Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **496,** 523–531 (2020)
Advance Access publication 2020 April 22

doi:10.1093/mnras/staa1077

Imprints of the redshift evolution of double neutron star merger rate on the signal-to-noise ratio distribution

Shilpa Kastha,[1,2]★ M. Saleem [3]★ and K. G. Arun[3,4]

[1]*The Institute of Mathematical Sciences, IV Cross Street, CIT Campus, Taramani, Chennai 600113, Tamil Nadu, India*
[2]*Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai 400094, Maharashtra, India*
[3]*Chennai Mathematical Institute, H1, SIPCOT IT Park, Kelambakkam, Siruseri 603103, Tamil Nadu, India*
[4]*Institute for Gravitation and the Cosmos, Pennsylvania State University, State College, PA 16802, USA*

Accepted 2020 April 14. Received 2020 March 24; in original form 2019 July 16

ABSTRACT

Proposed third-generation gravitational wave interferometers such as Cosmic Explorer (CE) will have the sensitivity to observe double neutron star mergers (DNS) up to a redshift of ∼5 with good signal-to-noise ratios (SNRs). We argue that the comoving spatial distribution of DNS mergers leaves a unique imprint on the statistical distribution of SNRs of the detected DNS mergers. Hence, the SNR distribution of DNS mergers will facilitate a novel probe of their redshift evolution independent of the luminosity distance measurements. We consider detections of DNS mergers by the CE and study the SNR distribution for different possible redshift evolution models of DNSs and employ Anderson Darling *p-value* statistic to demonstrate the distinguishability between these different models. We find that a few hundreds of DNS mergers in the CE era will allow us to distinguish between different models of redshift evolution. We further apply the method for various SNR distributions arising due to different merger delay-time and star formation rate (SFR) models and show that for a given SFR model, the SNR distributions are sensitive to the delay-time distributions. Finally, we investigate the effects of sub-threshold events in distinguishing between different merger rate distribution models.

Key words: Gravitational waves – star formation.

1 INTRODUCTION

The first two observation runs of advanced LIGO and Virgo interferometers have led to the detections of 10 binary black hole mergers (Abbott et al. 2016a, d, 2017b, c, k, 2018a, b) and a binary neutron star merger (Abbott et al. 2017d). The binary neutron star merger was also observed in various bands of the electromagnetic spectrum from gamma-rays to the radio (Abbott et al. 2017f; Goldstein et al. 2017; Valenti et al. 2017; D'Avanzo et al. 2018; Lazzati et al. 2018; Lyman et al. 2018; Margutti et al. 2018; Resmi et al. 2018; Ruan et al. 2018; Troja et al. 2018). These detections have given us unique insights about the astrophysics (Abbott et al. 2016c, e, f, 2017f, g, h, i, 2019; Albert et al. 2017), cosmology (Abbott et al. 2017e), and fundamental physics (Abbott et al. 2016b, c, 2017b, f, g, h, i, j, 2018c, 2019; Albert et al. 2017). With the planned upgrades of advanced LIGO and other similar interferometers [Virgo (Abbott et al. VIR), KAGRA (Aso et al. 2013), LIGO-India (Iyer et al. 2011)] joining the world-wide network of gravitational wave (GW) detectors, we are gearing up for exciting times in GW astronomy. There are ongoing research and development activities towards third-generation ground-based detectors such as Einstein Telescope (ET; Sathyaprakash et al. 2012) and Cosmic Explorer (CE; Dwyer et al. 2015). Following the success of LISA Pathfinder (Armano et al. 2016), the space-based LISA mission is now funded (Babak et al. 2017). With these developments, GW astronomy is going to be a very active field of research in the coming decades (Sathyaprakash & Schutz 2009).

One of the key science goals of third-generation detectors is to probe the redshift evolution of compact binary mergers in the Universe (Abernathy et al. 2010; Van Den Broeck 2014). The redshift evolution of compact binaries would shed light on several unknown facets of star formation and stellar evolution (Vitale et al. 2019). The conventional way to reconstruct the redshift evolution of compact binaries is by measuring the luminosity distance to each merger and use a cosmological model to translate the distance estimate to redshift. By repeating this for all the detected events, one can reconstruct the redshift evolution for compact binary mergers (Van Den Broeck 2014; Fishbach, Holz & Farr 2018). However, luminosity distance is one of the 14 parameters that describes the compact binary merger and hence the parameter estimation of each event involves multidimensional maximization of the likelihood func-

★ E-mail: shilpa.kastha@aei.mpg.de (SK); msaleem@cmi.ac.in (MS)

tion. This is computationally very demanding for third-generation detectors because of their low frequency sensitivity due to which the signals will last for several minutes to hours in the sensitivity band of the detector. This problem is even worse for low-mass systems, such as binary neutron star mergers, which merge in the kilo Hertz frequencies. As third-generation detectors are expected to detect tens of thousands of compact binary mergers, parameter estimation is going to pose challenges in terms of computing resources. Hence, detailed parameter estimation follow-up of GW events will be prioritized based on their astrophysical significance and low-significance events (also known as sub-threshold events) are most likely not to be followed up. Moreover, Huang et al. (2018) showed that the parameter estimation of low-significance events is subjected to noise systematics and often does not yield reliable estimates.

On the other hand, low-significance events, if astrophysical in origin, are mostly likely to be either from higher redshifts or a near-edge-on system. A population with high-redshift events is best suited for probing the redshift evolution. In some cases, such low-significance events might be identified to be in spatial and/or temporal coincidence with triggers from electromagnetic search that may boost the significance of being astrophysical in origin. Therefore, one may not want to ignore the low-significance events completely. Hence, developing methods to probe the redshift evolution of compact binary mergers that does not rely on the parameter estimation is going to be very useful. Here we propose a novel method that can be used to probe redshift evolution of binary neutron star mergers by using only the signal-to-noise ratio (SNR) of each event. This information is likely to be readily available as outcomes of detection algorithms, irrespective of whether the detailed parameter estimation follow-up of the signal is performed or not.

