

# Cosmology and dark energy from joint gravitational wave-GRB observations

Enis Belgacem,<sup>a</sup> Yves Dirian,<sup>b</sup> Stefano Foffa,<sup>a</sup> Eric J. Howell,<sup>c</sup>  
Michele Maggiore<sup>a</sup> and Tania Regimbau<sup>d,e</sup>

<sup>a</sup>Département de Physique Théorique and Center for Astroparticle Physics,  
Université de Genève, 24 quai Ansermet, CH-1211 Genève 4, Switzerland

<sup>b</sup>Center for Theoretical Astrophysics and Cosmology, Institute for Computational Science,  
University of Zürich, CH-8057 Zürich, Switzerland

<sup>c</sup>OzGrav-UWA, School of Physics and Astrophysics, University of Western Australia,  
Crawley WA 6009, Australia

<sup>d</sup>Laboratoire d'Annecy-le-Vieux de Physique des Particules (LAPP),  
Université Savoie Mont Blanc, CNRS/IN2P3, F-74941 Annecy, France

<sup>e</sup>Artemis, Université Côte d'Azur, Observatoire de la Côte d'Azur CNRS,  
CS 34229, F-06304 Nice Cedex 4, France

E-mail: [Enis.Belgacem@unige.ch](mailto:Enis.Belgacem@unige.ch), [yves.dirian@ics.uzh.ch](mailto:yves.dirian@ics.uzh.ch), [Stefano.Foffa@unige.ch](mailto:Stefano.Foffa@unige.ch),  
[eric.howell@uwa.edu.au](mailto:eric.howell@uwa.edu.au), [Michele.Maggiore@unige.ch](mailto:Michele.Maggiore@unige.ch),  
[tania.regimbau@lapp.in2p3.fr](mailto:tania.regimbau@lapp.in2p3.fr)

Received July 3, 2019

Accepted July 29, 2019

Published August 13, 2019

**Abstract.** Gravitational-wave (GW) detectors can contribute to the measurement of cosmological parameters and to testing the dark-energy sector of alternatives to  $\Lambda$ CDM, by using standard sirens. In this paper we focus on binary neutron stars with a counterpart detected through a gamma-ray burst (GRB), both at a second-generation network made by advanced LIGO+advanced Virgo+LIGO India+Kagra, and at third-generation (3G) detectors, discussing in particular the cases of a single Einstein Telescope (ET), and of a network of ET plus two Cosmic Explorer (CE). We construct mock catalogs of standard sirens, using different scenarios for the local merger rate and for the detection of the electromagnetic counterpart. For 3G detectors we estimate the coincidences with a GRB detector with the characteristics of the proposed THESEUS mission. We discuss how these standard sirens with a GRB counterpart can improve the determination of cosmological parameters (and particularly of  $H_0$ ) in  $\Lambda$ CDM, and we then study how to extract information on dark energy, considering both a non-trivial dark energy equation of state and modified GW propagation. We find that a 2G detector network can already reach, over several years of data taking, an interesting sensitivity to modified GW propagation, while a single ET detector would have a remarkable potential for discovery. We also find that, to fully exploit the potential of a

ET+CE+CE network, it is necessary a much stronger program of search for electromagnetic counterparts (or else to resort to statistical methods for standard sirens), and furthermore gravitational lensing can become a limiting factor.

**Keywords:** gravitational waves / experiments, gravitational waves / theory, modified gravity

**ArXiv ePrint:** [1907.01487](https://arxiv.org/abs/1907.01487)

---

**Contents**

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>1</b>  |
| <b>2</b> | <b>Construction of mock source catalogs</b>  | <b>4</b>  |
| 2.1      | GW events  | 4         |
| 2.2      | Electromagnetic counterpart  | 7         |
| 2.2.1    | GRB coincidences with the HLVKI network  | 7         |
| 2.2.2    | GRB coincidences with 3G detectors   | 9         |
| 2.3      | Events rates, redshift distributions and analytic estimates for $\Delta d_L(z)/d_L(z)$                           | 10        |
| 2.3.1    | Events at a single ET detector   | 10        |
| 2.3.2    | Events at a ET+CE+CE network   | 14        |
| <b>3</b> | <b>Constraints from standard sirens on <math>H_0</math> and <math>\Omega_M</math> in <math>\Lambda</math>CDM</b> | <b>16</b> |
| 3.1      | Results for the HLVKI network  | 18        |
| 3.2      | Results for ET   | 18        |
| 3.3      | Results for ET+CE+CE   | 20        |
| <b>4</b> | <b>Constraints from standard sirens on dark energy</b>   | <b>21</b> |
| 4.1      | Testing the dark energy sector: DE equation of state and modified GW propagation                                 | 21        |
| 4.2      | Results for the HLVKI network  | 25        |
| 4.3      | Results for ET   | 26        |
| 4.4      | Results for ET+CE+CE   | 31        |
| <b>5</b> | <b>Conclusions</b>   | <b>35</b> |
| <b>A</b> | <b>Catalogs of sources</b>   | <b>38</b> |

---

**1 Introduction**

The first observations of gravitational waves (GWs) from binary black-holes coalescences [1–6], as well as the first observation of a neutron star binary coalescence [7], together with the associated  $\gamma$ -ray burst (GRB) [8–10] and the follow-up studies of the electromagnetic counterpart (see [11] and references therein) have opened the era of GW astronomy. In the near future, with advanced LIGO and advanced Virgo reaching their target sensitivity, and other detectors such as KAGRA and LIGO-India joining the search, it is expected that such detections will take place routinely. On a longer timescale the space interferometer LISA [12], that is expected to fly in 2034, and third-generation (3G) ground-based interferometers currently under study, such as the Einstein Telescope (ET) in Europe [13, 14] and Cosmic Explorer (CE) in the US [15], will have the potential of detecting a large number of coalescing compact binaries at cosmological redshifts. The detection of the GWs from a coalescing binary allows a direct measurement of its luminosity distance  $d_L$  [16], so these sources are referred to as “standard sirens”, the GW analogue of standard candles. Much work has been devoted to investigating the cosmological informations that could be obtained

from such measurements, either when the redshift of the source is obtained thanks to the detection of an electromagnetic counterpart, or using statistical methods [17–34].

In this paper we contribute to the currently ongoing effort for exploring the scientific potential of 3G interferometers (see e.g. [35–37]), by performing an updated study of the cosmological information that can be obtained from the observation of standard sirens. We focus on binary neutron stars (BNS) with an electromagnetic counterpart. This type of analysis is appropriate for ground-based detectors, since in this case BNS merge within the bandwidth of the detector, and thus can in principle be detected in coincidence with an electromagnetic signal that can allow us to identify the host galaxy and therefore obtain the redshift.<sup>1</sup> Using the strategy presented in [39–44], we begin by constructing mock catalogs of BNS detections at GW detectors. We consider first a network of second-generation (2G) detectors composed by advanced LIGO-Hanford, advanced LIGO-Livingston, advanced Virgo, Kagra and LIGO India, assumed to be at their target sensitivity. We will refer to this as the HLVKI network. We will then consider third-generation (3G) detectors studying two different configurations, namely a single ET detector, and a network made of a single Einstein Telescope plus two Cosmic Explorers.<sup>2</sup>

We will then study the possibility of simultaneous detection of an electromagnetic counterpart, focusing on the case of a joint GW-GRB detection. A single GW detector, even in a triangular configuration as planned for ET, has limited localization capabilities.<sup>3</sup> However, the detection of a temporally coincident GRB can still allow for the measurement of the redshift of the source; indeed, GRB satellites such as Neil Gehrels Swift Observatory (*Swift*; <https://swift.gsfc.nasa.gov/>) regularly obtain redshifts of GRBs, without the need of a GW localization.<sup>4</sup> In particular, for 3G detectors we will estimate the expected number and the redshift distributions of coincidences between GW events and the electromagnetic signal observed at a GRB detector with the characteristics of the proposed THESEUS mission [46–48], that could be in operation at the same time as 3G detectors.

A network of third-generation GW detectors, such as the ET+CE+CE configuration, would instead localize the source, whose redshift could then be measured also by optical/IR telescopes. For instance, the electromagnetic signal from a kilonova associated to a BNS coalescence could be detected up to  $z \simeq 0.55$  by optical imaging at LSST (<https://www.lsst.org/>) and Subaru (<https://www.naoj.org/>), up to  $z \simeq 0.76$  with infrared imaging at WFIRST (<https://wfirst.gsfc.nasa.gov/>), and up to  $z \simeq 0.37$  by optical spectroscopy at E-ELT (<https://www.eso.org/sci/facilities/eelt/>) (for a recent compilation of kilo-

---

<sup>1</sup>For the space interferometer LISA the best studied example of standard siren with an expected electromagnetic counterpart is the coalescence of supermassive BH binaries at large redshift,  $z \lesssim 8$  (see [28] and references therein). Thus, the construction of the source catalog and the methodology are completely different from those used here. A study of the perspective for observing dark energy and modified gravity with LISA using supermassive BH binaries has been recently presented in [38].

<sup>2</sup>The ET+CE+CE network is the baseline configuration that is studied in the 3G Science Case document, that is currently being developed by GWIC 3G Committee <https://gwic.ligo.org/>, the GWIC 3G Science Case Team and the International 3G Science Team Consortium.

<sup>3</sup>Some localization can however be obtained by a single ET detector, especially in the ET-D configuration based on a xylophone design, because in this case the low-frequency cutoff  $f_{\text{low}}$  of the sensitivity curve is quite small and the duration of the signal in the detector bandwidth has a strong dependence on  $f_{\text{low}}$ , proportional to  $f_{\text{low}}^{-8/3}$ . For a low-mass system such as a BNS, the signal stays in the ET-D bandwidth for 0.29 days for  $f_{\text{low}} = 3$  Hz and for 5.44 days for  $f_{\text{low}} = 1$  Hz. The effect of the Earth rotation during such a time span then allows us to localize the source, with an average angular resolution around  $150^\circ$  for a BNS at  $z = 0.1$  [45].

<sup>4</sup>Swift attempts to obtain redshifts by using its UV/Optical Telescope, UVOT, or through arc-second accuracy positional information that is relayed to ground based telescopes.

nova detection ranges from a large number of facilities see [https://workshops.aei.mpg.de/3Gsubcomm/wp-content/uploads/sites/29/2018/11/3GMMO\\_Potsdam.pdf](https://workshops.aei.mpg.de/3Gsubcomm/wp-content/uploads/sites/29/2018/11/3GMMO_Potsdam.pdf) and ref. [36]).<sup>5</sup> However, it is currently difficult to estimate how much telescope time will be devoted by these facilities to the follow-up of GW events. In this paper we limit ourselves to the coincidence with GRB detectors, but we should keep in mind that, for a network with significant localization capabilities such as ET+CE+CE, at  $z \lesssim 0.5$  a significant number of coincidences with optical/IR telescopes is in principle possible, to the extent that the number of such coincidences could be much larger than those obtained from GRB detections, so for such a network our estimates will be conservative. To construct our mock source catalogs we will examine different possibilities for the local merger rate and for the probability of determining the redshift through the detection of an associated GRB. This will lead to different scenarios, more conservative or more optimistic, that will be presented in section 2. In particular, in section 2.3 we will present detailed and ready-to-use results for the redshift distribution of the GW events at 3G detectors and of the coincidences with GRBs, as well as expressions for the observational error  $\Delta d_L/d_L$  as a function of  $z$ .

In section 3, using a Markov Chain Monte Carlo (MCMC) analysis, we study how the detection of standard sirens from our mock catalogs would contribute to the knowledge of the cosmological parameters, in the context of  $\Lambda$ CDM, considering the HLVKI network in section 3.1 and 3G detectors in sections 3.2 and 3.3. For both 2G and 3G detectors, we combine the contribution from standard sirens with existing datasets from cosmic microwave background (CMB), baryon acoustic oscillation (BAO) and type Ia supernovae (SNe), to remove degeneracies and improve the accuracy of cosmological parameter reconstruction. For 3G detectors, we find that useful results can also sometimes be obtained already just from standard sirens, without external datasets.

In section 4 we extend the analysis performed for  $\Lambda$ CDM, by considering a more general dark energy (DE) sector. Beside the effect of the DE equation of state, we will study the effect induced by modified GW propagation. Indeed, as we will recall in section 4.1, it has been realized in recent years that modified gravity models that predict a non-trivial DE equation of state also predict deviations from general relativity in the GW propagation across cosmological distances, even in theories where the speed of gravity is equal to  $c$  [32, 33, 50–61]. As recently found in [32, 33], in a generic modified gravity model the effect on the luminosity distance induced by modified GW propagation dominates over that from the DE equation of state. Furthermore, it is specific to GW observations, making it a prime observable for the study of dark energy and modified gravity at 3G GW detectors. In ref. [33] has been proposed a simple parametrization of the effect on the luminosity distance induced by modified GW propagation, in terms of two parameters  $(\Xi_0, n)$ , that complements the  $(w_0, w_a)$  parametrization of the DE equation of state. Recently, in [38] it has been found that this parametrization covers almost all viable modified gravity models that have been considered in the literature. An analysis of the possibility of observing modified GW propagation at ET was already presented in [33]. In that paper, following the standard working hypothesis used to date in the literature, it was simply assumed that, given the detection rate at ET, which is estimated as  $\mathcal{O}(10^5)$  BNS/yr, in a few years one could collect  $\mathcal{O}(10^3)$  events with an electromagnetic counterpart. To go beyond this simple estimate, one needs a model for how the counterpart is detected, to see first of all whether the figure of  $\mathcal{O}(10^3)$  events with

<sup>5</sup>A recent detailed study of the best strategies for observing at LSST the kilonova associated to a GW event detected at advanced LIGO/Virgo is presented in [49]. In that paper, however, the authors are interested in 2G detectors, so the mock LSST data are generated using a maximal GW detection range of 450 Mpc.

counterpart is realistic. Furthermore, once one specifies how the counterpart is observed, one can study how this affects the redshift distribution of the standard sirens with observed electromagnetic counterpart which in [38], again following the standard assumption in the literature, was just assumed to follow a simple distribution determined by the star formation rate. In section 4 of the present paper we will improve on these results by using our more realistic scenarios for both the catalog of GW sources and the detection of the electromagnetic counterparts (restricted to the case of a GRB observation), and we will extend the study done for ET also to the 2G HLVKI network, and to the ET+CE+CE network. Section 5 contains our conclusions.

## 2 Construction of mock source catalogs

### 2.1 GW events

In order to simulate a catalog of binary neutron stars coalescences, we first produce an extra-galactic population of neutron star binaries using the Monte Carlo algorithm developed in [39–44]. We use the fiducial model of [62] for the distribution of the parameters and we proceed as follows for each source: the location in the sky  $\hat{\Omega}$ , the cosine of the orientation  $\iota$ , the polarization  $\psi$  and the phase of the signal at coalescence  $\phi_0$ , are drawn from uniform distributions. The redshift is drawn from a (normalized) probability distribution  $p(z)$ ,

$$p(z) = \frac{R_z(z)}{\int_0^{10} R_z(z) dz}, \quad (2.1)$$

where  $R_z(z)$  is the merger rate density per unit redshift, in the observer frame. It can be expressed as<sup>6</sup>

$$R_z(z) = \frac{R_m(z)}{1+z} \frac{dV(z)}{dz}, \quad (2.2)$$

where  $dV/dz$  is the comoving volume element and  $R_m$  is the rate per volume in the source frame. The latter is given by

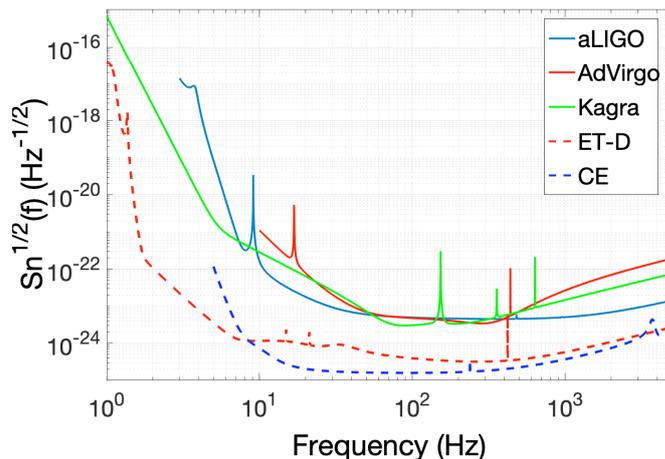
$$R_m(z) = \int_{t_{\min}}^{t_{\max}} R_f[t(z) - t_d] P(t_d) dt_d, \quad (2.3)$$

where  $R_f(t)$  is the formation rate of massive binaries,  $P(t_d)$  is the distribution of the time delay  $t_d$  between the formation of the massive progenitors and their merger, and  $t(z)$  is the age of the Universe at the time of merger. We assume that  $R_f(t)$  in eq. (2.3) follows the cosmic star formation rate, for which we use the recent model of [63]. We further assume that the time delay distribution follows  $P(t_d) \propto t_d^\alpha$ , with  $\alpha = -1$  for  $t_d > t_{\min}$ , where  $t_{\min} = 20$  Myr is the minimum delay time for a massive binary to evolve until coalescence, and  $t_{\max}$  is a maximum time delay, set equal to the Hubble time. The overall normalization is fixed by requiring that the value of  $R_m$  at  $z = 0$  agrees with the local rate estimated from the O1 LIGO observation run and the O2 LIGO/Virgo observation run [6], using the median rates obtained from the GstLAL pipeline. The result depends on the assumption for the mass distribution of the neutron stars. For a flat mass distribution

$$R_m(z = 0) = 662 \text{ Gpc}^{-3} \text{ yr}^{-1}, \quad (2.4)$$

---

<sup>6</sup>We correct here a typo in eq. (2) of [44], where a spurious  $\int dz$  appears on the right-hand side.



**Figure 1.** The strain sensitivities of advanced and 3G GW detectors. For ET we use the ET-D sensitivity curve.

while for a Gaussian mass distribution

$$R_m(z = 0) = 920 \text{ Gpc}^{-3} \text{ yr}^{-1}. \quad (2.5)$$

In the following we will refer to them just as the “O2 rates”, and we will give our results both for the flat distribution and for the Gaussian distribution. To have a quantitative measure of how the results depend on our astrophysical assumptions, we will also generate alternative catalogs of GW events by assuming a Madau-Dickinson star formation rate [64] and an exponential time delay between formation and merger with an e-fold time of 100 Myr [65]. This will also allow us to compare with the results presented in table 1 of ref. [36], which also computes the number of BNS detected per year at the HLVKI network and at the ET+CE+CE network under these astrophysical assumptions [and using a local comoving BNS merger rate of  $1000 \text{ Gpc}^{-3} \text{ yr}^{-1}$ , that is close to our rate for a Gaussian mass distribution, eq. (2.5)].

Next, for each BNS generated by this procedure, we determine if its resultant GW emission is detectable with a given GW detector network. We consider three cases: (1) a 2G network composed by advanced LIGO-Hanford+advanced LIGO-Livingston+advanced Virgo+Kagra+LIGO India (HLVKI). (2) A single 3G detector, chosen according to current estimates for the sensitivity of the Einstein Telescope. (3) A three-detector network made by ET and two CE. The sensitivity curves that we use are shown in figure 1.<sup>7</sup> The signal-to-noise ratio (SNR),  $\rho_a$ , detected by matched filtering with an optimum filter in the ideal case of Gaussian noise, in a detector labeled  $a$ , is

$$\rho_a^2 = 4 \int_0^\infty df \frac{|F_{+,a} \tilde{h}_+ + F_{\times,a} \tilde{h}_\times|^2}{S_{n,a}}, \quad (2.6)$$

where  $f$  is the GW frequency in the observer frame,  $\tilde{h}_+$  and  $\tilde{h}_\times$  the Fourier transforms of the GW strain amplitudes of + and  $\times$  polarizations,  $F_{+,a}$  and  $F_{\times,a}$  are the antenna response functions to the GW + and  $\times$  polarizations, and  $S_{n,a}(f)$  is the one-sided noise power spectral

<sup>7</sup>The ET and CE sensitivity curves, as well as the assumed locations of ET (in Europe) and two CE (in the US) correspond to the choices currently used to develop the Science Case for 3G detectors.

density (PSD) of detector  $a$ . The coherent SNR, assuming uncorrelated noises among the detectors, is simply given by the quadrature sum of the individual SNRs,  $\rho_{\text{tot}}^2 = \sum_a \rho_a^2$ . The triangular configuration of ET provides three independent differential signals between the arms, equivalent to three detectors, and again the coherent SNR is given by the quadrature sum of the individual SNRs for these three equivalent detectors.

For low-mass systems such as BNS the SNR in one detector is dominated by the inspiral part of the signal and is then given

$$\rho_a^2 = \frac{5}{6} \frac{[G\mathcal{M}(1+z)]^{5/3} \mathcal{F}_a^2}{c^3 \pi^{4/3} d_L^2(z)} \int_{f_{\text{min}}}^{f_{\text{insp}}(z)} df \frac{f^{-7/3}}{S_{n,a}(f)}. \quad (2.7)$$

Here  $\mathcal{M}$  is the intrinsic chirp mass, a combination of the two component masses,  $d_L(z)$  is the luminosity distance,  $G$  is the gravitational constant,  $c$  is the speed of light,  $f_{\text{min}}$  is the low frequency limit of the detector and  $f_{\text{insp}}(z) = f_{\text{insp}}/(1+z)$  is the observed (redshifted) gravitational-wave frequency at the end of the inspiral phase. The factor

$$\mathcal{F}_a^2 = \frac{(1 + \cos^2 \iota)^2}{4} F_{+,a}^2 + \cos^2 \iota F_{\times,a}^2, \quad (2.8)$$

characterizes the detector response. In order to decide which detectors contribute to the combined SNR, we assume that each detector has a duty cycle of 80%. We then classify the event as detectable if the combined SNR among the detectors in the network,  $\rho_{\text{tot}}$ , is larger than a SNR threshold level, that we take to be  $\rho_{\text{threshold}} = 12$ .<sup>8</sup> Both for the HLVKI network and for 3G detectors we assume for definiteness 10 years of running (which, given the 80% duty cycle for each detector, corresponds to a shorter stream of actual coincident data). This assumption should be taken as a limiting case (which, for the 2G case, is also necessary to have a sufficient sample of events to obtain the convergence of our MCMC); however, the results for a shorter time span  $T$  can be obtained basically by rescaling our results by a factor  $\sqrt{T/10 \text{ yr}}$ , corresponding to the fact that, with a large number of events  $N$ , the error scales roughly as  $1/\sqrt{N}$ .

