constant [25]. It is worth stressing that distance measurements with GWs are absolute, meaning that they do not depend on a specific cosmological model or other distance indicators.

---

[a] Electronic address: `lorenzo.speri@aei.mpg.de`

The standard siren method has already been used to produce constraints on the Hubble constant H_0 with real data collected by the LIGO-Virgo interferometers [25–28]. In particular GW170817 [29], a binary neutron star merger, was the first multi-messenger event for which both GW and EM signals were observed. This first standard siren provided a measurement of $H_0 = 70.0^{+12.0}_{-8.0}$ km s^{-1} Mpc^{-1} [30–32]. Other GW events observed without EM counterparts can also be used as standard sirens. If the host galaxy cannot be identified, one can still cross-correlate the sky localization region of the GW event with a galaxy catalog and assign to each galaxy within this region a non-zero probability of being the host galaxy [22, 33, 34]. This "statistical method" has less constraining power per event than the counterpart method because the redshift is not known precisely. However, combining information from many GW events allows the statistical significance to be increased. Combining all the events observed in the first two LIGO and Virgo observing runs gave a result $H_0 = 68^{+14}_{-7}$ km s^{-1} Mpc^{-1} [32].

Looking ahead, GW observations by LIGO/Virgo and proposed observatories, such as the Cosmic Explorer [35] or the Einstein Telescope [36], will yield orders of magnitude more candidate standard sirens. Perhaps the boldest of these future plans is the Laser Interferometer Space Antenna (LISA) [37], a GW observatory consisting of a fleet of three satellites in an Earth-trailing, heliocentric orbit, positioned at the vertices of a 2.5 million km triangle. Using laser interferometry, LISA will measure GWs in the mHz frequency range. The loudest objects in this band of the GW sky are expected to be mergers of massive black hole binaries (MBHBs). Observations of the expected GW signal, and an EM counterpart by ground-based telescopes, offer the exciting possibility to extend the Hubble diagram out to $z \sim 7$ or higher [38, 39]. This is the possibility we explore: the ability of deep-redshift standard sirens observed by LISA to validate or refute the expansion history inferred by the quasar Hubble diagram [13]. To do this, we generate mock LISA MBHB standard siren observations based on a flat ΛCDM cosmology and we adopt a Bayesian framework to test the deviation found in the quasar observations. We aim to estimate how many MBHB standard sirens are necessary to rule out a deviation of that size and to explore the power of LISA to constrain the matter density and Hubble parameter.

This paper is organized as follows. In Sec. II we introduce the quasar Hubble diagram constructed with quasar observations and we briefly summarise the main steps to reproduce the results found in Ref. [13]. In Sec. III we describe how the mock LISA standard sirens observations are generated and we investigate how such observations can constrain the Hubble parameter and the matter density. We present our method of Bayesian hypothesis testing in Sec. IV A, and our results, in terms of the number of LISA standard sirens needed to test the quasar Hubble diagram, in Sec. IV B. We conclude with a discussion of the results and assess future prospects in Sec. V.

The code used for the analysis is available at [40].

II. QUASAR HUBBLE DIAGRAM

Cosmologists have long hoped that quasars, observed at redshifts $z \sim 6$ and greater, could provide a reliable distance indicator with which to chart the cosmic expansion history. These aspirations took a significant step forward when a phenomenological relationship between the quasar x-ray and ultraviolet luminosities,

$$\log_{10} L_X = \gamma \log_{10} L_{UV} + \beta, \tag{1}$$

was obtained by Avni & Tananbaum [41]. Here L_X, L_{UV} are the monochromatic luminosities at 2 keV and 2500 Å. There is good physical reason to expect the UV and X-ray spectra are connected: the UV emission originates in the infalling accretion material onto the quasar, whereas the X-rays are emitted by the plasma of hot electrons surrounding the accretion disk that scatter off the UV photons. Following Eq. (1), measurement of the fluxes yields the luminosity distance, once the parameters γ, β are determined. For the sake of clarity, we summarize the calculation procedure leading from quasar data to luminosity distance. In terms of the observed fluxes, Eq. (1) becomes

$$\log_{10} F_X = \gamma \log_{10} F_{UV} + 2(\gamma - 1) \log_{10} H_0 D_L + \tilde{\beta}, \tag{2}$$

where $\tilde{\beta}$ absorbs the introduction of H_0 into the log of the luminosity distance, D_L. Hence, Eq. (1) is the relationship between the observed X-ray and UV fluxes, through which we constrain the luminosity distance - redshift relationship. As with SN magnitude-distances, the Hubble constant is a nuisance parameter until the magnitude offset is known. Then, given the parameters γ, β, one can extract the luminosity distance to each quasar. Alternatively, one can use the ensemble data to infer γ, β simultaneously with the parameters of a theoretical model D_L. Our procedure follows that of Ref. [16], whereby the likelihood is calculated as

$$\mathcal{L}_{QSO} = \prod_{i=1}^{N} \frac{e^{-\frac{1}{2}\Delta_i^2/\tilde{\sigma}_i^2}}{\sqrt{2\pi\tilde{\sigma}_i^2}}. \tag{3}$$
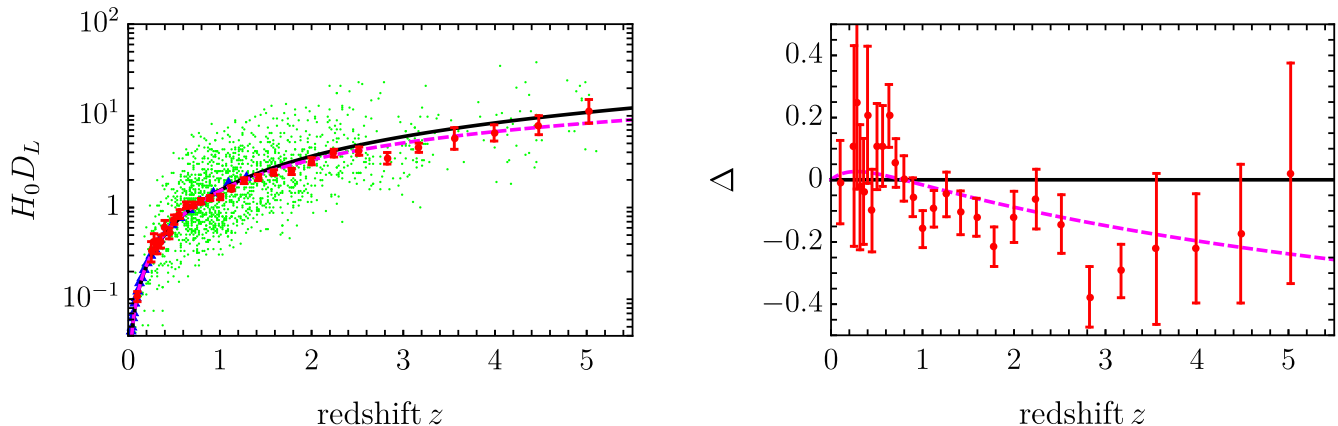
Figure 1. (Left) The quasar Hubble diagram based on the data set presented in Ref. [13], in the context of two cosmological models. Green dots are quasar data points, shown without error bars; red circles are the binned data, with errors. Blue triangles are the binned SN data, also shown without error bars. The solid black curve is the $\Omega_m = 0.3$ ΛCDM model that best fits the joint quasar and type Ia supernovae at $z < 1.4$. The dashed magenta curve is the best-fit phenomenological model proposed in Ref. [13]. (Right) The relative difference between the luminosity distances for the binned quasar data relative to ΛCDM is shown as a function of redshift. The solid black line represents the ΛCDM baseline. The dashed magenta curve is the best-fit phenomenological model of Ref. [13].

In the above equation, $\Delta_i = \log_{10} F_X|_i - (\gamma \log_{10} F_{UV} + 2(\gamma - 1) \log_{10} H_0 D_L + \tilde{\beta})|_i$, $\tilde{\sigma}_i^2 = \delta^2 + \delta(\log_{10} F_X)|_i^2$, where $\delta(\log_{10} F_X)$ is the observed uncertainty in the log of the X-ray flux, and δ introduces a new global variable to describe the intrinsic dispersion in the flux law, Eq. (2). By maximizing the likelihood, or minimizing $-2 \ln \mathcal{L}$, we constrain the luminosity distance - redshift relationship.

To demonstrate the utility of quasar standard candles, we perform a combined analysis of quasars with the SDSS-II/SNLS3 Joint Light-Curve Analysis (JLA) type Ia SNe data set [42]. We first focus on the redshift range spanned by the SNe, $z < 1.4$. The authors of Ref. [13] have kindly provided the tables of data, consisting of measured values of $\log_{10} F_X$, $\log_{10} F_{UV}$, z, and uncertainty $\delta(\log_{10} F_X)$ for $N = 1598$ quasars. For this first analysis, we use only those quasars in the same redshift range as the SNe, of which there are 1025. Following the procedure outlined in Ref. [13], we bin the quasar data to match the redshifts of the JLA data set. In each bin, the mean log-flux is determined; the inverse variance is obtained $\Sigma^{-2} = \sum_i 1/\delta(\log_{10} F_X)|_i^2$. The SNe likelihood is $\mathcal{L}_{SNe} \propto \exp\left(-\frac{1}{2}\chi^2_{SNe}\right)$, where

$$\chi^2_{SNe} = \sum_{ij} \Delta\mu_i C_{ij}^{-1} \Delta\mu_j, \tag{4}$$

$\Delta\mu = \mu_{obs} - \mu_{thy}$ is the difference in observed and theoretical distance modulus, and C is the covariance matrix. The total likelihood is the product of $\mathcal{L}_{QSO}$ and $\mathcal{L}_{SNe}$. We minimize the sum of the log-likelihoods; the cosmological terms are common to both. In this way we calibrate the quasars, and thereby determine parameters $\gamma = 0.6$, $\tilde{\beta} = -14.7$, $\delta = 0.25$; uncertainties in these parameters are subdominant to $\tilde{\sigma}$. Given these parameters, we obtain the data set $\vec{y} = (z^i, \log_{10}(H_0 D_L^i), \sigma[\log_{10}(H_0 D_L^i)])$. For this redshift range ($z < 1.4$), we recover a ΛCDM model best fit value consistent with $\Omega_m = 0.31 \pm 0.05$ (1σ) obtained in Ref. [13].