Recently, Schutz (Schutz 2011) pointed out, on very general grounds, that the observed SNR of the GW events detected by GW detectors should follow a universal distribution if the underlying spatial density of the source population is constant within the volume accessible to the detectors. This distribution is independent of the type of the sources and hence referred to as universal. As argued by Schutz (Schutz 2011), the universality arises from the fact that the SNRs of the GW events are inversely proportional to the luminosity distance ($\rho \propto \frac{1}{D_L}$) and at relatively low redshifts (say $z \lesssim 0.1$) the luminosity distance and comoving distance can be approximated to be the same. More precisely, following Chen and Holz (Chen & Holz 2014), the probability of a source (say a compact binary merger) to be found within a shell of thickness dD, at a comoving distance of D, goes as $f_D \mathrm{d}D \propto D^2\,\mathrm{d}D$, if the comoving number density of the source population is constant. Since $\rho \propto \frac{1}{D}$, the distribution of SNR corresponding to the particular source distribution can easily be shown to follow $p(\rho) = f_D \left|\frac{\mathrm{d}D}{\mathrm{d}\rho}\right| \propto \frac{1}{\rho^4}$. After normalization, we obtain

$$p(\rho) = \frac{3\rho_{\rm th}^3}{\rho^4}, \tag{1}$$

where $\rho_{\rm th}$ is the SNR threshold used for detection. The above derivation crucially assumes that the properties of the source population (such as mass distribution) do not evolve with redshift. Chen and Holz (Chen & Holz 2014) explored various implications of this universal distribution for the sources detectable by second-generation detectors such as advanced LIGO/Virgo. This universal distribution is also an ingredient used in Abbott et al. (2016f) to derive a bound on the rate of the binary black hole mergers from the first observation run of LIGO (Abbott et al. 2016d).

Motivated by Schutz (2011) and Chen & Holz (2014), in this paper, we study the SNR distribution of compact binary mergers but for cosmological sources. For binary black hole mergers, their mass distribution is likely to influence the SNR distribution as much as the cosmological evolution (see, for instance, Vitale 2016) which makes it difficult to disentangle the two effects. However, that is not the case with double neutron star (DNS) mergers as the masses are expected to vary over a relatively smaller range compared to binary black hole mergers. The planned CE will have enough sensitivity to detect DNS mergers with good SNR ($\sim$12) up to a redshift of $\sim$5 (Sathyaprakash et al. 2012; ET).

Considering CE sensitivity as representative of the third-generation GW detector, in this paper, we study how the SNR distribution of DNS mergers observed by CE gets affected by the redshift evolution of their rate density and hence we use the detected SNR distribution to probe the underlying redshift evolution of DNS mergers. Considering astrophysically motivated models for the redshift evolution of DNS merger rate density, we study how distinguishable are the resulting SNR distributions from each other. We find that observations of the order of a few hundreds of DNS mergers are sufficient to distinguish between different redshift evolution models. As the projected detection rate of DNS mergers by the third-generation GW detectors is of the order a few hundreds to thousands (Abernathy et al. 2010), 1 yr of observation by CE may itself be sufficient to track the redshift evolution of DNS using this method.

The paper is organized in the following way. In Section 2, we consider the cosmological effects on the optimal SNR of compact binary sources. In Section 3, we explore how the different DNS merger rate densities affect the SNR distributions. In Section 4, we discuss whether the distributions corresponding to all the merger rate densities are distinguishable from each other.

2 EFFECTS OF COSMOLOGICAL EXPANSION ON THE SNR OF COMPACT BINARIES

The data analysis technique of matched filtering (Wainstein & Zubakov 1962; Helström 1968) is usually employed to detect compact binary mergers using GW observations. Matched filtering involves cross-correlating various copies of the expected gravitational waveforms (templates) corresponding to different signal parameters (such as masses and spins), with the data, which potentially contains the signal (in addition to the noise). The template that maximizes the correlation is referred to as optimal template and the corresponding SNR is called optimal SNR that is defined as

$$\rho = \sqrt{4\int_0^\infty \frac{|\tilde{h}(f)|^2}{S_h(f)}\mathrm{d}f}, \tag{2}$$

where $S_h(f)$ is the detector's power spectral density and $\tilde{h}(f)$ is the frequency domain gravitational waveform [see, for instance, section 5.1 of Sathyaprakash & Schutz (2009) for details].

We employ restricted post-Newtonian (PN) waveform (RWF), $\tilde{h}(f) = \mathcal{A} f^{-7/6} e^{i\psi(f)}$, where $\mathcal{A}$ is the amplitude and $\psi(f)$ is the frequency domain GW phase. In RWF, the PN corrections to the amplitude of the gravitational waveform are ignored but the phase is accounted for to the maximum accuracy. Using the RWF, the optimal SNR for GW events of compact binary systems can be expressed as (Cutler & Flanagan 1994)

$$\rho(m_1, m_2, D_L, \theta, \phi, \psi, \iota) = \sqrt{4\frac{\mathcal{A}^2}{D_L^2}\left[F_+^2(\theta,\phi,\psi)(1+\cos^2\iota)^2 + 4F_\times^2(\theta,\phi,\psi)\cos^2\iota\right] I(M)}, \tag{3}$$

where $F_{+,\times}(\theta, \phi, \psi)$ are the antenna pattern functions for the 'plus' and 'cross' polarizations, and $\mathcal{A} = \sqrt{5/96}\,\pi^{-2/3}\mathcal{M}^{5/6}$, where $\mathcal{M}$ is the chirp mass, which is related to the total mass M by $\mathcal{M} = M\,\eta^{3/5}$, where $\eta = \frac{m_1 m_2}{M^2}$ is the symmetric mass ratio of the system and m_1, m_2 are the component masses. The four angles $\{\theta, \phi, \psi, \iota\}$ describe the location and orientation of the source with respect to the detector. $I(M)$ is the frequency integral defined as

$$I(M) = \int_0^\infty \frac{f^{-7/3}}{S_h(f)}\mathrm{d}f \simeq \int_{f_{\rm low}}^{f_{\rm LSO}} \frac{f^{-7/3}}{S_h(f)}\mathrm{d}f. \tag{4}$$

In the last step, we have replaced the lower and upper limits of the integral by the seismic cut-off, $f_{\rm low}$, of the detector and the frequency at the last stable orbit (LSO) of the black holes with masses m_1 and m_2, respectively. The GW frequency at the LSO up to which PN approximation is valid, as a function of the total mass M, is $f_{\rm LSO} = \frac{1}{6^{3/2}\,\pi\,M}$. This is the expression for the frequency at the LSO of a Schwarzschild BH with total mass M.

As we use CE as a proxy for third-generation detectors in this work, the SNR computations use the following fit for the CE wide-band (CE-wb) sensitivity curve (Abbott et al. 2017a)

$$\begin{aligned} S_h(f) = {} & 5.62\times10^{-51} + 6.69\times10^{-50} f^{-0.125} + 7.80 \\ & \times 10^{-31} f^{-20} + 4.35\times10^{-43} f^{-6} + 1.63\times10^{-53} f \\ & + 2.44\times10^{-56} f^{2} + 5.45\times10^{-66} f^{5}\ \mathrm{Hz}^{-1}. \end{aligned} \tag{5}$$

Next we discuss the effect of cosmology on the gravitational waveform and hence the expression for SNR.