In order to generate our mock catalogs of measured luminosity distances of standard sirens, we assume a fiducial  $\Lambda$ CDM model with  $\Omega_M = 0.3087$  and  $H_0 = 67.64 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , corresponding to the mean values obtained from the CMB+BAO+SNe dataset that we will use, which is presented in detail in section 3. Extracting randomly the redshift of the source from the theoretical distribution obtained from eqs. (2.1)–(2.3) and using our fiducial cosmological model, we obtain a value of  $d_L(z)$  for each source. To take into account the observational error in the reconstruction of the luminosity distance from the GW data, we scatter randomly the values of  $d_L(z)$  according to a Gaussian distribution with a width  $\Delta d_L(z)$  equal to the expected error in the reconstruction. For each generated event, this is estimated from  $\Delta d_L/d_L = 1/\text{SNR}$ , following e.g. ref. [18]. Note that, comparing with the result of an actual mock parameter reconstruction, one finds that, because of the degeneracy with the inclination angle, this can result in an underestimate of the actual value of  $\Delta d_L/d_L$ , by a factor which has a significant scatter from event to event, but is generically  $\sim 2$  [20]. However, for GW signals detected in coincidence with a GRB (which are the signals that we consider in this paper), assuming that the GRB is beamed within an angle of about  $25^\circ$  one

<sup>8</sup>Notice that we do not ask for the signal to be seen at least in two detectors, but we only use the SNR cutoff. In any case, for 5 detectors, the probability to have just one detector online is less than 1% [more precisely,  $(0.2)^4 \times 0.8 \times 5 = 0.64\%$ ], so our HLVKI catalogs are basically not affected by the inclusion of events seen in just one detector.

finds that the correlation between distance and inclination is substantially broken, and the above estimate becomes more accurate [20].

Beside the instrumental error, we must consider the error due to lensing. Following [22, 23], we model it as

$$\left(\frac{\Delta d_L(z)}{d_L(z)}\right)_{\text{lensing}} \simeq 0.05z, \quad (2.9)$$

and we add it in quadrature to the instrumental error. However, we will see below that, for the sources at  $z < 1.5$ , that will largely dominate our results, the lensing error is subdominant with respect to the instrumental error.

If the source is at very low redshift, once determined the measured redshift as discussed in section 2.2, to obtain the cosmological redshift we must correct for the peculiar Hubble flow. This is estimated adding an error on  $z$  corresponding to a recessional velocity of the host galaxy of 200 km/s, as in ref. [66].

## 2.2 Electromagnetic counterpart

In general, to identify the counterpart, one can consider two possible strategies. The first, that has been implemented successfully with GW170817, consists in having a network of GW detectors, that allows us to localize the source relatively well. Then, the follow-up with telescopes working, e.g., in the optical or IR can identify the host galaxy and determine its redshift. The second possibility, that can be applied even when no GW localization is available (as, for instance, with a single ET detector) is to use the temporal coincidence of the GW event with a short GRB; for many short GRBs, the redshift has indeed been determined from the X-ray afterglow, that can be accurately localized by *Chandra* or *Swift*/XRT. For instance, in the sample of 67 *Swift* short GRBs discussed in [67], 53 events were rapidly followed up with the on-board X-ray Telescope, leading to 47 detections of the source.

The estimate of the number of coincidences between GW events and electromagnetic observations depends crucially, of course, on the rate of expected GW events, as well as on the network of GRB satellites and telescopes available at the time. We therefore discuss the 2G and 3G cases separately.

### 2.2.1 GRB coincidences with the HLVKI network

We begin by investigating the possibility of detecting in coincidence a GW signal at the HLVKI network and a GRB with the current generation of GRB satellites. We assume here that the Fermi-GBM can make a coincident detection and that *Swift* can slew to the combined GW/GRB error box and identify an X-ray counterpart. We note here that for 170817A *Swift* was occulted by Earth at time of Fermi trigger, so imaging by the X-ray telescope (XRT) took place around 1 hr post trigger. At that time it was able to cover 90% of the GW skymap to rule out any bright sources [68].

For a GRB detected in coincidence with a GW signal we require that the peak flux is above the flux limit of the satellite. Based on the modeling of [69] we assume a Gaussian structured jet profile model of GRB170817A given by

$$L(\theta_V) = L_c \exp\left(-\frac{\theta_V^2}{2\theta_c^2}\right), \quad (2.10)$$

with  $L(\theta)$  the luminosity per unit solid angle,  $\theta_V$  the viewing angle and  $L_c$  and  $\theta_c$  structure parameters that define the angular profile. The structured jet parameter is given by  $\theta_c = 4.7^\circ$ .

The value of  $L_c$  is given by  $L_c = L_p/4\pi \text{ erg s}^{-1} \text{ sr}^{-1}$ , where  $L_p$  is the peak luminosity of each burst, which is obtained by sampling  $\Phi(L_p)dL_p$ . We assume the standard broken power-law distribution of the form

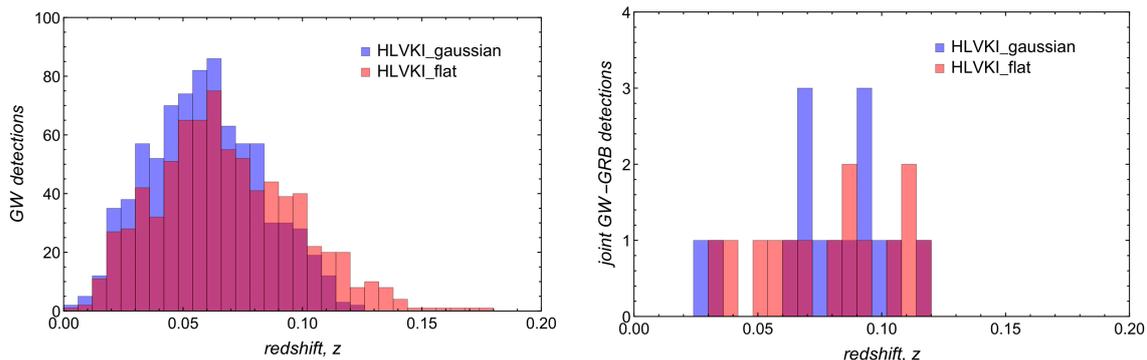
$$\Phi(L_p) \propto \begin{cases} (L_p/L_*)^\alpha, & L_p < L_* \\ (L_p/L_*)^\beta, & L_p \geq L_* \end{cases} \quad (2.11)$$

where  $L_p$  is the peak luminosity assuming isotropic emission in the rest frame in the 1–10000 keV energy range,  $L_*$  is a characteristic value separating the two regimes, and the slopes describing these regimes are given by  $\alpha$  and  $\beta$  respectively. Following [70] we use the values  $\alpha = -1.95$ ,  $\beta = -3$  and  $L_* = 2 \times 10^{52} \text{ erg sec}^{-1}$ . Given a source at luminosity distance  $d_L$  one can convert  $4\pi L(\theta_V)$  to an observed peak flux as a function of viewing angle,  $F_P(\theta_V)$ , obtained from the value of the GW inclination angle. A Fermi-GBM detection is recorded if the value of  $F_P(\theta_V)$  is greater than the flux limit of  $1.1 \text{ ph sec}^{-1} \text{ cm}^{-2}$  in the 50–300 keV band for Fermi-GBM [69], noting that 95% of the bursts detected in the 64 ms timescale are within this limit.<sup>9</sup> We further assume the total time-averaged observable sky fraction of the Fermi-GBM, which is 0.60 [71]. Using this procedure, among the events in the GW BNS catalog generated as discussed in section 2.1, we select those that have an observed GRB counterpart.

Table 1 shows the number of BNS sources along with the number of coincident GRB detections determined using the procedure above, for the HLVKI network. We see that 10 years of observation would yield of order 14–15 joint detections. In figure 2 we show the redshift distribution of the GW events (left panel) and of the GW-GRB coincidences (right panel), for a realization of our catalog. Observe that, while the peak in the left panel is at  $z \sim 0.05$ , in the right panel we have a flatter distribution with a peak that (within the small number statistics of the GW-GRB joint detections) is nearly at  $z \sim 0.10$ . This can be understood by observing that the GW detections peak at around  $z \sim 0.05$ , after which the detected more distant events tend to have inclination angles closer to face on; this effect is discussed in section 2.3.1 (see also [41]). The  $z$ -distribution of the joint detections is mainly dependent on the flux limited GRB instrument (in this case Fermi). Under a structured jet scenario, only emissions near to the jet axis will be detected; however, the fraction of such events will be small at low- $z$  but higher at larger cosmological volumes. Although the joint detection distribution has small numbers, the slight propensity of events at larger  $z$  is due to this effect (see [67] for a more detailed discussion of this effect). In appendix A we give, in table 23, the explicit values of  $z$ ,  $d_L$  and  $\Delta d_L$  for the events in the specific catalog corresponding to the right panel of figure 2, where  $d_L$  is the ‘measured’ luminosity distance reconstructed from  $z$  using our fiducial  $\Lambda$ CDM model and scattering randomly those fiducial values of  $d_L(z)$  according to a Gaussian distribution with a width equal to the error  $\Delta d_L(z)$  (as explained in section 2.1).

To test the impact of changing our astrophysical assumptions, we have also generated a catalog of GW events assuming a Madau-Dickinson star formation rate and an exponential time delay between formation and merger with an e-fold time of 100 Myr, as in ref. [36]. In this case, assuming again a duty cycle of 80% and a network SNR threshold level  $\rho_{\text{threshold}} = 12$ , we find that the number of BNS detected at the HLVKI network, for the Gaussian mass distribution, is 64/yr, to be compared with the value 48/yr reported in ref. [36].

<sup>9</sup>See the Fermi-GBM burst catalogue <https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigbrst.html>



**Figure 2.** Left panel: the redshift distribution of 10 yr of GW events from a realization of the mock catalog at the HLVKI network, for the flat and the Gaussian mass distribution. Right panel: the redshift distribution of 10-years of GW-GRB coincidences between the HLVKI network and the current generation of GRB satellites.

| Network | GW events |          | Joint GW-GRB events |          |
|---------|-----------|----------|---------------------|----------|
|         | Flat      | Gaussian | Flat                | Gaussian |
| HLVKI   | 768       | 814      | 14                  | 15       |

**Table 1.** Number of GW events detected by second generation (2G) networks in 10 years, and the expected GW-GRB coincidences obtained by assuming a GRB detector with the characteristics of Fermi-GBM. We show detection rates for BNS populations generated using O2 rates corresponding to both flat and Gaussian mass distributions.

## 2.2.2 GRB coincidences with 3G detectors

For 3G detectors the estimates are of course more uncertain. Indeed, the identification of the counterpart depends on the network of GRB satellites and of telescopes at the time when 3G detectors will operate, as well as on issues that are presently difficult to predict, such as the prioritization that will be given by various telescopes to the follow-up of GW signals.

The proposed THESEUS mission [46–48] could be particularly useful for performing coincidences between GW events and GRBs, even in the absence of localization from the GW signal. A crucial difference with the 2G case is that, at the sensitivity level of 3G detectors, there will be many more GW events compared to what GRB satellites could detect. The main reason for this is that the GRB instruments are limited by their flux sensitivity for the more distant GRB emissions at wider viewing angles. For instance, it was estimated in [22] that ET will be able to detect  $\mathcal{O}(10^5 - 10^6)$  BNS mergers per year. As we will see below our estimate, given in table 2, is that ET will detect about  $(6 - 7) \times 10^5$  events in 8 yr of actual data taking, corresponding to a rate of order  $(0.8 - 0.9) \times 10^5/\text{yr}$ , consistent with previous estimates, although somewhat smaller. In any case, according to the estimate in [47], only about 15 – 35 coincident short GRB (sGRB) per year will be detected by THESEUS with its X-Gamma ray Imaging Spectrometer (XGIS); we will see below that our results, using somewhat different assumptions for the GRB luminosity and BNS rate, gives a slightly higher number of coincidences, but still of this order of magnitude. Beside the collimated prompt GRB emission, more isotropic soft X emission is also expected from the afterglow. This could be detected by the Soft X-ray Imager (SXI) on board THESEUS, leading possibly to a few hundred more detections per year [47]. In any case, the number of joint GW-GRB detections will be a very small fraction of the number of GW events.

| Network  | GW events |           | Joint GW-GRB events |           |
|----------|-----------|-----------|---------------------|-----------|
|          | Flat      | Gaussian  | Flat                | Gaussian  |
| ET       | 621,700   | 688,426   | 389 (128)           | 511 (169) |
| ET+CE+CE | 5,420,656 | 7,077,131 | 644 (213)           | 907 (299) |

**Table 2.** Number of GW BNS events detected by third generation (3G) networks in 10 years of data taking (assuming a 80% duty cycle for each detector) and the corresponding GW-GRB coincidences obtained by assuming a GRB detector with the characteristics of THESEUS-XGIS; numbers in parenthesis show the number of sources with arcmin localisation. BNS populations are generated using the O2 rates corresponding to ‘flat’ and ‘Gaussian’ mass distributions.

With a network of at least three GW detectors, accurate localization of the GW signal becomes possible, allowing for electromagnetic follow-up observations, that could determine the redshift of the source. If the source localization is already available through GWs, LSST could detect the counterpart up to  $z \simeq 0.55$  and WFIRST up to  $z \simeq 0.76$ , and many more telescopes in the UV, optical, IR, radio could detect the counterpart at smaller redshifts, say  $z \sim 0.1 - 0.3$ . However, the follow-up of  $O(10^3)$  well localized GW events at  $z \sim 0.5$  would require the equivalent of 1 yr of dedicated LSST time, which is not realistic. Currently, a more realistic estimate is that LSST might use of order of 1% of its time for GW follow up, so it will be challenging for LSST to observe more than  $O(10)$  counterparts per year, at  $z \sim 0.5$ . The localization cost is much smaller at  $z \sim 0.1$ , where  $O(100)$  events per year could be a more realistic expectation, but this will depend on the science prioritization in the 2030s, when 3G detectors will hopefully operate.<sup>10</sup> Given these large uncertainties, in this paper we will limit ourselves to the coincidences with GRB detectors.

We repeat our simulations for the coincidence with a single ET detector and with a ET+CE+CE network, assuming that a THESEUS type satellite will be used for coincidence searches. For the GRB detection we assume a duty cycle of 80% due to a reduction of 20% as the satellite passes through the Southern Atlantic Anomaly, a flux limit of  $0.2 \text{ ph sec}^{-1} \text{ cm}^{-2}$  in the 50–300 keV band and a sky coverage fraction of 0.5 [48]. We note that the XGIS will be able to localise sources to around 5 arcmin only within the central 2 sr of its field of view (FOV); outside this central region localisation will be coarse at best.<sup>11</sup> We therefore consider two scenarios: one, that we will denote as ‘optimistic’, in which all the events detected by XGIS have a measured redshift, and one, that we will denote as ‘realistic’, where we assume that only around 1/3 of the sGRBs detected by XGIS could provide redshift estimates.

### 2.3 Events rates, redshift distributions and analytic estimates for $\Delta d_L(z)/d_L(z)$

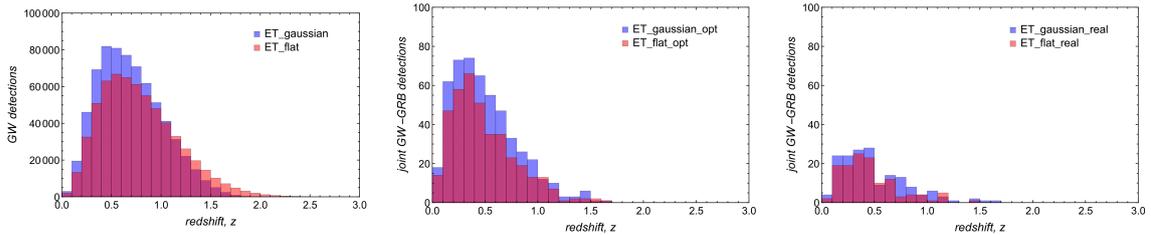
In our MCMCs we will use a given realization of the catalog of events obtained with the procedure discussed above. It is however useful to describe the qualitative features of these catalogs, such as the redshift distributions of the events and the average value of  $\Delta d_L(z)/d_L(z)$  as a function of redshift. This will provide a physical insight into which sources contribute most, to compare with previous works, and to provide ready-to-use formulas that can be applied to future studies.

#### 2.3.1 Events at a single ET detector

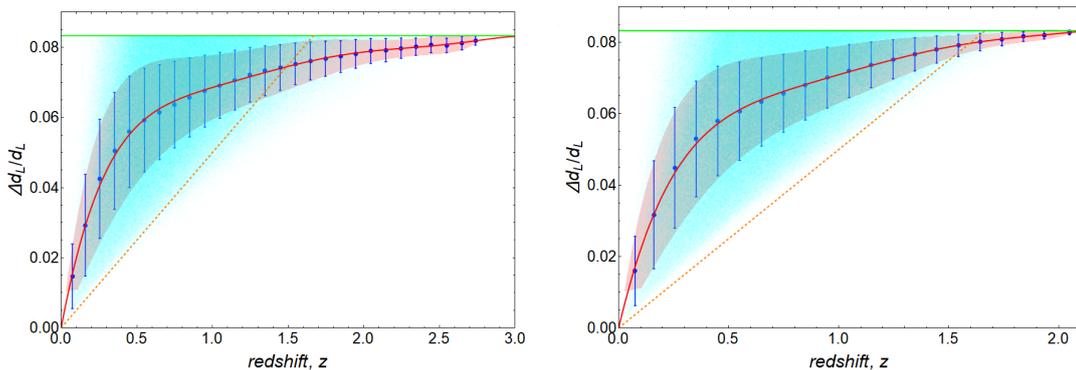
Table 2 shows the results of our simulations for the 3G era in terms of the number of GW signals from BNS, along with the number of joint GW/sGRB detections; the number of events with arcmin localisation are shown in parenthesis.

<sup>10</sup>We thank Matthew Bailes, Samaya Nissanke and Bangalore Sathyaprakash for discussions on these issues.

<sup>11</sup>We thank Giulia Stratta for valuable discussions on the localisation capabilities of THESEUS-XGIS.



**Figure 3.** Left panel: the redshift distributions of 10-years of BNS detections by a ET detector. Middle panel: the coincident detections made by THESEUS in the ‘optimistic’ scenario for the FOV. Right panel: the coincident detections in the ‘realistic’ scenario. Notice that the vertical scale for the left panel is very different from that in the middle and right panels.



**Figure 4.** The ET instrumental contribution to the relative error  $\Delta d_L/d_L$  for the specific realization of the catalog of BNS detections shown in figure 3. All the GW events, with or without a detected EM counterpart, are taken into account. Left panel: for the flat distribution of neutron star masses. Right panel: for the Gaussian distribution of neutron star masses. In each panel, the cyan shaded area corresponds to all the BNS events, while the coordinates of the blue points are given by the mean values of the redshift and of  $\Delta d_L/d_L$  in each redshift bin, with the bins chosen as in figure 3. The blue error bars are the standard deviations of  $\Delta d_L/d_L$  in each redshift bin. The horizontal green line at  $\Delta d_L/d_L = 1/12$  is fixed by the SNR threshold of 12. The red line and the light red region are the fits to the blue points and error bars, given explicitly in the text. The orange dotted line is the error on  $\Delta d_L/d_L$  induced by lensing. The redshift ranges shown in the two panels differ as a result of the different maximum values of redshift reached in the two corresponding catalogs.

For a single ET detector our estimate of the rate of BNS detection is between  $6.2 \times 10^5$  and  $6.9 \times 10^5$  events in 10 yr, having assumed a duty cycle of 80%, which, in the case of a single detector, corresponds to 8 yr of actual data. This corresponds to a rate (normalized to the actual time of data taking)

$$R \simeq (0.8 - 0.9) \times 10^5 \text{ BNS/yr}, \quad (2.12)$$

consistent with previous estimates, although somewhat smaller. This can be traced to the fact that we have used a threshold for the network SNR, obtained by combining the three arms of ET, given by  $\rho_{\text{threshold}} = 12$ , while previous work, e.g. ref. [23], used  $\rho_{\text{threshold}} = 8$ . We also see from table 2 that, with a single ET detector, we should expect around 39 – 51 coincident sGRB/GW events in one year of observation using the XGIS and the SXI detectors. These numbers differ from the 15 – 35 events quoted in [47] for two main reasons. Firstly, the assumed luminosity function and BNS rate differs from that assumed in [47]. Secondly,

our calculations assume a structured jet profile based on GRB170817A. From our sample of detections one could expect around  $13 - 17$  events  $\text{yr}^{-1}$  to have arcmin localisations.

Figure 3 shows the redshift distributions of 10-years of BNS detections using ET, along with the joint GW-GRB distributions with optimistic and realistic scenarios for the FOV of THESEUS. We see that the vast majority of the joint GW-GRB detections is at  $z \lesssim 1$ . In this realization of the catalog, the ET event with the highest redshift is at  $z \simeq 2.91$  for the flat mass distribution and at  $z \simeq 2.10$  for the Gaussian mass distribution, while the joint detection with the highest redshift is at  $z \simeq 1.63$  (found both in optimistic and realistic scenario for the THESEUS FOV, and for the Gaussian mass distribution). We find that the higher- $z$  sGRB detections have viewing angles close to the jet axis corresponding with almost face-on BNSs; this selection effect was highlighted in [41]. Note, however, that with respect to the 2G case shown in figure 2, the joint detection rate in figure 3 more closely tracks the GW distribution; this is mainly due to the improved sensitivity of THESEUS allowing detection at wider viewing angles.

