

Probing the redshift evolution and sub-populations of binary neutron stars with the Einstein Telescope

A. Toubiana^{1,2,*} and I. Dvorkin^{3,4}

¹ Dipartimento di Fisica “G. Occhialini”, Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

² INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

³ Institut d’Astrophysique de Paris, UMR 7095, CNRS and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France

⁴ Institut Universitaire de France, Ministère de l’Enseignement Supérieur et de la Recherche, 1 rue Descartes, 75231 Paris Cedex F-05, France

Received 4 December 2025 / Accepted 14 May 2026

ABSTRACT

Aims. The formation channels of binary neutron stars (BNSs) currently remain uncertain, but important information can be gathered by observing their mergers with gravitational-wave detectors. The processes that lead to BNS coalescence are encoded in the time-delay distribution between stellar binary formation and BNS coalescence, and therefore in the BNS merger rate. Moreover, the detection of GW190425 by LIGO/Virgo/KAGRA (LVK) suggests a sub-population of massive BNSs, possibly formed through unstable ‘case BB’ mass transfer with short merger delays. We investigate whether next-generation detectors such as the Einstein Telescope (ET) can constrain the time-delay distribution of BNSs and identify such sub-populations.

Methods. Using the latest LVK constraints, we generated mock ET catalogues that contain a mixture of light and heavy sub-populations. We modelled the redshift distribution of each sub-population as the convolution of the cosmic star formation rate with a time-delay distribution. We first considered a scenario where the time-delay distribution is common to all BNSs and follows a power law with indices $\alpha = -0.5, -1, -1.5$. In the second scenario, heavy BNSs have fixed short delays, while light BNSs follow power-law delays with the same set of indices. Hierarchical Bayesian analyses were then performed on catalogues of 100–5000 events.

Results. With thousands of events, ET will be able to accurately characterise the time-delay distribution for the $\alpha = -0.5$ and $\alpha = -1$ cases. We find that with hundreds of detections from ET, we will be able to establish that the total mass distribution is bimodal. A few thousand events are sufficient to disentangle the redshift distributions of the two sub-populations for moderate time-delay indices ($\alpha_L = -0.5$ or -1). For steeper indices ($\alpha_L = -1.5$), the differences are more subtle and require larger catalogues, which was beyond what we could explore given our computational resources.

Conclusions. Next-generation detectors should enable the detection of multiple BNS sub-populations and their redshift evolution, and provide valuable insights into their formation pathways.

Key words. gravitational waves – binaries: general – stars: evolution – stars: formation – stars: neutron

1. Introduction

Direct observations of binary neutron stars (BNSs) come primarily from radio observations of binary pulsars in the Milky Way (Özel & Freire 2016)¹, and thus probe the local population. The mass distribution of neutron stars (NSs) in these systems is found to be much narrower than that of the overall NS population (Alsing et al. 2018; Antoniadis et al. 2016; Farr & Chatziioannou 2020; Shao et al. 2020), and can be well described by a Gaussian distribution centred at $1.33 M_{\odot}$ with a standard deviation of $0.09 M_{\odot}$ (Özel et al. 2012; Özel & Freire 2016), although there is some evidence of bimodality between recycled and non-recycled pulsars, the former having slightly larger masses (Farrow et al. 2019). This particularity of double radio pulsars could be attributed to the fact that the components of BNSs undergo much less accretion than neutron stars in other types of binaries, or to different masses at birth (Tauris et al. 2017).

Gravitational waves (GWs) provide a new means to directly observe these systems. The LIGO/Virgo/KAGRA (LVK) collaboration has so far confidently reported the detection of two BNSs: GW170817 (Abbott et al. 2017) and GW190425 (Abbott et al. 2020), at $z \sim 0.009$ and $z \sim 0.03$, respectively. Next-generation ground-based observatories such as the Einstein Telescope (ET) (Branchesi et al. 2023; Abac et al. 2026) and Cosmic Explorer (CE) (Evans et al. 2021, 2023), expected to come online by the end of the next decade, should detect thousands to tens of thousands of BNSs out to redshift $z \sim 3$. The redshift distribution of BNS mergers encodes the convolution of the cosmic star formation rate (SFR) with the distribution of time delays between binary formation and coalescence. As such, it provides a direct probe of the physical processes that govern BNS formation. The orbital separation at the formation of the BNS, and therefore the time delay to coalescence, is determined by various stellar evolution processes, including mass transfer episodes, common envelope phases, and supernova explosions with associated natal kicks (e.g. Chruslinska et al. 2018; Giacobbo & Mapelli 2018; Vigna-Gómez et al. 2018; Neijssel et al. 2019; Santoliquido et al. 2021). Observations across a broad redshift range with next-generation detectors will therefore enable