Proceeding to higher redshift, however, it is not immediately apparent whether the ΛCDM model and quasars are still in agreement. Fig. 1 shows that the ΛCDM model does not provide a good fit to the high redshift data, as the binned quasar measurements depart from the ΛCDM prediction at high redshifts. To quantify this tension, an alternative (ALT) phenomenological model of the luminosity distance is introduced [13],

$$D_L^{\mathrm{ALT}}(z; a_2, a_3) = \frac{c \ln 10}{H_0}\left(x + a_2 x^2 + a_3 x^3\right), \qquad x = \log_{10}(1 + z). \tag{5}$$

By expanding the luminosity distance of a flat ΛCDM cosmology at low redshift, it is possible to relate the coefficients a_2, a_3 of the phenomenological model in Eq. (5) with the matter density Ω_m:

$$a_2 = 3\ln(10)(2 - \Omega_m)/4 \qquad a_3 = \ln(10)^2(7/6 - 2\Omega_m + 9\Omega_m^2/8) \tag{6}$$

Setting $\Omega_m = 0.3$, we find that, with $a_2 \simeq 2.95$ and $a_3 \simeq 3.54$, D_L^{ALT} differs at most 0.6% from the predictions of the flat ΛCDM for redshifts up to $z \sim 1$, and, at most 10% for redshifts up to $z \sim 5$. The redshift-luminosity-distance
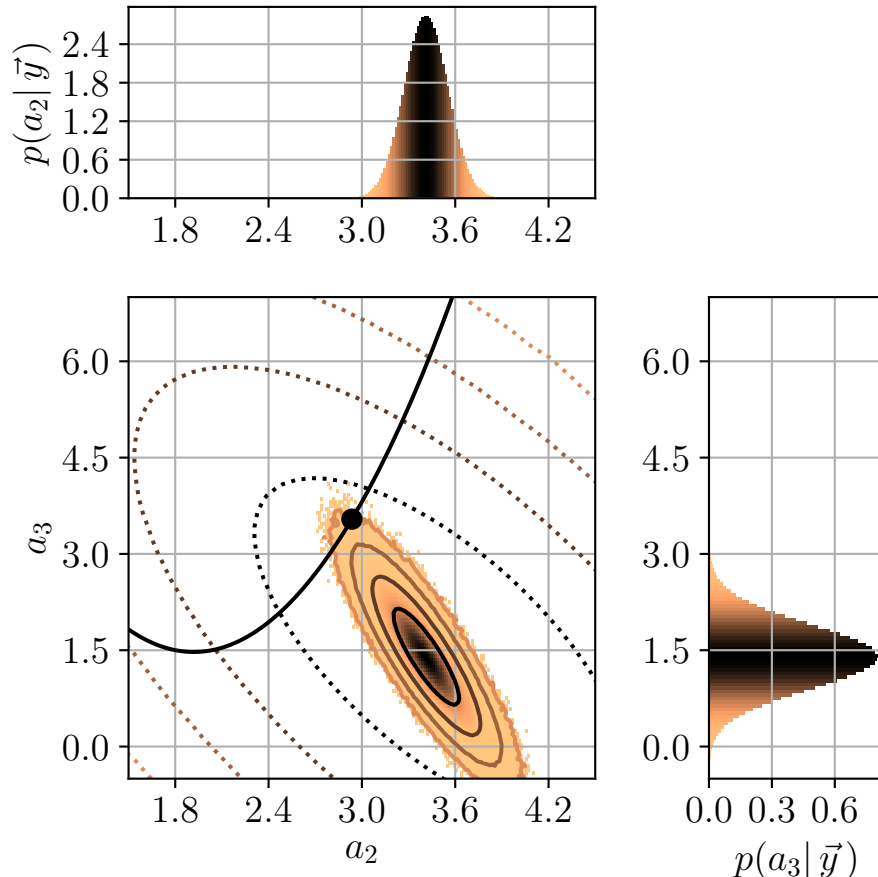
Figure 2. Posterior distribution of a_2, a_3 obtained with SNe and Quasar observations. The 1, 2, 3 and 4 σ contours of the multivariate Gaussian prior distribution chosen for the MCMC are shown as dotted lines, whereas the solid lines represent the respective contours for the posterior distribution.

relation D_L^{ALT} in Eq. (5) is useful to encode the predictions from different cosmological models that can be tested against ΛCDM.

Using the combined data set of SNe and quasars at all redshifts, Ref. [13] claims that the ALT model with $a_2 \simeq 3.5$ and $a_3 \simeq 1.5$ provides a much better fit to the data. We reproduce the results of Ref. [13] using the ensemble sampler for Markov Chain Monte Carlo (MCMC) *emcee* [43], and the data set $\vec{y}$ consisting of all 1598 quasars out to redshift $z \sim 5.5$ and the SNe observations to reproduce the posterior distribution of a_2, a_3. The likelihood used for the MCMC is given by the product of a normal likelihood with mean given by the prediction of D_L^{ALT} for the quasar data and the aforementioned multivariate normal distribution $\mathcal{L}_{SNe}$ for the SNe data. The prior on a_2, a_3 is chosen to be a multivariate normal distribution with $1, 2, 3, 4 - \sigma$ contours shown as dotted lines in Figure 2. The large contours of the prior let the MCMC sampler explore a wide region of the (a_2, a_3) plane. We also checked that, by shifting the center of the prior, the outcome of the MCMC sampling does not change. These choices allow us to recover the posterior distribution $p(a_i | \vec{y})$ of a_2, a_3 shown in Fig. 2, which is in good agreement with that of Figure 5 of Ref. [13].

It is possible to represent the flat ΛCDM model at low redshift in the $(a_2, a_3)-$parameter space by using the relation between the parameters a_2, a_3 and the matter density Ω_m in Eq. (6). The black dot in Figure 2 corresponds to the specific value of a_2, a_3 for $\Omega_m = 0.3$, whereas the black solid line corresponds to other values of matter density. Although the relation between a_2, a_3 and the ΛCDM parameters is only approximate and based on an expansion at low redshift, Fig. 2 suggests that the point corresponding to ΛCDM in the (a_2, a_3) parameter space is $\sim 4\sigma$ away from the best fit value in the (a_2, a_3) posterior retrieved from quasar data. Based mainly on this results, Ref. [13] claims that ΛCDM is no longer a good fit to the quasar Hubble diagram at $z \gtrsim 3$.

For the sake of argument, we can ask what does the phenomenological model have that ΛCDM does not? The arc of $H_0 D_L$ given by Eq. (5) passes closer to the quasars at $z \sim 1.8, 3$ and above, than ΛCDM with $\Omega_m = 0.3$. For visualization purposes, in Fig. 1 we have binned the data to emphasize this point. In particular, the quasars at

$2.7 \lesssim z \lesssim 3$ look like an outlier from the perspective of ΛCDM. This disagreement is made clearer where we plot the relative difference, $\Delta = (D_L^{QSO}/D_L^{LCDM}) - 1$, between the binned data and ΛCDM luminosity distance; the ΛCDM curve passes away from the binned data point by roughly 4 times its standard deviation. This difference is independent of binning in this redshift interval. However, we caution that the quasar data are sparse around both theoretical models at high redshift, suggesting that more data are needed. It is well to keep in mind that unresolved systematics, e.g. quasar evolution or accretion properties, may still affect the inferred luminosity-redshift relationship for quasars (see Ref. [44], but also Ref. [45].)

Now we formulate our questions. We would like to find some complementary check of the Hubble diagram at $z \gtrsim 3$. Can we determine the luminosity distance with sufficient resolution to test deviations from ΛCDM at the 10% level? Can we distinguish ΛCDM from the alternative illustrated in Fig. 1, at redshift $z \gtrsim 3$? These questions highlight a sort of new, high-redshift Hubble tension that can be addressed with GW standard sirens.

III. LISA MBHB STANDARD SIRENS

LISA is a space mission designed to open the millihertz frequency range to GW observations [37]. In this frequency range different populations of GW sources are expected to be detected, including Galactic binaries [46–49], stellar-origin black hole binaries [50], extreme mass ratio inspirals (EMRIs) [51], the mergers of massive black hole binaries (MBHBs) with masses between 10^4 to 10^7 $M_\odot$ [52] and perhaps stochastic GW backgrounds [53–55]. The population of MBHB sources is particularly interesting from a cosmological perspective, since MBHB mergers will be detected at high redshift (up to $z \sim 15 - 20$) and could have EM counterparts observable up to $z \sim 7$ [38].

As mentioned in Sec. I, obtaining complementary redshift information on GW events, through the identification of the corresponding host galaxy once an EM counterpart is spotted, is in fact one of the simplest ways to perform cosmological tests with GW observations. The GW events are standard sirens [22, 24] that provide a direct measurement of the luminosity distance D_L to the source. If this can be linked to a redshift z obtained from EM observations, the combined observation provides a point on the distance-redshift relation $D_L(z; \vec{\theta})$, which can be used to constrain the parameters, $\vec{\theta}$, of the cosmological model under consideration.

All GW sources that LISA will be able to observe at cosmological distances could be used as standard sirens, either with or without EM counterpart identification. For sources such as stellar-origin BHBs or EMRIs we do not expect any measurable EM counterpart (see however [56–58]). For these sources, cosmological inference will only be possible using the statistical method [59–62]. MBHBs on the other hand could produce observable EM emissions to be detected at higher redshift with respect to stellar-mass BHBs and EMRIs. Several studies found that MBHBs could emit radiation in different bands of the EM spectrum both at merger and during long lasting (ranging from weeks to months) afterglows (see e.g. [63–65]). Moreover pre-merger EM observational signatures could even be spotted during their inspiral phase [66–69]. Given a sufficiently accurate sky localization and a sufficiently powerful EM emission, future telescopes are expected to detect the counterpart of at least a few MBHB mergers per year [38, 39]. These high-redshift standard sirens will allow us to probe the expansion of the universe at unprecedented distances and to test possible deviations from ΛCDM at $z \gtrsim 3$ [70–72]. This latter topic is the focus of our investigation, for which we will need realistic catalogs of LISA MBHB standard sirens. In what follows we will explain how we construct such catalogs, building mainly on the results of [38, 39]. Afterwards, we will explore how the measurement precision of cosmological parameters is enhanced as we increase the number of observations.

