Cosmological model selection from standard siren detections by third generation gravitational wave observatories

Josiel Mendonça¹ and Riccardo Sturani²

¹Departamento de Física Teórica e Experimental, Universidade Federal do Rio Grande do Norte, Natal-RN 59072-970, Brazil
²International Institute of Physics, Universidade Federal do Rio Grande do Norte, Campus Universitario, Lagoa Nova, Natal-RN 59078-970, Brazil

The multi-messenger observation of GW170817 enabled the first historic measure of the Hubble-Lemaître constant via a standard siren, so called in analogy to standard candles that enabled the measure of the luminosity distance versus redshift relationship at small redshift. In the next decades third generation observatories are expected to detect hundreds to thousand gravitational wave events from compact binary coalescences with potentially a joint electromagnetic counterpart. In the present work we show how future standard siren detections can be used within the framework of Bayesian model selection to discriminate between cosmological models differing by the parametrization of the late time acceleration. In particular we show that the standard LambdaCDM model can be confirmed or ruled out in case the uncertainty in the gravitational determination of the luminosity distance is reduced with respect to current expectations, e.g. by combining detections from multiple detectors.

I. INTRODUCTION

Coalescing binary systems in which at least one of the component is a neutron star have long been considered the most likely candidate for a simultaneous source of gravitational and electromagnetic radiation and indeed GW170817, GRB170817A and SSS17a/AT 2017gfo [1,2] represented the first historic multi-messenger detection involving gravitational waves and it was originated by the coalescence of two neutron stars. As first suggested in [3], see also [4], such coincidence detection and the subsequent localization of the host galaxy can be used to determine the Hubble-Lemaître constant $H_0$ by short-circuiting the luminosity distance estimated via the gravitational channel and the redshift measured electromagnetically.

The characteristic chirping signal of coalescing binary in the gravitational enables an accurate determination of the intrinsic parameters of the source (intrinsic luminosity) which combined with measured signal amplitude (apparent luminosity) enable a determination of the luminosity distance, hence the name of standard sirens for coalescing binaries, in analogy to standard candles, the supernovae type Ia that enabled the measure of the Hubble-Lemaître constant and gave convincing evidence of the late time cosmological acceleration, see [5] for recent observations.

The source redshift affects the gravitational wave signal degenerately with the binary constituent masses, making impossible its determination from the gravitational signal alone. However this degeneracy is not perfect and different approaches have been tried to obtain a luminosity distance versus redshift relationship with [6] and without electromagnetic counterparts, see [7].

Third generation detectors like Einstein Telescope (ET) [8] and Cosmic Explorer [9] are planned earth-based gravitational wave detectors building on and improving the technology and sensitivity of currently operating detectors Advanced LIGO [10] and advanced Virgo [11] to reach sources up to redshift $z \sim$ few. Given present rate estimates of binary neutron star coalescences of of $O(10^3)\text{Gpc}^{-3}\text{year}^{-1}$ detection rates of $O(10^3)$ events per year or larger are expected.

With such a plethora of future data we want to test the late time dynamics of the Universe via the redshift versus luminosity distance relationship. At the moment the standard cosmological model is $\Lambda$CDM accommodates the observed late time acceleration ($z \lesssim 5$) as well as other observations at larger redshift, however with a small but significant tension in the value of the Hubble-Lemaître constant, which is estimated to be $73.48 \pm 1.66$ and $67.66 \pm 0.42\text{km s}^{-1}\text{Mpc}^{-1}$ respectively by standard candles [12] and by Cosmic Microwave Background -based [13] measures. See also [14] for an alternative determination of $H_0$.

The dark energy component required to explain current cosmological acceleration is parametrized in the $\Lambda$CDM model by a bare cosmological constant: a perfect fluid with negative pressure equal in modulus to its energy density. However other phenomenological parametrizations are possible, involving an equal or larger number of parameters than in the $\Lambda$CDM case, and the goal of the present paper is to test if standard sirens detected by third generation gravitational detectors can discriminate among different dark energy models, with possibly different number of parameters, giving a new handle to to solve the long-standing puzzle of what is the origin of the late time cosmological acceleration.