* Corresponding author: alexandre.toubiana@unimib.it

¹ In this paper, BNS refers to binary systems where both components are neutron stars.

a reconstruction of the merger rate density and place strong constraints on the underlying time-delay distribution, offering a powerful means to discriminate between formation channels. These observations will also enable tests for the presence of multiple BNS sub-populations, as current data may already suggest. While the component masses of GW170817 are consistent with those of the Galactic population, GW190425 appears significantly heavier. This discrepancy is particularly evident in terms of the total mass: $3.4^{+0.03}_{-0.01} M_{\odot}$ for GW190425, compared to $2.66 \pm 0.12 M_{\odot}$ for Galactic BNSs, which potentially points to the existence of a distinct, more massive sub-population.

This intriguing observation naturally raises the question of how GW190425 formed and why such heavy BNSs are not observed in the Milky Way. Dynamical formation in stellar clusters could, in principle, produce heavier BNSs if a massive neutron star exchanges its stellar companion with another neutron star; however, this channel is expected to contribute negligibly to the overall BNS merger rate (Ye et al. 2020; Belczynski et al. 2018). Focusing on isolated binary evolution, some studies suggest that GW190425 could have formed through pathways similar to those of Galactic BNSs (Qin et al. 2024; Nair & Stevenson 2025), and that the spins and associated magnetic fields of heavy BNSs may render them effectively radio-invisible (Safarzadeh et al. 2020; Chu et al. 2025). Alternatively, Romero-Shaw et al. (2020) proposed that massive binaries like GW190425 may originate from a distinct evolutionary channel characterised by much shorter delay times between the formation of the progenitor stars and BNS merger, potentially explaining their scarcity in the Galaxy. In this scenario, the system undergoes unstable ‘case BB’ mass transfer from a helium star onto the first-formed neutron star (Belczynski et al. 2002; Dewi & Pols 2003; Tauris et al. 2017; Vigna-Gómez et al. 2018; Iorio et al. 2023), triggering a second common-envelope phase between the neutron star and the remaining CO core. The resulting tighter orbital separation allows the binary to survive the second supernova kick and merge within a few million years. Following this idea, Galaudage et al. (2021) jointly fitted the radio and GW BNS populations and found mild evidence that GW190425 could belong to a fast-merging population comprising 8–79% of all BNSs, with associated delay times between 5 and 401 Myr (assumed to be the same for all heavy BNSs). We stress that other studies also find that such a discrepancy between the Milky Way and extragalactic populations can naturally arise in population synthesis, without the need for a distinct time-delay distribution (Kruckow 2020). As the exact pathway to produce GW190425 is still open to debate, in this work we explore the consequence of the scenario proposed by Romero-Shaw et al. (2020) for future observations. If this hypothesis is correct, we could expect to observe two sub-populations of BNSs with GWs: a light population and a heavy one, potentially exhibiting distinct redshift evolutions, which ET and CE could be able to probe.

Using the latest constraints from the LVK on the local BNS merger rate (The LIGO Scientific Collaboration 2025), we constructed mock catalogues of BNS observations on which we performed hierarchical Bayesian analyses to reconstruct the astrophysical distributions. First, considering the case of a single population, we illustrate the accuracy with which the time-delay distribution can be reconstructed. We then turn to the possibility of multiple sub-populations. We find that approximately 500 events are sufficient to identify the presence of two sub-populations, while a few thousand detections allow us to distinguish their redshift distributions; this highlights the

potential of next-generation facilities to constrain BNS formation channels.