A. Construction of standard siren catalogs

We start by drawing plausible redshift values for MBHB mergers from an interpolated redshift distribution based on the data presented in Fig. 1 of [38], where the redshift distribution of LISA MBHB standard sirens with identified EM counterparts are reported. These data are based on the LISA sensitivity curve adapted to the LISA design proposed in [37] and have been constructed using the procedure outlined in [39], where MBHB merger histories have been reproduced using semi-analytical merger-tree simulations and the emission and detection of GW and EM signals have been estimated with standard astrophysical and data analysis techniques. We choose to work with the "popIII" model, which produces average results with respect to the astrophysical populations considered in [38, 39]. In any case we checked that our final results do not change appreciably if the other astrophysical populations are considered, as the redshift distributions are similar (cf. Fig. 1 of [38]). We moreover set to zero the probability of drawing MBHB redshift values at $z < 0.1$, in order to match our expectations of not finding MBHBs at very low redshift. We do not expect our results to be sensitive to this cut because the probability to have an event below $z = 0.1$ from our redshift distribution is particularly low, approximately 0.05%.

To each redshift value we then associate a unique value of the luminosity distance D_L by using the flat ΛCDM distance-redshift relation

$$D_L^{\Lambda\text{CDM}}(z; H_0, \Omega_m) = \frac{c\,(1+z)}{H_0} \int_0^z \frac{dx}{\sqrt{\Omega_m(1+x)^3 + 1 - \Omega_m}}\,, \tag{7}$$

with fiducial cosmological parameters set to $H_0 = 70\,\text{km s}^{-1}\,\text{Mpc}^{-1}$ and $\Omega_m = 0.3$. For ease of notation, we drop the label $^{\Lambda\text{CDM}}$ in this section. Next we associate a realistic 1σ distance error to each MBHB event by combining the following uncertainties:

- A weak lensing contribution analytically estimated by the following fitting formula [39, 73]

$$\frac{\sigma_{\text{lens}}(z)}{D_L(z)} = 0.066 \left(\frac{1 - (1+z)^{-0.25}}{0.25} \right)^{1.8}. \tag{8}$$

  We will consider the possibility of delensing, i.e. the use of dedicated matter surveys along the line of sight of the GW event in order to estimate the lensing magnification distribution and thus remove part of the uncertainty due to weak lensing. Following [74] we will realistically assume that up to a 30% delensing will be achievable at redshift 2 (note that in the most optimistic scenario this reduction could be as large as 50% [74]). Furthermore the delensing power should grow linearly from redshift zero, until it reaches a constant value at around $z = 2$ (see Fig. 4 of [74]). For these reasons we will employ a delensing factor provided by the following phenomenological formula

$$F_{\text{delens}}(z) = 1 - \frac{0.3}{\pi/2} \arctan\left(\frac{z}{z_*}\right), \tag{9}$$

  where $z_* = 0.073$ has been fixed requiring that $F_{\text{delens}}(z)$ be within 1% of 0.7 (the $z \to \infty$ asymptotic value) at $z = 2$. The final lensing uncertainty on D_L will thus be

$$\sigma_{\text{delens}}(z) = F_{\text{delens}}(z)\sigma_{\text{lens}}(z). \tag{10}$$

- A peculiar velocity uncertainty contribution as estimated by the following fitting formula [39, 75]

$$\frac{\sigma_v(z)}{D_L(z)} = \left[1 + \frac{c(1+z)^2}{H(z)D_L(z)} \right] \frac{\sqrt{\langle v^2 \rangle}}{c}, \tag{11}$$

  where the r.m.s. peculiar velocity value is set to 500 km/s, in agreement with average values observed in galaxy catalogs.

- A LISA instrument error estimated as $\sigma_{\text{LISA}}/D_L \propto 2/\text{SNR}$ [76]. If we further assume that, once marginalised over all other waveform parameters, the SNR scales as $1/D_L$, we can take the LISA measurement relative error on D_L to be $\sigma_{\text{LISA}}/D_L \propto 2D_L$. To fix the constant of proportionality we rely on recent parameter estimation results [77], which yield a 1σ error on D_L of roughly 2-3% at redshift 4, by employing a full Bayesian approach over a MBHB signal with SNR = 945. The LISA measurement error on D_L that we use is thus be given by

$$\frac{\sigma_{\text{LISA}}(z)}{D_L(z)} = 0.05 \left(\frac{D_L(z)}{36.6\,\text{Gpc}} \right), \tag{12}$$

  where 36.6 Gpc corresponds to $z = 4$ in our fiducial cosmology.

- A redshift error for all events at $z > 2$ associated to photometric measurements estimated as [78–80]

$$\sigma_{\text{photo}}(z) = 0.03(1+z) \quad \text{if } z > 2. \tag{13}$$

  For the sake of simplicity we propagate this redshift uncertainty to the distance uncertainty by assuming our fiducial cosmology, similarly to [39]. We assume redshift measurements at $z < 2$ are taken spectroscopically and, for our purposes, with negligible errors (see e.g. [81, 82]).

The total distance uncertainty σ_D is then obtained by adding in quadrature all contributions listed above. Therefore, the total distance uncertainty depends not only on the source redshift but also on the assumed cosmological model.
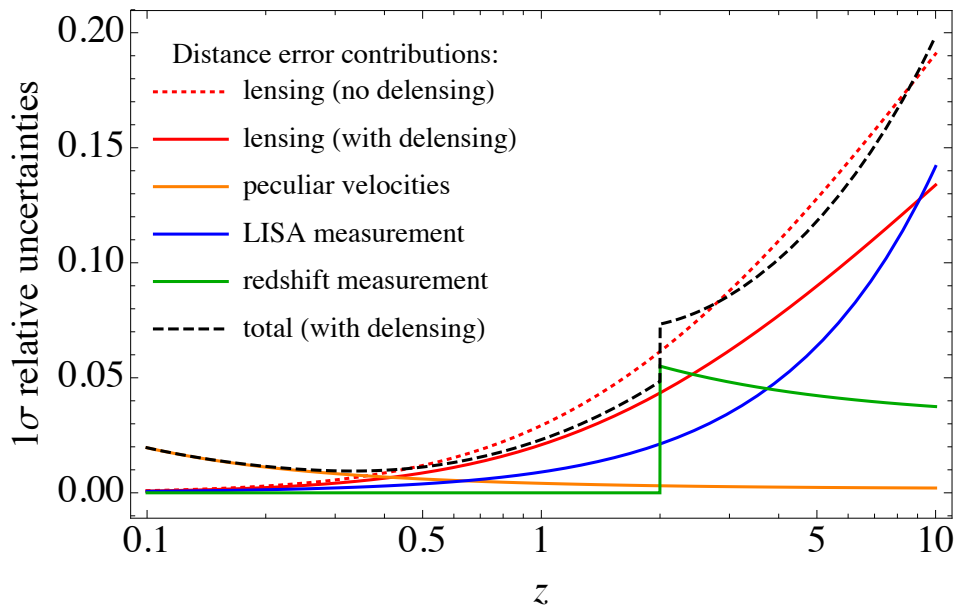
Figure 3. Relative uncertainty contributions to the luminosity distance of standard sirens (σ_D/D_L) as a function of redshift (see Sec. III A for details). For comparison, the average relative distance uncertainty of quasar data in [13] is 0.14.

Fig. 3 shows the contribution of all these sources of error as a function of redshift, together with the total estimated error on the luminosity distance. Finally the value of the luminosity distance of each MBHB event is randomly scattered around the value given by our fiducial cosmology according to a Gaussian probability with standard deviation σ_D. Each LISA MBHB event that will be used in the rest of the paper has been constructed according to the procedure outlined above. Note however that this procedure cannot estimate the actual number of cosmologically useful events that LISA will detect. This information can only be obtained by simulating the full formation and evolution history of MBHBs, by performing parameter estimation over the GW signal and by estimating the emission and detectability of EM counterparts [39, 83]. Because of this all our results are given in terms of the number of standard siren that LISA will observe, with fiducial values taken from the existing literature.

B. Testing ΛCDM with LISA MBHB standard sirens

We can now use the catalogs defined above to check how well ΛCDM, the standard cosmological model, can be constrained by a given number of MBHB standard sirens. In particular we focus on how well the matter density parameter and the Hubble constant are constrained. For this reason, we express Eq. (7) in terms of the Hubble parameter h defined through $H_0/c = h/2997.9\,\mathrm{Mpc}^{-1}$, which for our fiducial flat ΛCDM is equal to $h = 0.7$. Since no more than ~ 50 standard sirens are optimistically expected to be detected by LISA over its maximal extended mission duration (10 years) [38, 39, 72], we will analyse the constraints arising from different possible observational outcomes, with the number of total detected standard sirens ranging from a minimum of 10 to a maximum of 50 events. Our fiducial scenario will assume 15 standard sirens, in agreement with a realistically expected outcome over a nominal mission of 4 years [38, 72].

Given a set of standard sirens observations with associated luminosity distance errors, namely $\vec{y} = (z_j, D_j, \sigma_{D_j})$, we choose to work with the following likelihood,

$$p(\vec{y}|\vec{\theta}, M_A) = \prod_{j=1}^{N_{SS}} \exp\left(-\frac{\left(D_j - D_L^{M_A}(z_j; \vec{\theta}_A)\right)^2}{2\,\sigma_{D_j}^2}\right) \frac{1}{\sqrt{2\pi\sigma_{D_j}^2}}, \qquad (14)$$

to conduct our statistical analysis. This specific choice implicitly assumes that the luminosity distance measurements are normally distributed around the cosmological predictions $D_L^{M_A}(z_j; \vec{\theta}_A)$ of a given cosmological model M_A with parameters $\vec{\theta}_A$. In addition, the standard deviation of the normal likelihood is assumed to be the luminosity distance uncertainty $\sigma_{D_j} = \sigma_D(z_j)$ of the respective measurement, which in turn depends on the redshift and on the luminosity
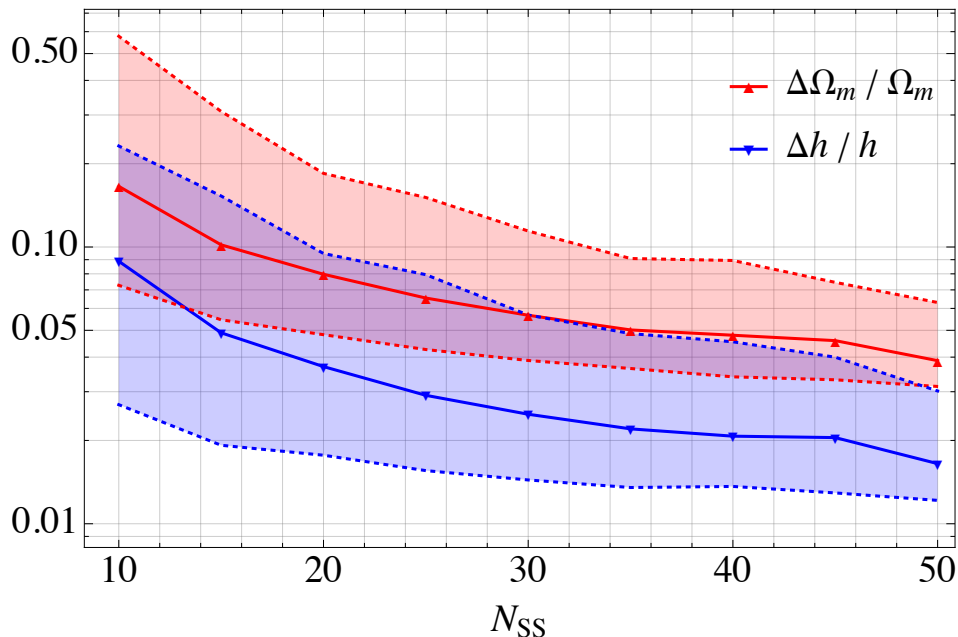
Figure 4. 90% C.I. relative constraints on Ω_m and h as a function of the number of standard sirens observed by LISA for 300 catalogue realizations. The triangles (and solid-line) indicate the median of the 90% C.I. relative constraint distribution of the 300 catalogs, whereas the shaded regions show the 90% range of the 300 confidence intervals.