2.1 Effects of cosmological expansion

Assuming a flat ΛCDM cosmological model (Ade et al. 2014; Aghanim et al. 2018) for the universe, we explore the modification to the SNR for compact binary systems at cosmological distances. Cosmological expansion of the universe affects the gravitational waveforms in two ways. According to general relativity, GW amplitude is inversely proportional to the comoving distance D, which is no longer same as the luminosity distance D_L but is related by $D_L = D\,(1+z)$, where z is the redshift to the source. Due to the expansion of universe, there is an additional factor $(1+z)$ in the denominator of the amplitude. This is already accounted in equation (3) as it is written in terms of D_L. Secondly, due to the cosmological expansion, the GW frequency gets redshifted. This results in redefining the chirp mass in such a way that the observed chirp mass ($\mathcal{M}$) is related to the corresponding chirp mass in the source frame ($\mathcal{M}_{\rm source}$), by $\mathcal{M} = \mathcal{M}_{\rm source}\,(1+z)$. This happens due to the fact that the only time-scale of the problem $\frac{G\mathcal{M}}{c^3}$ is redshifted, which is completely degenerate with mass and hence leading to the notion of redshifted mass (see section 4.1.4 of Maggiore 2007). In order to explicitly incorporate these effects, we rewrite the expression for SNR in equation (3) as

$$\rho = \frac{\mathcal{M}_{\rm source}^{5/6}}{D(1+z)^{1/6}}\; g(\theta, \phi, \psi, \iota)\,\sqrt{I(M)}, \tag{6}$$

where all the angular dependences in the waveform are captured in the definition of $g(\theta, \phi, \psi, \iota)$ and other variables have their usual meanings.

In a flat FLRW cosmology, the comoving distance (following $c = G = 1$ units), corresponding to a redshift z (Hogg 1999; Wright 2006) is given by

$$D(z) = \frac{1}{H_0}\int_0^z \frac{\mathrm{d}z'}{E(z')}, \tag{7}$$

where H_0 is the Hubble constant and

$$E(z) = \sqrt{\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda}, \tag{8}$$

with the total density parameter (Ω) consisting of matter (dark and baryonic) density ($\Omega_{\rm m}$) and cosmological constant (Ω_Λ). Throughout this work, we consider a cosmology with $\Omega_\Lambda = 0.7$ and $\Omega_{\rm m} = 0.3$ and $H_0 = 72\,\mathrm{km\,Mpc^{-1}\,s^{-1}}$ (Ade et al. 2014; Aghanim et al. 2018).

Given that z is a function of D (equation 7), it is clear from equation (6) that the simple scaling relation for SNR ($\rho \propto 1/D$) would no longer hold. Hence, it is obvious that the universal SNR distribution given in equation (1) does not apply any more. In the next section, we discuss the effect of redshift evolution on the SNR distribution and how the SNR distribution encodes the signature of merger rate density evolution.

3 IMPRINTS OF COMOVING MERGER RATE DENSITY EVOLUTION OF DNS SYSTEMS ON THE SNR DISTRIBUTION

Usually, it is assumed that the DNS formation rate follows the star formation rate (SFR) while their merger rate will also depend on the delay-time distribution: the distribution of the time delay between the formation and the merger. Hence, following Regimbau & Hughes (2009), binary merger rate density can be written as

$$R(z) \propto \int_{t_{\rm d}^{\rm min}}^\infty \frac{\dot\rho_*(z_f(z, t_{\rm d}))}{1 + z_f(z, t_{\rm d})}\,P(t_{\rm d})\,\mathrm{d}t_{\rm d}, \tag{9}$$

where $\dot\rho_*$ is the SFR, $t_{\rm d}$ is the delay time, and $t_{\rm d}^{\rm min}$ is the minimum delay time for a binary to merge since its formation. The redshift z describes the epoch at which the compact binary merges whereas $z_{\rm f}$ is the redshift at which its progenitor binary forms and they are related by a delay time $t_{\rm d}$. The factor $P(t_{\rm d})$ is the distribution of the delay time. According to various population synthesis models (Tutukov & Yungelson 1994; Lipunov et al. 1995; Ando 2004; de Freitas Pacheco et al. 2006; Belczynski et al. 2008; O'Shaughnessy, Kalogera & Belczynski 2008), the delay time follows a power-law (PL) distribution, $P(t_{\rm d}) \propto 1/t_{\rm d}$, with $t_{\rm d} > t_{\rm d}^{\rm min}$. The factor $(1+z_{\rm f})^{-1}$ takes into account the cosmological time dilation between the star formation and the merger.

For our analysis in this paper, we use two merger rate models, following the two SFR models proposed by Hopkins & Beacom (2006) and Wilkins et al. (2008) and denote them as $M_{\rm HB}$ and $M_{\rm Wilkins}$, respectively. In both the cases, we consider (Belczynski & Kalogera 2001; Belczynski et al. 2008) $t_{\rm d}^{\rm min} \sim 20$ Myr. As discussed in Dominik et al. (2013), the redshift evolution of the host galaxy affects the merger rate of DNS binaries (see their top panel of figs 3 and 4). For higher metallicities, the peak of the merger rate density shifts towards lower redshifts. From this perspective, our $M_{\rm Wilkins}$ is representative of the case where the DNS mergers dominantly happen in high-metallicity environments, shifting the peak towards lower redshifts. We also consider another model of rate density evolution, obtained by Wanderman & Piran (2015),

$$R_{\rm WP}(z) = 45\,\mathrm{Mpc^{-3}\,Gyr^{-1}}.\begin{cases} e^{(z-0.9)/0.39} & z \le 0.9 \\ e^{-(z-0.9)/0.26} & z > 0.9 \end{cases}$$

This is a model (denoted as $\mathrm{M_{WP}}$) derived based on the short GRBs observed by the gamma-ray satellites accounting for the effect of beaming. Though, somewhat indirect, we use this model to have enough diversity in the set of models we compare against.

Along with these models, we also consider a case with constant comoving rate density M_0 characterized by $R(z) = R_0$