Figure 4 shows the ET instrumental contribution to the relative error on luminosity distance, considering all the events in the specific realization of the catalog of BNS detections presented in the left panel of figure 3. The events are organized in the same redshift bins as figure 3 and, for the events belonging to each bin, the mean value of  $\Delta d_L/d_L$  and its standard deviation  $\sigma(\Delta d_L/d_L)$  are evaluated. Of course, the mean value of  $\Delta d_L/d_L$  increases with distance, until we reach the threshold at  $\text{SNR}=12$ , and therefore the value  $\Delta d_L/d_L = 1/12$ , beyond which we no longer record the triggers as detections. In contrast, its variance eventually decreases; this is due to the fact that, in a given bin at some intermediate redshift, we have events with different possible orientations with respect to the detector, and therefore different SNR. As the redshift increases toward the horizon of the detector, only the events with optimal orientation can go above the threshold.

It is useful to provide a fit for the mean value and for the standard deviation of  $\Delta d_L/d_L$ , as a function of the redshift. In order to increase the significance of the fit, in the case of flat mass distribution the two highest redshift bins containing events of the catalog shown in figure 3 have been excluded (they only contain a very low number of events and we do not show them in the figure). Using sixth degree polynomials intercepting zero, for the flat distribution (f) of neutron star masses, the fit to the mean value of  $\Delta d_L/d_L$  and to its standard deviation are given by

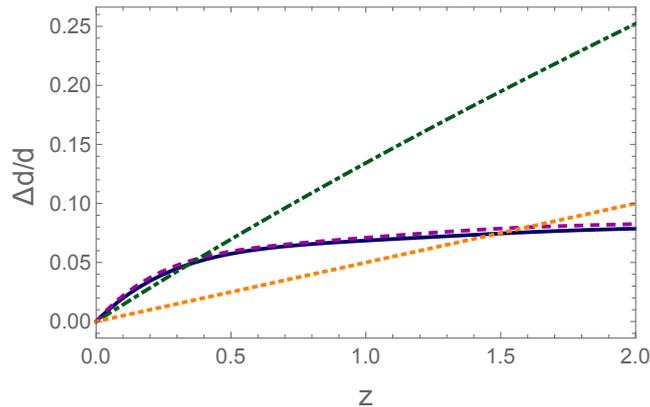
$$\overline{\left(\frac{\Delta d_L}{d_L}\right)}_{\text{f}} = 0.230896z - 0.345265z^2 + 0.282128z^3 - 0.124715z^4 + 0.028027z^5 - 0.002508z^6, \quad (2.13)$$

$$\sigma\left[\left(\frac{\Delta d_L}{d_L}\right)_{\text{f}}\right] = 0.129548z - 0.334210z^2 + 0.369258z^3 - 0.205939z^4 + 0.056642z^5 - 0.006100z^6. \quad (2.14)$$

In the case of Gaussian distribution (G) the corresponding fits are given by

$$\overline{\left(\frac{\Delta d_L}{d_L}\right)}_{\text{G}} = 0.261577z - 0.456248z^2 + 0.442636z^3 - 0.232171z^4 + 0.061693z^5 - 0.006502z^6, \quad (2.15)$$

$$\sigma\left[\left(\frac{\Delta d_L}{d_L}\right)_{\text{G}}\right] = 0.154942z - 0.490698z^2 + 0.683342z^3 - 0.490253z^4 + 0.175418z^5 - 0.024721z^6. \quad (2.16)$$



**Figure 5.** Our fits to  $\Delta d_L/d_L$  given in eq. (2.13) (blue solid line) and in eq. (2.15) (magenta dashed line) compared to the error due to lensing given in eq. (2.9) (orange dotted line) and to the fit (2.17) to the instrumental error given in ref. [23] (green dot-dashed line).

Tables 24, 25, 26 and 27, in appendix A, list the ET instrumental contribution to the relative error  $\Delta d_L/d_L$  in each redshift bin, for the realizations of the catalogs of joint GW-GRB detections displayed in the middle and right panels of figure 3.

It is instructive to compare our fits to the instrumental error on  $\Delta d_L/d_L$ , with the fit to the instrumental error found in [23] (see also figure 2 of [72]), which is

$$\overline{\left(\frac{\Delta d_L}{d_L}\right)} = 0.1449z - 0.0118z^2 + 0.0012z^3, \quad (2.17)$$

as well as with the lensing error (2.9). The comparison is shown in figure 5. We see that our estimate of the average instrumental error is consistent with that of ref. [23] at low redshifts, say  $z \lesssim 0.3$ , but becomes smaller at large redshift. The sensitivity curve for ET used in ref. [23] [see their eq. (19)] is relatively close to the one that we use, shown in figure 1. The main reason for the difference is rather that we use a SNR threshold  $\rho_{\text{threshold}} = 12$ , while ref. [23] uses  $\rho_{\text{threshold}} = 8$ . On the one hand, this implies that we get less events, and up to smaller redshift, compared to ref. [23]. On the other hand, since  $\Delta d_L/d_L$  is estimated as  $1/\text{SNR}$ , the events that we retain are those with smaller value of  $\Delta d_L/d_L$  and therefore also the average value of  $\Delta d_L/d_L$  over the events in a redshift bin is smaller. This effect becomes more important as  $z$  increases, as there are many more events that are sub-threshold with respect to  $\rho_{\text{threshold}} = 12$ , but above threshold with respect to  $\rho_{\text{threshold}} = 8$ . It should also be observed that in ref. [23]  $\Delta d_L/d_L$  was estimated using a Fisher matrix approach, that takes into account the degeneracy of  $d_L$  with the other parameters, and in particular with the inclination angle of the orbit, while here we have used the simpler estimate  $\Delta d_L/d_L \simeq 1/\text{SNR}$ . As we discussed above eq. (2.9) following ref. [20], our approximation can result in an underestimate of the actual value of  $\Delta d_L/d_L$  by a factor  $\sim 2$  (with a significant scatter from event to event), which however substantially decreases for GW signals detected in coincidence with a GRB, because the latter are preferentially face-on, especially at large  $z$ .

It is also interesting to compare the lensing contribution to  $\Delta d_L/d_L$  to the observational error, in the light of this understanding of the dependence of the average observational error on the threshold. Comparing the fit obtained in ref. [23] with  $\rho_{\text{threshold}} = 8$ , given by the green dot-dashed line in figure 5, to the lensing error, given by the orange dotted line, one would be tempted to conclude that, at ET, the error induced by lensing is negligible

at all redshifts. In fact, this statement needs some qualification. The green dot-dashed line represents the average of  $\Delta d_L/d_L$  over an ensemble of events, in the same redshift bin, selected by requiring  $\text{SNR} \geq \rho_{\text{threshold}} = 8$ . However, within each bin, the events that are most useful for cosmological studies are those with the smallest errors on  $\Delta d_L/d_L$ , i.e. with the highest SNR. Already selecting only the events with  $\text{SNR} \geq \rho_{\text{threshold}} = 12$  reduces significantly the instrumental error averaged over such events, and we see that for this ensemble of events lensing becomes larger than the mean value of  $\Delta d_L/d_L$  at  $z \gtrsim 1.5$ .

We have seen above that the vast majority of the GW-GRB coincidences are at  $z < 1$ . Therefore, in our analysis, lensing will indeed be subdominant. However, it must be kept in mind that, if one restricts the analysis to the ‘golden events’, i.e. the loudest and best characterized events in each redshift bin (corresponding to the lower edge of the cyan shaded area in figure 4), the effect of lensing will become more and more important. The contribution to the error on  $\Delta d_L/d_L$  from lensing, estimated as in eq. (2.9), is shown as the orange dotted line in figure 4. We see that, for  $z < 1$ , it is comparable to the lower edge of the distribution of events given by the cyan shaded area, and therefore its inclusion, in quadrature with the observational error, at most degrades by a factor  $\sqrt{2}$  the error on  $\Delta d_L/d_L$  on these very few events at the edge of the distribution. These events are interesting because they are those with the smallest error, but for  $z < 1$  they are extremely rare. For larger redshifts the situation is different and we see that, say at  $z \simeq 1.5$ , for a non-negligible fraction of the events the error from lensing can be the limiting factor.

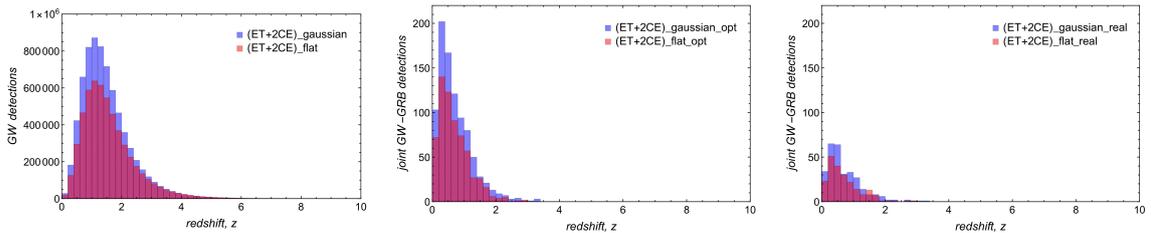
### 2.3.2 Events at a ET+CE+CE network

In the second line of table 2 we show the results for the ET+CE+CE network. This configuration, featuring three 3G detectors, can be considered an extreme case. Other cases, such as one ET detector and two advanced 2G detectors, will be intermediate between this case and a single 3G detector.<sup>12</sup> The result for the GW events in table 2 in the Gaussian case corresponds to a detection rate of 710k events/yr. To understand the dependence of these results on the astrophysical assumptions, we have also generated a catalog of GW events, for ET+CE+CE, assuming a Madau-Dickinson star formation rate and an exponential time delay between formation and merger with an e-fold time of 100 Myr. In that case we find 840k events/yr, in good agreement with the value 990k events/yr found, under similar hypothesis, but a slightly larger rate  $R_m(z=0) = 1000 \text{ Gpc}^{-3} \text{ yr}^{-1}$ , in ref. [36].

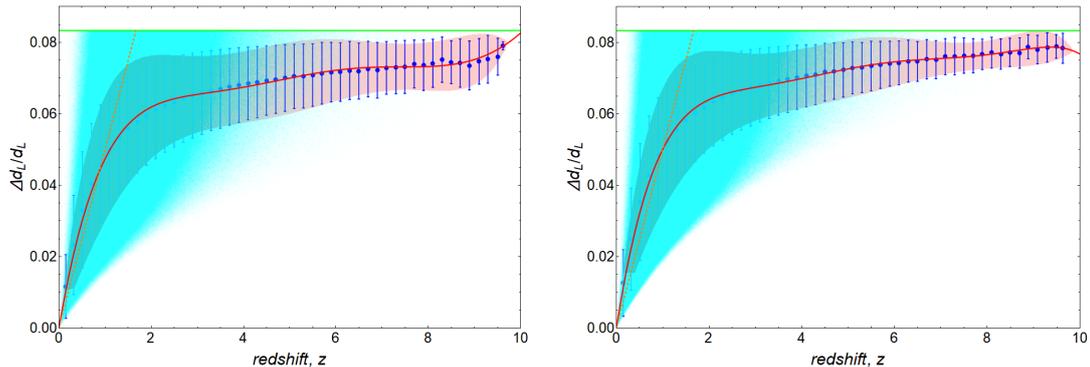
For the coincidences with GRBs we find that, for an ET+CE+CE network, we would get of order 64 – 90 coincident events per year, and around 20 – 30 events per year will have arcmin localisations. Observe that, even if the ET+CE+CE network has a number of GW detections larger than a single ET by a factor  $\mathcal{O}(10)$ , the number of coincidences with GRBs is higher only by a factor less than 2. This already tells us that the bottleneck, for joint GW-GRB detections, is on the GRB side, that cannot keep pace with the GW detection rate of 3G detectors.

Figure 6 shows the redshift distributions of BNS detections using ET+CE+CE, over 10 yr, along with the joint GW-GRB distributions with optimistic and realistic scenarios for the FOV of THESEUS. In this realization of the catalog, the ET+CE+CE event with the highest redshift is at  $z \simeq 9.63$  for the flat mass distribution and at  $z \simeq 9.66$  for the Gaussian mass distribution; this is due to the fact that a single CE already has a reach to BNS of order

<sup>12</sup>Currently, are under study a series of improvements of the advanced LIGO detectors, that should eventually lead to the Voyager detectors. The Voyager phase in the US could happen partly in temporal overlap with ET in Europe.



**Figure 6.** As in figure 3 for the ET+CE+CE network and its coincidences with THESEUS. For uniformity, we use the same redshift range in the three panels.



**Figure 7.** As in figure 4, for the ET+CE+CE network.

$z \simeq 9$  (while we have seen above that, for BNS, ET alone reaches  $z \simeq 2 - 3$ ). For the joint GW-GRB detections we find that the one with the highest redshift is at  $z \simeq 3.38$  (found both in optimistic and realistic scenario for the THESEUS FOV, and for the Gaussian mass distribution). Notice that, in the left panel, events with  $z \gtrsim 6$  are not visible on the vertical scale used in the figure, but are indeed present.

Figure 7 shows the instrumental error for the luminosity distance in our catalog of events for ET+CE+CE, with the same meaning of the lines and shaded areas as in figure 4. The distribution now extends to much higher redshifts, because, as we have seen, CE (assuming the current design configuration, with 40 km arms) has a much larger horizon to BNS. For the flat distribution of neutron star masses the fit for the mean value of the instrumental error and to its standard deviation is given by

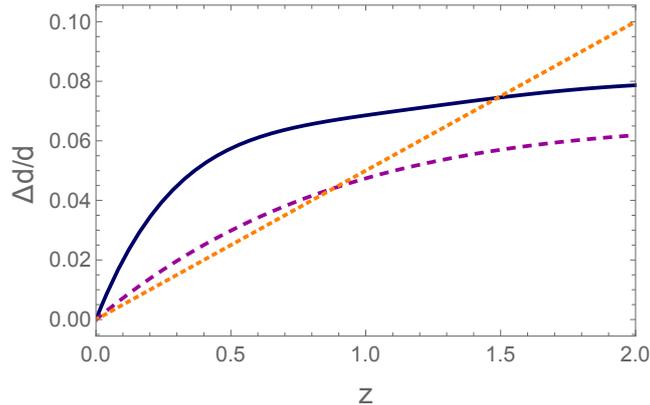
$$\overline{\left(\frac{\Delta d_L}{d_L}\right)}_f = 0.0756734z - 0.0360303z^2 + 0.00882417z^3 - 0.00113238z^4 + 7.18676 \times 10^{-5}z^5 - 1.75818 \times 10^{-6}z^6, \quad (2.18)$$

$$\sigma\left[\left(\frac{\Delta d_L}{d_L}\right)_f\right] = 0.0411611z - 0.0334084z^2 + 0.0115623z^3 - 0.00198248z^4 + 1.65371 \times 10^{-4}z^5 - 5.35185 \times 10^{-6}z^6, \quad (2.19)$$

while for the Gaussian distribution

$$\overline{\left(\frac{\Delta d_L}{d_L}\right)}_G = 0.0828382z - 0.0427639z^2 + 0.0115373z^3 - 0.00166096z^4 + 1.21048 \times 10^{-4}z^5 - 3.50769 \times 10^{-6}z^6, \quad (2.20)$$

$$\sigma\left[\left(\frac{\Delta d_L}{d_L}\right)_G\right] = 0.0400958z - 0.0323344z^2 + 0.0109567z^3 - 0.00183302z^4 + 1.48911 \times 10^{-4}z^5 - 4.687178 \times 10^{-6}z^6, \quad (2.21)$$



**Figure 8.** The fit to the average value of the instrumental error on  $\Delta d_L/d_L$  for ET (blue solid line) and for ET+CE+CE (magenta dashed line) (both, for definiteness, for the flat mass distribution), compared to the error due to lensing (orange dotted line).

However, we also see from figure 7 that the contribution to the error from lensing, given by the dotted orange line, becomes quickly dominant at  $z \gtrsim 1.5$  [assuming that the linear extrapolation (2.9) is still correct at large redshifts]. For the study of joint GW-GRB detections that we have performed this has a limited impact, as for ET+CE+CE the bulk of the joint GW-GRB detections is at  $z \lesssim (1 - 1.5)$ . Given the localization capability of the ET+CE+CE network, one could still hope to extract cosmological information from the very large number of standard sirens at larger redshifts, through statistical methods. However, at these redshifts the dominant contribution to  $\Delta d_L/d_L$  will come from lensing (similarly to what happens for LISA). This is clearly seen from figure 8, where we plot the instrumental error  $\Delta d_L/d_L$  at ET and at ET+CE+CE, in the range  $z < 2$  where they can be compared, and the error due to lensing. We see that, even if the instrumental error from the ET+CE+CE network is obviously better than for a single ET, above  $z \simeq 1$  the error in ET+CE+CE starts to be dominated by lensing, so that at  $z \gtrsim 1.5$  both ET and ET+CE+CE are dominated by lensing and therefore eventually this becomes the limiting factor in both configurations.

This limitation could however be turned into a virtue. The situation is indeed similar to the one that was discussed a decade ago in ref. [21], in the context of a study for the Big Bang Observer (BBO). It was found that also in BBO the error is almost entirely dominated by lensing. This means that, once one has determined the dependence of  $d_L$  on  $z$  from the full ensemble of sources (and possibly by combining standard sirens with CMB, BAO and SNe), the scatter around this mean value for each single BNS event is basically a measurement of the gravitational lensing magnification along that line of sight. Given the very large number of sources, this will produce a map of the lensing magnification across the sky; in particular, the corresponding two-point correlation function gives a measurement of the convergence power spectrum, and provides important information on cosmological structure formation.

### 3 Constraints from standard sirens on $H_0$ and $\Omega_M$ in $\Lambda$ CDM

In this section we study how the addition of the standard sirens with GRB counterpart that could be observed with the HLVKI network or with 3G detectors, would contribute to the knowledge of cosmological parameters in  $\Lambda$ CDM. In  $\Lambda$ CDM the luminosity distance to

standard sirens is given by

$$d_L(z) = \frac{1+z}{H_0} \int_0^z \frac{d\tilde{z}}{\sqrt{\Omega_M(1+\tilde{z})^3 + \Omega_\Lambda}}, \quad (3.1)$$

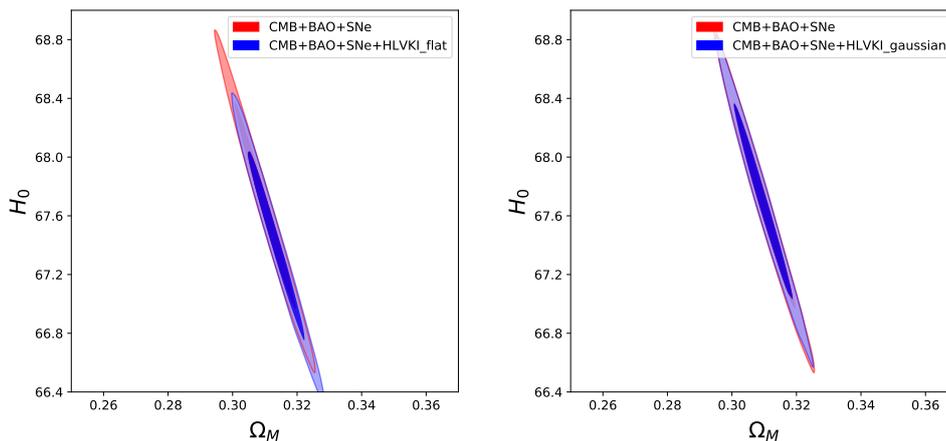
where, as usual,  $\Omega_M$  and  $\Omega_\Lambda$  are the present fraction of matter energy density and of vacuum energy density and, assuming flatness,  $\Omega_\Lambda \simeq 1 - \Omega_M$  (we do not write explicitly the contribution of radiation, which is completely negligible at the redshifts relevant for standard sirens). The measurement of the luminosity distances from a set of coalescing binaries therefore gives constraints on  $H_0$  and  $\Omega_M$ .

The most accurate results are obtained by combining the constraints from standard sirens with other cosmological datasets such as CMB, BAO and SNe, to remove the degeneracies between cosmological parameters. A priori it would be interesting to study also the constraints that emerge using only standard sirens, that, even if less constraining, are conceptually interesting because they have systematics completely different from those of electromagnetic observations. However, for 2G detectors with a counterpart identified through a GRB, we find that the number of sources is too small to obtain significant results from standard sirens alone, and we will present only the results obtained by combining standard sirens with CMB+BAO+SNe. For 3G detectors, we will also show the separate results from standard sirens. When combining standard sirens with CMB, BAO and SNe, we use the following datasets:

- *CMB*. We use the 2015 *Planck* [73] measurements of the angular (cross-)power spectra, including full-mission lowTEB data for low multipoles ( $\ell \leq 29$ ) and the high- $\ell$  Plik TT,TE,EE (cross-half-mission) ones for the high multipoles ( $\ell > 29$ ) of the temperature and polarization auto- and cross- power spectra [74]. We also include the temperature+polarization (T+P) lensing data, using only the conservative multipole range  $\ell = 40 - 400$  [75, 76].
- *Type Ia supernovae*. We use the JLA data for SN Ia provided by the SDSS-II/SNLS3 Joint Light-curve Analysis [77].
- *Baryon Acoustic Oscillations* (BAO). We use the isotropic constraints provided by 6dFGS at  $z_{\text{eff}} = 0.106$  [78], SDSS-MGS DR7 at  $z_{\text{eff}} = 0.15$  [79] and BOSS LOWZ at  $z_{\text{eff}} = 0.32$  [80], as well as the anisotropic constraints from CMASS at  $z_{\text{eff}} = 0.57$  [80].