distance prediction of the cosmological model used to generate the data, i.e. ΛCDM. In the following inference analysis we study the constraining power on the cosmological parameters of the ΛCDM model through the luminosity distance relation $D_L^{M_A}(z_j; \vec{\theta}_A) = D_L^{\Lambda\text{CDM}}(z; h, \Omega_m)$. Therefore, we do not include the dependency of the luminosity distance uncertainty σ_{D_j} on the cosmological model in our inference analysis and we take each σ_{D_j} as an estimate for the measurement uncertainty (in other words the ΛCDM parameters over which σ_{D_j} depends will not be varied, but fixed to their fiducial values).

In Fig. 4 we report the ΛCDM relative constraints at 90% Confidence Interval (C.I.) as a function of the number N_{SS} of standard siren events. To obtain these results, for each value of N_{SS} we have simulated 300 standard siren catalogs. For each catalog, we find the best-fit ΛCDM cosmology through maximum likelihood estimation, which is equivalent to a least squares-fit to the D_L-z measurements. Confidence intervals on the best-fit parameters are found by assuming Gaussian errors in the measurements, consistent with the model above. The numbers reported in Fig. 4 correspond to the median of the distribution of all the 90% C.I. obtained in this way from all the 300 catalogs with the same N_{SS}. In other words, the reported numbers represent the median of the corresponding quantity over 300 realizations with a fixed number of standard siren events, but different redshift and distance values for each event. The shaded regions show the central 90% range of the 300 confidence intervals, for each choice of N_{SS}. From Fig. 4 it is clear that roughly 25 MBHB standard sirens will suffice to provide a $\sim$3% error on H_0 and $\sim$6% error on Ω_m, while 40 MBHB standard sirens will improve these results to $\sim$2% and $\sim$5%, respectively. For our fiducial case however ($N_{SS} = 15$) we can reach only a relative error of 5% on H_0 and of 10% on Ω_m. These results are broadly in agreement with previous analyses [39, 72].

The constraining power in the ΛCDM parameter space depends on the specific realization. We estimate the constraining power of the measurements by computing a Gaussian approximation to the likelihood, which is

$$\Gamma_{ij} = \sum_n^{N_{SS}} \frac{1}{\sigma_{D_n}^2} \frac{\partial D_L^{\Lambda\text{CDM}}(z_n; \vec{\theta})}{\partial \theta_i} \frac{\partial D_L^{\Lambda\text{CDM}}(z_n; \vec{\theta})}{\partial \theta_j}, \tag{15}$$

where $\vec{\theta} = (h, \Omega_m)$. The determinant of the matrix Γ_{ij} determines the volume of the localization region. For each N_{SS}, we found the realization that had the median localization volume among all those simulated and took this to be "representative" of that number of standard sirens. In Fig. 5 we use these representative realizations to construct contour regions at the 90% confidence level in the Ω_m-h parameter space, for $N_{SS} = 10, 15, 20$. The contour regions in Fig. 5 are based on the Gaussian approximation of the likelihood, and thus well characterize the latter only for a relatively large number of observations, which in our case means $N_{SS} \gtrsim 15$. The contour regions (ellipses) are centered
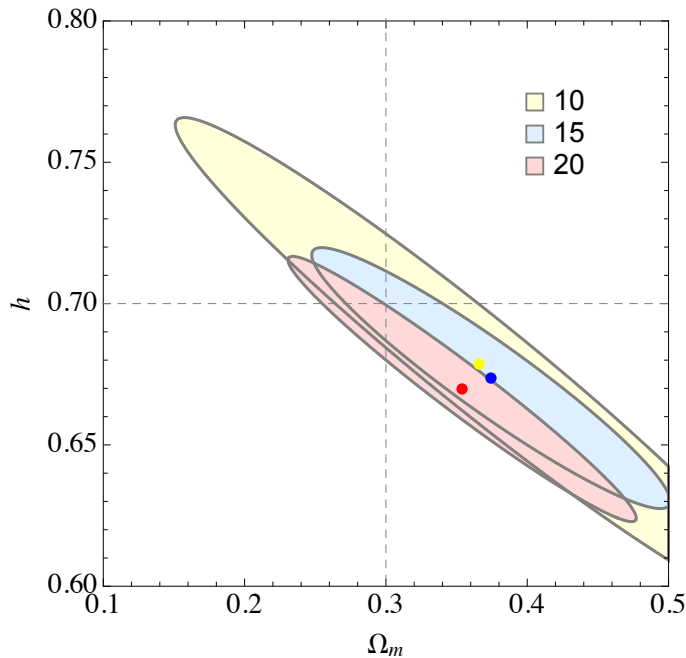
Figure 5.   90% C.I. contour regions (Gaussian approximation) and best-fit values of the representative realizations of 10 (yellow), 15 (blue) and 20 (red) standard sirens, respectively, in the ΛCDM parameter space.

	68% Credible Interval	68% Confidence Interval	Median	Best-Fit
Ω_m	$[0.327, 0.453]$	$[0.313, 0.436]$	0.384	0.374
h	$[0.648, 0.692]$	$[0.651, 0.696]$	0.670	0.674

Table I. Summary statistics of the frequentist and Bayesian analysis of the 15 representative standard sirens shown in Fig. 7

on the best fit value estimated from the representative realization through the maximum likelihood estimator. Note that since in each dataset the data points have been scattered randomly around our fiducial cosmology, the actual values of the best fit parameters are not significant, as the best fit parameters scatter around the injected values depending on which redshift and measurement uncertainty are randomly drawn in the data generation procedure (the fact that all three best fit values lie on the bottom-right with respect to the true injected value is just a coincidence).

In order to further check the results for our fiducial scenario, and thus cross-check the Gaussian approximation, we have further performed a Bayesian analysis of the representative realization with 15 events. Therefore, we obtain the posterior distribution $p(\Omega_m, h|\vec{y})$ using *emcee* [43] and we show the results of the MCMC sampling in Fig. 6. We also provide the MCMC implementation in Ref. [40]. For the MCMC, we set flat priors for both parameters Ω_m and h respectively with support $\Omega_m \in [0, 1]$ and $h \in [0.2, 2.2]$ and we use the likelihood reported in Eq. (14).

In Table I we show some summary statistics of the Bayesian and frequentist analysis of the 15 representative standard sirens. The median of the marginalized posterior shown in Fig. 6 are in agreement with the best-fit values obtained with the aforementioned frequentist analysis at 68% confidence level and both Bayesian and frequentist analysis yield 3% and 16% constraints on the relative precision of h and Ω_m, respectively.

Fig. 5 shows that the higher the number of standard sirens, the better the ΛCDM constraints, as expected. By comparing Fig. 5 and Fig. 6 we can also notice that the Gaussian approximation cannot fully catch all the features of the true posterior for 15 events, although the approximation seems reasonably correct at least to define 1 and 2σ confidence regions.

Finally, in order to better visualize the uncertainty in the distance-redshift space as well, the representative realization with 15 standard sirens has been used to produce a regression plot in the z-D_L space (left panel of Fig. 7), which shows the allowed ΛCDM curves at the 90% C.I. (solid blue lines) together with the best fit from the dataset realization (dotted blue line), the injected ΛCDM curve (black solid line) and the data points of that particular realization with their associated 1σ uncertainties in the distance (in red). The right panel of Fig. 7 shows the residuals of each curve and data point with respect to the fiducial ΛCDM curve. From Fig. 7 we deduce that LISA will test deviations from ΛCDM in the distance-redshift space at the 10% level or better up to redshift $\sim$6, roughly.
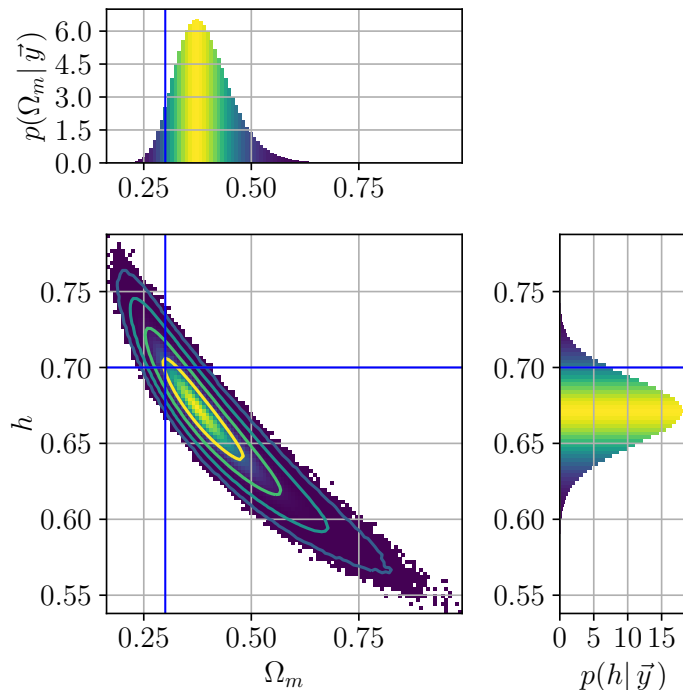
Figure 6. Posterior distribution of the Hubble parameter h and matter density Ω_m obtained with the 15 representative standard siren observations $\vec{y}$. The median of the marginalized posteriors for matter density and Hubble parameter are respectively 0.384 and 0.670. The colored solid lines correspond to the $1, 2, 3, 4\sigma$ contours, whereas the blue solid line shows the injected values $\Omega_m = 0.3$ and $h = 0.7$ used to construct the data set. We used flat priors for both parameters Ω_m and h with support $\Omega_m \in [0, 1]$ and $h \in [0.2, 2.2]$, respectively.
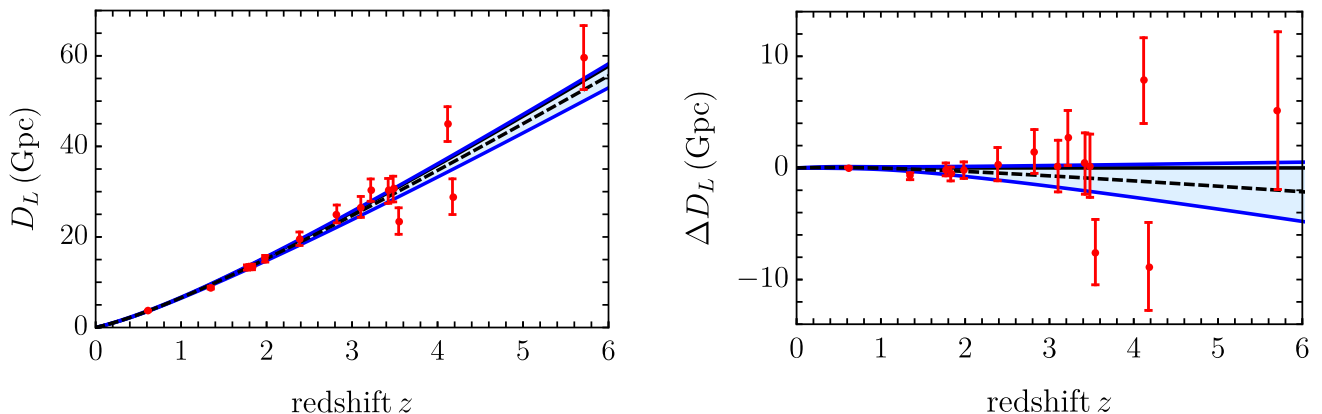