2. Binary neutron star population

Our population model builds upon the work of Farrow et al. (2019) and Galaudage et al. (2021). We emphasise that it is not directly informed by binary stellar evolution calculations, but is designed to capture the phenomenology relevant to the case studied in this work. In Galaudage et al. (2021), BNS systems are described as pairs composed of a recycled and a slowly spinning NS, each drawn from either a heavy (H) or a light (L) sub-population. Here, we further assumed that BNSs form preferentially as light and light or heavy and heavy, as appears to be the case for GW170817 and GW190425. Under this assumption, the individual NS masses, m_1 and m_2 , associated with the heavy/light sub-populations are sampled from Gaussian distributions with means and standard deviations of $(\mu_{1,H/L}, \sigma_{1,H/L})$ and $(\mu_{2,H/L}, \sigma_{2,H/L})$, respectively. The numerical values of these parameters, adopted from Galaudage et al. (2021), are listed in Table 1.

In this formulation, no explicit ordering between m_1 and m_2 is imposed. To prevent labelling ambiguities that may arise in the GW context, we reparametrised the model in terms of the total mass $m_t = m_1 + m_2$ and the mass ratio $q = \min(m_1, m_2)/\max(m_1, m_2) \leq 1$. The total mass is assumed to follow a normal distribution with a mean and standard deviation of $\mu_{t,H/L} = \mu_{1,H/L} + \mu_{2,H/L}$ and $\sigma_{t,H/L} = \sqrt{\sigma_{1,H/L}^2 + \sigma_{2,H/L}^2}$, respectively. The mass ratio is modelled as a truncated log-normal distribution between 0 and 1, characterised by parameters $\mu_{q,H/L}$ and $\sigma_{q,H/L}$. The parameters of the distribution were chosen to qualitatively reproduce the population of Galaudage et al. (2021), as we show that it is indeed the case in Appendix A. If we denote the fraction of light BNS with λ and use $\Lambda = (\lambda, \Lambda_{t,L}, \Lambda_{q,L}, \Lambda_{z,L}, \Lambda_{t,H}, \Lambda_{q,H}, \Lambda_{z,H})$ to denote the full set of population parameters, i.e. the hyperparameters, our population model reads

$$p(m_t, q, z|\Lambda) = \lambda \times p(m_t|\Lambda_{t,L})p(q|\Lambda_{q,L})p(z|\Lambda_{z,L}) \\ + (1 - \lambda) \times p(m_t|\Lambda_{t,H})p(q|\Lambda_{q,H})p(z|\Lambda_{z,H}). \quad (1)$$

The redshift distribution is given by

$$p(z|\Lambda_{z,H/L}) \propto \frac{1}{1+z} \frac{dV_c}{dz} \mathcal{R}_{H/L}(z). \quad (2)$$

Distributions were normalised between $z_{\min} = 0$ and $z_{\max} = 3$, as we found that almost no BNS event is detected past redshift 3 with ET. In the above equation, $\frac{dV_c}{dz}$ is the differential comoving volume, and $\mathcal{R}_{H/L}(z)$ is the volumetric rate of events. The latter depends on the cosmic SFR, the BNS formation efficiency, $\eta_{H/L}$, and the time-delay distribution, $p_{H/L}(\Delta t)$:

$$\mathcal{R}_{H/L}(z) = \int \eta_{H/L}(t(z) + \Delta t) \text{SFR}(t(z) + \Delta t) p_{H/L}(\Delta t) d\Delta t. \quad (3)$$

We considered two distinct scenarios for the time-delay distribution. First, we considered the case of a single BNS population, where we assumed a unique time-delay distribution that follows a power law with index α , bounded between $\Delta t_{\min} = 10$ Myr and $\Delta t_{\max} = 13$ Gyr. Although our mass model assumed a bimodal distribution to capture the high-mass end, our conclusions on the reconstruction of the time-delay distribution

Table 1. Parameters of our population models.