We underline that the focus of the present work is not on parameter determination with standard sirens, that
has been pursued in several recent publications, see e.g. \[15\text{–}19\] but rather model discrimination, see \[20\] for a model independent attempt to reconstruct the distance versus redshift relationship via Gaussian process methods with simulated LISA data.

In the present work we use Bayesian model selection framework to rank one model versus another, hence we compute the evidence of each model to rank them.

The outline of the paper is as follows: in sec. II we detail the method used to simulate data and rank models, in sec. III the results of Bayesian evidence computation are reported and finally we conclude in sec. IV.

II. METHOD

The Hubble law relating redshift \(z\) and luminosity distance \(d_L\) via the Hubble-Lemaître constant \(H_0\)

\[
d_L H_0 = z + a(z),
\]

(1)

can be interpreted within the standard cosmological model based on General Relativity as the first order expansion of a more general relationship between \(z\) and \(d_L\):

\[
d_L = \frac{1 + z}{H_0} \int_0^z \frac{dz'}{E(z')}
\]

(2)

where the inverse of the integrand

\[
E(z) \equiv \sqrt{\Omega_M (1 + z)^3 + \Omega_R (1 + z)^4 + \Omega_{DE}}
\]

(3)
is expressed in terms of the normalized present energy densities \(\Omega_X\) in generic species \(X\)

\[
\Omega_X \equiv \frac{8\pi G N}{3H_0^2} \rho_0X,
\]

(4)

with \(\sum X \Omega_X = 1\) by the time-time component of the Einstein’s equation. The equation of state relating pressure \(p_X\) and energy density \(\rho_X\) of each species is assumed \(p_X = w_X \rho_X\), implying \(\Omega_X \propto a^{-3(1+w_X)}\) (when no inter-species interactions are present) and we have assumed that the only species present in the Universe are non-relativistic matter \(w_m = 0\), radiation \(w_r = 1/3\) and the cosmological constant \(w_\Lambda = -1\).

Beside \(\Lambda\)CDM, we will use three additional parametrizations of the dark energy in the rest of the paper

- Model \(w\)CDM, with dark energy free parameter \(w_{DE} = p_{DE}/\rho_{DE}\) constant in time.
- Model \(w_0w_a\)CDM, with dark energy \(w_{DE} = p_{DE}/\rho_{DE} = w_0 + w_a z/(1 + z)\), with both \(w_0\) and \(w_a\) constant free parameters, as suggested in \[21\text{–}22\].
- The non-local massive gravity model, henceforth mass\(G\), described in \[23\], whose modified dynamics results in an identical luminosity distance for electromagnetic waves as in General Relativity and in a modified one for gravitational waves \(d_{L}^{mG-gw}\) which can be phenomenologically parametrized as

\[
\frac{d_{L}^{mG-gw}(z)}{d_{L}^{m}} = \Xi_0 + \frac{1 - \Xi_0}{(1 + z)^n},
\]

(5)

with \(n = 5/2\) and \(\Xi_0 = 0.97\ \[24\].

The \(\Lambda\)CDM, \(w\)CDM, \(w_0w_a\)CDM have the feature of being nested, i.e. one can go from the more complex to the simplest by fixing one or more parameters to specific values. On the other hand the non-local mass\(G\) model gives a different description of late time Universe dynamics, still consistent with the data. Since the fundamental origin of the cosmic acceleration is presently unknown, it may well be that \(\Lambda\)CDM or its \(w\)-variants will not be able to match at all redshifts the dynamics resulting from the fundamental cosmological theory, hence we find useful to have a different toy model to extract our simulated data from, to verify how different nested parametrizations perform on data from a model none of them cannot match exactly, see \[24\] for the effectiveness of standard sirens in discriminating between \(\Lambda\)CDM and the mass\(G\) model.
A. Nested Models treatment: toy example

In the case of nested models, the more general model always gives a better fit by construction, but it can be disfavored as dictated by the *Ocamm razor* in case it uses unnecessary extra parameters. Let us see how nested model selection works in a simplified example that be fully treated analytically \cite{25}. The simplest case of 2 nested models consists of model $\mathcal{M}_0$ having no free parameter and $\mathcal{M}_1$ having one free parameter, say $\theta$, with $\mathcal{M}_1$ reducing to $\mathcal{M}_0$ for $\theta \to 0$.