Figure 7. Representative LISA MBHB standard siren dataset with 15 events (red data points - 1σ uncertainties reported). The left panel shows the distance-redshift space, while the right panel shows the relative difference with respect to our fiducial ΛCDM cosmology (black solid line). The black dashed line identifies our best fit to the dataset, while the solid blue lines delimit the 90% C.I. bands around it.

This clearly shows the potential of LISA to provide unique cosmological constraints, not only for ΛCDM but also for other cosmological models, in a regime not yet efficiently probed by standard EM observations. These results already indicate that LISA will have the required power to test the deviations from ΛCDM reported with quasar data at $z \gtrsim 3$. In what follows we will quantify this statement using Bayesian hypothesis testing, and by looking in particular at how well LISA will differentiate between ΛCDM and the high-z deviation found in Ref. [13].

IV. BAYESIAN COSMOLOGICAL HYPOTHESIS TESTING

A. Methodology

The scope of this section is to forecast the number N_{SS} of MBHB standard sirens that we would need to observe to test the deviation from the ΛCDM model inferred from the quasar luminosity distances [13]. Bayesian hypothesis testing is a suitable framework for this goal, because the parameters of different cosmological models are not known precisely and are therefore well represented as distributions which can depend on our belief in the model.

If we have two cosmological models M_1 and M_2 that describe the expansion history of the Universe and predict a luminosity distance relation $D_L^{M_A}(z; \vec{\theta}_A)$ with parameters of the model $\vec{\theta}_A$, we can compute the Bayes factor

$$O_{12} = \frac{p(\vec{y}|M_1)}{p(\vec{y}|M_2)} \qquad p(\vec{y}|M_A) = \int p(\vec{y}|\vec{\theta}_A, M_A)\, p(\vec{\theta}_A\,|M_A)\, \mathrm{d}\vec{\theta}_A \qquad A = 1, 2 \qquad (16)$$

to understand which model is favored by a set $\vec{y}$ of observations, assuming that the two models are equally probable a priori. The Bayes factor is the ratio of the evidences of the two models, where each evidence $p(\vec{y}|M_A)$ is the probability of observing the data $\vec{y}$ under the assumptions of model M_A. The evidence is a quantity that combines the information given by the observed data and the prior belief on the parameters of a model. Given a set of observations $\vec{y}$, the likelihood $p(\vec{y}|\vec{\theta}_A, M_A)$ expresses which parameter regions are preferred by the data, whereas the prior $p(\vec{\theta}_A\,|M_A)$ tells which regions of the parameter space are more probable according to the model M_A.

The Bayes factor can be used to forecast how many measurements $\vec{y} = (z_j, D_j, \sigma_{D_j})$ of MBHB standard sirens are needed to conclude that the data strongly favor ΛCDM with respect to the deviation found in Ref. [13], assuming the underlying model providing the data is indeed ΛCDM. As observations, we use catalogues of luminosity distances $\vec{y} = (z_j, D_j, \sigma_{D_j})$ generated using the prescription explained in Sec. III. Whereas, the other two ingredients to calculate Bayes factors, the likelihood $p(\vec{y}|\vec{\theta}_A, M_A)$ and the priors $p(\vec{\theta}_A\,|M_A)$, are chosen according to the following prescriptions. In our case, the likelihood $p(\vec{y}|\vec{\theta}_A, M_A)$ is given by Eq. (14) and the expected value $D_L^{M_A}(z_j; \vec{\theta}_A)$ is given by the prediction of luminosity distance of the following two cosmological models:

- Model 1: the flat ΛCDM model predicts the luminosity distance relationship given in Eq. (7). Therefore we set $D_L^{M_1}(z; \theta_1) = D_L^{\Lambda\mathrm{CDM}}(z; \Omega_m)$ where we consider the single parameter of the model to be the matter density Ω_m.

- Model 2: the ALT best fit-model[1] predicts the luminosity distance relationship given in Eq. (5). Thus, we set $D_L^{M_2}(z; \vec{\theta}_2) = D_L^{\mathrm{ALT}}(z; a_2, a_3)$, where the parameters of the model are a_2 and a_3.

We fixed in both models $H_0 = 70\,\mathrm{km/s\ Mpc}$ and $c = 2.9979 \times 10^8 \mathrm{m/s}$ as done in Ref. [13]. We do not include the Hubble constant as an additional parameter to Model 2 because we aim to test exclusively the predictions of the Universe expansion history of the two cosmological models, and because it is a calibration constant for the ALT best-fit model. Even including the Hubble constant as an additional parameter to Model 2, we find that our results are not considerably affected.

The priors $p(\vec{\theta}_A|M_A)$ represent the current state of knowledge about the parameter values and our belief in the model parameters. The prior specification strongly influences the value of Bayes factors, especially if we have a low number of observations and peaked priors. On the other hand, generic or wide priors do not really specify the parameter space preferred by the given model and they do not allow any model comparison. Therefore, we adopt priors based on the analysis in Ref. [13].

For Model 1 (flat ΛCDM model) a suitable prior is the beta p.d.f. $\Omega_m \sim \beta(a, b)$ because of its support $\Omega_m \in [0, 1]$. Notwithstanding the tight constraints on the value of Ω_m from different cosmological measurements (see e.g. [11]), we choose a prior peaked on the value of matter density of $\Omega_m = 0.31 \pm 0.05$ found in Ref. [13] at low redshifts, where good agreement is obtained between the ΛCDM model and the quasar data. Therefore, we choose $\beta(a = 22.65, b = 50)$ as the prior on the ΛCDM model in order to have a median of 0.31 ± 0.05, where the uncertainty is the 68% credible interval. For comparison, the analysis of the CMB power spectrum combined with lensing and BAO give the following constraint on the matter density $\Omega_m = 0.315 \pm 0.007$ [11].

For Model 2 (ALT best-fit model) we approximate the posterior of a_2, a_3 of Fig. 2 (Fig. 5 in Ref. [13]) with a Multivariate Normal distribution $(a_2, a_3) \sim \mathcal{N}(\vec{\mu}, \Sigma)$:

$$p(\vec{\theta}_2|M_2) = p(a_2, a_2|M_2) = \exp\left(-\frac{1}{2}\left(\vec{\mathcal{A}} - \vec{\mu}\right)^{\mathrm{T}} \Sigma^{-1} \left(\vec{\mathcal{A}} - \vec{\mu}\right)\right) \frac{1}{2\pi\sqrt{\det\Sigma}} \qquad \vec{\mathcal{A}} = \begin{pmatrix} a_2 \\ a_3 \end{pmatrix}. \qquad (17)$$

---

[1] This is not truly a cosmological model, but it can represent a class of possible cosmological models that agree with the observations of Ref. [13].

And we set the prior of Model 2 to be the multivariate Gaussian distribution given by Eq. (17) with covariance matrix Σ and expected value $\vec{\mu}$ estimated from the MCMC posterior samples a_2, a_3 of Fig. 2 (Fig. 5 in Ref. [13]):

$$\vec{\mu} = \begin{pmatrix} 3.41309063 \\ 1.40248073 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 0.01961402 & -0.06005197 \\ -0.06005197 & 0.24775043 \end{pmatrix}. \tag{18}$$

The goodness of this approximation is confirmed by the very low Kullback-Leibler divergence ($\approx 4 \times 10^{-7}$) between the approximate multivariate Gaussian distribution and true posterior distribution of a_2, a_3. An advantage of using a multivariate normal distribution as the prior for Model 2 is that we can solve the double integral of the evidence $p(\vec{y}|M_2)$ analytically as shown in the Appendix. Having specified all the tools needed to calculate the Bayes factor, we can use several MBHB standard siren catalogues to study how many observations are necessary to test the claimed deviation from ΛCDM.

B. Results

Even if we expect ΛCDM to be favored because the data are generated using this model, our aim is to study the Bayes factor O_{12} as we add more observations. However, the Bayes factor itself depends on the redshifts of the observations, and measurements with different redshifts can favor one model over the other. Therefore we decided to study the distribution of Bayes factors by generating 10^4 realizations of catalogues of N_{SS} standard sirens.

For each value of N_{SS} we obtain the distribution of O_{12} as follows

- We draw a set of N_{SS} redshifts from the expected distribution of MBHB standard sirens and assign uncertainties σ_{D_j}, as described in Sec. III.