Parameter	Value
$\mu_{1,L}$	$1.34 M_{\odot}$
$\sigma_{1,L}$	$0.02 M_{\odot}$
$\mu_{2,L}$	$1.29 M_{\odot}$
$\sigma_{2,L}$	$0.09 M_{\odot}$
$\mu_{1,H}$	$1.47 M_{\odot}$
$\sigma_{1,H}$	$0.15 M_{\odot}$
$\mu_{2,H}$	$1.80 M_{\odot}$
$\sigma_{2,H}$	$0.15 M_{\odot}$
λ	0.50
$\lambda_{\mathcal{R}}$	0.69; 0.56; 0.49
$\Lambda_{t,L}$	
$\mu_{t,L}$	$2.63 M_{\odot}$
$\sigma_{t,L}$	$0.092 M_{\odot}$
$\Lambda_{t,H}$	
$\mu_{t,H}$	$3.27 M_{\odot}$
$\sigma_{t,H}$	$0.21 M_{\odot}$
$\Lambda_{q,L}$	
$\mu_{q,L}$	1.00
$\sigma_{q,L}$	0.08
$\Lambda_{q,H}$	
$\mu_{q,H}$	0.82
$\sigma_{q,H}$	0.20
$\Lambda_{z,L}$	
$\Delta t_{\min,L}$	10 Myr
$\Delta t_{\max,L}$	13 Gyr
α_L	-0.5; -1; -1.5
$\Lambda_{z,H}$	
Δt_H	30 Myr

Notes. The first eight lines list the parameters of the population model of Galaudage et al. (2021), which inspired our own. We then report the values of the hyperparameters used to generate the mock populations in our study, grouped by category. For completeness, we also provide the values of $\lambda_{\mathcal{R}}$, when considering two distinct sub-populations, that correspond to the three choices of α_L . When assuming a common redshift distribution, we use the same parameters as for the light sub-population.

are largely insensitive to the precise form of the mass model. For instance, adopting a skewed distribution with an extended high-mass tail would yield comparable results, provided that the abundance of heavy systems is similar. For simplicity, we therefore adopted a single model for the mass distribution, namely the one described above.

In the second scenario, following Romero-Shaw et al. (2020) and Galaudage et al. (2021), we assumed that the heavy sub-population merges on systematically shorter timescales than the light one. In this case, heavy binaries were assigned a fixed, short delay time of $\Delta t_H = 30$ Myr, modelled as $p_H(\Delta t) = \delta(\Delta t - \Delta t_H)$, while light binaries were assumed to follow a power-law delay-time distribution with index α_L , bounded between $\Delta t_{L,\min} = 10$ Myr and $\Delta t_{L,\max} = 13$ Gyr. In both cases we found that values of the lower cutoff of the power-law time-delay distribution, $\Delta t_{L,\min}$, and/or the fixed delay of the heavy population, Δt_H ,

below ~ 100 Myr have a negligible impact on the resulting redshift distributions. Hence, the results presented below are largely insensitive to the precise choice of these parameters.

We adopt the cosmic SFR from Madau & Dickinson (2014). The formation efficiency of each sub-population was assumed to be constant with redshift and thus acted as a normalisation factor. The overall normalisation was constrained by the total BNS merger rate inferred by the LVK Collaboration (The LIGO Scientific Collaboration 2025), while the relative contributions of each sub-population were set by the choice of the mixing fraction λ . We emphasise that the mixing fraction entering the probability density function, λ , differs from the one entering the volumetric rate, $\lambda_{\mathcal{R}}$. If we define the latter such that $\mathcal{R}(z) = \lambda_{\mathcal{R}} \mathcal{R}_L(z) + (1 - \lambda_{\mathcal{R}}) \mathcal{R}_H(z)$, the two are related by

$$\lambda_{\mathcal{R}} = \frac{\lambda \int_{z_{\min}}^{z_{\max}} \frac{\mathcal{R}_H(z)}{1+z} \frac{dV_c}{dz} dz}{\int_{z_{\min}}^{z_{\max}} \frac{\lambda \mathcal{R}_L(z) + (1-\lambda) \mathcal{R}_H(z)}{1+z} \frac{dV_c}{dz} dz}. \quad (4)$$

Here, we assumed an equal contribution from both populations, i.e. $\lambda = 0.5$. We summarise the parameters of our population in Table 1. For completeness, we also report there the corresponding values of $\lambda_{\mathcal{R}}$. Our values are compatible with the estimates of Galaudage et al. (2021) that 8–79% of BNS at birth are rapidly merging (in terms of $\lambda_{\mathcal{R}}$).