If $\mathcal{M}_0$ describes the distribution of a Gaussian variable $x$ centered in 0, then experimental measures $\{x_1, \ldots, x_n\}$ of mean $\lambda$ and standard deviation $\sigma$ should result in a probability distribution for $\lambda$ given by

$$P(\lambda|d, \mathcal{M}_0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\lambda^2/(2\sigma^2)}.$$ 

$\mathcal{M}_1$ on the other hand predicts that the measure of $x$ should be described by a Gaussian centered in $\theta$, with $\theta$ a free parameter, to which an a priory knowledge could be applied: e.g. we assume it is distributed as a Gaussian with standard deviation $\Sigma$. $\mathcal{M}_1$ then predicts that the experimental measures enable the determination of a probability distribution for $\lambda$ follow a $\theta$-dependent Gaussian distribution:

$$P(\lambda|\theta, \mathcal{M}_1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\lambda-\theta)^2/(2\sigma^2)}.$$ 

According to standard Bayesian inference the probability distribution of $\theta$ is given by

$$p(\theta|d, \mathcal{M}_1) = p(d|\theta, \mathcal{M}_1) \frac{p(\theta)}{p(d)} \propto p(d|\theta, \mathcal{M}_1)p(\theta),$$

where $p(d|\theta, \mathcal{M}_1)$ is the likelihood of the data given parameter $\theta$ and model $\mathcal{M}_1$, $p(\theta)$ is the prior on $\theta$ and in the last passage we have dropped $p(d)$, which is uninteresting for $p(\theta|d, \mathcal{M}_1)$ since it does not depend on $\theta$ and thus can be absorbed in the normalization factor of $p(\theta|d, \mathcal{M}_1)$.

In this work we are interested in model comparison rather than parameter estimation, hence we want to compare $p(d|\mathcal{M}_0)$ to $p(d|\mathcal{M}_1)$ irrespectively of the parameter values, i.e. we want to know if data favor model $\mathcal{M}_0$ or $\mathcal{M}_1$. This question can be addressed quantitatively by considering the ratio of the evidences $Z_{0,1}$

$$Z_0 = \frac{p(d|\mathcal{M}_0)}{p(d|\mathcal{M}_1)} = \frac{p(d|\mathcal{M}_0)}{\int d\theta p(\theta)p(d|\theta, \mathcal{M}_1)} = \frac{(2\pi\sigma^2)^{-1/2} e^{-\lambda^2/(2\sigma^2)}}{(2\pi\Sigma^2)^{-1/2} (2\Sigma \sigma^2)^{-1/2} \int d\theta e^{-\theta^2/(2\Sigma^2)} e^{-(\lambda-\theta)^2/(2\sigma^2)}};$$

where as mentioned earlier we assumed a Gaussian prior on $\theta$

$$p(\theta) = (2\pi\Sigma^2)^{-1/2} e^{-\theta^2/(2\Sigma^2)}.$$ 

After performing the integration in $\theta$ one gets

$$Z_0 = e^{-\lambda^2/(2\sigma^2(1+\sigma^2/\Sigma^2))} \sqrt{1 + \frac{\Sigma^2}{\sigma^2}},$$

showing that for $\lambda / \sigma \ll 1$ $\mathcal{M}_1$ is disfavored for $\Sigma > \sigma$ and the models have similar evidences for $\Sigma \approx \sigma$.

On the other hand for $\lambda / \sigma \gg 1$ (and $\Sigma > \sigma$) $\mathcal{M}_1$ quickly gains over $\mathcal{M}_0$ despite the $\theta$ prior may disfavor large values of $\theta$. Note that $\mathcal{M}_1$ having more parameters and including $\mathcal{M}_0$ as a particular case will always give a better fit to the data, but will not necessarily have better evidence.

B. Merger rates and uncertainties

Following \cite{26} we consider events for third generation gravitational wave detectors up to redshift $z \sim 2$. Actually gravitationally signals could be seen up to much $z \sim 10$ \cite{27}, but we focus on a smaller range of redshift as electromagnetic counterpart detections necessary for redshift determination would be too difficult to observe from such large distances.