- We scatter the luminosity distance measurements around the flat ΛCDM model with a Normal distribution $D_j \sim \mathcal{N}(D_L^{\Lambda\text{CDM}}, \sigma_{D_j})$, as mentioned in Sec. III.

- We calculate the Bayes factor for the N_{SS} simulated standard sirens $\vec{y} = (z_j, D_j, \sigma_{D_j})$ using the likelihood in Eq. (14) and the priors of the two cosmological models with the aforementioned values of a, b, $\vec{\mu}$, Σ.

We then obtain 10^4 realizations of Bayes factors per each N_{SS}, which gives us a distribution of Bayes factors of possible future observations of MBHBs standard sirens. We can define the $\alpha\%$ *Evidence Interval* $I = [O_L, O_R]$ as the symmetric interval of Bayes factors which covers the population of Bayes factor with probability of $\alpha\%$:

$$\alpha\% = \int_{O_L}^{O_R} p(O_{12}) \mathrm{d}O_{12}.$$

Note that an evidence interval is neither a confidence interval nor a credible interval, because we are not estimating any parameters, but it tells us how likely we are to observe a realization of the Universe which allows us to distinguish between the two models. The distribution of the Bayes factors for the aforementioned priors is shown in Fig. 8.

As we add more observations, the distribution of Bayes factors initially widens, but the median, represented by the vertical line, shifts to higher values. The shift in the median is due to the data favoring the ΛCDM model (Model 1) with respect to the ALT best-fit model (Model 2) as we increase the number of observations. This is of course expected since the standard siren data are generated according to the ΛCDM model.

The widening of the distribution is a consequence of the larger space of realizations of Bayes factor per N_{SS}. In fact, if we draw a large number of N_{SS} we can obtain different redshift realizations that can favor more or less one model over the other one. Those catalogues with many high-redshift standard sirens easily discriminate between the two models and have high Bayes factors, whereas catalogues with many low-redshift standard sirens have lower Bayes factors.

To interpret how high the Bayes factor should be to discriminate between the two models, we refer to Kass and Raftery [84]. For Bayes factors above ~ 20 Model 1 is strongly favored with respect to Model 2, while for Bayes factors above ~ 150 the data very strongly favor Model 1 with respect to Model 2.

In order to assess how many standard sirens we need to distinguish between the two models, we plot in Fig. 9 the median, the 90% and 50% evidence interval of the Bayes factor distribution as a function of the number of standard sirens. In 50% of realizations of the Universe, the Bayes factor will lie above the red line, under the aforementioned assumptions. From Fig. 9 it is clear that already with four standard sirens we have a 50% chance that the Bayes factor would lie above the very strong evidence threshold, ruling out the deviation found in Ref. [13]. With 14 MBHB standard sirens we have a 95% chance that the two models will be distinguished by LISA.
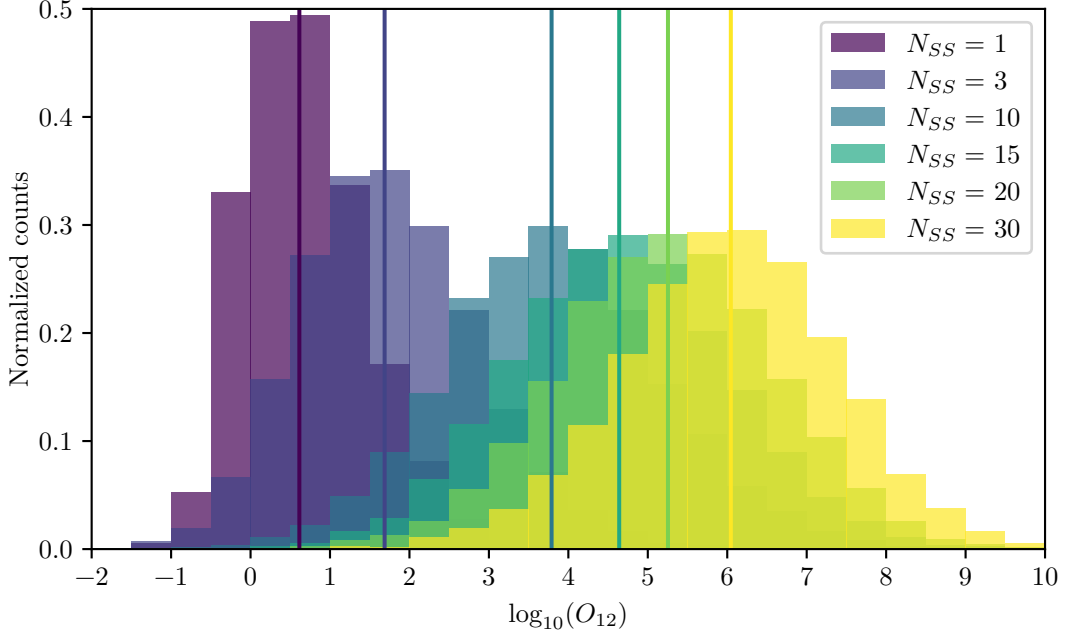
Figure 8. Normalized distribution of 10^4 Bayes factor O_{12} realizations for given value of N_{SS} obtained setting a beta prior on the ΛCDM of $\Omega_m \sim \beta(a = 22.65, b = 50)$ and a multivariate Gaussian prior $(a_2, a_3) \sim \mathcal{N}(\vec{\mu}, \Sigma)$ for the ALT best-fit model with $(\vec{\mu}, \Sigma)$ given in Eq. (18). The solid vertical lines indicate the median of the distribution.
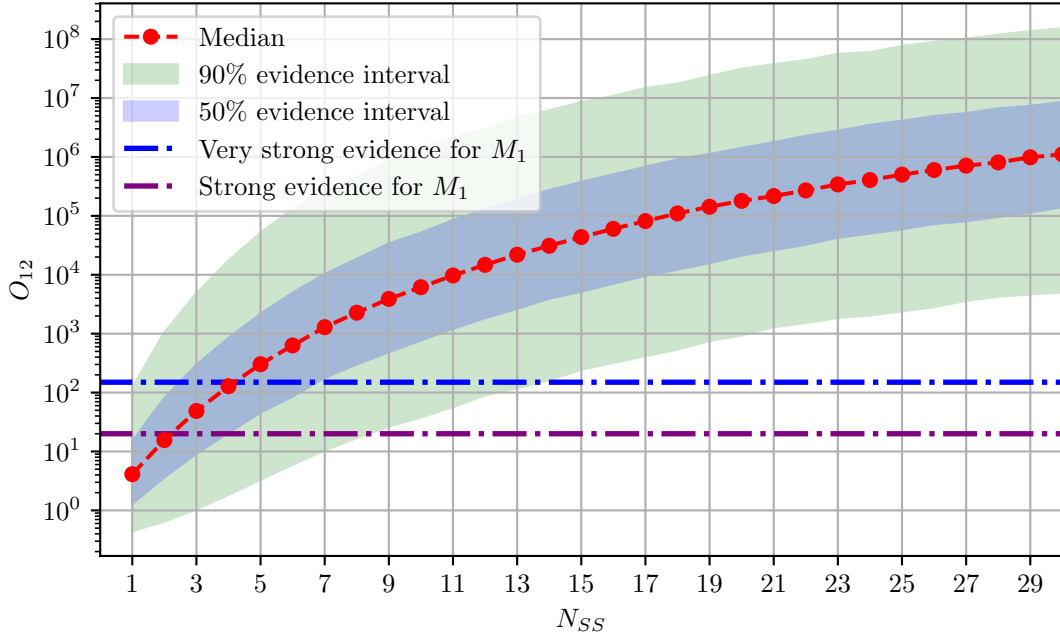

Figure 9. Bayes factor evolution as a function of the number of standard sirens N_{SS} obtained setting a beta prior on the ΛCDM of $\Omega_m \sim \beta(a = 22.65, b = 50)$ and multivariate Gaussian prior $(a_2, a_3) \sim \mathcal{N}(\vec{\mu}, \Sigma)$ for the ALT best-fit model with $(\vec{\mu}, \Sigma)$ given in Eq. (18). The red dots represent the median of the distribution of Bayes factor for given N_{SS}, and the shaded regions represent respectively the 90% and 50% evidence interval. The horizontal dashed lines are the threshold above which the ΛCDM model is favored with respect to the ALT best-fit model. As a reference, we report that the Bayes factor of the 15 representative SS dataset of Section III is 6387.

V.   DISCUSSION AND CONCLUSION

Constructing the Hubble diagram is a fundamental step to study the expansion history of the Universe, but it requires accurate and precise measurements of the luminosity distance. The extension of the Hubble diagram with quasar observations has not only opened up the possibility to explore higher redshift regions but it has also suggested a possible new tension with the current cosmological model, the ΛCDM model. In this work we addressed the crucial question of whether and when we will be able to rule out or confirm this tension. Observations of GWs from MBHB mergers with associated EM counterparts will be critical to explore the high redshift regions of the Hubble diagram and, thus, to test the claimed deviation from ΛCDM.

Assuming the validity of the ΛCDM model, we described how to generate catalogs of mock MBHB standard siren observations by LISA. We used a Bayesian approach to determine the relative favor of ΛCDM and the best fit model of Ref. [13], as determined by the data. This is expressed in terms of the Bayes factor, which can be calculated given priors of the model parameters, the likelihood function and the observed data. In the process, we reproduced some of the results of Ref. [13], in particular the posterior distribution of the phenomenological model, which we used as a basis for the prior beliefs on the deviation from ΛCDM. Given a fixed number of simulated standard sirens N_{SS} we used 10^4 catalog realizations to study the behavior of the Bayes factor and we found a 50%, 75% and 95% chance to rule out the deviation found by Ref. [13] with $N_{SS} = 4, 7, 14$, respectively. These numbers are within the range of the expected rates for LISA MBHB standard sirens [39], suggesting that LISA will indeed be able to discriminate the observed tension from the standard ΛCDM model.

Furthermore, we evaluated the constraints on the matter density and Hubble constant as a function of the number of MBHB standard sirens and we found that the Hubble parameter can be more precisely constrained than the matter density on average. The relative precision of the Hubble parameter ranges from $3 - 20\%$ for 10 standard sirens up to $1 - 4\%$ for 50. Whereas, the matter density can be constrained with relative precision ranging from $7 - 50\%$ for 10 standard sirens up to $4 - 7\%$ for 50. The combination of MBHB events with other standard sirens detectable by LISA, in particular stellar-origin black hole binaries and EMRIs, will allow to improve the constraints on H_0 found in this paper, perhaps to below-percent accuracy in optimistic scenarios. These results will help elucidate the current tension on the measured value of the Hubble constant, which may well not be solved by the time LISA will launch.