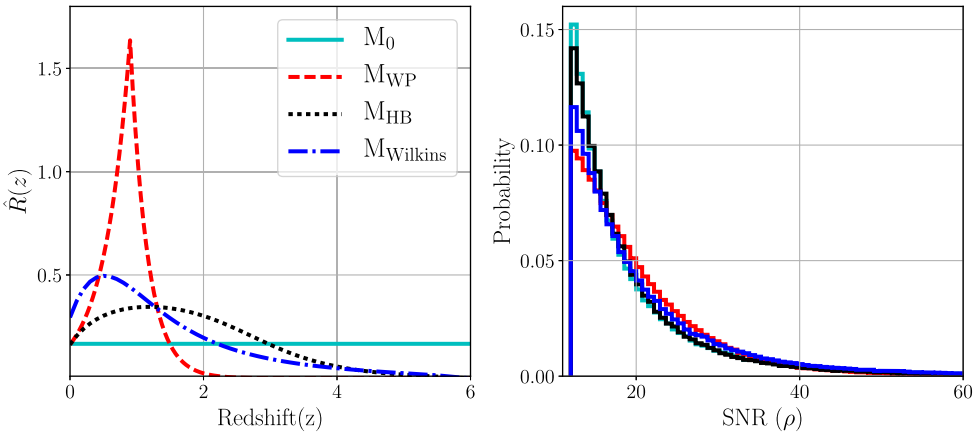

Figure 1. Figure on the left-hand panel shows the evolution of comoving merger rate density with redshift for four different models, M_0 stands for the constant comoving merger rate, $M_{\rm WP}$ represents the model for rate density evolution obtained by Wanderman & Piran (2015), and $M_{\rm HB}$ and $M_{\rm Wilkins}$ denote the merger rate models obtained in Regimbau & Hughes (2009) and Van Den Broeck (2014) following the SFRs given in Hopkins & Beacom (2006) and Wilkins, Trentham & Hopkins (2008), respectively. Figure on the right most panel contains the corresponding normalized SNR distributions.

$= 1\,{\rm Mpc}^{-3}\,{\rm Myr}^{-1}$. The left-hand panel of Fig. 1 shows the normalized forms of all the merger rate density models discussed above.

Given the merger rate density $R(z)$ (in units of ${\rm Mpc}^{-3}\,{\rm Myr}^{-1}$), the total number of sources (in units of ${\rm Myr}^{-1}$) in a comoving volume of radius D is

$$N(D) \propto \int_0^D \frac{R(z(D'))}{1+z(D')} D'^2 {\rm d}D', \tag{10}$$

where z can be numerically inverted to obtain the corresponding comoving distance D. The $(1 + z)$ factor in the denominator accounts for the time dilation between the source frame and the observer frame.

Considering the proposed models above to be the underlying source distribution within the comoving volume and assuming isotropy, we obtain the optimal SNR distribution of DNS mergers for CE (right-hand panel of Fig. 1). To generate the source population for obtaining the optimal SNR distributions corresponding to different $R(z)$ (left-hand panel of Fig. 1), first the sources are assumed to be uniformly located and oriented on the sphere parametrized by the comoving distance. This is achieved by making sure that the azimuth angles ϕ, ψ are drawn from a uniform distribution $[0, 2\pi]$ and the polar angles θ, ι are chosen such that their cosines are uniformly distributed between $[-1, 1]$. These choices ensure that at any radius, the source population is uniformly located and oriented on the surface of the sphere. Further, we need to distribute the sources *within* the detection volume specified by the maximum radius $D_{\rm max}(\rho_{\rm th}, M)$ (or $z_{\rm max}(\rho_{\rm th}, M)$), which depends on the SNR threshold (=12). Hence, we choose $N(D)$ to be uniformly distributed between $N(1)$ and $N(D_{\rm max})$ and for each realization, we numerically solve equation (10) to obtain corresponding D value from $N(D)$.

Using the procedure outlined above, we compute the optimal SNR distributions for the different models by imposing the SNR threshold of 12 that is shown in the right-hand panel of Fig. 1. Next we discuss how many detections are required for these models to be statistically distinguishable from each other.

4 STATISTICAL TESTS OF DISTINGUISHABILITY OF VARIOUS MODELS

In this section, we demonstrate the distinguishability of the SNR distributions corresponding to all the merger rate density evolution models discussed in the previous section. First of all, we discuss how to account for the error bars on the SNRs associated with GW detections and then we discuss the distinguishability of different models.

4.1 Error bars on the SNRs

In reality, GW detections are made applying certain detection threshold on the matched filter SNRs that are calculated by matching the observed data and a template bank of pre-computed GW waveforms. However, the SNR distributions in the right-hand panel of Fig. 1 are produced using optimal SNR for each source, where optimal SNR is a point estimate (SNR of the best-fitting template from matched filtering). Therefore, it is important to fold in the error bars to the point estimates of SNRs to account for the usage of matched-filtering in the process of GW detections.

Under the assumption of zero mean Gaussian random noise in the detectors, the matched filter SNR (σ) follows the Rice distribution (Rice 1945) of the following form:

$$f(\sigma, \rho) = \sigma \exp\left(-\frac{\sigma^2+\rho^2}{2}\right) I_0(\rho\sigma), \tag{11}$$

where ρ is the optimal SNR and I_0 is the zeroth-order modified Bessel function of the first kind (see Moore, Gerosa & Klein 2019, for a detailed discussion). In order to account for the errors on SNRs in our distributions, we first calculate the optimal SNR (say ρ_i) for each source and then replace it with a number chosen at random from the distribution $f(\sigma, \rho_i)$ (equation 11).

4.2 Statistical tests

Now we quantify the distinguishability of the different SNR distributions by employing the Anderson–Darling (AD; Anderson & Darling 1952) test. The AD test is a well-known tool used to assess whether a sample data belongs to a reference distribution. The test returns a probability value (p-value) for the 'null' hypothesis that *sample belongs to the reference distribution*. If the null hypothesis is true, the p-value distribution obtained by performing the experiment multiple time is uniform between 0 and 1 with a median p-value of 0.5. If the sample does not belong to the reference distribution, the p-value distribution will sharply peak around 0. A p-value distribution

weighted more towards 0 implies a stronger evidence of rejecting the null hypothesis or ability to distinguish the two distributions.

In order to quantify the distinguishability between two arbitrary merger rate density models M and N, we follow the procedure below. First, we synthesize a fiducial data of SNR distribution of size n (number of detections) assuming that the model M is the true distribution. As the data contain noise, we will account for the errors on the SNRs in the synthesized data along the lines mentioned earlier. The *data* are labelled as data_M, where the subscript refers to the underlying model. Then, we carry out the AD test between data_M and the reference distribution $p_N(\rho)$ that is the predicted SNR distribution corresponding to the model N. In the above step, since $p_N(\rho)$ is the theoretical prediction, it is always free of errors, which is ensured by using sufficiently large number of samples to generate that.

The test returns a p-value that is denoted as $\mathcal{P}(M|N)$. Due to limited number (n) of synthesized samples, the data_M may not capture the essence of the model M and hence affects the p-value. To account for this, we repeat the p-value estimation 100 times, each time synthesizing the data_M randomly and then compute the median of the resulting p-value distribution. The median of p-values is denoted as $\bar{\mathcal{P}}(M|N)$.

4.3 Weighted *p*-values

As mentioned earlier, we have used Rice distribution to model the errors in SNR. The presence of these errors in the data will affect the p-values, which in turn can lead to false detection or false rejection. For example, the median of the p-value distribution resulting from AD test of data_M with the model M, in principle, should be 0.5. However, due to the errors, the test may return a lower median that may even lead to the rejection of the null hypothesis when it is actually true. In our case, we have multiple models {N} to be tested against the data_M and p-value for each model ($\bar{\mathcal{P}}(M|N)$) will decrease due to the errors, thereby reducing the ability to distinguish between various models.