We then run a MCMC, using the CLASS Boltzmann code [81]. For the baseline  $\Lambda$ CDM model we use the standard set of six independent cosmological parameters: the Hubble parameter today  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ , the physical baryon and cold dark matter density fractions today  $\omega_b = \Omega_b h^2$  and  $\omega_c = \Omega_c h^2$ , respectively, the amplitude  $A_s$  and tilt  $n_s$  of the primordial scalar perturbations, and the reionization optical depth  $\tau_{\text{re}}$ . We keep the sum of neutrino masses fixed, at the value  $\sum_\nu m_\nu = 0.06 \text{ eV}$ , as in the *Planck* baseline analysis [82].

Since standard sirens, within  $\Lambda$ CDM, are only sensitive to  $H_0$  and  $\Omega_M$ , we focus on the two-dimensional likelihoods in the  $(\Omega_M, H_0)$  plane (although, of course, when we combine standard sirens with CMB+BAO+SNe, the fact that the addition of standard sirens allows a more accurate determination of  $H_0$  and  $\Omega_M$  also has a beneficial effects on the determination of the other parameters, because it helps to reduce the degeneracies).



**Figure 9.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(\Omega_M, H_0)$  plane, in  $\Lambda$ CDM, from CMB+BAO+SNe (red), and the result obtained by combining standard sirens at the HLVKI network with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution.

### 3.1 Results for the HLVKI network

The result is shown in figure 9, where we compare the likelihood in the  $(\Omega_M, H_0)$  plane obtained from a MCMC using the above CMB+BAO+SNe dataset (red contours) with those obtained from the combined datasets CMB+BAO+SNe+standard sirens (blue contours), for the two distributions of neutron star masses. The contours from standard sirens only are not shown since the MCMC fails to converge due to the small number of sources.

In table 3 we show the relative errors  $\Delta H_0/H_0$  and  $\Delta\Omega_M/\Omega_M$  obtained from the corresponding one-dimensional marginalized likelihood. We see that, for the HLVKI network, the addition of joint GW-GRB detections to the current cosmological dataset does not improve substantially the accuracy on  $H_0$  and  $\Omega_M$ . This should be traced to the fact that, as we see from table 1, if we perform coincidences of the GW events with GRBs, we only have of order 15 joint detections, even over a 10-yr period. On the other hand, at the typical redshifts of the events seen by 2G detectors, the network of optical and infrared telescopes is expected to provide many more electromagnetic counterparts. In the present paper, even for 2G detectors we have restricted ourselves to a study of coincidences with GRBs only, also as a benchmark for the study of GW-GRB coincidences with 3G detectors. However, it is clear that for 2G detectors the contribution of optical/IR telescopes will be crucial and could raise substantially the number of standard sirens with observed counterpart.

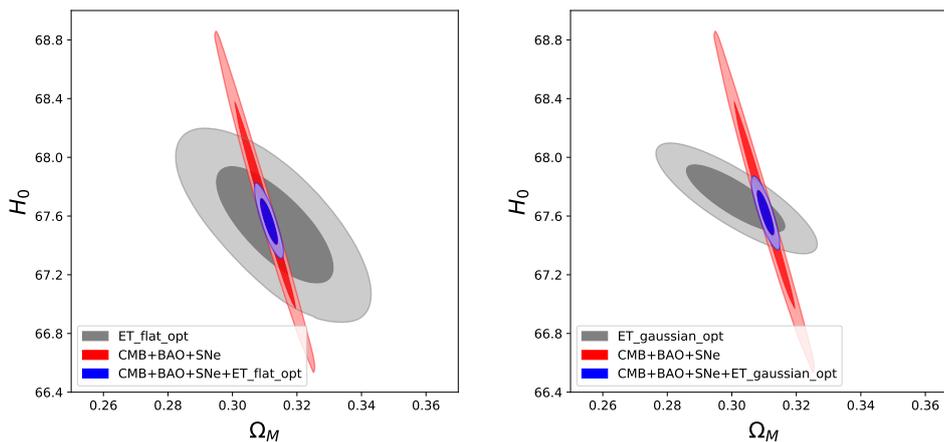
Our results are broadly consistent with the analysis of [66, 83], which show that, to obtain a measurement of  $H_0$  below 1% at 2G detectors with standard sirens only,  $\mathcal{O}(50-100)$  standard sirens with counterpart are needed. This would allow to address the discrepancy between the local measurement of  $H_0$  and the value inferred by *Planck* and BAO observations assuming  $\Lambda$ CDM, which has now reached the  $4.4\sigma$  level [84].

### 3.2 Results for ET

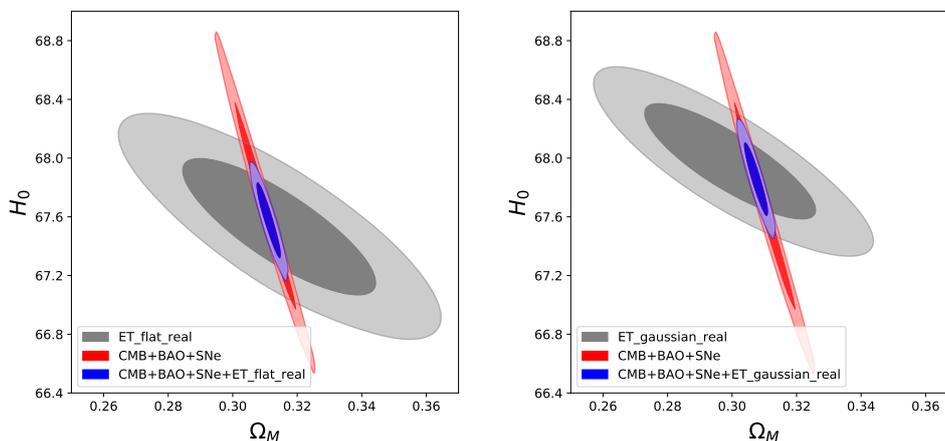
We next consider the case of joint detections between a single ET detector and a GRB detected by THESEUS, again restricting at first to  $\Lambda$ CDM. The result is shown in figures 10 and 11, where we compare the likelihood in the  $(\Omega_M, H_0)$  plane obtained from a MCMC using our CMB+BAO+SNe dataset (red contours) with those obtained from standard sirens

|                            | CMB+BAO+SNe | combined, flat | combined, gaussian |
|----------------------------|-------------|----------------|--------------------|
| $\Delta H_0/H_0$           | 0.72%       | 0.65%          | 0.66%              |
| $\Delta \Omega_M/\Omega_M$ | 2.11%       | 1.91%          | 1.96%              |

**Table 3.** Accuracy ( $1\sigma$  level) in the reconstruction of  $H_0$  and  $\Omega_M$  with CMB+BAO+SNe only, and the combined result CMB+BAO+SNe+standard sirens, using the HLVKI detector network and the flat and gaussian mass distributions.



**Figure 10.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(\Omega_M, H_0)$  plane, in  $\Lambda$ CDM, from CMB+BAO+SNe (red), standard sirens at ET with a GRB counterpart determined by THESEUS (gray), and the result obtained by combining standard sirens with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution. We use the optimistic estimate for the FOV of THESEUS.



**Figure 11.** As in figure 10, with the realistic estimate for the FOV of THESEUS.

only (gray contour) and the combined datasets CMB+BAO+SNe+standard sirens (blue contours). In particular, in figure 10 we show the result for the two distribution of neutron star masses in the optimistic scenario for the FOV of THESEUS, while in figure 11 we show the result for the two distribution of neutron star masses in the realistic scenario for the FOV.

|                           | CMB+BAO+SNe | ET,<br>flat | ET,<br>gaussian | combined,<br>flat | combined,<br>gaussian |
|---------------------------|-------------|-------------|-----------------|-------------------|-----------------------|
| $\Delta H_0/H_0$          | 0.72%       | 0.28%       | 0.23%           | 0.16%             | 0.15%                 |
| $\Delta\Omega_M/\Omega_M$ | 2.11%       | 3.68%       | 3.38%           | 0.59%             | 0.57%                 |

**Table 4.** Accuracy ( $1\sigma$  level) in the reconstruction of  $H_0$  and  $\Omega_M$  with only CMB+BAO+SNe, with only standard sirens (with the flat and gaussian mass distributions, respectively) and the combined results CMB+BAO+SNe+standard sirens, using ET and THESEUS and assuming the optimistic FOV of THESEUS.

|                           | CMB+BAO+SNe | ET,<br>flat | ET,<br>gaussian | combined,<br>flat | combined,<br>gaussian |
|---------------------------|-------------|-------------|-----------------|-------------------|-----------------------|
| $\Delta H_0/H_0$          | 0.72%       | 0.42%       | 0.39%           | 0.26%             | 0.25%                 |
| $\Delta\Omega_M/\Omega_M$ | 2.11%       | 6.17%       | 5.88%           | 0.82%             | 0.82%                 |

**Table 5.** As in table 4, assuming the realistic FOV of THESEUS.

First of all observe that, despite the fact that the mock catalog of standard sirens has been generated by taking as fiducial cosmological model  $\Lambda$ CDM with the values of  $H_0$  and  $\Omega_M$  obtained from these CMB+BAO+SNe data, the contour obtained from standard sirens only is not always centered on the mean values given by the CMB+BAO+SNe contour. This is an unavoidable consequence of the fact that, in order to simulate the result from actual observations, we have scattered the values of  $d_L(z)$  with a variance  $\Delta d_L(z)$  given by the expected observational error, see the discussion in section 2. Of course, this is the situation that would take place in an actual observation (although the actual position of the gray contour will depend on the particular realization of the random scattering of the data around their mean value), and all that we should expect is that the contours are consistent at, say, the  $(1 - 2)\sigma$  level, which is indeed the case. However, one should be aware of the fact that, if we combine the standard sirens and the CMB+BAO+SNe dataset in a realization where the two contours do not overlap well, we get a larger error on the parameters with respect to what is obtained in a realization where the contours happens to overlap well. Once again, this is exactly the situation that will be faced in the actual experiment.

The accuracy on  $H_0$  and  $\Omega_M$  from the corresponding one-dimensional likelihood are given in table 4 (for the optimistic FOV of THESEUS) and table 5 (for the realistic FOV). From these tables we see that standard sirens at ET, already before combining them with other cosmological datasets, give an accuracy on  $H_0$  between 0.2% and 0.4%, depending on the scenarios considered. This is a very interesting accuracy, that would allow to conclusively arbitrate the tension between the local  $H_0$  measurement and the *Planck*- $\Lambda$ CDM value, with totally different systematic compared to SNe. In particular, this measurement would have the potential of falsifying  $\Lambda$ CDM.

If instead the value of  $H_0$  from standard sirens should agree with the CMB+BAO+SNe value obtained using  $\Lambda$ CDM, it would then make sense to combine these datasets. As we see from the tables, in that case the overall accuracy on  $H_0$  could reach  $(0.15 - 0.25)\%$ .

### 3.3 Results for ET+CE+CE

We finally consider the ET+CE+CE network. The two-dimensional likelihoods in the  $(\Omega_M, H_0)$  plane are shown in figures 12 and 13, and the corresponding  $1\sigma$  accuracies from the one-dimensional likelihoods are shown in tables 6 (for the optimistic FOV of THESEUS) and

|                           | CMB+BAO+SNe | ET+CE+CE,<br>flat | ET+CE+CE,<br>gaussian | combined,<br>flat | combined,<br>gaussian |
|---------------------------|-------------|-------------------|-----------------------|-------------------|-----------------------|
| $\Delta H_0/H_0$          | 0.72%       | 0.20%             | 0.22%                 | 0.07%             | 0.07%                 |
| $\Delta\Omega_M/\Omega_M$ | 2.11%       | 1.43%             | 1.31%                 | 0.43%             | 0.42%                 |

**Table 6.** Accuracy ( $1\sigma$  level) in the reconstruction of  $H_0$  and  $\Omega_M$  with only CMB+BAO+SNe, with only standard sirens (with the flat and gaussian mass distributions, respectively) and the combined results CMB+BAO+SNe+standard sirens, using ET+CE+CE and THESEUS, and assuming the optimistic FOV of THESEUS.

|                           | CMB+BAO+SNe | ET+CE+CE,<br>flat | ET+CE+CE,<br>gaussian | combined,<br>flat | combined,<br>gaussian |
|---------------------------|-------------|-------------------|-----------------------|-------------------|-----------------------|
| $\Delta H_0/H_0$          | 0.72%       | 0.24%             | 0.23%                 | 0.12%             | 0.11%                 |
| $\Delta\Omega_M/\Omega_M$ | 2.11%       | 2.12%             | 2.09%                 | 0.51%             | 0.52%                 |

**Table 7.** As in table 4, assuming the realistic FOV of THESEUS.

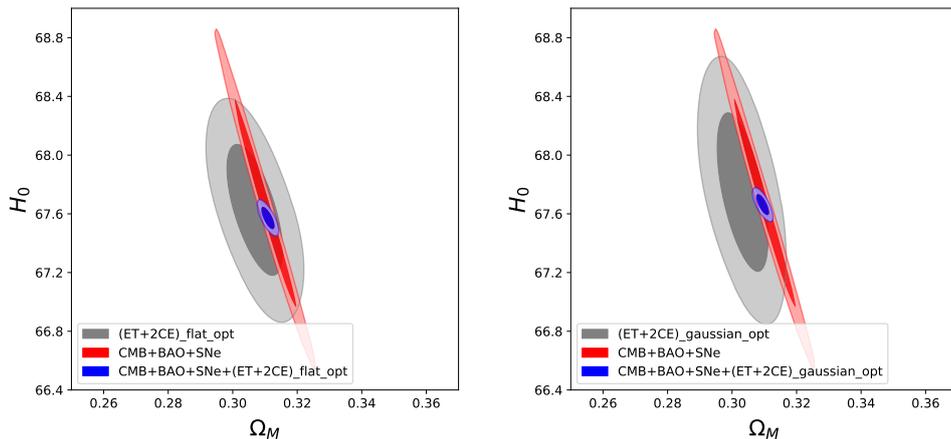
table 7 (for the realistic FOV). We see that, in this case, with standard sirens only we get an accuracy of  $H_0$  of about 0.2%, while, combining with the other cosmological datasets we can reach an accuracies of order (0.07 – 0.12)%. While in itself this would be a remarkable accuracy, still it is not significantly better than that reached with a single ET detector, as we see comparing with tables 4 and 5. This result can be understood by looking at the number of events in our catalogs, shown in table 2. Despite the fact that the ET+CE+CE network has a number of GW detections higher by a factor  $\mathcal{O}(10)$  compared to a single ET, when we look at joint GW-GRB detections the increase in the number of events is less than a factor of 2. In other words, the bottleneck here is on the GRB side. It is crucial to observe, however, that a three detector network such as ET+CE+CE will have excellent localization capabilities. Thus, at least for the events at  $z \lesssim 0.5$ , the follow-up with optical and IR telescope will be possible, and will probably lead to a significant increase in the number of standard sirens with electromagnetic counterpart. As we discussed in section 2.2.2, realistic estimates are currently difficult because they also depend on choices such as the amount of telescope time that will be devoted by the various facilities to the follow-up of GW events, and in this paper we will not attempt such an estimate. However, it should be borne in mind that, for a ET+CE+CE network, the joint GW-GRB detections that we are studying in this paper might provide just a fraction of the whole sample of GW signals with electromagnetic counterpart. Correspondingly, the accuracies that can be obtained on  $H_0$  and  $\Omega_M$  at ET+CE+CE could be significantly better, compared to the figures that we find.

## 4 Constraints from standard sirens on dark energy

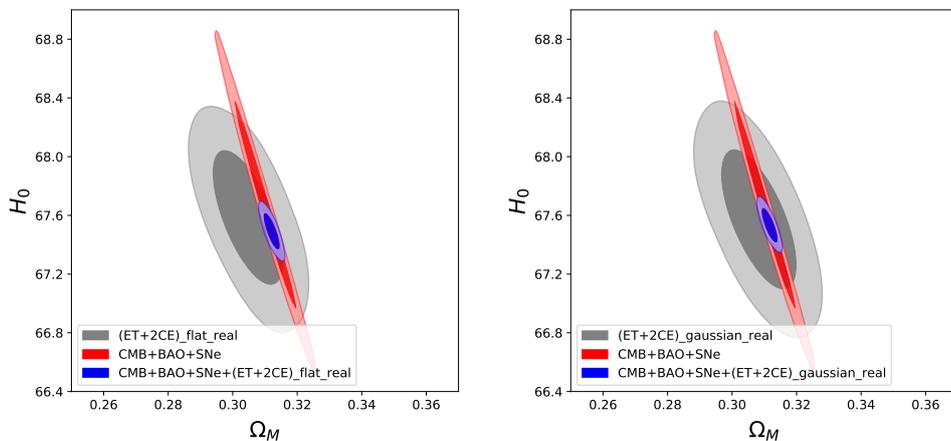
### 4.1 Testing the dark energy sector: DE equation of state and modified GW propagation

We next extend  $\Lambda$ CDM by considering a non-trivial DE sector. A generic modified gravity theory, that generates a dynamical DE density  $\rho_{\text{DE}}(z)$ , will produce a different evolution at the background level, so that eq. (3.1) is now replaced by

$$d_L(z) = \frac{1+z}{H_0} \int_0^z \frac{d\tilde{z}}{\sqrt{\Omega_M(1+\tilde{z})^3 + \Omega_{\text{DE}}(\tilde{z})}}, \quad (4.1)$$



**Figure 12.** As in figure 10, for the ET+CE+CE network and optimistic FOV of THESEUS.



**Figure 13.** As in figure 11, for the ET+CE+CE network and realistic estimate for the FOV of THESEUS.

where

$$\Omega_{\text{DE}}(z) = \Omega_{\text{DE}} \exp \left\{ 3 \int_0^z \frac{d\tilde{z}}{1+\tilde{z}} [1 + w_{\text{DE}}(\tilde{z})] \right\}, \quad (4.2)$$

$\Omega_{\text{DE}} \equiv \Omega_{\text{DE}}(z=0) \simeq 1 - \Omega_M$ , and  $w_{\text{DE}}(z)$  is the DE equation of state. At the phenomenological level, the DE equation of state is usually parametrized as [85, 86]

$$w_{\text{DE}}(z) = w_0 + w_a \frac{z}{1+z}. \quad (4.3)$$

On top of this, any concrete modified gravity model will produce deviations at the level of cosmological perturbations, both in the scalar sector and in the tensor sector. Modified perturbations in the scalar sectors affect the growth of cosmological structures or weak lensing, but are not directly relevant for standard sirens. In contrast, modifications in the tensor perturbations can affect the luminosity distance measured by standard sirens [32, 33, 50–61]. Recall that, in GR, the free propagation of tensor perturbations over a Friedmann-Robertson-Walker (FRW) background is governed by the equation

$$\tilde{h}''_A + 2\mathcal{H}\tilde{h}'_A + k^2\tilde{h}_A = 0, \quad (4.4)$$

where  $\tilde{h}_A(\eta, \mathbf{k})$  are the Fourier modes of the GW amplitude,  $A = +, \times$  labels the two polarizations, the prime denotes the derivative with respect to cosmic time  $\eta$ ,  $a(\eta)$  is the scale factor,  $\mathcal{H} = a'/a$ , and we use units  $c = 1$ . In a generic modified gravity model both the coefficient of the term  $k^2\tilde{h}_A$  and of the term  $2\mathcal{H}\tilde{h}'_A$  can in principle be modified. A modification of the  $k^2\tilde{h}_A$  term would affect the speed of propagation of GWs and is by now excluded, at the level  $|c_{\text{gw}} - c|/c < O(10^{-15})$ , by the observation of GW170817/GRB 170817A [10]. However, one can find explicit examples of viable modified gravity models where the equation for the free propagation of tensor perturbation takes the form

$$\tilde{h}_A'' + 2\mathcal{H}[1 - \delta(\eta)]\tilde{h}'_A + k^2\tilde{h}_A = 0, \quad (4.5)$$

where  $\delta(\eta)$  is a function of conformal time or, equivalently, of redshift. In this case the speed of GWs, that is determined by the coefficient of the  $k^2\tilde{h}_A$  term, is not affected. However, the coefficient of the ‘friction term’, i.e. of the term  $2\mathcal{H}\tilde{h}'_A$ , changes with respect to its value in GR. In fact, the recent analysis in [38] shows that all the best studied models of modified gravity, such as Horndeski or the more general degenerate higher order scalar-tensor (DHOST) theories, non-local infrared modifications of gravity, or bigravity theories, have a propagation equation with a non-trivial function  $\delta(\eta)$ , even when they predict a speed of GWs  $c_{\text{gw}} = c$ .<sup>13</sup> In that case, it is possible to show that standard sirens do not measure the usual “electromagnetic” luminosity distance  $d_L^{\text{em}}(z)$  given by eq. (4.1), but rather a “GW luminosity distance”  $d_L^{\text{gw}}(z)$ , related to  $d_L^{\text{em}}(z)$  by [32, 88] (see also section 19.6.3 of [89])

$$d_L^{\text{gw}}(z) = d_L^{\text{em}}(z) \exp \left\{ - \int_0^z \frac{dz'}{1+z'} \delta(z') \right\}. \quad (4.6)$$

Thus, standard sirens can in principle test the DE sector of a cosmological model through the functions  $w_{\text{DE}}(z)$  and  $\delta(z)$ . To perform phenomenological studies, it is convenient to have simple parametrization of these functions. The function  $w_{\text{DE}}(z)$  is usually parametrized as in eq. (4.3), in terms of the two parameters  $(w_0, w_a)$ . For modified GW propagation a convenient parametrization, in terms of two parameters  $(\Xi_0, n)$ , has been proposed in [33],

$$\frac{d_L^{\text{gw}}(z)}{d_L^{\text{em}}(z)} = \Xi_0 + \frac{1 - \Xi_0}{(1+z)^n}. \quad (4.7)$$

This expression correctly reproduces the fact that, as  $z \rightarrow 0$ ,  $d_L^{\text{gw}}/d_L^{\text{em}} = 1$ , since as the redshift of the source goes to zero, there is no effect from modified gravity propagation. On the opposite limit of large redshifts, in eq. (4.7)  $d_L^{\text{gw}}/d_L^{\text{em}}$  goes to a constant value  $\Xi_0$ . This is

<sup>13</sup>A note of caution, however, is that in the bigravity case discussed in [38] the prediction was obtained in a region of parameter space where the model does not give accelerated expansion, and therefore must be supplemented by a cosmological constant. Another concern, valid for Horndeski-type theories, is that sizable deviations from GR might only be possible in regions of parameter space where the model lacks screening and therefore is not consistent with Solar System tests, as recently claimed in [87]. In contrast, the ‘RT’ nonlocal modification of gravity that we will mention below gives an interesting prediction for  $\Xi_0$  and at the same time is fully viable observationally, both from the point of view of large-scale cosmology and solar system tests. The difference can be traced to the fact that in some scalar-tensor theories, such as Horndeski, modified GW propagation is due to a time dependence of the effective Newton’s constant of the theory. However, this is by no means the most general situation, see the discussion in section III of ref. [33]. For instance, in the RT nonlocal model there is no relation between modified GW propagation and the effective Newton’s constant, and this model gives sizable deviations from  $\Xi_0 = 1$ , while being fully viable. Similarly, for models with extra dimensions, modified GW propagation is a consequence of the propagation of gravitons into a higher-dimensional bulk [50, 59] and a priori is unrelated to the effective Newton’s constant.

motivated by the fact that, in typical models where the modifications from GR only appear close to the recent cosmological epoch, one expects that  $\delta(z)$  would go to zero at large redshift so, from eq. (4.6),  $d_L^{\text{gw}}(z)/d_L^{\text{em}}(z)$  should saturate to a constant value at large  $z$ .