For the distribution of merger events we take the *star formation rate* per unit of comoving volume and redshift $\psi(z)$ computed in \cite{28}

$$\psi(z) \propto \frac{(1+z)^{2.7}}{1 + (\frac{1+z}{2.9})^{5.8}},$$
allowing a delay between star formation and merger which is distributed following a Poisson law with characteristic delay $\tau$. Following [29] one can find that the differential rate of mergers $R_m$ happening at redshift $z_m$ is given by

$$R_m(z) = \frac{dN_m}{dt_o dz} = \frac{dV_c}{dz} R_m(z_m) \frac{1}{1 + z_m},$$

(11)

in terms of merger rate $R_m$ per unit of comoving volume $V_c$ and redshift $z$, and we have introduced the observer time $t_o$ related to source time $t_s$ by $dt_o/dt_s = (1 + z)$ and denoted with $N_m$ the number of mergers. $R_m$ can in turn be modeled by convolving $\psi(z)$ with a stochastic delay between star formation epoch characterized by redshift $z_{sf}$ and binary merger, which we assume for simplicity to be Poisson distributed, leading to

$$R_m(z_m) = \frac{1}{\tau} \int_{z_{sf}}^\infty dz_{sf} \frac{dt}{dz_{sf}} \psi(z_{sf}) \exp \left[ -t(z_{sf}) - t(z_m) / \tau \right].$$

(12)

Fig. 1 shows the resulting merger rate for different delays compared with a histogram of the redshifts of the supernova data [30]. Clearly gravitational waves give access to larger values of the redshift even in the extreme case of large delay (10 Gyr) between star formation and binary mergers.

We assume that redshifts can be determined with negligible uncertainty in the presence of an electromagnetic counterpart (see [7] for results of implementing a measure of the Hubble-Lemaître constant $H_0$ without electromagnetic counterpart) and the uncertainty in $dL$ has usually two main contributions: an instrumental one intrinsic to GW observatories here denoted as $(\Delta d_L)_{\text{inst}}$ [26], and another one $(\Delta d_L)_{\text{lens}}$ due to lensing, see e.g. [31, 32]

$$\frac{\Delta d_L(z)}{d_L(z)} = \left[ \frac{\Delta d_L(z)}{d_L(z)} \right]_{\text{inst}}^2 + \left[ \frac{\Delta d_L(z)}{d_L(z)} \right]_{\text{lens}}^2)^{1/2},$$

(13)

with

$$\left( \frac{\Delta d_L(z)}{d_L(z)} \right)_{\text{inst}} \approx 0.1449 z - 0.0118 z^2 + 0.0012 z^3,$n

(14)

$$\left( \frac{\Delta d_L(z)}{d_L(z)} \right)_{\text{lens}} \approx 0.066 \left[ 4 \left( 1 - (1 + z)^{-1/4} \right) \right]^{1.8},$$

which are shown in fig. 2.
FIG. 2: Uncertainty budget in the determination of the luminosity distance by third generation detector like ET [26, 32], see eq. (14). Instrumental effects give the leading uncertainty for all redshift of interest.

In the next section we consider the possibility that estimate uncertainty in the distance determination can be reduced with respect to this formula by correlating detection from multiple detectors [27] and by instrument improvement, which can act on the instrumental part of the uncertainty budget, which is the dominant one.

III. RESULTS

This section reports the result obtained by simulating 1000 detections of electromagnetic bright coalescing binaries up to redshift \( z = 2 \) with realistic distributions of events, and comparing phenomenological models for late time cosmological acceleration within Bayesian model selection framework. We use two different sets of data for simulating the standard sirens luminosity distance and redshift:

1. the standard \( \Lambda \)CDM which we also use at recovery, with parameters \( \Omega_m = 0.3111, \ H_0 = 67.66 \text{km/s/Mpc}, \ \Omega_L = 1 - \Omega_m \) taken from [13],

2. the non-local massive gravity (massG) model [23], useful as testing ground for different models at recovery, none of which include the massG model used in this second set of injections. Note that since the background evolution in this model is different than any of the \( \Lambda \)CDM and \( \omega \)CDM, the best fit background parameter value are slightly different than in the previous case: \( \Omega_m = 0.2989, \ H_0 = 69.49 \text{km/s/Mpc} \) [34].