Since the time-delay distribution is poorly constrained, we explored different assumptions for the delay-time distribution of the slowly merging (light) population. Specifically, we fixed the minimum and maximum values of the time-delay range and vary the power-law index, both in the single population and distinct sub-population cases. Population synthesis models typically find $\alpha = -1$ (e.g. Chruslinska et al. 2018), although at specific progenitor metallicities the distribution can be much shallower (de Sá et al. 2024; Pellouin et al. 2025), while observational-based estimates that use gamma-ray bursts or Galactic pulsars find much steeper values of $\alpha = -1.5$ to -2 for some sub-populations (Zevin et al. 2022; Maoz & Nakar 2025). In this work we considered $\alpha = -0.5, -1$, and -1.5 . The corresponding volumetric merger rates are shown in Fig. 1.

In each case, the total rate is normalised to the mean of the 90% credible interval on the local BNS merger-rate range reported by the LVK (The LIGO Scientific Collaboration 2025), between 7.6 and 250 $\text{Gpc}^{-3} \text{ yr}^{-1}$, i.e. 129 $\text{Gpc}^{-3} \text{ yr}^{-1}$. Using the value of $1.16 \times 10^{-2} \text{ Mpc}^{-3}$ for the number density of Milky Way-equivalent galaxy (MWEG) (Kopparapu et al. 2008), this rate corresponds to a merger rate per MWEG of 0.65–22 $\text{MWEG}^{-1} \text{ Myr}^{-1}$. The grey band at $z = 0$ shows the credible interval reported by the LVK. The short-delay population is rescaled by a different normalisation factor for each model, since for a fixed λ the corresponding volumetric-rate mixing fraction $\lambda_{\mathcal{R}}$ depends on α_L . Larger values of α_L assign greater weight to long delays, thereby enhancing the contrast with the short-delay population.

3. Mock catalogues and data analysis

We considered a triangular configuration for ET with a 10 km arm length and adopt the ET-10 (10 km-xylophone) sensitivity curve (Danilishin & Zhang 2023) implemented in GWBENCH (Borhanian 2021). Our mock detection and parameter estimation pipeline was based on the approach outlined in Fishbach et al. (2020) and Farah et al. (2023). Differences arising in our implementation are discussed in Appendix B.