The great potential of MBHB mergers for cosmology lies however beyond the results one can obtain on H_0. As we showed in this paper, by using these events as standard sirens LISA will probe the expansion history at $z \gtrsim 2$ with an accuracy possibly not attained by future EM observations, even as a new suite of large-scale structure surveys [7–10, 85] in the post-reionisation matter-dominated regime, $2 \lesssim z \lesssim 6$, are poised to search for clues to the nature of dark energy and chart the cosmic acceleration. Quasars have recently been proposed anew as a standard candles in this redshift range [13]. Other methods may also open up this frontier in the coming years, such as line intensity mapping [86–88]. Reconciling the Hubble diagram among all these different methods will be a crucial activity in the future. LISA standard sirens offer a proven and complementary method to independently test the Hubble diagram at $z \gtrsim 2$. This synergy between GW and EM observations will provide the necessary confidence to reliably test the cosmic expansion at high redshift.

APPENDIX

The evidence is the integral of the product between the likelihood and the model prior. Here we show how it is possible to calculate explicitly the evidence for the ALT best-fit model (Model 2) with the likelihood and model prior given respectively in Eqs. (14) and (17).

Let us expand the argument of the exponential of Eq. (14):

$$\ln \tilde{\mathcal{L}} = \ln p(\vec{y}|a_2, a_3) = -\frac{\left(D_j - D_L^{\mathrm{ALT}}(z_j; a_2, a_3)\right)^2}{2\,\sigma_{D_j}^2} = -\frac{\xi^2}{2} \sum_j \frac{(D_j/\xi - x_j - x_j^2 a_2 - x_j^3 a_3)^2}{\sigma_{D_j}^2}\,, \qquad (19)$$

where we defined $\xi = c \ln 10/H_0$ and the sum over j is the sum over the SS measurements expressed in terms of

$x_j = \log_{10}(1 + z_j)$. We now define $\psi_j = D_j/\xi - x_j$ and we express the above sum as

$$\ln \tilde{\mathcal{L}} = -\sum_j \left\{ \frac{\psi_j^2}{\sigma_{D_j}^2} + a_2^2 \frac{x_j^4}{\sigma_{D_j}^2} + a_3^2 \frac{x_j^6}{\sigma_{D_j}^2} - 2a_2 \frac{\psi_j x_j^2}{\sigma_{D_j}^2} - 2a_3 \frac{\psi_j x_j^3}{\sigma_{D_j}^2} + 2a_2 a_3 \frac{x_j^5}{\sigma_{D_j}^2} \right\}. \tag{20}$$

If we express the above equation in terms of:

$$\vec{\mathcal{A}} = \begin{pmatrix} a_2 \\ a_3 \end{pmatrix} \qquad \mathcal{C} = \begin{pmatrix} \sum_j \frac{x_j^4}{\sigma_{D_j}^2} & \sum_j \frac{x_j^5}{\sigma_{D_j}^2} \\ \sum_j \frac{x_j^5}{\sigma_{D_j}^2} & \sum_j \frac{x_j^6}{\sigma_{D_j}^2} \end{pmatrix} \qquad \vec{\mathcal{B}} = \begin{pmatrix} -2\sum_j \frac{\psi_j x_j^2}{\sigma_{D_j}^2} \\ -2\sum_j \frac{\psi_j x_j^3}{\sigma_{D_j}^2} \end{pmatrix} \tag{21}$$

we obtain that the likelihood Eq. (14) can be expressed as:

$$\tilde{\mathcal{L}} = \exp\left\{ -\frac{\xi^2}{2} \left[\sum_j \frac{\psi_j^2}{\sigma_{D_j}^2} + \vec{\mathcal{A}}^{\mathrm{T}} \mathcal{C} \vec{\mathcal{A}} + \vec{\mathcal{B}}^{\mathrm{T}} \vec{\mathcal{A}} \right] \right\} \prod_j \frac{1}{\sqrt{2\pi\sigma_{D_j}^2}} \tag{22}$$

In order to obtain the evidence for the ALT model we need to multiply the likelihood by the model prior in Eq. (17):

$$p(\vec{\mathcal{A}}|M_2) = p(a_2, a_3|M_2) = \exp\left\{ -\frac{1}{2}\left(\vec{\mathcal{A}}\right)^{\mathrm{T}} \Sigma^{-1} \vec{\mathcal{A}} - \frac{1}{2}(\vec{\mu})^{\mathrm{T}} \Sigma^{-1} \vec{\mu} + (\vec{\mu})^{\mathrm{T}} \Sigma^{-1} \vec{\mathcal{A}} \right\} \frac{1}{2\pi\sqrt{\det\Sigma}} \tag{23}$$

So the evidence for the ALT model is

$$\begin{aligned} p(\vec{y}|M_2) =& \frac{1}{2\pi\sqrt{\det\Sigma}} \left(\prod_j \frac{1}{\sqrt{2\pi\sigma_{D_j}^2}} \right) \exp\left(-\frac{1}{2}(\vec{\mu})^{\mathrm{T}} \Sigma^{-1} \vec{\mu} - \frac{\xi^2}{2} \sum_j \frac{\psi_j^2}{\sigma_{D_j}^2} \right) \times \\ & \int \exp\left\{ \left(-\frac{\xi^2}{2}\vec{\mathcal{B}}^{\mathrm{T}} + (\vec{\mu})^{\mathrm{T}} \Sigma^{-1} \right) \vec{\mathcal{A}} - \frac{1}{2}\left(\vec{\mathcal{A}}\right)^{\mathrm{T}} \left(\Sigma^{-1} + \xi^2 \mathcal{C}\right) \vec{\mathcal{A}} \right\} \mathrm{d}a_2 a_3 = \\ =& \frac{1}{2\pi\sqrt{\det\Sigma}} \left(\prod_j \frac{1}{\sqrt{2\pi\sigma_{D_j}^2}} \right) \exp\left(-\frac{1}{2}(\vec{\mu})^{\mathrm{T}} \Sigma^{-1} \vec{\mu} - \frac{\xi^2}{2} \sum_j \frac{\psi_j^2}{\sigma_{D_j}^2} \right) \times \\ & \exp\left(\frac{1}{2}\vec{\mathcal{V}}^{\mathrm{T}} \left[\Sigma^{-1} + \xi^2 \mathcal{C}\right]^{-1} \vec{\mathcal{V}} \right) \frac{2\pi}{\sqrt{\det\left(\Sigma^{-1} + \xi^2 \mathcal{C}\right)}} \end{aligned} \tag{24}$$

where we introduced $\vec{\mathcal{V}} = -\frac{\xi^2}{2}\vec{\mathcal{B}} + \Sigma^{-1}\vec{\mu}$ in the last line. The above equation shows how it is possible to directly evaluate the evidence of the ALT best fit model using the observations $\vec{y}$.

---

[1] A. G. Riess *et al.* (Supernova Search Team), Astron. J. **116**, 1009 (1998), arXiv:astro-ph/9805201 [astro-ph].
[2] S. Perlmutter *et al.* (Supernova Cosmology Project), Astrophys. J. **517**, 565 (1999), arXiv:astro-ph/9812133 [astro-ph].
[3] R. R. Caldwell and M. Kamionkowski, Ann. Rev. Nucl. Part. Sci. **59**, 397 (2009), arXiv:0903.0866 [astro-ph.CO].
[4] D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo, Phys. Rept. **530**, 87 (2013), arXiv:1201.2434 [astro-ph.CO].
[5] A. G. Riess *et al.*, Astrophys. J. **826**, 56 (2016), arXiv:1604.01424 [astro-ph.CO].
[6] N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. F. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly, K. S. Dawson, A. Dey, M. Doi, M. Donahue, S. Deustua, P. Eisenhardt, E. Ellingson, L. Faccioli, V. Fadeyev, H. K. Fakhouri, A. S. Fruchter, D. G. Gilbank, M. D. Gladders, G. Goldhaber, A. H. Gonzalez, A. Goobar, A. Gude, T. Hattori, H. Hoekstra, E. Hsiao, X. Huang, Y. Ihara, M. J. Jee, D. Johnston, N. Kashikawa, B. Koester, K. Konishi, M. Kowalski, E. V. Linder, L. Lubin, J. Melbourne, J. Meyers, T. Morokuma, F. Munshi, C. Mullis, T. Oda, N. Panagia, S. Perlmutter, M. Postman, T. Pritchard, J. Rhodes, P. Ripoche, P. Rosati, D. J. Schlegel, A. Spadafora, S. A. Stanford, V. Stanishev, D. Stern, M. Strovink, N. Takanashi, K. Tokita, M. Wagner, L. Wang, N. Yasuda, H. K. C. Yee, and T. Supernova Cosmology Project, ApJ **746**, 85 (2012), arXiv:1105.3470 [astro-ph.CO].
[7] https://www.desi.lbl.gov/.
[8] https://euclid.jpl.nasa.gov/.
[9] https://www.lsst.org/.