Each of the two sets of simulated data (\( \Lambda \)CDM and massG) is produced for two different distributions of merger events:

1. one following the star formation rate proposed in [28],

2. the other allowing a Poisson distribution of delays between star formation and binary merger with average delay \( \tau = 10 \) Gyr,

resulting in four different type of injections. Fig. 3 displays explicitly one of the four type of injections we use, and in fig. 4 the cumulative distributions of the \( 10^3 \) injections are reported for the 4 cases, showing little difference between the \( \Lambda \)CDM and massG case, but a notable difference if a 10 Gyr delay between merger and star formation is allowed.

We use these four type of simulated data for comparing \( \Lambda \)CDM versus \( \omega \)CDM and \( \Lambda \)CDM versus \( \omega \)CDM, with the results for model comparisons displayed in respectively figs. 5-6 for \( \Lambda \)CDM injections, and in figs. 7-8 for massG

---

1 A direct measure of the mass distribution on the line of sight, which may be available by the time GW data are collected, will reduce the lensing part of the error budget [33].
FIG. 3: One thousand events simulated according to the ΛCDM model for a merger rate following the star formation rate, with 1σ error band given by eq. (14) [26, 33]. For reference the luminosity distance versus redshift curve for ΛCDM and massG models are also shown, for best fit parameters respectively given in [13] and [34].

FIG. 4: Cumulative distribution of injections for ΛCDM and the non-local massive gravity model [23] for the case of no delay between star formation and binary merger (solid curves) and assuming a Poisson distribution of delay between formation and merger given by eq. (12) with τ = 10 Gyr (dashed curves). Injections. For each graph 50 lines (in grey) are shown, corresponding to different ordering of the 50 catalogs, with the blue thick line giving their average.

Results are obtained via the Nestle implementation [35] of the nested sampling algorithm [36] using 200 live points. Priors for parameters has all be chosen flat for simplicity and they are as follows: \( H_0 \) : [60, 80] km/s/Mpc, \( \Omega_m \) : [0.2, 0.4], \( w_0 \) : [-2, 0], \( w_a \) : [-1, 1].

Fig. 5 shows the evidence comparison for ΛCDM versus \( \omega \)-CDM, with simulated data taken from a distribution of merger events following the star formation rate of eq. (10) [28] or the same star formation rate convoluted with a delay with a Poisson distribution with average \( \tau = 10 \text{ Gyr} \), with luminosity distance uncertainties as in eq. (13) and with uncertainties reduced to 20% that value.
FIG. 5: Evidence for $\Lambda$CDM versus $w_0\Lambda$CDM model with $10^3$ events (divided in 50 catalogs of 20 events each) simulated according to the $\Lambda$CDM model with standard uncertainty in the luminosity distance given by eq. (14) (top line) and uncertainty reduced to 20% (bottom line), with merger rate equal to star formation rate (left column) and a Poisson distribution with a $\tau = 10$ Gyr delay in eq. (12) between formation and merger (right column). Grey lines refers to 50 different ordering of the same 50 catalogs, the blue tick line is their average. Moderate and strong evidence reference lines ($\ln Z_{\Lambda\text{CDM}}/Z_{w\Lambda\text{CDM}} = \pm 2.5, \pm 5$) according to Jeffreys’ scale [37] are also shown for reference.

All evidence ratios mildly but correctly favor $\Lambda$CDM vs. $w\Lambda$CDM, as expected since injections follow the $\Lambda$CDM model, and also unsurprisingly reducing observational uncertainty strengthen the discriminating power of the Bayes model selection test, even if the logarithm of evidence ratio never crosses even the moderate evidence line $\ln Z_{\Lambda\text{CDM}}/Z_{w\Lambda\text{CDM}} = 2.5$, see Jeffreys’ scale [37].

Analogously in fig. 6 $\Lambda$CDM is ranked against $w_0w_a\Lambda$CDM, i.e. against a model with two extra parameters that contains $\Lambda$CDM for $w_0 = -1$, $w_a = 0$ for two sets of $d_L$ uncertainties and two sets of signal distributions as above, again with the result of the $\Lambda$CDM model being favored. Here in particular we see that the strong evidence threshold $\ln Z_{\Lambda\text{CDM}}/Z_{w_0w_a\Lambda\text{CDM}} = 5$ is crossed only in the case of reduced (20%) error and when most the detections are at sufficiently low redshifts ($\tau = 10$ Gyr), both conditions tending to minimize the dilution of results into large observational error. We then note that when the event distribution is shifted to lower redshift the model discriminating power is also stronger, due to the smaller luminosity distance observational error at smaller redshift, making event distribution also play an important role.