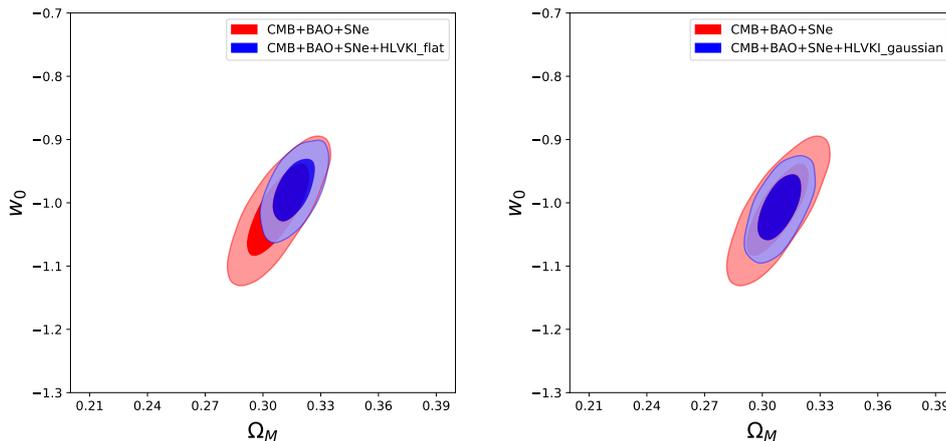
An explicit example of models with modified GW propagation of this type is given by the ‘RT’ and ‘RR’ nonlocal modifications of gravity that were proposed in [90] and [91], respectively, and whose cosmological consequences have been studied in detail in [88, 92–97] (see [88, 98] for review and further references). In both cases the predictions for  $d_L^{\text{gw}}(z)/d_L^{\text{em}}(z)$  are very accurately reproduced by eq. (4.7), with  $\{\Xi_0 \simeq 0.970, n \simeq 2.5\}$  for the RR model [88], and  $\{\Xi_0 \simeq 0.934, n \simeq 2.6\}$  for the ‘minimal’ RT model [38].<sup>14</sup> Actually, recent results [97] have shown that the RR model is ruled out by Lunar Laser Ranging, while the RT model is fully viable: it has a background evolution with accelerated expansion in the present epoch, stable cosmological perturbations, passes solar system and Lunar Laser Ranging constraints, fits the cosmological CMB, BAO and SNa data at a level statistically indistinguishable from  $\Lambda$ CDM, and predicts  $c_{\text{gw}} = c$ . It therefore gives an example of a phenomenologically viable alternative to  $\Lambda$ CDM, which predicts a value of  $\Xi_0$  that differs from the GR value  $\Xi_0 = 1$  by about 6.6%. As we will see below, such a value is well within the reach of ET, and in fact even of the HLVKI network. Several other modified gravity models have been explored in [38], and they all give a modified GW luminosity distance, with an expression that is very well fitted by eq. (4.7) [with the only exception of bigravity, where the interaction between the two metrics induces an interesting phenomenon of oscillations in  $d_L^{\text{gw}}(z)/d_L^{\text{em}}(z)$ ].

In the following we will explore the potential of standard sirens to constrain the DE sector of a modified gravity theory, using either  $(w_0, w_a)$  or  $(\Xi_0, w_0)$  as the extra parameters with respect to  $\Lambda$ CDM.<sup>15</sup>

As before, we will combine standard sirens with CMB, BAO and SNe. Observe that we use  $\Lambda$ CDM as our fiducial cosmological model (and GR as our fiducial description of gravity), so in particular our fiducial values for the extra parameters in the DE sector are  $w_0 = -1$ ,  $w_a = 0$  and  $\Xi_0 = 1$ , and we are studying to what accuracy we can get back these values from the data. For this reason, it is correct to use  $\Lambda$ CDM (and GR) to compute the cosmological perturbations when comparing with CMB, BAO and SNe. A generic modified gravity theory might also a priori induce modifications in the GW production mechanism, beside the modification in the GW propagation. The two effects are however largely decoupled, with modified GW propagation affecting the amplitude of the GW signal through a modification of the luminosity distance, in a way that is incremental with the distance to the source and therefore redshift dependent [at least up to some redshift where the effect saturates, for models reproduced by the  $(\Xi_0, n)$  parametrization], while modified GW production affects the phase of the signal, at some post-Newtonian order. On top of this, models that provide interesting modifications of GR at cosmological scales must have a form of screening at short distances in order not to spoil the successes of GR at the Solar System and laboratory scales, so they typically do not induce any observable effect on GR production. This is in particular the case for the RT nonlocal model, whose effects at scales small compared to  $H_0^{-1}$  are

<sup>14</sup>The ‘minimal’ RT model is defined as the RT model with vanishing initial conditions set during radiation dominance. An analysis with more general initial conditions set during inflation shows that a much larger result, up to  $\Xi_0 \simeq 1.6$ , can be obtained [99].

<sup>15</sup>As discussed in [33], the parameter  $n$  plays in general a lesser role, since it just determines the precise way in which the function  $d_L^{\text{gw}}(z)/d_L^{\text{em}}(z)$  interpolates between its asymptotic value  $\Xi_0$  at large redshift, and the value  $d_L^{\text{gw}}/d_L^{\text{em}} = 1$  at  $z = 0$ . Most sources relevant for ET are already in, or close to, the asymptotic regime. For definiteness, we will then hold  $n$  fixed to the value  $n = 5/2$ .



**Figure 14.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(\Omega_M, w_0)$  plane, in  $w$ CDM, from CMB+BAO+SNe (red), and the result obtained by combining standard sirens at the HLVKI network with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution.

totally negligible [90, 97, 100]. It therefore makes sense to study modified GW propagation in isolation, without worrying about possible corrections on the accuracy on  $\Delta d_L/d_L$  that could be induced by modifications in the GW production mechanism.

## 4.2 Results for the HLVKI network

We begin by presenting the constraints that can be obtained on the DE sector by combining standard sirens at the HLVKI network with the CMB+BAO+SNe dataset described in section 3.

Figure 14 shows the likelihood in the  $(\Omega_M, w_0)$  plane in  $w$ CDM, i.e. when we introduce  $w_0$  as the only new parameter that describes the DE sector, while setting  $w_a = 0$  and excluding also modified GW propagation, i.e setting  $\Xi_0 = 1$ , while Table 8 shows the error on  $w_0$  (at  $1\sigma$ , as in all our tables) from the corresponding one-dimensional likelihood. We give the result from CMB+BAO+SNe only, and that obtained by combining CMB+BAO+SNe with standard sirens with flat mass distribution or with the gaussian mass distribution.

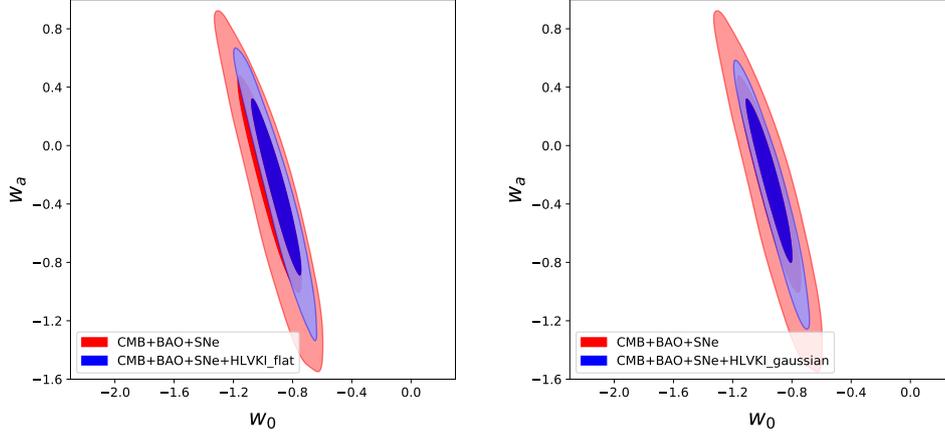
The results for the  $(w_0, w_a)$  parametrization are shown in figure 15 and table 9. Of course, enlarging the parameter space with one more parameter  $w_a$  results in a larger error on  $w_0$ , compared to the results in table 8.

We finally consider the  $(\Xi_0, w_0)$  extension of the DE sector. The results are shown in figure 16 and table 10. The most interesting result is the one for  $\Xi_0$ , which (with our rather extreme assumption of 10 yr of data taking) can be measured to an accuracy of order  $\Delta\Xi_0 \simeq 0.1$ , i.e. (given that our fiducial value has been taken to be the  $\Lambda$ CDM value  $\Xi_0 = 1$ ), a relative accuracy  $\Delta\Xi_0/\Xi_0 \simeq 10\%$ , which is already in the ballpark of the predictions of interesting modified gravity models [38, 99]. With a shorter but more realistic time of data taking, say 3-4 yr, we still expect to get  $\Delta\Xi_0/\Xi_0$  at the level of about 20%.

Of course, CMB, BAO and SNe are blind to modified GW propagation, and the corresponding contour is flat in the  $\Xi_0$  direction. Standard sirens lift this flat direction. In contrast, we see that the improvement on  $w_0$  or  $w_a$  from the inclusion of standard sirens is quite modest.

|              | CMB+BAO+SNe | combined, flat | combined, gaussian |
|--------------|-------------|----------------|--------------------|
| $\Delta w_0$ | 0.045       | 0.033          | 0.035              |

**Table 8.** Accuracy ( $1\sigma$  level) in the reconstruction of  $w_0$  with CMB+BAO+SNe only, and the combined result CMB+BAO+SNe+standard sirens, using the HLVKI detector network and the flat and gaussian mass distributions.



**Figure 15.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(w_0, w_a)$  plane, in the  $(w_0, w_a)$  extension of the DE sector, from CMB+BAO+SNe (red), and the result obtained by combining standard sirens at the HLVKI network with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution.

|              | CMB+BAO+SNe | combined, flat | combined, gaussian |
|--------------|-------------|----------------|--------------------|
| $\Delta w_0$ | 0.140       | 0.113          | 0.106              |
| $\Delta w_a$ | 0.483       | 0.406          | 0.380              |

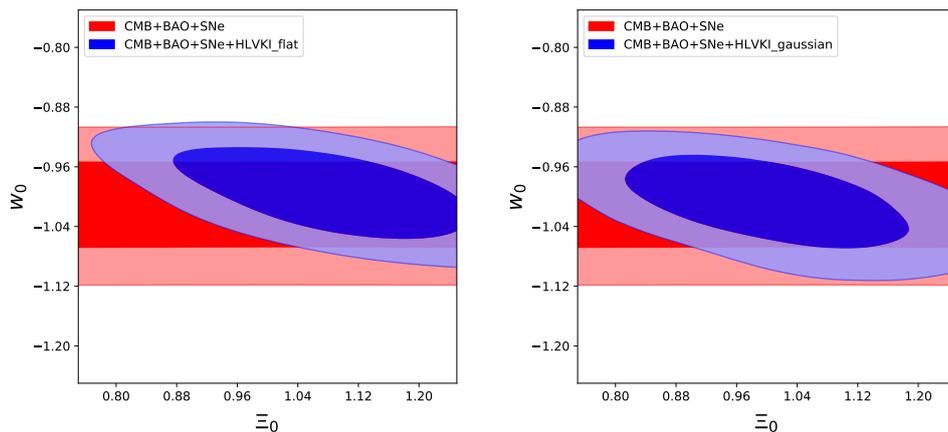
**Table 9.** Accuracy ( $1\sigma$  level) in the reconstruction of  $w_0$  and  $w_a$  with CMB+BAO+SNe only, and the combined result CMB+BAO+SNe+standard sirens, using the HLVKI detector network and the flat and gaussian mass distributions.

|                | CMB+BAO+SNe | combined, flat | combined, gaussian |
|----------------|-------------|----------------|--------------------|
| $\Delta w_0$   | 0.045       | 0.042          | 0.042              |
| $\Delta \Xi_0$ | —           | 0.130          | 0.125              |

**Table 10.** Accuracy ( $1\sigma$  level) in the reconstruction of  $w_0$  and  $\Xi_0$  with CMB+BAO+SNe only, and the combined result CMB+BAO+SNe+standard sirens, using the HLVKI detector network and the flat and gaussian mass distributions.

### 4.3 Results for ET

We next consider a single ET in coincidence with a GRB detector with the characteristics of THESEUS. Proceeding as before, we show first the results of the  $w_0$  extension, in tables 11 and 12 and figures 17 and 18, displaying separately the results for the optimistic and realistic FOV of THESEUS. Observe that, contrary to the HLVKI case, now we can obtain some bounds already using standard sirens alone. However, we see from the figures that the central value of the contour of the standard sirens can happen to be displaced with respect



**Figure 16.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(\Xi_0, w_0)$  plane, in the  $(\Xi_0, w_0)$  extension of the DE sector, from CMB+BAO+SNe (red), and the result obtained by combining standard sirens at the HLVKI network with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution.

|              | CMB+BAO+SNe | ET, flat | ET, gaussian | combined, flat | combined, gaussian |
|--------------|-------------|----------|--------------|----------------|--------------------|
| $\Delta w_0$ | 0.045       | 0.109    | 0.116        | 0.020          | 0.021              |

**Table 11.** Accuracy ( $1\sigma$  level) in the reconstruction of  $w_0$  with only CMB+BAO+SNe, with only standard sirens (with the flat and gaussian mass distributions, respectively) and the combined results CMB+BAO+SNe+standard sirens, using ET and THESEUS and assuming the optimistic FOV of THESEUS.

|              | CMB+BAO+SNe | ET, flat | ET, gaussian | combined, flat | combined, gaussian |
|--------------|-------------|----------|--------------|----------------|--------------------|
| $\Delta w_0$ | 0.045       | 0.301    | 0.158        | 0.023          | 0.024              |

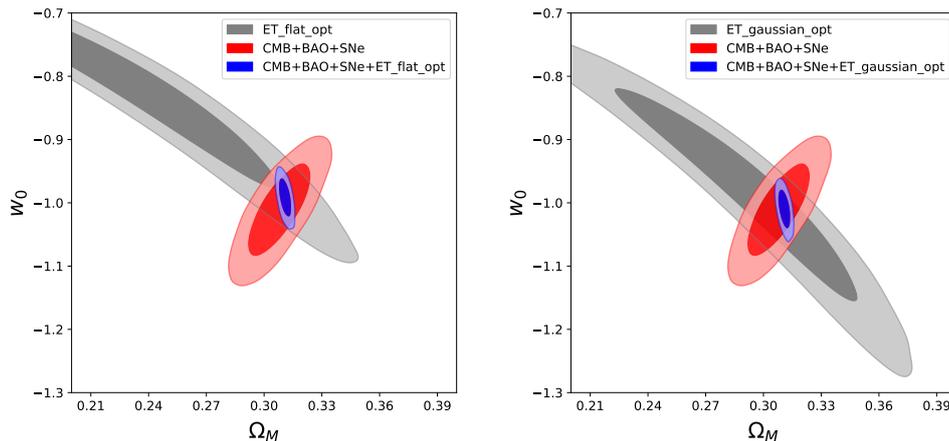
**Table 12.** As in table 11, assuming the realistic FOV of THESEUS.

to that from CMB+BAO+SNe. As we already discussed in section 3.2, this is a statistical effect due to the random scattering of the mock GW data according to the error estimate that, depending on the specific realization, can induce a more or less significant displacement between the two contours.

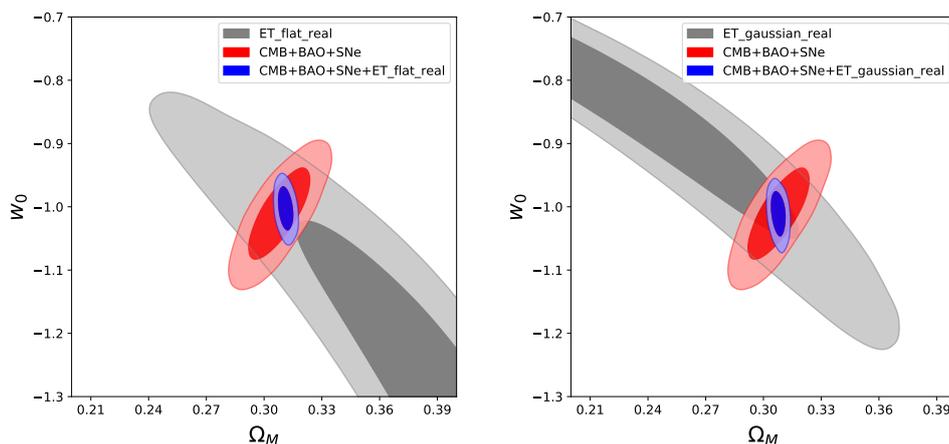
We see that, adding the joint GW-GRB events to the CMB+BAO+SNe dataset, we can improve the accuracy on  $w_0$  by about a factor of 2. This is interesting, although certainly not spectacular. We will see below that the most interesting contribution of a 3G detector such as ET to the exploration of the DE sector rather comes from modified GW propagation.

We next include  $(w_0, w_a)$  as extra parameters. The results are shown in tables 13 and 14 and in figures 19 and 20. In this case, with one extra parameter  $w_a$ , for some scenarios the MCMC does not converge well with standard sirens only, and we just show the combined result for CMB+BAO+SNe+standard sirens, for all scenarios.

We finally include modified GW propagation, adding to the baseline  $\Lambda$ CDM model the parameters  $(\Xi_0, w_0)$ , and writing the GW luminosity distance as in eq. (4.7). We set for simplicity  $n = 5/2$ , which is of the order of the value predicted by the RT and RR nonlocal gravity model. However, the precise value of  $n$  is of limited relevance for the analysis. Note



**Figure 17.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(\Omega_M, w_0)$  plane, in  $w$ CDM, from CMB+BAO+SNe (red), joint detection of standard sirens at ET and THESEUS (gray) and the result obtained by combining standard sirens with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution. We use the optimistic estimate for the FOV of THESEUS.



**Figure 18.** As in figure 17, with the realistic estimate for the FOV of THESEUS.

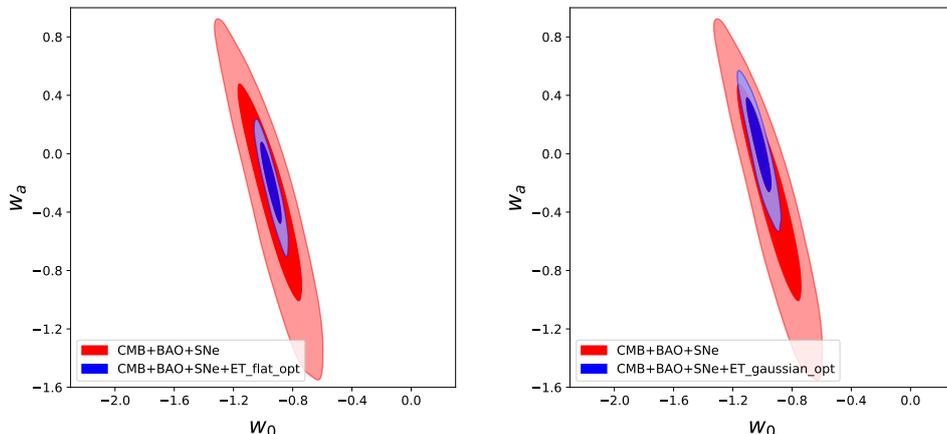
|              | CMB+BAO+SNe | combined,<br>flat | combined,<br>gaussian |
|--------------|-------------|-------------------|-----------------------|
| $\Delta w_0$ | 0.140       | 0.050             | 0.058                 |
| $\Delta w_a$ | 0.483       | 0.193             | 0.224                 |

**Table 13.** Accuracy ( $1\sigma$  level) in the reconstruction of  $(w_0, w_a)$  with only CMB+BAO+SNe and the combined results CMB+BAO+SNe+standard sirens, using ET and THESEUS and assuming the optimistic FOV of THESEUS.