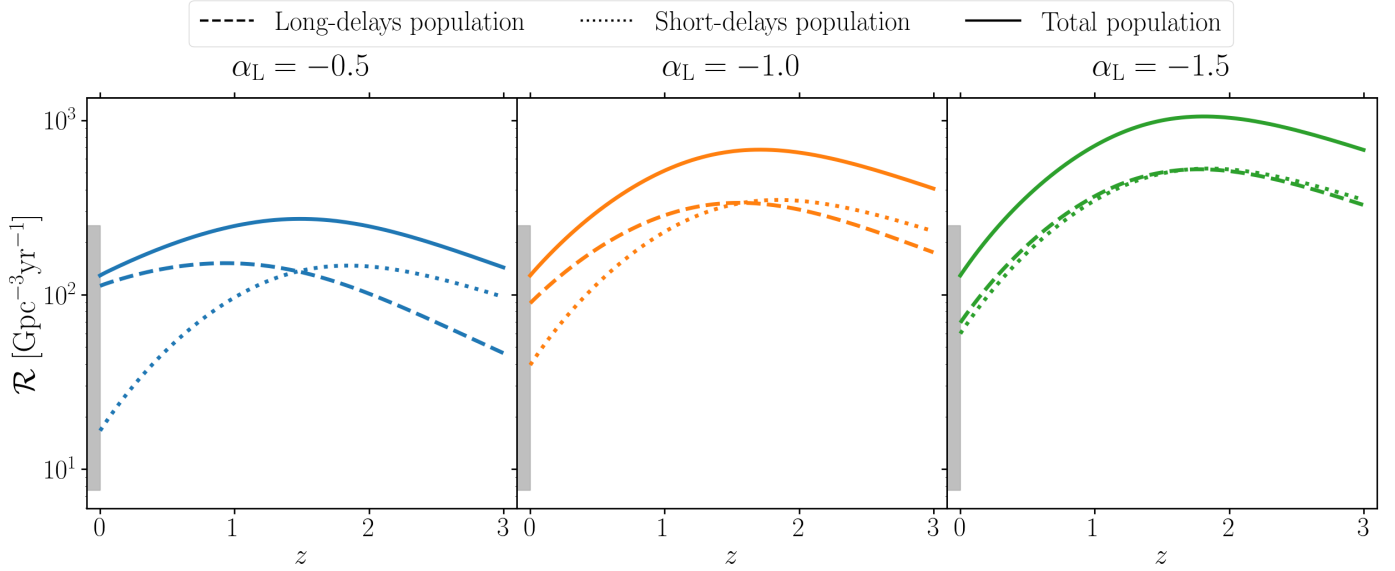


Fig. 1. Merger rate of the long delays population (dashed lines), short delays population (dotted lines), and total population (full line) for three different hypotheses of the time-delays distribution. The rate is normalised at $z = 0$ to the mean of the interval reported by the LVK following GWTC-4 (The LIGO Scientific Collaboration 2025). The grey band at $z = 0$ shows the 90% credible interval reported by the LVK. In the case of a single population, the merger rate follows the long-delay curve, with the appropriate normalisation to match the total event rate.

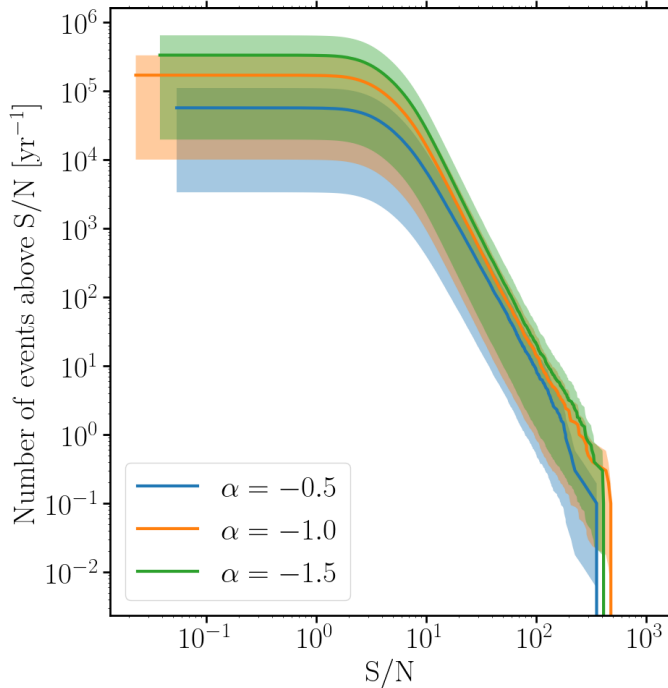


Fig. 2. Number of events above a given S/N threshold per year in the single population scenario for the three values of α considered in this work.

Figure 2 shows the cumulative number of detections as a function of signal-to-noise ratio (S/N) threshold for the three values of α in the single population case. The shaded regions represent the uncertainty associated with the local BNS merger rate reported by the LVK (The LIGO Scientific Collaboration 2025), corresponding to the lower and upper bounds of the 90% credible interval. If

we assume a S/N threshold of eight in the single population scenario, we estimate that ET will detect between approximately 680–22 310, 1710–56 280, and 3090–101 630 BNS mergers per year for $\alpha = -0.5$, -1 , and -1.5 , respectively. In the case of two sub-populations, we estimate these numbers to be 1180–38 920, 2370–77 950, and 3320–109 150 BNS mergers per year for $\alpha_L = -0.5$, -1 , and -1.5 , respectively. In all cases, $\sim 93\%$ of detected BNS are within redshift 2. These estimates are compatible with Branchesi et al. (2023) after accounting for the fact that they assumed a local merger rate of $250 \text{ Gpc}^{-3} \text{ yr}^{-1}$, which lies at the higher end of our range.

Based on these projections, we generated mock catalogues containing 100, 500, 1000, and 5000 detected events, with ten realisations for each case. We then performed hierarchical Bayesian inference to recover the population hyperparameters. These catalogue sizes are consistent with the expected ET detection rate, towards its lower end, while remaining computationally tractable. Current computational resources, in terms of both processing time and memory requirements, prevent us from extending the analysis to the tens of thousands of detections that ET is expected to observe. For this reason, we restricted our study to catalogues of up to 5000 events. The hierarchical Bayesian framework is described in detail in Appendix C; here, we summarise the population modelling choices relevant to this work.