In order to quantify the distinguishability between the data and a model along the lines described earlier, we introduce the notion of *weighted* p-values. For a given data_M, we define a weighting function $\mathcal{W}$ as

$$\mathcal{W} = \frac{1}{\bar{\mathcal{P}}(M|M)}, \tag{12}$$

where $\bar{P}(M|M)$ is the median of p-values between data_M and the underlying model M (which, in the absence of noise, is 0.5). We now define the weighted p-value between models data_M and N as

$$\bar{\mathcal{P}}_w(M|N) = \mathcal{W} \times \bar{\mathcal{P}}(M|N). \tag{13}$$

The weighting factor $\mathcal{W}$ is chosen in such a way that the weighted p-value for data_M with the model M itself always returns unity (i.e. $\bar{\mathcal{P}}_w(M|M) = 1$). Weighted p-values have been extensively discussed in literature in the context of testing multiple hypotheses (for example, see the references Holm 1979; Benjamini & Yekutieli 2001; Durand 2017). The definition we use here may be thought of as an adaptation of this generic definition to our problem.

Based on our previous discussion, it is clear that if two distributions are distinguishable, $\bar{\mathcal{P}}_w(M|N)$ will always be smaller than 1. Using a threshold on the median of the p-value distribution of 0.05 while performing the AD test (i.e. 95 per cent of the time the model is rejected), we set a rejection threshold on $\bar{\mathcal{P}}_w(M|N)$ to be 0.05/0.5=0.1.

4.4 Effect of different SFR models on SNR distribution

Here we discuss the distinguishability of different resulting SNR distributions from four different SFR models and present our results in Fig. 2, where, in the x-axis, we show the number of detections n (for n = 20, 50, 100, 200, 500, 1000, 2000, 5000, and 10 000) and in the y-axis, we show the distinguishability of each of the four rate models (M_0, $M_{\rm HB}$, $M_{\rm Wilkins}$, and $M_{\rm WP}$) from each other by computing the *weighted* p-values $\bar{\mathcal{P}}_w(M|N)$ among them. Each panel corresponds to a particular model for the data and the different curves in each panel correspond to $\bar{\mathcal{P}}_w(M|N)$ estimated for all the four models. For example, in the top-left panel of Fig. 2, we synthesize the data following constant comoving rate density and compare against the theoretical distributions of all the four models. By construction, the *weighted* p-value $\bar{\mathcal{P}}_w(M_0|M_0)$, when the data containing M_0 are compared with model M_0 itself, represented by the cyan line, is constant and is 1. As opposed to this scenario, all the other ratios fall off as a function of the number of detections. Hence, the data can be distinguished from the other models. In the top-left panel, we also see that a low number of detections ($\sim$500) is sufficient to distinguish between data_{M_0} and the model $M_{\rm WP}$ or $M_{\rm Wilkins}$ whereas we need at least thousands of detections to differentiate between the data_{M_0} and the model $M_{\rm HB}$.

In the remaining three panels, we perform the same exercise for the rest of the three models. In the top right panel, the data ($\mathrm{data}_{\rm M_{WP}}$) are generated from the merger rate density model, $M_{\rm WP}$. As expected, $\bar{\mathcal{P}}_w(M_{\rm WP}|M_{\rm WP})$ is unity (red curve). The cyan, blue, and the black curves represent the comparison with models M_0, $M_{\rm HB}$, and $M_{\rm Wilkins}$, respectively. We find that for a few hundreds of detections, all the three models are distinguishable from $M_{\rm WP}$. This is not surprising given how different this distribution is from others in the left-hand panel of Fig. 1.

Similarly, in the bottom left panel the data are generated following the merger rate density model, $M_{\rm HB}$ and in the bottom right panel the data are generated following the merger rate density model, $M_{\rm Wilkins}$. In case of data containing $M_{\rm HB}$ (bottom left panel), we find that larger number of detections ($\sim$few thousands) of DNS mergers are needed in order to distinguish this model especially from M_0. As opposed to this scenario, in the bottom right panel, $M_{\rm Wilkins}$ is distinguishable from other models given a few hundreds of detections.

Therefore, it is evident that $M_{\rm WP}$ and $M_{\rm Wilkins}$ can be distinguished from all other models with high confidence with a few hundreds of detections, whereas $M_{\rm HB}$ is difficult to distinguish from the M_0 using this method with a low number of detections. However, with a large number of detections (say 10 000) $M_{\rm HB}$ is distinguishable from the other models. Given a sufficiently large number of detections, we expect $\bar{\mathcal{P}}_w(M|N)$ to be either 0 or 1 given that the two distributions are different or the same, respectively. Hence, we do not show any $\bar{\mathcal{P}}_w(M|N) < 10^{-4}$ in Fig. 2 and treat them as a scenario where the two distributions are completely distinguishable.

As shown in Abernathy et al. (2010) and the most recent work by Baibhav et al. (2019), the forecasted DNS detection rates by the third-generation detectors ET-B and CE range from one thousand to tens of thousands per year. Given this rate, it is clear that the SNR data collected from less than an year of observation will be sufficient to test various merger rate density models.

4.5 Effect of different delay-time models on SNR distributions

In the previous section, we have shown the effects of the different merger rate distribution models on the observed SNR distributions. In this section, we investigate how the distribution of delay time