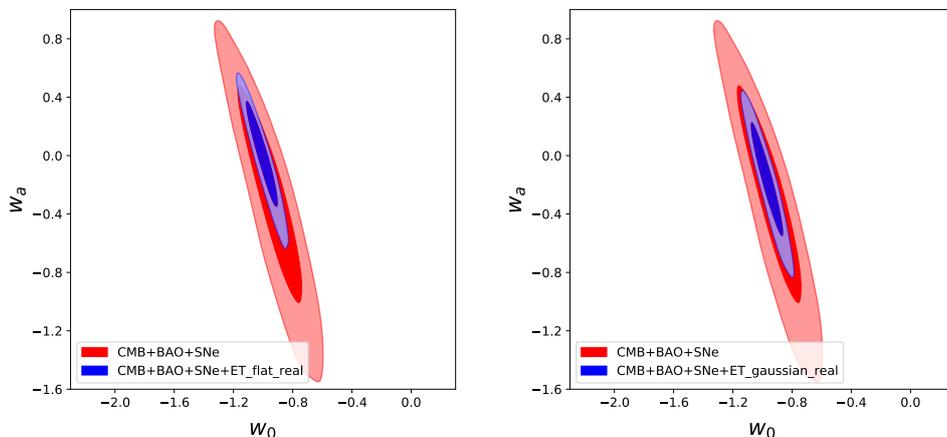
that, since our catalog of sources has been generated assuming  $\Lambda$ CDM as our fiducial model, our fiducial values for these parameters are  $\Xi_0 = 1$  and  $w_0 = -1$ , and we compute the accuracy  $\Delta \Xi_0$  and  $\Delta w_0$  to which we can find back these values. The results are shown in tables 15 and 16 and in figures 21 and 22.

|              | CMB+BAO+SNe | combined,<br>flat | combined,<br>gaussian |
|--------------|-------------|-------------------|-----------------------|
| $\Delta w_0$ | 0.140       | 0.073             | 0.072                 |
| $\Delta w_a$ | 0.483       | 0.246             | 0.260                 |

**Table 14.** As in table 13, assuming the realistic FOV of THESEUS.



**Figure 19.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(w_0, w_a)$  plane from CMB+BAO+SNe (red) and the result obtained by combining standard sirens at ET with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution. We use the optimistic estimate for the FOV of THESEUS.



**Figure 20.** As in figure 19, with the realistic estimate for the FOV of THESEUS.

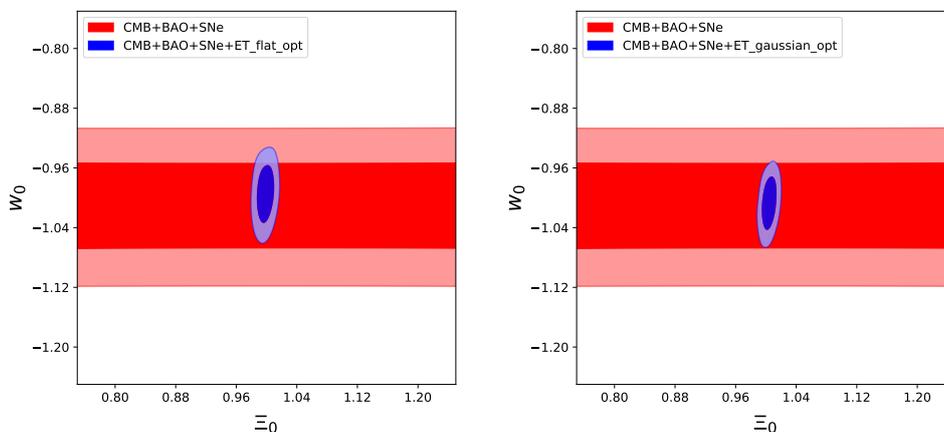
It is quite remarkable that, by combining the joint GW-GRB detections from a single ET detector and a GRB detector such as THESEUS, with the current CMB+BAO+SNe dataset, we can reach an accuracy on  $\Xi_0$  that, depending on the scenario, is between 0.7% and 1.1%. By comparison, the ‘minimal’ RT nonlocal model predicts a deviation from  $\Xi_0 = 1$  at the level of 6.6%, almost an order of magnitude larger than this observational sensitivity (and much large values can be obtained with initial conditions set during inflation [99]). This shows that, while the sensitivity of a single ET detector to  $w_0$  will not allow us to obtain

|                | CMB+BAO+SNe | combined,<br>flat | combined,<br>gaussian |
|----------------|-------------|-------------------|-----------------------|
| $\Delta w_0$   | 0.045       | 0.026             | 0.024                 |
| $\Delta \Xi_0$ | —           | 0.008             | 0.007                 |

**Table 15.** Accuracy ( $1\sigma$  level) in the reconstruction of  $(w_0, \Xi_0)$  with only CMB+BAO+SNe and the combined results CMB+BAO+SNe+standard sirens, using ET and THESEUS and assuming the optimistic FOV of THESEUS.

|                | CMB+BAO+SNe | combined,<br>flat | combined,<br>gaussian |
|----------------|-------------|-------------------|-----------------------|
| $\Delta w_0$   | 0.045       | 0.026             | 0.026                 |
| $\Delta \Xi_0$ | —           | 0.011             | 0.010                 |

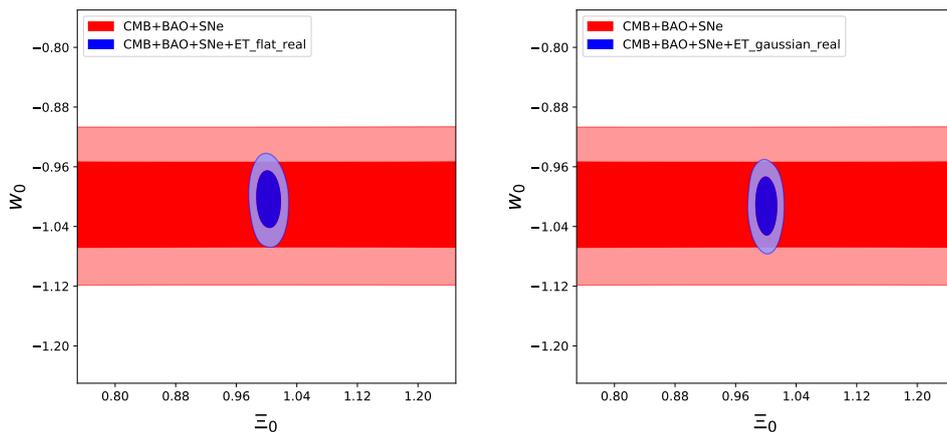
**Table 16.** As in table 15, assuming the realistic FOV of THESEUS.



**Figure 21.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(\Xi_0, w_0)$  plane from CMB+BAO+SNe (red) and the result obtained by combining joint detections of standard sirens at ET and THESEUS with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution. We use the optimistic estimate for the FOV of THESEUS.

a dramatic improvement on the current knowledge of  $w_0$ , the sensitivity to  $\Xi_0$  is extremely interesting and well within the prediction of viable modified gravity models. Furthermore, modified GW propagation, as encoded for instance in the  $\Xi_0$  parameter [or, more generally, in the  $(\Xi_0, n)$  parameters] is an observable specific to GW detectors, to which electromagnetic observations are blind.

The results of this section can be compared to those in [33], which, following [22, 23, 30], were obtained under the working hypothesis that ET, over a few years of data taking, will eventually be able to collect  $\mathcal{O}(10^3)$  BNS with counterpart (without specifying how the counterpart is actually detected), and assuming a redshift distribution proportional to a simple model for the star formation rate (and neglecting the effect of the delay between binary formation and merger). For the  $(w_0, \Xi_0)$  extension of the DE sector, the analysis of ref. [33] then led to the forecast  $\Delta \Xi_0 = 0.008$  and  $\Delta w_0 = 0.032$ , which happens to be very close to the results in table 15.



**Figure 22.** As in figure 21, with the realistic estimate for the FOV of THESEUS.

Comparing with our present results, first of all we see from table 2 that the assumption of  $10^3$  standard sirens with counterpart, while optimistic, was not unrealistic; in particular, for a gaussian neutron-star mass distribution (and with the optimistic estimate for the FOV of THESEUS), in 10 yr of data we could have  $\mathcal{O}(500)$  joint GW-GRB events. Furthermore, as we have already mentioned in section 2.2.2, in [23] was used a threshold for the network SNR obtained combining the three arms of ET given by  $\rho_{\text{threshold}} = 8$  while we use  $\rho_{\text{threshold}} = 12$ . Lowering the threshold would lead to an increase in the number of GW events. Still, at first it could be surprising that the result for  $\Delta\xi_0$  that we find in this paper happens to be practically identical to that of ref. [33], given that the number of sources that we are using here is smaller by a factor  $\simeq 2$  compared to the  $10^3$  sources used in ref. [33]. However, this can be traced to the fact that also the redshift distribution of the sources is different. Indeed, we see from the right-panel in figure 3 that most of the joint GW-GRB detections are at  $z < 0.5$ , while in ref. [33] it was assumed that the catalog of sources followed a distribution in redshift determined by the star formation rate; that catalog was peaked at  $z \simeq 1$ , with long tails at larger  $z$ , see figure 8 of ref. [33]. On the other hand, it was also found in ref. [33] that the main contribution to the determination of  $\xi_0$  was given by the sources at  $z < 0.7$ , that were about one half of the total, so in the end it is not surprising that our catalog, with about a factor of 2 less sources, but almost all concentrated at  $z < 0.7$ , gives basically the same results as the catalog used in ref. [33].

#### 4.4 Results for ET+CE+CE

We finally give the corresponding results for the ET+CE+CE network. For  $w$ CDM the results are shown in figures 23 and 24 and tables 17 and 18. For the  $(w_0, w_a)$  extension the results are shown in figures 25 and 26 and tables 19 and 20. For the  $(\xi_0, w_0)$  extension the results are shown in figures 27 and 28 and tables 21 and 22.

As we already observed in section 3.3 when discussing parameter estimation in  $\Lambda$ CDM, the improvement in the accuracy of the cosmological parameters, compared to the ET-only case, is not very large, because, despite the fact that a network ET+CE+CE detects a number of source larger by an order of magnitude compared to a single ET (and to much larger redshift, see table 2 and figure 3), the corresponding joint GW-GRB detections do not follow the same increase, because of intrinsic limitations in the GRB detections. In particular, as we discussed in section 2.2.2, despite the fact that the ET+CE+CE network can detect BNS up to  $z \simeq 10$  (in our catalog the source with the largest redshift has  $z \simeq 9.66$ ), the joint

|              | CMB+BAO+SNe | ET+CE+CE,<br>flat | ET+CE+CE,<br>gaussian | combined,<br>flat | combined,<br>gaussian |
|--------------|-------------|-------------------|-----------------------|-------------------|-----------------------|
| $\Delta w_0$ | 0.045       | 0.041             | 0.034                 | 0.014             | 0.013                 |

**Table 17.** Accuracy ( $1\sigma$  level) in the reconstruction of  $w_0$  with only CMB+BAO+SNe, with only standard sirens (with the flat and gaussian mass distributions, respectively) and the combined results CMB+BAO+SNe+standard sirens, using ET+CE+CE and THESEUS and assuming the optimistic FOV of THESEUS.

|              | CMB+BAO+SNe | ET+CE+CE,<br>flat | ET+CE+CE,<br>gaussian | combined,<br>flat | combined,<br>gaussian |
|--------------|-------------|-------------------|-----------------------|-------------------|-----------------------|
| $\Delta w_0$ | 0.045       | 0.074             | 0.063                 | 0.020             | 0.018                 |

**Table 18.** As in table 17, assuming the realistic FOV of THESEUS.

|              | CMB+BAO+SNe | combined,<br>flat | combined,<br>gaussian |
|--------------|-------------|-------------------|-----------------------|
| $\Delta w_0$ | 0.140       | 0.027             | 0.025                 |
| $\Delta w_a$ | 0.483       | 0.139             | 0.137                 |

**Table 19.** Accuracy ( $1\sigma$  level) in the reconstruction of  $(w_0, w_a)$  with only CMB+BAO+SNe and the combined results CMB+BAO+SNe+standard sirens, using ET+CE+CE and THESEUS and assuming the optimistic FOV of THESEUS.

|              | CMB+BAO+SNe | combined,<br>flat | combined,<br>gaussian |
|--------------|-------------|-------------------|-----------------------|
| $\Delta w_0$ | 0.140       | 0.041             | 0.037                 |
| $\Delta w_a$ | 0.483       | 0.160             | 0.145                 |

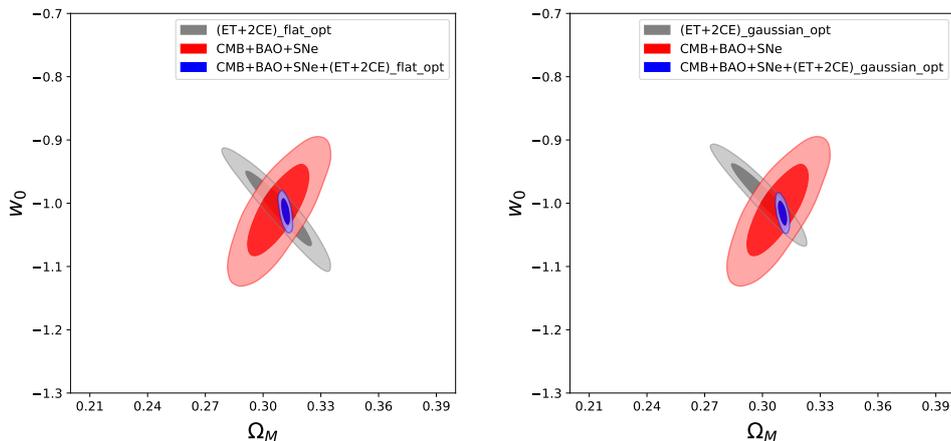
**Table 20.** As in table 19, assuming the realistic FOV of THESEUS.

|                | CMB+BAO+SNe | combined,<br>flat | combined,<br>gaussian |
|----------------|-------------|-------------------|-----------------------|
| $\Delta w_0$   | 0.045       | 0.038             | 0.042                 |
| $\Delta \Xi_0$ | —           | 0.007             | 0.007                 |

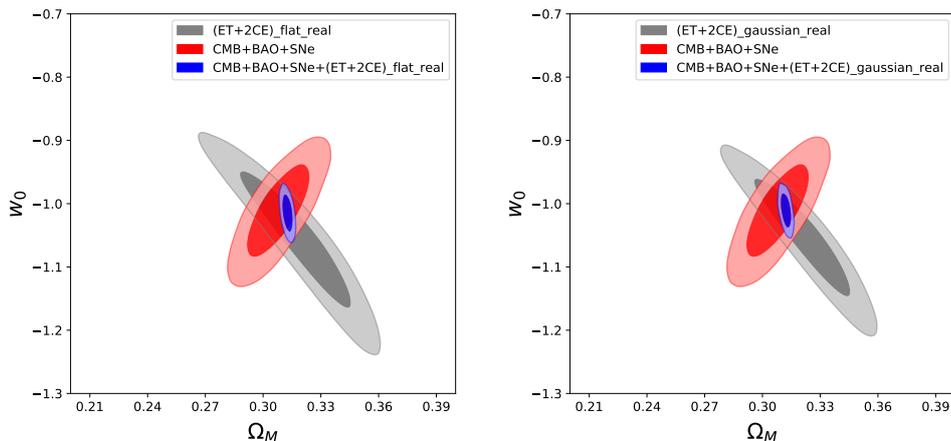
**Table 21.** Accuracy ( $1\sigma$  level) in the reconstruction of  $(w_0, \Xi_0)$  with only CMB+BAO+SNe and the combined results CMB+BAO+SNe+standard sirens, using ET+CE+CE and THESEUS and assuming the optimistic FOV of THESEUS.

detections with GRB only reach, in our catalog,  $z \simeq 3.38$ ; all higher redshift sources are lost because their GRB is beyond the flux limit for detection.

To fully exploit the potential of a ET+CE+CE network for standard sirens it is therefore crucial either to have a more powerful network of multi-messenger observations, for instance with IR/optical telescopes (that, guided by the localization capability of the ET+CE+CE network, could provide many more counterparts, at least the region  $z < 0.5 - 0.7$  corresponding to their reach), or else one should resort to statistical methods for the determination of the host galaxy of standard sirens without electromagnetic counterpart. This is also an important message of our analysis.



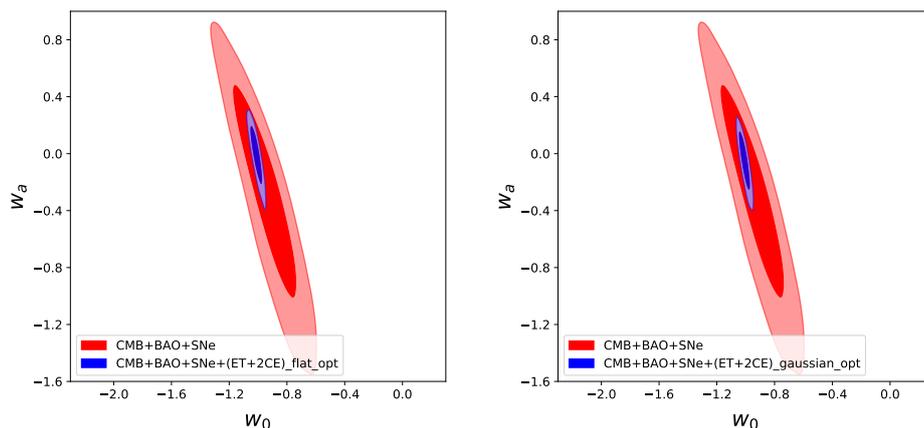
**Figure 23.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(\Omega_M, w_0)$  plane, in  $w$ CDM, from CMB+BAO+SNe (red), joint detection of standard sirens at ET+CE+CE and THESEUS (gray) and the result obtained by combining standard sirens with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution. We use the optimistic estimate for the FOV of THESEUS.



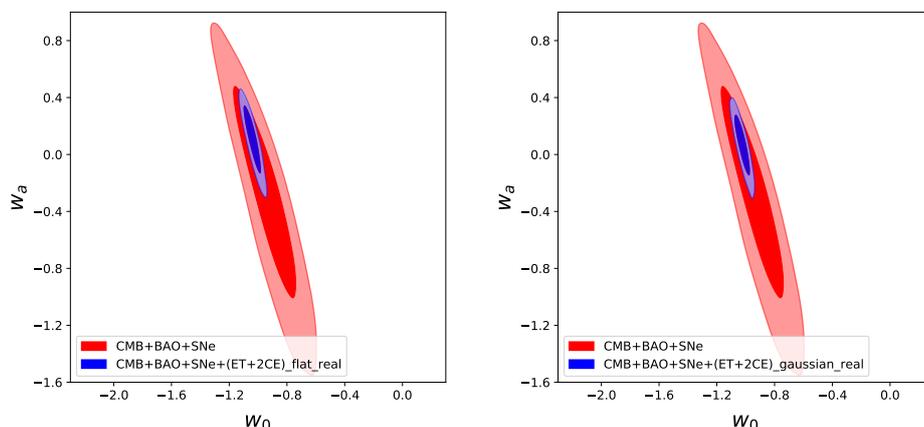
**Figure 24.** As in figure 23, with the realistic estimate for the FOV of THESEUS.

|                | CMB+BAO+SNe | combined,<br>flat | combined,<br>gaussian |
|----------------|-------------|-------------------|-----------------------|
| $\Delta w_0$   | 0.045       | 0.030             | 0.033                 |
| $\Delta \Xi_0$ | —           | 0.006             | 0.007                 |

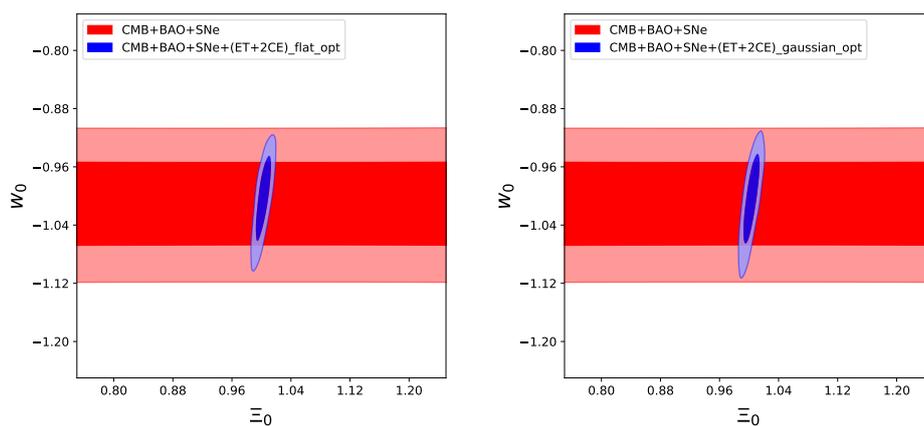
**Table 22.** As in table 21, assuming the realistic FOV of THESEUS.



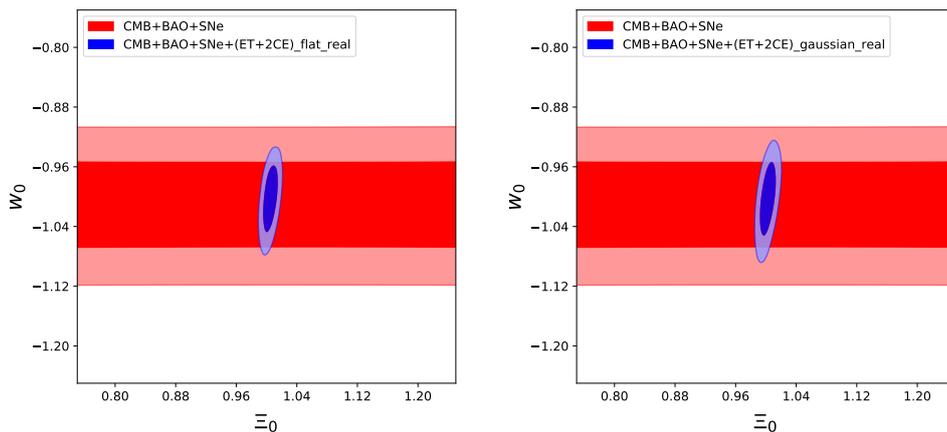
**Figure 25.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(w_0, w_a)$  plane from CMB+BAO+SNe (red) and the result obtained by combining standard sirens at ET+CE+CE with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution. We use the optimistic estimate for the FOV of THESEUS.