In figs. 7, 8 the same exercise is replayed, with the difference of using injections belonging to the non-local massive gravity model [23], hence not being described by any of the nested models used at recovery stage, which are again $\Lambda$CDM and $w\Lambda$CDM in fig. 7 and $\Lambda$CDM and $w_0w_a\Lambda$CDM in 8. In this case in which none of the recovery models coincide with the one used for injections, irrespectively of the assumptions on the error on the luminosity distance and event distribution, the outcome of evidence comparison is inconclusive, showing that future inconclusive results in comparing models may be due to the lack of appropriate parametrization to describe data.
Motivated by the advent of gravitational wave astronomy and by the first measure of the Hubble-Lemaître constant via standard sirens, we have performed a numerical exercise simulating future measures of the luminosity distance versus redshift relationship via combined detections of gravitational and electromagnetic waves to test their power in discriminating among cosmological models of late time acceleration.

There are two main conclusions that we can draw from our simulations. The first is that the error in luminosity distance as estimated in eq. (13) should be decreased to enable model comparison with future gravitational wave detectors. This should be possible by correlating the output of several observatories [27]: in particular for the case of Einstein Telescope combining several detectors can lead to substantial improvement (a factor of few) in the luminosity distance error with respect to single detector.

The second main conclusions is that the intrinsic event distribution also plays an important role: distributing events at smaller distances concentrate events where observational error are smaller and hence convey more information, even if different models tend all to reproduce the same dynamics for small redshifts \((z \lesssim 1)\) and disagree more at larger redshifts \((z \gtrsim 2)\).

Finally we underline that in case the dynamics underlying observations is only approximately described by models used to analyze data, different models may have comparable performance on data, leaving open the search for a better model able to catch the right physics conveyed by the observations.

FIG. 6: Analog to fig. [5] for \(\Lambda\)CDM versus \(w_0w_a\)CDM model: simulated \(10^3\) events following \(\Lambda\)CDM model divided in 50 catalogs, standard \(d_L\) uncertainty (top line) and reduced uncertainty, 20\% of eq. (14) (bottom line). Left column refers to equal merger and star formation rate, right column to a Poisson distribution delay with \(\tau = 10\) Gyr delay in eq. (12). Grey lines refers to 50 different ordering of the same 50 catalogs, the blue tick line is their average. Moderate and strong evidence reference lines (\(\ln Z_1/Z_2 = \pm 2.5, \pm 5\) according to Jeffreys’ scale) are also shown for reference.
FIG. 7: Analog to fig. 6 for simulated data following non-local massive gravity cosmology \cite{23}: evidence for ΛCDM versus wCDM models with $10^3$ events, standard $d_L$ uncertainty eq. (14) (top line) and uncertainty reduced to 20% (bottom line). Left column refers to equal merger and star formation rate, right column to a Poisson distribution delay with $\tau = 10$ Gyr delay, see eq. (12). Grey lines refers to 50 different ordering of the same 50 catalogs, the blue tick line is their average. Moderate and strong evidence reference lines ($\ln Z_1/Z_2 = \pm 2.5, \pm 5$ according to Jeffreys’ scale) are also shown for reference.

Acknowledgments

The authors thank Luciano Casarini, Jailson Alcaniz and Rodrigo de Hollanda for useful discussions. This work is based on the thesis prepared by JMSdS to obtain the master degree in physics at the Federal University of Rio Grande de Norte in Natal (Brazil).