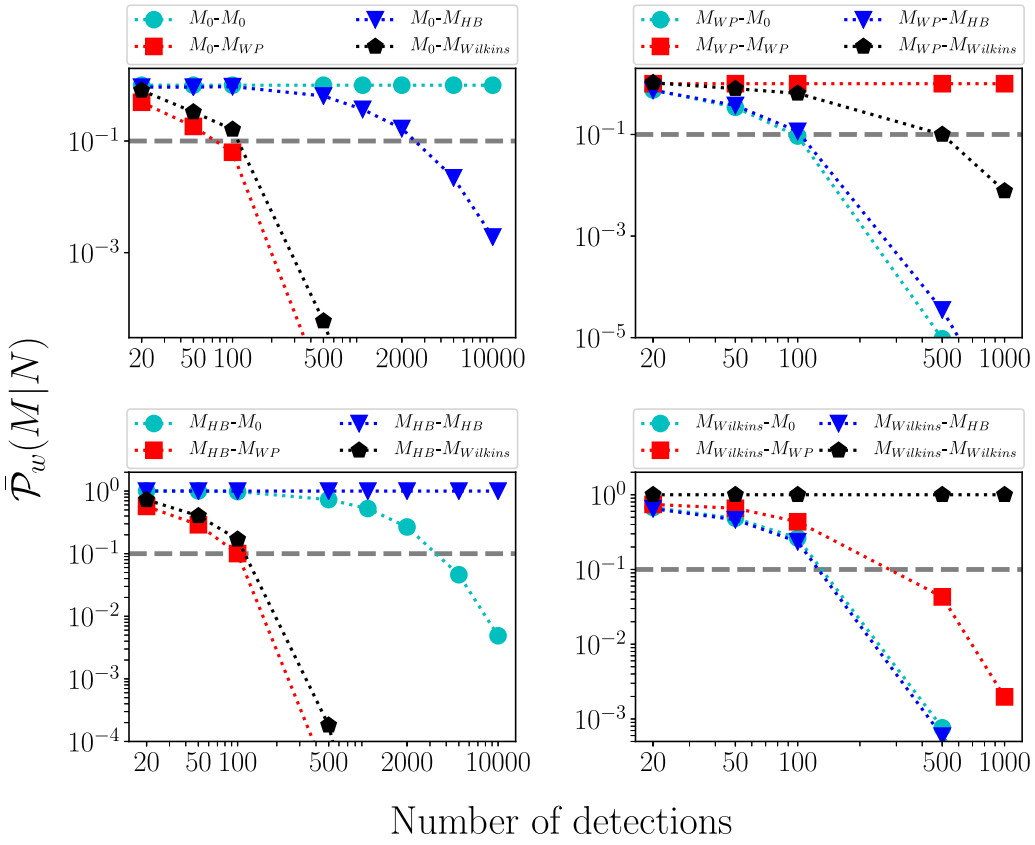

Figure 2. Weighted p-values ($\bar{\mathcal{P}}_w(M|N)$) from AD test performed on the data obtained from the four models as a function of the number of detections. The first argument in each legend represents the data generated by following a particular model M (denoted as data$_M$ in the main text), whereas the second argument is the theoretical model with which the data are compared to. We put a grey horizontal line in every panel corresponding to the threshold on $\bar{\mathcal{P}}_w(M|N)$ (see the main text for details).

affects the observed SNR distributions. As is evident from equation (9), the merger rate density is a convolution of the SFR and the delay-time distribution. For the analysis in the preceding section, we have used a PL model for the delay-time distribution, $p(t_d) \propto t_d^{-1}$, with four different SFRs. However, there are different models proposed in literature to represent the actual underlying delay-time distribution. In this section, we investigate the ability of SNR distributions to distinguish between different delay-time models.

We consider two models for delay-time distributions. The first is a PL distribution characterized by a PL coefficient α_t,

$$p(t_d) \propto t_d^{-\alpha_t} \tag{14}$$

and the second is a lognormal (LN) distribution given as

$$p(t_d) = \exp\left[-\frac{(\ln t_d - \ln \tau)^2}{2{\sigma_t}^2}\right] \Big/ \left(\sqrt{2\pi}\sigma_t\right), \tag{15}$$

where τ and σ_t are model parameters. These models have been considered in Wanderman & Piran (2015) for estimating the short-GRB rate distribution. To be consistent with their estimated 1σ error bars on the model parameters, we have fixed $\tau = 3.9$ Gyr, $\sigma_t = 0.05$, and $\alpha_t = 0.81$ for our study. We have used these two models together with the two SFR models proposed in Hopkins & Beacom (2006) (HB) and Wilkins et al. (2008) (W). Using equation (9), we obtain the resulting four different merger rate distributions labelled as $\mathrm{HB_{PL}}$, $\mathrm{HB_{LN}}$, $\mathrm{W_{PL}}$, and $\mathrm{W_{LN}}$. Following the same procedure as in the previous section, in Fig. 3, we show the distinguishability of the SNR distributions corresponding to these four models (see Fig. 2 and the corresponding texts for figure description).

We find that for a given SFR, the two delay-time models considered here are distinguishable with a few tens of events while for a given delay-time model, the two SFR considered here require about hundred events to distinguish from each other. This feature could be specific to the particular models we considered here, but conveys the broader message that the SNR distributions are sensitive to the variations in delay-time models and indicate that under the assumption of a given SFR model, the SNR distributions can be used to perform model selection between different delay-time models and one can constrain their model parameters. Further, though not extensively studied in this work, the results also point to the possible degenerate combinations of SFR and delay-time models that can give identical merger rates and hence identical SNR distributions. Therefore, it might be difficult to constrain both SFR and delay-time models simultaneously unless we have a large number of detected events as expected with third-generation GW detectors.

4.6 Effect of sub-threshold events

In Fig. 2, we have seen that the distinguishability increases as the number of detected sources increases. For a given detector sensitivity, the number of detections can be increased only by lowering the threshold, which means by the inclusion of sub-threshold events. However, unlike in Fig. 2, the increased number of sources by lower-