**Figure 26.** As in figure 25, with the realistic estimate for the FOV of THESEUS.



**Figure 27.** The  $1\sigma$  and  $2\sigma$  contours of the two-dimensional likelihood in the  $(\Xi_0, w_0)$  plane from CMB+BAO+SNe (red) and the result obtained by combining standard sirens at ET+CE+CE with CMB+BAO+SNe (blue). Left: in the case of flat neutron star mass distribution. Right: in the case of gaussian neutron star mass distribution. We use the optimistic estimate for the FOV of THESEUS.



**Figure 28.** As in figure 27, with the realistic estimate for the FOV of THESEUS.

## 5 Conclusions

One of the main motivations for developing third-generation GW detectors, such as the Einstein Telescope in Europe and Cosmic Explorer in the US, is the possibility of using them for cosmology studies, in a way that will be complementary to what is done with electromagnetic probes such as the CMB, type Ia SNe, BAO, or large-scale structures. In particular, the measurement of the luminosity distance to coalescing binaries through the observation of their GWs, combined with an electromagnetic measurement of the redshift (or with statistical methods), gives access to the Hubble parameter  $H_0$  and, for sources at sufficiently large redshift, also to the dark-energy equation of state.

On top of this, a very interesting example of the potential contributions of advanced GW detectors to cosmology is given by the possibility of studying modified GW propagation. A specific realization of this effect was already discussed more than a decade ago, in terms of leakage of GWs into extra dimensions in the context of the DGP model [50]. More recently, several studies of modified GW propagation have been performed [32, 33, 38, 52–61], and this effect is now being appreciated as one of the most promising tools for testing modified gravity and the dark energy sector of alternatives to  $\Lambda$ CDM, both with third-generation ground-based GW detectors, and with the space interferometer LISA. There are several reasons for this:

- Modified GW propagation emerges very generally in all modified gravity theories, typically as a modification of the friction term in the propagation over a FRW background, and in this form it appears for instance in scalar-tensor theories of the Horndeski and DHOST type [52, 54, 56, 57], in non-local modifications of gravity [32, 33, 38, 88], or in bigravity theories [38], and also emerges naturally in the effective field theory of dark energy [53]. In all these models (with the exception of bigravity, where there is an interesting phenomenon of oscillations due to the beatings between the two metrics), the effect is very neatly captured by the parametrization (4.7), in terms of two parameters  $(\Xi_0, n)$ , which provides a simple and convenient way of studying it.
- In a generic modified gravity theory, the effect of modified GW propagation on the luminosity distance of a GW source dominates over that due to the dark energy equation of state. This is due to the fact that, when one changes the equation of state of dark energy with respect to the value  $w = -1$  of  $\Lambda$ CDM, the Bayesian parameter

estimation of the other cosmological parameters, such as  $H_0$  and  $\Omega_M$ , as obtained e.g. by comparison with CMB, BAO and SNe data, of course also gives different values, with respect to their values  $\Lambda$ CDM, and the change goes in the direction of reducing the change of the luminosity distance at large redshift. Thus, in the luminosity distance, the effect due to the change in the equation of state of DE is largely compensated by that in the cosmological parameters. In contrast, modified GW propagation is an extra effect on top of this, that is not partially compensated by Bayesian parameter estimation [32, 33]. Thus, the sensitivity of GW detectors to the parameter  $\Xi_0$  that characterizes modified GW propagation is better than to  $w_0$ , as has also been verified with the explicit MCMC analysis of [33, 38] and of the present work.

- Modified GW propagation can only be revealed with GW observations, and is therefore specific to GW detectors. Furthermore, as found in [33] for ET and in [38] for LISA, the sensitivity of these experiments to  $\Xi_0$  is well within the prediction of some of the most studied alternatives to  $\Lambda$ CDM.

The original motivation of the present paper was to improve on the results of ref. [33] for ET by using a more realistic catalog of sources, in the context of a current effort for building the Science Case for ET and, more generally, for 3G detectors. Indeed, in ref. [33] the sensitivity of ET to  $\Xi_0$  was computed assuming that, over a few years of data taking, ET will be able to collect  $10^3$  standard sirens with observed electromagnetic counterpart [out of the  $\mathcal{O}(10^5)$  BNS/yr that ET is expected to detect]. This is the assumption that has been usually made in the literature in this context but, in the absence of a concrete study of how to detect the counterpart, it is nothing more than a reasonable working hypothesis. Furthermore, the actual redshift distribution of the joint GW-electromagnetic detections depends not only on the specifications of ET, but also of the detectors used to observe the electromagnetic counterpart.

To go beyond these simple assumptions one needs a concrete scenario for the detection of the electromagnetic counterpart. In this paper we have focused on the possibility of a joint GW-GRB detection using, as an electromagnetic partner to 3G detector, the proposed THESEUS mission. As we have repeatedly stressed, this represents only a subset of the possible electromagnetic counterparts that could be observed, particularly for a network with significant localization capabilities such as HLVKI or ET+CE+CE (in which case it will probably not even provide the dominant contribution). However, given the large field of view of GRB satellites, in this case it is possible to give at least some estimates based uniquely on the characteristics of the detector. In contrast, the estimate for the number of counterparts detected at optical/IR telescope also strongly depend on issues, such as the fraction of telescope time that will be devoted to GW follow up, that are currently more difficult to predict.

Our main results are as follows. For the HLVKI network, the number of joint GW-GRB detections (computed assuming that the Fermi-GBM can make a coincident detection and that *Swift* can identify an X-ray counterpart) is quite small, of order 1.5/yr, and is not sufficient to obtain significant cosmological results using standard sirens only. Even when combined with the CMB+BAO+SNe dataset that we have used throughout the paper, it only provides a modest improvement in the accuracy on  $H_0$ . Indeed, to get interesting accuracy on  $H_0$  with standard sirens, say of order 1% or better, as needed to solve the discrepancy between *Planck*- $\Lambda$ CDM value and the local measurement of  $H_0$ , a much higher number of events with counterparts are indeed needed,  $\mathcal{O}(50-100)$  [66, 83]. However, a 2G network such

as HLVKI (or, indeed, already HLV) has significant localization accuracy. This number of sources with counterpart could therefore in principle be obtained through optical/IR follow up. Similarly, joint GW-GRB detections at the HLVKI network cannot be expected to improve significantly the accuracy on the DE equation of state, compared to what already know from electromagnetic observation, see tables 8 and 9. The situation is however very different for what concern modified GW propagation, as encoded in the parameter  $\Xi_0$ . Since electromagnetic observations are blind to it, any result from 2G detectors will be potentially interesting, and we have found that (after combining with CMB+BAO+SNe to reduce the degeneracy with  $H_0$  and  $\Omega_M$ ), the HLVKI network can measure  $\Xi_0$  to about 13%. This results assumes 10 yr of data, which is probably a very optimistic assumption; however, taking e.g. 5 yr of data the result should be approximately worse by about a factor  $\sqrt{2}$ , so that a measurement at the 20% level should still be possible. This is already in the range of the predictions from some modified gravity models [38, 99] and is therefore already a very interesting sensitivity.

For a single ET detector, our results for  $H_0$  are given in tables 4 and 5 and show that, already using only standard sirens, and limiting ourselves to the coincidences with GRBs (estimated using a detector with the characteristic of THESEUS), ET can reach an extremely interesting accuracy, with  $\Delta H_0/H_0$  between 0.2% and 0.4%, depending on the assumptions on the event rate (gaussian vs. flat neutron star mass distribution) and on the FOV of the GRB detector. For the dark energy sector, including both  $w_0$  and  $\Xi_0$ , the results are given in table 15. In particular (again, upon combining with CMB+BAO+SNe to reduce the degeneracies),  $\Xi_0$  can be measured to better than 1%, more precisely to (0.7 – 0.8)%. This is an exciting result, since several modified gravity models give predictions significantly higher than this. This result also strengthens the science case for THESEUS. The corresponding result on  $w_0$  is still valuable but less exciting, since it just brings the error down from 4.5% to 2.5%.

We have finally studied the configuration ET+CE+CE. In this case, despite the huge increase in GW detections, by a factor  $\mathcal{O}(10)$  compared to a single ET, the final results on the cosmological parameters, limiting ourselves to GW-GRB coincidences, are not significantly better than with a single ET. For instance  $\Xi_0$  can now be measured to (0.6 – 0.7)%. This is due to the fact that the number of joint GW-GRB detections does not increase correspondingly; in fact, from table 2 we see that it does not even increase by a factor of 2. In other words, there is a bottleneck, due to the fact that a single GRB detector cannot keep the pace, and does not have the reach, of GW detections from a ET+CE+CE network. However, this network would have remarkable localization accuracy, and therefore could benefit from the detection of counterparts from optical/IR/radio telescopes. It is clear from our analysis that a strong effort in the follow up of GW signals would be necessary to really exploit the remarkable potential for cosmology of a network of 3G detectors, whether in the ET+CE+CE configuration that we have studied, or in intermediate configurations involving e.g. the planned Voyager upgrade of the advanced LIGO detectors.

## Acknowledgments

We thank Lorenzo Amati, Matthew Bailes, Carlo Ferrigno, Bruce Gendre, Vuk Mandic, Samaya Nissanke, Bangalore Sathyaprakash, Volodymyr Savchenko and Giulia Stratta for very useful discussions. The work of E.B., S.F. and M.M. is supported by the Swiss National Science Foundation and by the SwissMap National Center for Competence in Research.

The work of Y.D. is supported by Swiss National Science Foundation and by a Consolidator Grant of the European Research Council (ERC-2015-CoG grant 680886). E.J.H acknowledges support from a Australian Research Council DECRA Fellowship (DE170100891), and parts of his research were supported by the Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), through project number CE170100004.

## A Catalogs of sources

In this appendix we provide some explicit tables with informations on the realizations of the catalogs that we have used, in order to facilitate comparisons with our work. For the 2G HLVKI network, the explicit realization of the catalog of joint GW-GRB detections is given in table 23.

For 3G detectors, given the large number of events, we rather give their distribution in bins of frequency as well as the mean value and the variance of the ET instrumental contribution to  $\Delta d_L/d_L$ , averaged over the events in the given redshift bin, for the specific realization of the catalog of joint GW-GRB detections shown in figure 3. The four case that we consider, corresponding to flat or gaussian mass distribution and optimistic or realistic FOV of THESEUS, are given in tables 24-27.

| $z$      | $d_L$ (Mpc) | $\Delta d_L$ (Mpc) |
|----------|-------------|--------------------|
| 0.030550 | 142.195     | 4.321              |
| 0.036110 | 178.107     | 7.666              |
| 0.053624 | 250.543     | 9.899              |
| 0.055359 | 256.893     | 9.889              |
| 0.064093 | 310.823     | 11.951             |
| 0.069841 | 337.359     | 17.002             |
| 0.082936 | 372.050     | 29.253             |
| 0.087941 | 413.191     | 29.292             |
| 0.089227 | 418.742     | 35.200             |
| 0.092182 | 422.677     | 27.394             |
| 0.104137 | 510.429     | 35.753             |
| 0.108121 | 487.845     | 36.081             |
| 0.108558 | 555.629     | 38.370             |
| 0.119288 | 587.063     | 46.405             |

| $z$      | $d_L$ (Mpc) | $\Delta d_L$ (Mpc) |
|----------|-------------|--------------------|
| 0.029271 | 134.815     | 4.000              |
| 0.035195 | 157.475     | 5.636              |
| 0.060585 | 283.567     | 18.706             |
| 0.066283 | 316.373     | 14.509             |
| 0.071053 | 327.381     | 20.085             |
| 0.071730 | 342.952     | 16.957             |
| 0.076180 | 341.595     | 22.360             |
| 0.081819 | 418.469     | 30.238             |
| 0.088698 | 396.734     | 25.757             |
| 0.091869 | 402.590     | 34.170             |
| 0.094237 | 406.423     | 31.472             |
| 0.095288 | 432.996     | 36.423             |
| 0.099956 | 491.071     | 31.721             |
| 0.102531 | 461.627     | 36.858             |
| 0.114869 | 626.939     | 43.010             |

**Table 23.** The events in the specific realization of the catalog of joint GW-GRB detections for the HLVKI network, shown in figure 2, for the flat mass distribution (left table) and the Gaussian mass distribution (right table).

| redshift bin | number of joint GW-GRB events | mean redshift | mean $\Delta d_L/d_L$ | standard deviation of $\Delta d_L/d_L$ |
|--------------|-------------------------------|---------------|-----------------------|--|
| (0 , 0.1)    | 14                            | 0.07539       | 0.00746               | 0.00306                                |
| (0.1 , 0.2)  | 47                            | 0.15233       | 0.01590               | 0.00667                                |
| (0.2 , 0.3)  | 58                            | 0.24660       | 0.02372               | 0.00818                                |
| (0.3 , 0.4)  | 66                            | 0.35216       | 0.03337               | 0.01247                                |
| (0.4 , 0.5)  | 51                            | 0.44213       | 0.04290               | 0.01439                                |
| (0.5 , 0.6)  | 35                            | 0.54520       | 0.05069               | 0.01422                                |
| (0.6 , 0.7)  | 35                            | 0.64652       | 0.05424               | 0.01391                                |
| (0.7 , 0.8)  | 23                            | 0.74763       | 0.05267               | 0.01265                                |
| (0.8 , 0.9)  | 19                            | 0.84150       | 0.05823               | 0.01161                                |
| (0.9 , 1.0)  | 13                            | 0.95992       | 0.05893               | 0.01139                                |
| (1.0 , 1.1)  | 13                            | 1.03980       | 0.06516               | 0.01056                                |
| (1.1 , 1.2)  | 7                             | 1.15798       | 0.06457               | 0.00814                                |
| (1.2 , 1.3)  | 1                             | 1.27448       | 0.07567               | 0                                      |
| (1.3 , 1.4)  | 2                             | 1.35285       | 0.07305               | 0.00350                                |
| (1.4 , 1.5)  | 2                             | 1.45724       | 0.07000               | 0.00482                                |
| (1.5 , 1.6)  | 2                             | 1.57550       | 0.06693               | 0.00042                                |
| (1.6 , 1.7)  | 1                             | 1.62090       | 0.06890               | 0                                      |

**Table 24.** The mean value and the variance of the ET instrumental contribution to  $\Delta d_L/d_L$ , averaging over the events in the given redshift bin, for the specific realization of the catalog of joint GW-GRB detections shown in figure 3, for the flat distribution of neutron star masses and the ‘optimistic’ scenario for the FOV of THESEUS.

| redshift bin | number of joint GW-GRB events | mean redshift | mean $\Delta d_L/d_L$ | standard deviation of $\Delta d_L/d_L$ |
|--------------|-------------------------------|---------------|-----------------------|--|
| (0 , 0.1)    | 18                            | 0.07076       | 0.00853               | 0.00342                                |
| (0.1 , 0.2)  | 62                            | 0.15528       | 0.01791               | 0.00659                                |
| (0.2 , 0.3)  | 73                            | 0.24902       | 0.02826               | 0.00877                                |
| (0.3 , 0.4)  | 74                            | 0.35218       | 0.03670               | 0.01168                                |
| (0.4 , 0.5)  | 65                            | 0.44866       | 0.04737               | 0.01509                                |
| (0.5 , 0.6)  | 55                            | 0.54196       | 0.05672               | 0.01726                                |
| (0.6 , 0.7)  | 47                            | 0.64370       | 0.05495               | 0.01398                                |
| (0.7 , 0.8)  | 33                            | 0.74970       | 0.06027               | 0.01339                                |
| (0.8 , 0.9)  | 26                            | 0.84755       | 0.06457               | 0.01007                                |
| (0.9 , 1.0)  | 22                            | 0.94282       | 0.06533               | 0.01079                                |
| (1.0 , 1.1)  | 12                            | 1.05950       | 0.06478               | 0.00645                                |
| (1.1 , 1.2)  | 10                            | 1.14518       | 0.06993               | 0.00508                                |
| (1.2 , 1.3)  | 3                             | 1.27270       | 0.07187               | 0.00153                                |
| (1.3 , 1.4)  | 3                             | 1.34741       | 0.07152               | 0.00292                                |
| (1.4 , 1.5)  | 6                             | 1.46689       | 0.07687               | 0.00418                                |
| (1.5 , 1.6)  | 1                             | 1.58407       | 0.07577               | 0                                      |
| (1.6 , 1.7)  | 1                             | 1.62843       | 0.07947               | 0                                      |

**Table 25.** As in table 24, for the Gaussian distribution of neutron star masses and the ‘optimistic’ scenario for the FOV of THESEUS.

| redshift bin | number of joint GW-GRB events | mean redshift | mean $\Delta d_L/d_L$ | standard deviation of $\Delta d_L/d_L$ |
|--------------|-------------------------------|---------------|-----------------------|--|
| (0 , 0.1)    | 2                             | 0.09183       | 0.00582               | 0.00023                                |
| (0.1 , 0.2)  | 19                            | 0.14579       | 0.01675               | 0.00731                                |
| (0.2 , 0.3)  | 19                            | 0.24404       | 0.02630               | 0.00789                                |
| (0.3 , 0.4)  | 25                            | 0.35527       | 0.03324               | 0.01065                                |
| (0.4 , 0.5)  | 23                            | 0.44039       | 0.04143               | 0.01561                                |
| (0.5 , 0.6)  | 10                            | 0.54527       | 0.05305               | 0.00887                                |
| (0.6 , 0.7)  | 12                            | 0.64379       | 0.05181               | 0.01299                                |
| (0.7 , 0.8)  | 3                             | 0.74185       | 0.05141               | 0.00496                                |
| (0.8 , 0.9)  | 4                             | 0.85262       | 0.06450               | 0.00922                                |
| (0.9 , 1.0)  | 4                             | 0.95465       | 0.05803               | 0.00518                                |
| (1.0 , 1.1)  | 1                             | 1.03492       | 0.06103               | 0                                      |
| (1.1 , 1.2)  | 5                             | 1.16308       | 0.06442               | 0.00667                                |
| (1.2 , 1.3)  | —                             | —             | —                     | —                                      |
| (1.3 , 1.4)  | —                             | —             | —                     | —                                      |
| (1.4 , 1.5)  | 1                             | 1.41882       | 0.07482               | 0                                      |

**Table 26.** As in table 24, for the flat distribution of neutron star masses and the ‘realistic’ scenario for the FOV of THESEUS.

| redshift bin | number of joint GW-GRB events | mean redshift | mean $\Delta d_L/d_L$ | standard deviation of $\Delta d_L/d_L$ |
|--------------|-------------------------------|---------------|-----------------------|--|
| (0 , 0.1)    | 4                             | 0.07108       | 0.00868               | 0.00244                                |
| (0.1 , 0.2)  | 24                            | 0.15001       | 0.01784               | 0.00692                                |
| (0.2 , 0.3)  | 24                            | 0.24043       | 0.02558               | 0.00680                                |
| (0.3 , 0.4)  | 27                            | 0.35355       | 0.03529               | 0.01004                                |
| (0.4 , 0.5)  | 28                            | 0.44966       | 0.04843               | 0.01528                                |
| (0.5 , 0.6)  | 9                             | 0.53785       | 0.05646               | 0.01807                                |
| (0.6 , 0.7)  | 14                            | 0.64540       | 0.05329               | 0.01318                                |
| (0.7 , 0.8)  | 13                            | 0.73793       | 0.05493               | 0.01368                                |
| (0.8 , 0.9)  | 8                             | 0.85497       | 0.06413               | 0.00746                                |
| (0.9 , 1.0)  | 4                             | 0.93702       | 0.06257               | 0.01228                                |
| (1.0 , 1.1)  | 6                             | 1.05334       | 0.06494               | 0.00651                                |
| (1.1 , 1.2)  | 3                             | 1.15162       | 0.06749               | 0.00246                                |
| (1.2 , 1.3)  | 1                             | 1.25943       | 0.07373               | 0                                      |
| (1.3 , 1.4)  | —                             | —             | —                     | —                                      |
| (1.4 , 1.5)  | 2                             | 1.45375       | 0.07851               | 0.00398                                |
| (1.5 , 1.6)  | 1                             | 1.58407       | 0.07577               | 0                                      |
| (1.6 , 1.7)  | 1                             | 1.62843       | 0.07947               | 0                                      |

**Table 27.** As in table 24, for the Gaussian distribution of neutron star masses and the ‘realistic’ scenario for the FOV of THESEUS.