\begin{thebibliography}{99}
\bibitem{1} B.P. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. \textit{Phys. Rev. Lett.}, 119(16):161101, 2017. doi: 10.1103/PhysRevLett.119.161101.
\bibitem{2} B. P. Abbott et al. Multi-messenger Observations of a Binary Neutron Star Merger. \textit{Astrophys. J.}, 848(2):L12, 2017. doi: 10.3847/2041-8213/aa91c9.
\bibitem{3} Bernard F. Schutz. Determining the Hubble Constant from Gravitational Wave Observations. \textit{Nature}, 323:310–311, 1986. doi: 10.1038/323310a0.
\bibitem{4} Daniel E. Holz and Scott A. Hughes. Using gravitational-wave standard sirens. \textit{Astrophys. J.}, 629:15–22, 2005. doi: 10.1086/431341.
\bibitem{5} Adam G. Riess et al. A 2.4% Determination of the Local Value of the Hubble Constant. \textit{Astrophys. J.}, 826(1):56, 2016. doi: 10.3847/0004-637X/826/1/56.
\bibitem{6} C. Messenger and J. Read. Measuring a cosmological distance-redshift relationship using only gravitational wave observations of binary neutron star coalescences. \textit{Phys. Rev. Lett.}, 108:091101, 2012. doi: 10.1103/PhysRevLett.108.091101.
\end{thebibliography}
FIG. 8: Same as in fig. 5 for simulated data following massive gravity cosmology [23]: evidence for $\Lambda$CDM versus $w_0w_a$CDM models with $10^3$ events. Standard deviation uncertainty on the top line and reduced uncertainty in the bottom line. Left column refers to equal merger and star formation rate, right column to a Poisson distribution delay with $\tau = 10$ Gyr delay. Grey lines refers to 50 different ordering of the same 50 catalogs, the blue tick line is their average. Moderate and strong evidence reference lines ($\ln Z_1/Z_2 = \pm 2.5, \pm 5$ according to Jeffreys’ scale) are also shown for reference.

[7] Walter Del Pozzo. Inference of the cosmological parameters from gravitational waves: application to second generation interferometers. *Phys. Rev.*, D86:043011, 2012. doi: 10.1103/PhysRevD.86.043011.
[8] M. Punturo et al. The Einstein Telescope: A third-generation gravitational wave observatory. *Class. Quant. Grav.*, 27: 194002, 2010. doi: 10.1088/0264-9381/27/19/194002.
[9] Benjamin P. Abbott et al. Exploring the Sensitivity of Next Generation Gravitational Wave Detectors. *Class. Quant. Grav.*, 34(4):044001, 2017. doi: 10.1088/1361-6382/aa51f4.
[10] Gregory M. Harry. Advanced LIGO: The next generation of gravitational wave detectors. *Class. Quant. Grav.*, 27:084006, 2010. doi: 10.1088/0264-9381/27/8/084006.
[11] F. Acernese et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.*, 32(2):024001, 2015. doi: 10.1088/0264-9381/32/2/024001.
[12] A. G. Riess, S. Casertano, W. Yuan, L. Macri, J. Anderson, J. W. MacKenty, J. B. Bowers, K. I. Clubb, A. V. Filippenko, D. O. Jones, and B. E. Tucker. New parallaxes of galactic cepheids from spatially scanning the hubble space telescope: Implications for the hubble constant. *Astrophys. J.*, 136:855, 2018.
[13] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. 2018.
[14] S. Birrer et al. H0LiCOW - IX. Cosmographic analysis of the doubly imaged quasar SDSS 1206+4332 and a new measurement of the Hubble constant. *Mon. Not. Roy. Astron. Soc.*, 484:4726, 2019. doi: 10.1093/mnras/stz200.
[15] Neal Dalal, Daniel E. Holz, Scott A. Hughes, and Bhuvnesh Jain. Short grb and binary black hole standard sirens as a probe of dark energy. *Phys. Rev.*, D74:063006, 2006. doi: 10.1103/PhysRevD.74.063006.
[16] B. S. Sathyaprakash, B. F. Schutz, and C. Van Den Broeck. Cosmography with the Einstein Telescope. *Class. Quant. Grav.*, 27:215006, 2010. doi: 10.1088/0264-9381/27/21/215006.
[17] Rong-Gen Cai and Tao Yang. Estimating cosmological parameters by the simulated data of gravitational waves from the Einstein Telescope. *Phys. Rev.*, D95(4):044024, 2017. doi: 10.1103/PhysRevD.95.044024.
[18] Xuan-Neng Zhang, Ling-Feng Wang, Jing-Fei Zhang, and Xin Zhang. Improving cosmological parameter estimation with the future gravitational-wave standard siren observation from the Einstein Telescope. *Phys. Rev.*, D99(6):063510, 2019. doi: 10.1103/PhysRevD.99.063510.