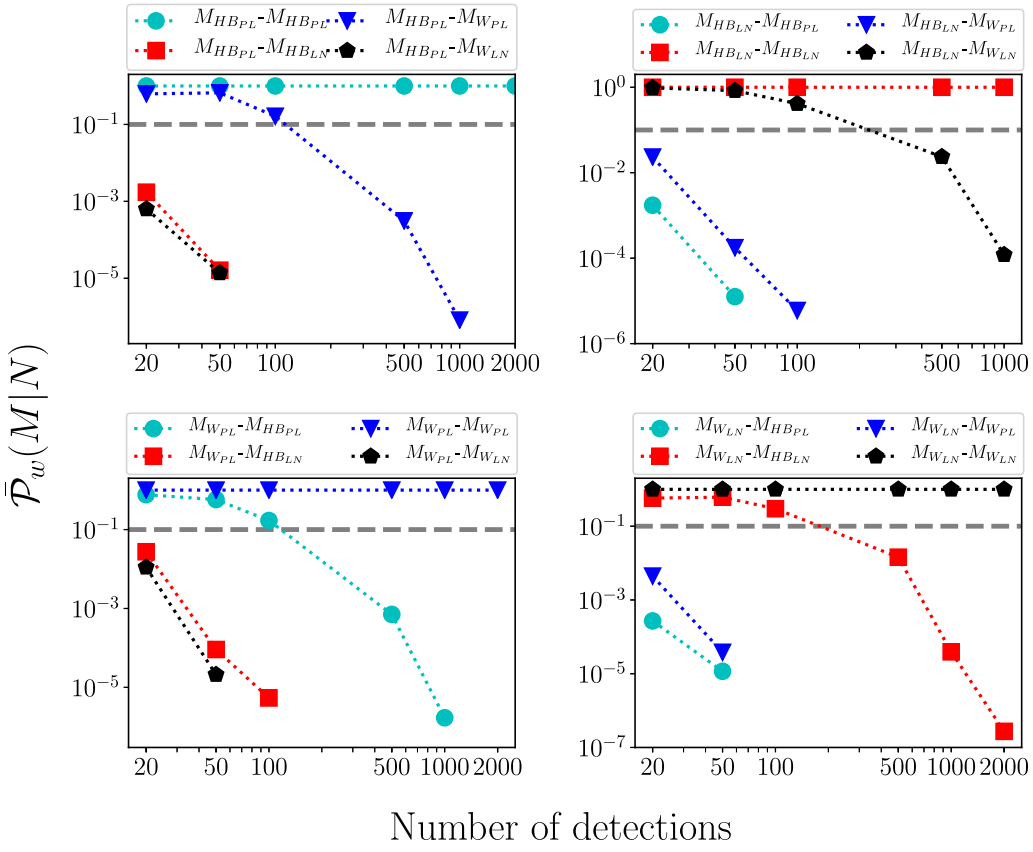

Figure 3. Same as Fig. 2 but performed between models that are obtained from various combinations of delay-time distributions and SFR models. This is to demonstrate how different delay-time models can be captured by the SNR distributions and also to show how the uncertainty in delay-time models can contaminate the star formation models. See the main text for a quantitative discussion.

ing the threshold are associated with larger uncertainties as given by equation (11) and this might result in reducing the discriminatory power of the detected distribution. In this section, we investigate this in detail to see which of these two complementary effects is dominant and hence to assess whether sub-threshold events might be of use.

In order to demonstrate this, we repeat one of the tests shown in Fig. 2 using different detection thresholds. To generate representative detections, we take 1000 sources distributed as per the model $M_{\rm HB}$ and apply the detection thresholds 8, 16, 32, and 64 to get the respective detected sources where the number of detections are highest for threshold 8 and lowest for 64. The results are shown in Fig. 4 where these four data sets are tested against two models $M_{\rm Wilkins}$ and $M_{\rm HB}$, which is the underlying true model itself. The black diamonds show the weighted p-values ($\bar{\mathcal{P}}_w(M|N)$) for the data against the model $M_{\rm Wilkins}$. We find that as we decrease the detection threshold, the median of the weighted p-values consistently decreases. This shows that the discriminatory power between two models increases with the inclusion of more number of sources coming from lowering the threshold despite their higher uncertainties.

In the exercise above, the data points are made up of only sources of astrophysical origin that follow an underlying distribution even at low thresholds. However, in a real scenario, things are more complicated as there will be many noise transients mimicking as signals that can contaminate the detected distributions. Lowering the threshold to extremely low values will make the detections dominated by false alarms and hence the test will fail to achieve the goal. Hence, in order to take the advantage of sub-threshold events, one will have to choose the threshold by fine-tuning the false alarm rates so that noise transients do not contaminate the detections.

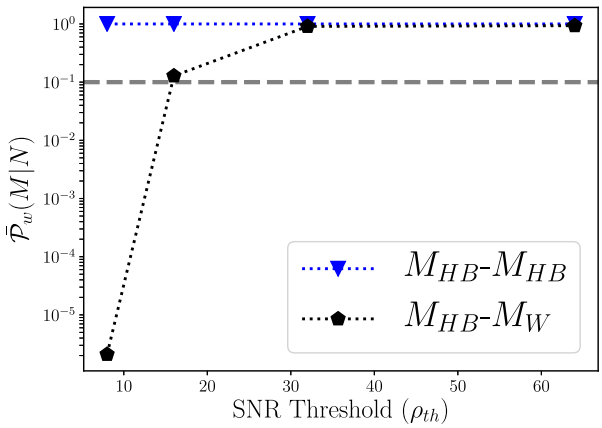

Figure 4. Weighted p-values ($\bar{\mathcal{P}}_w(M|N)$) shown as a function of the detection threshold. The data are generated as per the model $M_{\rm HB}$ and tested against the model $M_{\rm Wilkins}$ (shown as black diamonds). With lower detection thresholds, the weighted p-value decreases, which indicates higher discriminatory power by the inclusion of sub-threshold events. The blue triangles are the tests against same model as also shown in the previous figure. We put a grey horizontal line corresponding to the threshold on $\bar{\mathcal{P}}_w(M|N)$ (see Section 4.6 for details).

5 CONCLUSIONS AND OUTLOOK

In this work, we have shown that the SNR distribution obtained from the GW detections of DNS mergers carries the imprints of merger rate evolution and hence can be used to probe the redshift evolution of binary neutron star mergers observed by third-generation GW observatories. Consequently, SNR distributions can also be used to probe the two ingredients of merger rates: the SFR model and the delay-time distribution model for the time delay between the binary formation and merger. Unlike the traditional approach that uses luminosity distance estimates from the detailed follow-up studies of each detected DNS event, the proposed method only uses the SNRs of the DNS mergers that are very likely to be available in real time by the time the third-generation detectors are operational. Hence, a method based on SNR distribution will be of interest. The importance of such a method is further emphasized in the context of studies using the numerous sub-threshold events for which the parameter estimation follow-up may not be performed (due to systematics; Huang et al. 2018) while such events might still be of interest when there are sub-threshold triggers from EM counterpart searches in temporal and/or spatial coincidence. In fact, it would be interesting to test if the inclusion of sub-threshold DNS triggers would alter our inference about the merger rate evolution models obtained using only highly significant events (sub-threshold events are expected to have high redshifts – if they are astrophysical). We also studied how many binary neutron star detections would be required to rule in or rule out different known redshift evolution models. Using AD test between the observed SNR distribution with various model predictions, we show that a few hundreds of binary neutron star detections should be adequate to track the redshift evolution of binary neutron stars. Further, our results indicate that the effects of SFRs and delay-time distributions can mimic each other and it might be challenging to constrain both from a given DNS merger distribution with a few hundreds of events. However, for a given SFR model assumption, a few tens to few hundred DNS detections will be enough to constrain the delay-time models.

ACKNOWLEDGEMENTS

KGA and MS are partially supported by a grant from Infosys Foundation. KGA acknowledges the support by the Indo-US Science and Technology Forum through the Indo-US Center for the Exploration of Extreme Gravity (Grant No. IUSSTF/JC-029/2016). We would like to thank Nathan Johnson-McDaniel, B. S. Sathyaprakash, P. Ajith, Chris Van Den Broeck, Archisman Ghosh, Anuradha Gupta, Ghanashyam Date, Alok Laddha, and N. V. Krishnendu for very valuable discussions. We thank Maya Fishbach for useful comments on an earlier version of the manuscript. This document has LIGO preprint number P1800002.