## References

- [1] LIGO SCIENTIFIC, VIRGO collaboration, *Observation of gravitational waves from a binary black hole merger*, *Phys. Rev. Lett.* **116** (2016) 061102 [[arXiv:1602.03837](#)] [[INSPIRE](#)].
- [2] LIGO SCIENTIFIC, VIRGO collaboration, *GW151226: observation of gravitational waves from a 22-solar-mass binary black hole coalescence*, *Phys. Rev. Lett.* **116** (2016) 241103 [[arXiv:1606.04855](#)] [[INSPIRE](#)].
- [3] LIGO SCIENTIFIC, VIRGO collaboration, *GW170104: observation of a 50-solar-mass binary black hole coalescence at redshift 0.2*, *Phys. Rev. Lett.* **118** (2017) 221101 [Erratum *ibid.* **121** (2018) 129901] [[arXiv:1706.01812](#)] [[INSPIRE](#)].
- [4] LIGO SCIENTIFIC, VIRGO collaboration, *GW170608: observation of a 19-solar-mass binary black hole coalescence*, *Astrophys. J.* **851** (2017) L35 [[arXiv:1711.05578](#)] [[INSPIRE](#)].
- [5] LIGO SCIENTIFIC, VIRGO collaboration, *GW170814: a three-detector observation of gravitational waves from a binary black hole coalescence*, *Phys. Rev. Lett.* **119** (2017) 141101 [[arXiv:1709.09660](#)] [[INSPIRE](#)].
- [6] LIGO SCIENTIFIC, VIRGO collaboration, *GWTC-1: a gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs*, [arXiv:1811.12907](#) [[INSPIRE](#)].
- [7] LIGO SCIENTIFIC, VIRGO collaboration, *GW170817: observation of gravitational waves from a binary neutron star inspiral*, *Phys. Rev. Lett.* **119** (2017) 161101 [[arXiv:1710.05832](#)] [[INSPIRE](#)].
- [8] A. Goldstein et al., *An ordinary short gamma-ray burst with extraordinary implications: Fermi-GBM detection of GRB 170817A*, *Astrophys. J.* **848** (2017) L14 [[arXiv:1710.05446](#)] [[INSPIRE](#)].
- [9] V. Savchenko et al., *INTEGRAL detection of the first prompt gamma-ray signal coincident with the gravitational-wave event GW170817*, *Astrophys. J.* **848** (2017) L15 [[arXiv:1710.05449](#)] [[INSPIRE](#)].
- [10] B.P. Abbott et al., *Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A*, *Astrophys. J.* **848** (2017) L13 [[arXiv:1710.05834](#)] [[INSPIRE](#)].
- [11] B.P. Abbott et al., *Multi-messenger observations of a binary neutron star merger*, *Astrophys. J.* **848** (2017) L12 [[arXiv:1710.05833](#)] [[INSPIRE](#)].
- [12] LISA collaboration, *Laser interferometer space antenna*, [arXiv:1702.00786](#) [[INSPIRE](#)].
- [13] M. Punturo et al., *The Einstein telescope: a third-generation gravitational wave observatory*, *Class. Quant. Grav.* **27** (2010) 194002 [[INSPIRE](#)].
- [14] B. Sathyaprakash et al., *Scientific objectives of Einstein telescope*, *Class. Quant. Grav.* **29** (2012) 124013 [Erratum *ibid.* **30** (2013) 079501] [[arXiv:1206.0331](#)] [[INSPIRE](#)].
- [15] S. Dwyer et al., *Gravitational wave detector with cosmological reach*, *Phys. Rev. D* **91** (2015) 082001 [[arXiv:1410.0612](#)] [[INSPIRE](#)].
- [16] B.F. Schutz, *Determining the Hubble constant from gravitational wave observations*, *Nature* **323** (1986) 310 [[INSPIRE](#)].
- [17] D.E. Holz and S.A. Hughes, *Using gravitational-wave standard sirens*, *Astrophys. J.* **629** (2005) 15 [[astro-ph/0504616](#)] [[INSPIRE](#)].
- [18] N. Dalal, D.E. Holz, S.A. Hughes and B. Jain, *Short GRB and binary black hole standard sirens as a probe of dark energy*, *Phys. Rev. D* **74** (2006) 063006 [[astro-ph/0601275](#)] [[INSPIRE](#)].

- [19] 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** (2008) 043512 [[arXiv:0712.0618](#)] [[INSPIRE](#)].
- [20] S. Nissanke et al., *Exploring short gamma-ray bursts as gravitational-wave standard sirens*, *Astrophys. J.* **725** (2010) 496 [[arXiv:0904.1017](#)] [[INSPIRE](#)].
- [21] C. Cutler and D.E. Holz, *Ultra-high precision cosmology from gravitational waves*, *Phys. Rev. D* **80** (2009) 104009 [[arXiv:0906.3752](#)] [[INSPIRE](#)].
- [22] B.S. Sathyaprakash, B.F. Schutz and C. Van Den Broeck, *Cosmography with the Einstein Telescope*, *Class. Quant. Grav.* **27** (2010) 215006 [[arXiv:0906.4151](#)] [[INSPIRE](#)].
- [23] W. Zhao, C. Van Den 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** (2011) 023005 [[arXiv:1009.0206](#)] [[INSPIRE](#)].
- [24] W. Del Pozzo, *Inference of the cosmological parameters from gravitational waves: application to second generation interferometers*, *Phys. Rev. D* **86** (2012) 043011 [[arXiv:1108.1317](#)] [[INSPIRE](#)].
- [25] A. Nishizawa, K. Yagi, A. Taruya and T. Tanaka, *Cosmology with space-based gravitational-wave detectors — dark energy and primordial gravitational waves*, *Phys. Rev. D* **85** (2012) 044047 [[arXiv:1110.2865](#)] [[INSPIRE](#)].
- [26] S.R. Taylor and J.R. Gair, *Cosmology with the lights off: standard sirens in the Einstein Telescope era*, *Phys. Rev. D* **86** (2012) 023502 [[arXiv:1204.6739](#)] [[INSPIRE](#)].
- [27] S. Camera and A. Nishizawa, *Beyond concordance cosmology with magnification of gravitational-wave standard sirens*, *Phys. Rev. Lett.* **110** (2013) 151103 [[arXiv:1303.5446](#)] [[INSPIRE](#)].
- [28] N. Tamanini et al., *Science with the space-based interferometer eLISA. III: probing the expansion of the Universe using gravitational wave standard sirens*, *JCAP* **04** (2016) 002 [[arXiv:1601.07112](#)] [[INSPIRE](#)].
- [29] C. Caprini and N. Tamanini, *Constraining early and interacting dark energy with gravitational wave standard sirens: the potential of the eLISA mission*, *JCAP* **10** (2016) 006 [[arXiv:1607.08755](#)] [[INSPIRE](#)].
- [30] R.-G. Cai and T. Yang, *Estimating cosmological parameters by the simulated data of gravitational waves from the Einstein Telescope*, *Phys. Rev. D* **95** (2017) 044024 [[arXiv:1608.08008](#)] [[INSPIRE](#)].
- [31] W. Del Pozzo, A. Sesana and A. Klein, *Stellar binary black holes in the LISA band: a new class of standard sirens*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 3485 [[arXiv:1703.01300](#)] [[INSPIRE](#)].
- [32] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, *Gravitational-wave luminosity distance in modified gravity theories*, *Phys. Rev. D* **97** (2018) 104066 [[arXiv:1712.08108](#)] [[INSPIRE](#)].
- [33] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, *Modified gravitational-wave propagation and standard sirens*, *Phys. Rev. D* **98** (2018) 023510 [[arXiv:1805.08731](#)] [[INSPIRE](#)].
- [34] J. Mendonça and R. Sturani, *Cosmological model selection from standard siren detections by third generation gravitational wave observatories*, [arXiv:1905.03848](#) [[INSPIRE](#)].
- [35] B.S. Sathyaprakash et al., *Cosmology and the early universe*, [arXiv:1903.09260](#) [[INSPIRE](#)].
- [36] B.S. Sathyaprakash et al., *Multimessenger universe with gravitational waves from binaries*, [arXiv:1903.09277](#) [[INSPIRE](#)].
- [37] B.S. Sathyaprakash et al., *Extreme gravity and fundamental physics*, [arXiv:1903.09221](#) [[INSPIRE](#)].

- [38] LISA COSMOLOGY WORKING GROUP collaboration, *Testing modified gravity at cosmological distances with LISA standard sirens*, *JCAP* **07** (2019) 024 [[arXiv:1906.01593](#)] [[INSPIRE](#)].
- [39] T. Regimbau et al., *Mock data challenge for the Einstein Gravitational-Wave Telescope*, *Phys. Rev. D* **86** (2012) 122001 [[arXiv:1201.3563](#)] [[INSPIRE](#)].
- [40] T. Regimbau, D. Meacher and M. Coughlin, *Second Einstein Telescope mock science challenge: Detection of the gravitational-wave stochastic background from compact binary coalescences*, *Phys. Rev. D* **89** (2014) 084046 [[arXiv:1404.1134](#)] [[INSPIRE](#)].
- [41] T. Regimbau et al., *Revisiting coincidence rate between gravitational wave detection and short gamma-ray burst for the advanced and third generation*, *Astrophys. J.* **799** (2015) 69 [[arXiv:1410.2739](#)] [[INSPIRE](#)].
- [42] D. Meacher et al., *Mock data and science challenge for detecting an astrophysical stochastic gravitational-wave background with Advanced LIGO and Advanced Virgo*, *Phys. Rev. D* **92** (2015) 063002 [[arXiv:1506.06744](#)] [[INSPIRE](#)].
- [43] D. Meacher et al., *Second Einstein Telescope mock data and science challenge: low frequency binary neutron star data analysis*, *Phys. Rev. D* **93** (2016) 024018 [[arXiv:1511.01592](#)] [[INSPIRE](#)].
- [44] T. Regimbau et al., *Digging deeper: observing primordial gravitational waves below the binary black hole produced stochastic background*, *Phys. Rev. Lett.* **118** (2017) 151105 [[arXiv:1611.08943](#)] [[INSPIRE](#)].
- [45] 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** (2018) 064031 [[arXiv:1710.05325](#)] [[INSPIRE](#)].
- [46] THESEUS collaboration, *The THESEUS space mission concept: science case, design and expected performances*, *Adv. Space Res.* **62** (2018) 191 [[arXiv:1710.04638](#)] [[INSPIRE](#)].
- [47] THESEUS collaboration, *THESEUS: a key space mission concept for multi-messenger astrophysics*, *Adv. Space Res.* **62** (2018) 662 [[arXiv:1712.08153](#)] [[INSPIRE](#)].
- [48] G. Stratta, L. Amati, R. Ciolfi and S. Vinciguerra, *THESEUS in the era of multi-messenger astronomy*, *Mem. Soc. Ast. It.* **89** (2018) 205 [[arXiv:1802.01677](#)] [[INSPIRE](#)].
- [49] P.S. Cowperthwaite, V.A. Villar, D.M. Scolnic and E. Berger, *LSST target-of-opportunity observations of gravitational wave events: essential and efficient*, *Astrophys. J.* **874** (2019) 88 [[arXiv:1811.03098](#)] [[INSPIRE](#)].
- [50] C. Deffayet and K. Menou, *Probing gravity with spacetime sirens*, *Astrophys. J.* **668** (2007) L143 [[arXiv:0709.0003](#)] [[INSPIRE](#)].
- [51] N. Yunes, R. O’Shaughnessy, B.J. Owen and S. Alexander, *Testing gravitational parity violation with coincident gravitational waves and short gamma-ray bursts*, *Phys. Rev. D* **82** (2010) 064017 [[arXiv:1005.3310](#)] [[INSPIRE](#)].
- [52] I.D. Saltas, I. Sawicki, L. Amendola and M. Kunz, *Anisotropic stress as a signature of nonstandard propagation of gravitational waves*, *Phys. Rev. Lett.* **113** (2014) 191101 [[arXiv:1406.7139](#)] [[INSPIRE](#)].
- [53] J. Gleyzes, D. Langlois and F. Vernizzi, *A unifying description of dark energy*, *Int. J. Mod. Phys. D* **23** (2015) 1443010 [[arXiv:1411.3712](#)] [[INSPIRE](#)].
- [54] L. Lombriser and A. Taylor, *Breaking a dark degeneracy with gravitational waves*, *JCAP* **03** (2016) 031 [[arXiv:1509.08458](#)] [[INSPIRE](#)].
- [55] A. Nishizawa, *Generalized framework for testing gravity with gravitational-wave propagation. I. Formulation*, *Phys. Rev. D* **97** (2018) 104037 [[arXiv:1710.04825](#)] [[INSPIRE](#)].

- [56] S. Arai and A. Nishizawa, *Generalized framework for testing gravity with gravitational-wave propagation. II. Constraints on Horndeski theory*, *Phys. Rev. D* **97** (2018) 104038 [[arXiv:1711.03776](#)] [[INSPIRE](#)].
- [57] L. Amendola, I. Sawicki, M. Kunz and I.D. Saltas, *Direct detection of gravitational waves can measure the time variation of the Planck mass*, *JCAP* **08** (2018) 030 [[arXiv:1712.08623](#)] [[INSPIRE](#)].
- [58] E.V. Linder, *No slip gravity*, *JCAP* **03** (2018) 005 [[arXiv:1801.01503](#)] [[INSPIRE](#)].
- [59] K. Pardo, M. Fishbach, D.E. Holz and D.N. Spergel, *Limits on the number of spacetime dimensions from GW170817*, *JCAP* **07** (2018) 048 [[arXiv:1801.08160](#)] [[INSPIRE](#)].
- [60] M. Lagos, M. Fishbach, P. Landry and D.E. Holz, *Standard sirens with a running Planck mass*, *Phys. Rev. D* **99** (2019) 083504 [[arXiv:1901.03321](#)] [[INSPIRE](#)].
- [61] A. Nishizawa and S. Arai, *Generalized framework for testing gravity with gravitational-wave propagation. III. Future prospect*, *Phys. Rev. D* **99** (2019) 104038 [[arXiv:1901.08249](#)] [[INSPIRE](#)].
- [62] LIGO SCIENTIFIC, VIRGO collaboration, *GW170817: implications for the stochastic gravitational-wave background from compact binary coalescences*, *Phys. Rev. Lett.* **120** (2018) 091101 [[arXiv:1710.05837](#)] [[INSPIRE](#)].
- [63] E. Vangioni et al., *The impact of star formation and gamma-ray burst rates at high redshift on cosmic chemical evolution and reionization*, *Mon. Not. Roy. Astron. Soc.* **447** (2015) 2575 [[arXiv:1409.2462](#)] [[INSPIRE](#)].
- [64] P. Madau and M. Dickinson, *Cosmic star formation history*, *Ann. Rev. Astron. Astrophys.* **52** (2014) 415 [[arXiv:1403.0007](#)] [[INSPIRE](#)].
- [65] S. Vitale and W.M. Farr, *Measuring the star formation rate with gravitational waves from binary black holes*, [arXiv:1808.00901](#) [[INSPIRE](#)].
- [66] H.-Y. Chen, M. Fishbach and D.E. Holz, *A two per cent Hubble constant measurement from standard sirens within five years*, *Nature* **562** (2018) 545 [[arXiv:1712.06531](#)] [[INSPIRE](#)].
- [67] E. Berger, *Short-duration gamma-ray bursts*, *Ann. Rev. Astron. Astrophys.* **52** (2014) 43 [[arXiv:1311.2603](#)] [[INSPIRE](#)].
- [68] P.A. Evans et al., *Swift and NuSTAR observations of GW170817: detection of a blue kilonova*, *Science* **358** (2017) 1565.
- [69] E.J. Howell, K. Ackley, A. Rowlinson and D. Coward, *Joint gravitational wave — Gamma-ray burst detection rates in the aftermath of GW170817*, [arXiv:1811.09168](#) [[INSPIRE](#)].
- [70] D. Wanderman and T. Piran, *The rate, luminosity function and time delay of non-collapsar short GRBs*, *Mon. Not. Roy. Astron. Soc.* **448** (2015) 3026 [[arXiv:1405.5878](#)] [[INSPIRE](#)].
- [71] E. Burns et al., *Do the Fermi gamma-ray burst monitor and swift burst alert telescope see the same short gamma-ray bursts?*, *Astrophys. J.* **818** (2016) 110 [[arXiv:1512.00923](#)] [[INSPIRE](#)].
- [72] W. Zhao, B.S. Wright and B. Li, *Constraining the time variation of Newton's constant  $G$  with gravitational-wave standard sirens and supernovae*, *JCAP* **10** (2018) 052 [[arXiv:1804.03066](#)] [[INSPIRE](#)].
- [73] PLANCK collaboration, *Planck 2015 results. I. Overview of products and scientific results*, *Astron. Astrophys.* **594** (2016) A1 [[arXiv:1502.01582](#)] [[INSPIRE](#)].
- [74] PLANCK collaboration, *Planck 2015 results. XIV. Dark energy and modified gravity*, *Astron. Astrophys.* **594** (2016) A14 [[arXiv:1502.01590](#)] [[INSPIRE](#)].
- [75] PLANCK collaboration, *Planck 2015 results. XI. CMB power spectra, likelihoods and robustness of parameters*, *Astron. Astrophys.* **594** (2016) A11 [[arXiv:1507.02704](#)] [[INSPIRE](#)].

- [76] PLANCK collaboration, *Planck 2015 results. XV. Gravitational lensing*, *Astron. Astrophys.* **594** (2016) A15 [[arXiv:1502.01591](#)] [[INSPIRE](#)].
- [77] SDSS collaboration, *Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples*, *Astron. Astrophys.* **568** (2014) A22 [[arXiv:1401.4064](#)] [[INSPIRE](#)].
- [78] F. Beutler et al., *The 6dF galaxy survey: baryon acoustic oscillations and the local Hubble constant*, *Mon. Not. Roy. Astron. Soc.* **416** (2011) 3017 [[arXiv:1106.3366](#)].
- [79] A.J. Ross et al., *The clustering of the SDSS DR7 main Galaxy sample — I. A 4 per cent distance measure at  $z = 0.15$* , *Mon. Not. Roy. Astron. Soc.* **449** (2015) 835 [[arXiv:1409.3242](#)] [[INSPIRE](#)].
- [80] BOSS collaboration, *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples*, *Mon. Not. Roy. Astron. Soc.* **441** (2014) 24 [[arXiv:1312.4877](#)] [[INSPIRE](#)].
- [81] D. Blas, J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) II: approximation schemes*, *JCAP* **07** (2011) 034 [[arXiv:1104.2933](#)] [[INSPIRE](#)].
- [82] PLANCK collaboration, *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [[arXiv:1502.01589](#)] [[INSPIRE](#)].
- [83] S.M. Feeney et al., *Prospects for resolving the Hubble constant tension with standard sirens*, *Phys. Rev. Lett.* **122** (2019) 061105 [[arXiv:1802.03404](#)] [[INSPIRE](#)].
- [84] A.G. Riess, S. Casertano, W. Yuan, L.M. Macri and D. Scolnic, *Large Magellanic Cloud Cepheid standards provide a 1% foundation for the determination of the Hubble constant and stronger evidence for physics beyond  $\Lambda$ CDM*, *Astrophys. J.* **876** (2019) 85 [[arXiv:1903.07603](#)] [[INSPIRE](#)].
- [85] M. Chevallier and D. Polarski, *Accelerating universes with scaling dark matter*, *Int. J. Mod. Phys. D* **10** (2001) 213 [[gr-qc/0009008](#)] [[INSPIRE](#)].
- [86] E.V. Linder, *Exploring the expansion history of the universe*, *Phys. Rev. Lett.* **90** (2003) 091301 [[astro-ph/0208512](#)] [[INSPIRE](#)].
- [87] C. Dalang and L. Lombriser, *Limitations on standard sirens tests of gravity from screening*, [arXiv:1906.12333](#) [[INSPIRE](#)].
- [88] E. Belgacem, Y. Dirian, S. Foffa and M. Maggiore, *Nonlocal gravity. Conceptual aspects and cosmological predictions*, *JCAP* **03** (2018) 002 [[arXiv:1712.07066](#)] [[INSPIRE](#)].
- [89] M. Maggiore, *Gravitational Waves. Vol. 2. Astrophysics and Cosmology*, Oxford University Press, Oxford U.K. (2018).
- [90] M. Maggiore, *Phantom dark energy from nonlocal infrared modifications of general relativity*, *Phys. Rev. D* **89** (2014) 043008 [[arXiv:1307.3898](#)] [[INSPIRE](#)].
- [91] M. Maggiore and M. Mancarella, *Nonlocal gravity and dark energy*, *Phys. Rev. D* **90** (2014) 023005 [[arXiv:1402.0448](#)] [[INSPIRE](#)].
- [92] S. Foffa, M. Maggiore and E. Mitsou, *Cosmological dynamics and dark energy from nonlocal infrared modifications of gravity*, *Int. J. Mod. Phys. A* **29** (2014) 1450116 [[arXiv:1311.3435](#)] [[INSPIRE](#)].
- [93] Y. Dirian et al., *Cosmological perturbations and structure formation in nonlocal infrared modifications of general relativity*, *JCAP* **06** (2014) 033 [[arXiv:1403.6068](#)] [[INSPIRE](#)].
- [94] Y. Dirian et al., *Non-local gravity and comparison with observational datasets*, *JCAP* **04** (2015) 044 [[arXiv:1411.7692](#)] [[INSPIRE](#)].

- [95] Y. Dirian et al., *Non-local gravity and comparison with observational datasets. II. Updated results and Bayesian model comparison with  $\Lambda$ CDM*, *JCAP* **05** (2016) 068 [[arXiv:1602.03558](#)] [[INSPIRE](#)].
- [96] Y. Dirian, *Changing the Bayesian prior: absolute neutrino mass constraints in nonlocal gravity*, *Phys. Rev. D* **96** (2017) 083513 [[arXiv:1704.04075](#)] [[INSPIRE](#)].
- [97] E. Belgacem, A. Finke, A. Frassino and M. Maggiore, *Testing nonlocal gravity with Lunar Laser Ranging*, *JCAP* **02** (2019) 035 [[arXiv:1812.11181](#)] [[INSPIRE](#)].
- [98] M. Maggiore, *Nonlocal infrared modifications of gravity. A review*, *Fundam. Theor. Phys.* **187** (2017) 221 [[arXiv:1606.08784](#)].
- [99] E. Belgacem et al., *Nonlocal gravity and gravitational-wave observations*, [arXiv:1907.02047](#) [[INSPIRE](#)].
- [100] A. Kehagias and M. Maggiore, *Spherically symmetric static solutions in a nonlocal infrared modification of general relativity*, *JHEP* **08** (2014) 029 [[arXiv:1401.8289](#)] [[INSPIRE](#)].