[19] Hsin-Yu Chen, Maya Fishbach, and Daniel E. Holz. A two per cent Hubble constant measurement from standard sirens within five years. *Nature*, 562(7728):545–547, 2018. doi: 10.1038/s41586-018-0606-0.

[20] Rong-Gen Cai, Nicola Tamanini, and Tao Yang. Reconstructing the dark sector interaction with LISA. *JCAP*, 1705(05):031, 2017. doi: 10.1088/1475-7516/2017/05/031.

[21] Michel Chevallier and David Polarski. Accelerating universes with scaling dark matter. *Int. J. Mod. Phys.*, D10:213–224, 2001. doi: 10.1142/S021827180000822.

[22] Eric V. Linder. Exploring the expansion history of the universe. *Phys. Rev. Lett.*, 90:091301, 2003. doi: 10.1103/PhysRevLett.90.091301.

[23] Enis Belgacem, Yves Dirian, Stefano Foffa, and Michele Maggiore. Nonlocal gravity. Conceptual aspects and cosmological predictions. *JCAP*, 1803(03):002, 2018. doi: 10.1088/1475-7516/2018/03/002.

[24] Enis Belgacem, Yves Dirian, Stefano Foffa, and Michele Maggiore. Modified gravitational-wave propagation and standard sirens. *Phys. Rev.*, D98(2):023510, 2018. doi: 10.1103/PhysRevD.98.023510.

[25] Roberto Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.*, 49:71–104, 2008. doi: 10.1080/00107510802066753.

[26] 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 SN Ia Observations. *Phys. Rev.*, D83:023005, 2011. doi: 10.1103/PhysRevD.83.023005.

[27] Salvatore Vitale and Matthew Evans. Parameter estimation for binary black holes with networks of third generation gravitational-wave detectors. *Phys. Rev.*, D95(6):064052, 2017. doi: 10.1103/PhysRevD.95.064052.

[28] Piero Madau and Mark Dickinson. Cosmic Star Formation History. *Ann. Rev. Astron. Astrophys.*, 52:415–486, 2014. doi: 10.1146/annurev-astro-081811-125615.

[29] Salvatore Vitale and Will M. Farr. Measuring the star formation rate with gravitational waves from binary black holes. 2018.

[30] D. M. Scolnic et al. The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. *Astrophys. J.*, 859(2):101, 2018. doi: 10.3847/1538-4357/aab9bb.

[31] Camille Bonvin, Ruth Durrer, and M. Alice Gasparini. Fluctuations of the luminosity distance. *Phys. Rev.*, D73:023523, 2006. doi: 10.1103/PhysRevD.73.023523. [Erratum: Phys. Rev.D85,029901(2012)].

[32] Christopher M. Hirata, Daniel E. Holz, and Curt Cutler. Reducing the weak lensing noise for the gravitational wave Hubble diagram using the non-Gaussianity of the magnification distribution. *Phys. Rev.*, D81:124046, 2010. doi: 10.1103/PhysRevD.81.124046.

[33] Nicola Tamanini, Chiara Caprini, Enrico Barausse, Alberto Sesana, Antoine Klein, and Antoine Petiteau. Science with the space-based interferometer eLISA. III: Probing the expansion of the Universe using gravitational wave standard sirens. *JCAP*, 1604(04):002, 2016. doi: 10.1088/1475-7516/2016/04/002.

[34] Yves Dirian. Changing the Bayesian prior: Absolute neutrino mass constraints in nonlocal gravity. *Phys. Rev.*, D96(8):083513, 2017. doi: 10.1103/PhysRevD.96.083513.

[35] Pia Mukherjee, David Parkinson, and Andrew R. Liddle. A nested sampling algorithm for cosmological model selection. *Astrophys. J.*, 638:L51–L54, 2006. doi: 10.1086/501068.

[36] John Skilling. Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4):833–859, 2006. doi: 10.1214/06-BA127.

[37] H. Jeffreys. *Theory of probability*. Clarendon Press, Oxford, 1961, 3rd edn 1998.