REFERENCES

Abbott B. P. et al., 2016a, Phys. Rev. Lett., 116, 061102
Abbott B. P. et al., 2016b, Phys. Rev. Lett., 116, 221101
Abbott B. P. et al., 2016c, Phys. Rev. Lett., 116, 241102
Abbott B. P. et al., 2016d, Phys. Rev. Lett., 116, 241103
Abbott B. P. et al., 2016e, ApJ, 818, L22
Abbott B. P. et al., 2016f, ApJ, 833, L1
Abbott B. P. et al., 2017a, Class. Quantum Gravity, 34, 044001
Abbott B. P. et al., 2017b, Phys. Rev. Lett., 118, 221101
Abbott B. P. et al., 2017c, Phys. Rev. Lett., 119, 141101
Abbott B. P. et al., 2017d, Phys. Rev. Lett., 119, 161101
Abbott B. P. et al., 2017e, Nature, 551, 85
Abbott B. P. et al., 2017f, ApJ, 848, L12
Abbott B. P. et al., 2017g, ApJ, 848, L13
Abbott B. P. et al., 2017h, ApJ, 850, L39
Abbott B. P. et al., 2017i, ApJ, 850, L40
Abbott B. P. et al., 2017j, ApJ, 851, L16
Abbott B. P. et al., 2017k, ApJ, 851, L35
Abbott B. P. et al., 2018a, preprint (arXiv:1811.12907) .
Abbott B. P. et al., 2018b, preprint (arXiv:1811.12940)
Abbott B. P. et al., 2018c, Phys. Rev. Lett., 121, 161101
Abbott B. P. et al., 2019, Phys. Rev., X9, 011001
Abernathy M. et al., 2010, Einstein Gravitational Wave Telescope: Conceptual Design Study (Document Number ET-0106A-10)
Ade P. A. R. et al., 2014, A&A, 571, A16
Aghanim N. et al., 2018, preprint (arXiv:1807.06209)
Albert A. et al., 2017, ApJ, 850, L35
Anderson T. W., Darling D. A., 1952, Ann. Math. Stat., 23, 193
Ando S., 2004, J. Cosmol. Astropart. Phys., 2004, 007
Armano M. et al., 2016, Phys. Rev. Lett., 116, 231101
Aso Y., Michimura Y., Somiya K., Ando M., Miyakawa O., Sekiguchi T., Tatsumi D., Yamamoto H., 2013, Phys. Rev. D, 88, 043007
Babak S. et al., 2017, Phys. Rev. D, 95, 103012
Baibhav V., Berti E., Gerosa D., Mapelli M., Giacobbo N., Bouffanais Y., Di Carlo U. N., 2019, Phys. Rev. D, 100, 064060
Belczynski K., Kalogera V., 2001, ApJ, 550, L183
Belczynski K., Taam R. E., Rantsiou E., van der Sluys M., 2008, ApJ, 682, 474
Benjamini Y., Yekutieli D., 2001, Ann. Stat., 29, 1165
Chen H.-Y., Holz D. E., 2014, preprint (arXiv:1409.0522)
Cutler C., Flanagan E., 1994, Phys. Rev. D, 49, 2658
D'Avanzo P. et al., 2018, A&A, 613, L1
de Freitas Pacheco J. A., Regimbau T., Vincent S., Spallicci A., 2006, Int. J. Mod. Phys. D, 15, 235
Dominik M., Belczynski K., Fryer C., Holz D. E., Berti E., Bulik T., Mandel I., O'Shaughnessy R., 2013, ApJ, 779, 72
Durand G., 2017, preprint (arXiv:1710.01094)
Dwyer S., Sigg D., Ballmer S. W., Barsotti L., Mavalvala N., Evans M., 2015, Phys. Rev. D, 91, 082001
Fishbach M., Holz D. E., Farr W. M., 2018, ApJ, 863, L41
Goldstein A. et al., 2017, ApJ, 848, L14
Helström C., 1968, Statistical Theory of Signal Detection, 2nd edn. International Series of Monographs in Electronics and Instrumentation, Vol. 9. Pergamon Press, Oxford, UK
Hogg D. W., 1999, preprint (arXiv:astro-ph/9905116)
Holm S., 1979, Scand. J. Stat., 6, 65
Hopkins A. M., Beacom J. F., 2006, ApJ, 651, 142
http://www.et-gw.eu/
Huang Y., Middleton H., Ng K. K. Y., Vitale S., Veitch J., 2018, Phys. Rev. D, 98, 123021
Iyer B. et al., 2011, LIGO-India, Technical Report No. LIGO-M1100296
Lazzati D., Perna R., Morsony B. J., López-Cámara D., Cantiello M., Ciolfi R., Giacomazzo B., Workman J. C., 2018, Phys. Rev. Lett., 120, 241103
Lipunov V. M., Postnov K. A., Prokhorov M. E., Panchenko I. E., Jorgensen H. E., 1995, ApJ, 454, 593
Lyman J. D. et al., 2018, Nat. Astron., 2, 751
Maggiore M., 2007, Gravitational Waves, Vol. 1: Theory and Experiments (Oxford Master Series in Physics). Oxford Univ. Press
Margutti R. et al., 2018, ApJ, 856, L18
Moore C. J., Gerosa D., Klein A., 2019, MNRAS, 488, L94
O'Shaughnessy R. W., Kalogera V., Belczynski K., 2008, ApJ, 675, 566
Regimbau T., Hughes S. A., 2009, Phys. Rev. D, 79, 062002
Resmi L. et al., 2018, ApJ, 867, 57
Rice S. O., 1945, Bell Syst. Tech. J., 24, 46
Ruan J. J., Nynka M., Haggard D., Kalogera V., Evans P., 2018, ApJ, 853, L4
Sathyaprakash B., Schutz B., 2009, Living Rev. Relativ., 12, 2
Sathyaprakash B. et al., 2012, Class. Quantum Gravity, 29, 124013
Schutz B. F., 2011, Class. Quantum Gravity, 28, 125023
Troja E. et al., 2018, MNRAS, 478, L18

Tutukov A. V., Yungelson L. R., 1994, MNRAS, 268, 871
Valenti S. et al., 2017, ApJ, 848, L24
Van Den Broeck C., 2014, J. Phys. Conf. Ser., 484, 012008
Vitale S., 2016, Phys. Rev. D, 94, 121501
Vitale S., Farr W. M., Ng K., Rodriguez C. L., 2019, Astrophys. J. Lett., 886, L1
Wainstein L. A., Zubakov V. D., 1962, Extraction of Signals from Noise. Prentice-Hall, Englewood Cliffs
Wanderman D., Piran T., 2015, MNRAS, 448, 3026
Wilkins S. M., Trentham N., Hopkins A. M., 2008, in Kodama T., Yamada T., Aoki K., eds, ASP Conf. Ser. Vol. 399, Panoramic Views of Galaxy Formation and Evolution. Astron. Soc. Pac., San Francisco, p. 178
Wright E. L., 2006, Publ. Astron. Soc. Pac., 118, 1711
http://www.virgo.infn.it

This paper has been typeset from a TEX/LATEX file prepared by the author.