[10] https://wfirst.gsfc.nasa.gov/.
[11] N. Aghanim *et al.* (Planck), Astron. Astrophys. **641**, A6 (2020), arXiv:1807.06209 [astro-ph.CO].
[12] S. Bisogni, G. Risaliti, and E. Lusso (2017) arXiv:1712.07515 [astro-ph.GA].
[13] G. Risaliti and E. Lusso, Nat. Astron. **3**, 272 (2019), arXiv:1811.02590 [astro-ph.CO].
[14] E. Lusso and G. Risaliti, ApJ **819**, 154 (2016), arXiv:1602.01090 [astro-ph.GA].
[15] E. Lusso, E. Piedipalumbo, G. Risaliti, M. Paolillo, S. Bisogni, E. Nardini, and L. Amati, Astron. Astrophys. **628**, L4 (2019), arXiv:1907.07692 [astro-ph.CO].
[16] F. Melia, Mon. Not. Roy. Astron. Soc. **489**, 517 (2019), arXiv:1907.13127 [astro-ph.CO].
[17] N. Khadka and B. Ratra, Mon. Not. Roy. Astron. Soc. **492**, 4456 (2020), arXiv:1909.01400 [astro-ph.CO].
[18] T. Yang, A. Banerjee, and E. O. Colgáin, (2019), arXiv:1911.01681 [astro-ph.CO].
[19] H. Velten and S. Gomes, Phys. Rev. **D101**, 043502 (2020), arXiv:1911.11848 [astro-ph.CO].
[20] W. L. Freedman, Nat. Astron. **1**, 0121 (2017), arXiv:1706.02739 [astro-ph.CO].
[21] A. G. Riess, Nature Rev. Phys. **2**, 10 (2019), arXiv:2001.03624 [astro-ph.CO].
[22] B. F. Schutz, Nature **323**, 310 (1986).
[23] B. F. Schutz, in *Proceedings, MPA / ESO / MPE / USM Conference on Lighthouses of the Universe: The Most Luminous Celestial Objects and their use for Cosmology: Garching, Munich, Germany, August 6-9, 2001* (2001) arXiv:gr-qc/0111095 [gr-qc].
[24] D. E. Holz and S. A. Hughes, Astrophys. J. **629**, 15 (2005), arXiv:astro-ph/0504616 [astro-ph].
[25] B. P. Abbott *et al.* (LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, MASTER), Nature **551**, 85 (2017), arXiv:1710.05835 [astro-ph.CO].
[26] M. Fishbach *et al.* (LIGO Scientific, Virgo), Astrophys. J. **871**, L13 (2019), arXiv:1807.05667.
[27] M. Soares-Santos *et al.* (DES, LIGO Scientific, Virgo), Astrophys. J. **876**, L7 (2019), arXiv:1901.01540.
[28] B. P. Abbott *et al.* (LIGO Scientific, Virgo), (2019), arXiv:1908.06060.
[29] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832.
[30] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, Nat **551**, 85 (2017), arXiv:1710.05835.
[31] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, Phys. Rev. X **9**, 011001 (2019), arXiv:1805.11579 [gr-qc].
[32] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, , arXiv:1908.06060 (2019).
[33] W. Del Pozzo, Phys. Rev. **D86**, 043011 (2012), arXiv:1108.1317 [astro-ph.CO].
[34] R. Gray *et al.*, (2019), arXiv:1908.06050 [gr-qc].
[35] D. Reitze *et al.*, (2019), arXiv:1907.04833 [astro-ph.IM].
[36] B. Sathyaprakash *et al.*, *Gravitational waves. Numerical relativity - data analysis. Proceedings, 9th Edoardo Amaldi Conference, Amaldi 9, and meeting, NRDA 2011, Cardiff, UK, July 10-15, 2011*, Class. Quant. Grav. **29**, 124013 (2012), [Erratum: Class. Quant. Grav.30,079501(2013)], arXiv:1206.0331 [gr-qc].
[37] P. Amaro-Seoane *et al.* (LISA), (2017), arXiv:1702.00786.
[38] N. Tamanini, C. Caprini, E. Barausse, A. Sesana, A. Klein, and A. Petiteau, JCAP **1604**, 002 (2016), arXiv:1601.07112 [astro-ph.CO].
[39] N. Tamanini, *Proceedings, 11th International LISA Symposium: Zurich, Switzerland, September 5-9, 2016*, J. Phys. Conf. Ser. **840**, 012029 (2017), arXiv:1612.02634 [astro-ph.CO].
[40] https://github.com/lorenzsp/StandardSirensVSQuasars.
[41] Y. Avni and H. Tananbaum, ApJ **262**, L17 (1982).
[42] M. Betoule *et al.* (SDSS), Astron. Astrophys. **568**, A22 (2014), arXiv:1401.4064 [astro-ph.CO].
[43] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, PASP **125**, 306 (2013), 1202.3665.
[44] E. Sonbas, K. S. Dhuga, and E. Göğüş, ApJ **873**, L12 (2019).
[45] F. Salvestrini, G. Risaliti, S. Bisogni, E. Lusso, and C. Vignali, Astron. Astrophys. **631**, A120 (2019), arXiv:1909.12309 [astro-ph.GA].
[46] K. Breivik, K. Kremer, M. Bueno, S. L. Larson, S. Coughlin, and V. Kalogera, Astrophys. J. **854**, L1 (2018), arXiv:1710.08370 [astro-ph.SR].
[47] V. Korol, E. M. Rossi, P. J. Groot, G. Nelemans, S. Toonen, and A. G. A. Brown, Mon. Not. Roy. Astron. Soc. **470**, 1894 (2017), arXiv:1703.02555 [astro-ph.HE].
[48] V. Korol, O. Koop, and E. M. Rossi, Astrophys. J. **866**, L20 (2018), arXiv:1808.05959 [astro-ph.HE].
[49] M. Y. M. Lau, I. Mandel, A. Vigna-Gómez, C. J. Neijssel, S. Stevenson, and A. Sesana, Mon. Not. Roy. Astron. Soc. **492**, 3061 (2020), arXiv:1910.12422 [astro-ph.HE].
[50] A. Sesana, Phys. Rev. Lett. **116**, 231102 (2016), arXiv:1602.06951 [gr-qc].
[51] S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. L. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein, Phys. Rev. **D95**, 103012 (2017), arXiv:1703.09722 [gr-qc].
[52] A. Klein *et al.*, Phys. Rev. **D93**, 024003 (2016), arXiv:1511.05581 [gr-qc].
[53] C. Caprini *et al.*, JCAP **1604**, 001 (2016), arXiv:1512.06239 [astro-ph.CO].
[54] N. Bartolo *et al.*, JCAP **1612**, 026 (2016), arXiv:1610.06481 [astro-ph.CO].
[55] C. Caprini and D. G. Figueroa, Class. Quant. Grav. **35**, 163001 (2018), arXiv:1801.04268 [astro-ph.CO].
[56] M. Eracleous, S. Gezari, A. Sesana, T. Bogdanovic, M. MacLeod, N. Roth, and L. Dai, (2019), arXiv:1902.06612 [astro-ph.HE].
[57] Y. Wang, F. Wang, Y. Zou, and Z. Dai, Astrophys. J. Lett. **886**, L22 (2019), arXiv:1911.04117 [astro-ph.HE].
[58] B. Zhang, Astrophys. J. **873**, L9 (2019), arXiv:1901.11177 [astro-ph.HE].
[59] K. Kyutoku and N. Seto, Phys. Rev. **D95**, 083525 (2017), arXiv:1609.07142 [astro-ph.CO].

[60] W. Del Pozzo, A. Sesana, and A. Klein, Mon. Not. Roy. Astron. Soc. **475**, 3485 (2018), arXiv:1703.01300 [astro-ph.CO].
[61] C. L. MacLeod and C. J. Hogan, Phys. Rev. **D77**, 043512 (2008), arXiv:0712.0618 [astro-ph].
[62] D. Laghi *et al.*, (2020), *in preparation*.
[63] C. Palenzuela, L. Lehner, and S. L. Liebling, Science **329**, 927 (2010), arXiv:1005.1067 [astro-ph.HE].
[64] M. Dotti, A. Sesana, and R. Decarli, Advances in Astronomy **2012**, 940568 (2012), arXiv:1111.0664 [astro-ph.CO].
[65] B. Giacomazzo, J. G. Baker, M. C. Miller, C. S. Reynolds, and J. R. van Meter, Astrophys. J. **752**, L15 (2012), arXiv:1203.6108 [astro-ph.HE].
[66] B. Kocsis, Z. Haiman, and K. Menou, Astrophys. J. **684**, 870 (2008), arXiv:0712.1144 [astro-ph].
[67] R. O'Shaughnessy, D. L. Kaplan, A. Sesana, and A. Kamble, Astrophys. J. **743**, 136 (2011), arXiv:1109.1050 [astro-ph.CO].
[68] D. L. Kaplan, R. O'Shaughnessy, A. Sesana, and M. Volonteri, ApJ **734**, L37 (2011), arXiv:1105.3653 [astro-ph.HE].
[69] Z. Haiman, Phys. Rev. **D96**, 023004 (2017), arXiv:1705.06765 [astro-ph.HE].
[70] C. Caprini and N. Tamanini, JCAP **10**, 006 (2016), arXiv:1607.08755 [astro-ph.CO].
[71] R.-G. Cai, N. Tamanini, and T. Yang, JCAP **05**, 031 (2017), arXiv:1703.07323 [astro-ph.CO].
[72] E. Belgacem *et al.* (LISA Cosmology Working Group), JCAP **07**, 024 (2019), arXiv:1906.01593 [astro-ph.CO].
[73] C. M. Hirata, D. E. Holz, and C. Cutler, Phys. Rev. D **81**, 124046 (2010), arXiv:1004.3988 [astro-ph.CO].
[74] C. Shapiro, D. Bacon, M. Hendry, and B. Hoyle, Mon. Not. Roy. Astron. Soc. **404**, 858 (2010), arXiv:0907.3635 [astro-ph.CO].
[75] B. Kocsis, Z. Frei, Z. Haiman, and K. Menou, Astrophys. J. **637**, 27 (2006), arXiv:astro-ph/0505394.
[76] T. G. F. Li, Springer Thesis (2015), 10.1007/978-3-319- 19273-4.
[77] S. Marsat, J. G. Baker, and T. Dal Canton, (2020), arXiv:2003.00357 [gr-qc].
[78] T. Dahlen *et al.*, Astrophys. J. **775**, 93 (2013), arXiv:1308.5353 [astro-ph.CO].
[79] O. Ilbert *et al.*, Astron. Astrophys. **556**, A55 (2013), arXiv:1301.3157 [astro-ph.CO].
[80] T. Kruhler *et al.*, Astron. Astrophys. **526**, A153 (2011), arXiv:1011.1205 [astro-ph.HE].
[81] K. S. Dawson *et al.*, Astron. J. **151**, 44 (2016), arXiv:1508.04473 [astro-ph.CO].
[82] L. Amendola *et al.*, Living Rev. Rel. **21**, 2 (2018), arXiv:1606.00180 [astro-ph.CO].
[83] LISA-Science-Group (LISA), in preparation (2021).
[84] R. E. Kass and A. E. Raftery, Journal of the American Statistical Association **90**, 773.
[85] P. Bull, M. White, and A. Slosar, (2020), arXiv:2007.02865 [astro-ph.CO].
[86] J. L. Bernal, P. C. Breysse, and E. D. Kovetz, Phys. Rev. Lett. **123**, 251301 (2019), arXiv:1907.10065 [astro-ph.CO].
[87] K. S. Karkare and S. Bird, Phys. Rev. D **98**, 043529 (2018), arXiv:1806.09625 [astro-ph.CO].
[88] E. D. Kovetz *et al.*, (2017), arXiv:1709.09066 [astro-ph.CO].