Our population prior is given by a mixture model as in Eq. (1). The component distributions in total mass and mass ratio are modelled as Gaussian and truncated log-normal distributions, respectively. We adopted flat priors on the mixture fraction, as well as on the means and standard deviations of these distributions. For the redshift dependence, we explore two complementary approaches.

We assumed that the redshift evolution of each sub-population follows Eqs. (2) and (3). In the inference, we modelled the delay-time distribution of both sub-populations as a power law, i.e. we do not assume any of them to have a fixed

delay time. The delta-function limit is recovered when $\Delta t_{\max} \rightarrow \Delta t_{\min}$. The priors adopted are:

- $\Delta t_{\min}$: log-flat between 1 Myr and 1 Gyr;
- $\Delta t_{\max} - \Delta t_{\min}$: log-flat between 10^{-3} Myr and 13.5 Gyr²;
- α : flat between -3 and 0 .

When investigating the ability of ET to identify distinct BNS sub-populations, after performing the hierarchical Bayesian inference, we computed the probability that the mass distribution is bimodal, p_{bimodal} , as follows. For each hyperposterior sample, we draw 2000 (m_t, q) samples and apply Hartigan's dip test (Hartigan & Hartigan 1985) using the implementation of Urlus (2025). This test quantifies whether the empirical probability distribution function of the samples exhibits a 'dip,' which would indicate the presence of two distinct modes. The test returns a p value that can be interpreted as the probability that the distribution is unimodal. Repeating this for all hyperposterior samples yields a distribution of p values for each hierarchical Bayesian analysis. We then computed p_{bimodal} as the fraction of p values below 0.05, which provides an estimate of the probability that the mass distribution is bimodal with more than 95% confidence. We verified that the number of (m_t, q) samples drawn per hyperparameter sample was sufficient, in the sense that increasing it does not affect the results. Since Hartigan's dip test is defined for one-dimensional data, we rotated the (m_t, q) samples for each hyperparameter sample using a rotation matrix, and retained the rotation that yields the lowest p value (i.e. the highest probability of bimodality). This procedure allowed us to identify which linear combination of m_t and q exhibits the strongest bimodality. In practice, we find that the angle that maximises the bimodality was approximately zero, which indicates that the total mass distribution itself is the main source of bimodality.

The probability that the two redshift distributions are different, p_{dist} , was estimated as follows. We randomly drew pairs of hyperparameters from the posterior samples and computed the Kolmogorov–Smirnov (KS) statistic between the resulting redshift distributions of the two sub-populations. This 'between-population' KS distribution captures the typical differences between the two sets of distributions. To construct a background distribution, we also computed two 'within-population' KS distributions by comparing different samples from the same sub-population, which provides a reference for variations expected from measurement uncertainty (detector noise and a finite number of events). We then determined the 5% quantile of the between-population KS distribution and find its corresponding quantiles in the within-population KS distributions. The maximum of these two quantiles is defined as p_{dist} . It measures whether the difference between the sub-populations exceeds the typical measurement-induced variations in at least one sub-population, indicating that the distributions can be reliably distinguished.

4. Results

4.1. Single population

We first discuss the results in the case of a common redshift distribution, focusing on how well the parameters of the time-delay distribution can be constrained. Figure 3 shows the uncertainty on the power-law index, α , as a function of the number of events.

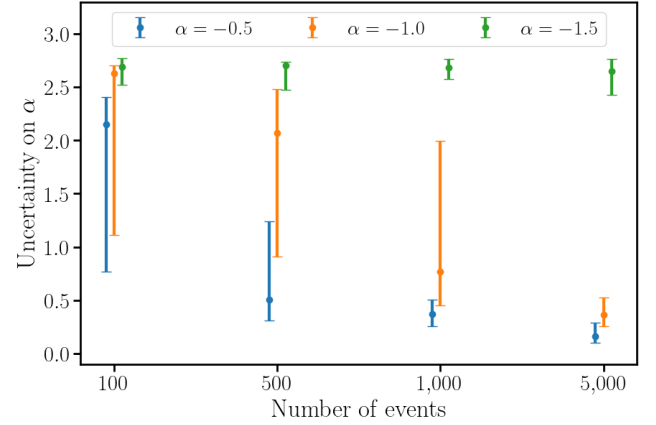


Fig. 3. Credible intervals on the width of the 90% credible interval for the power-law exponent, as a function of the number of events and for different values of α .

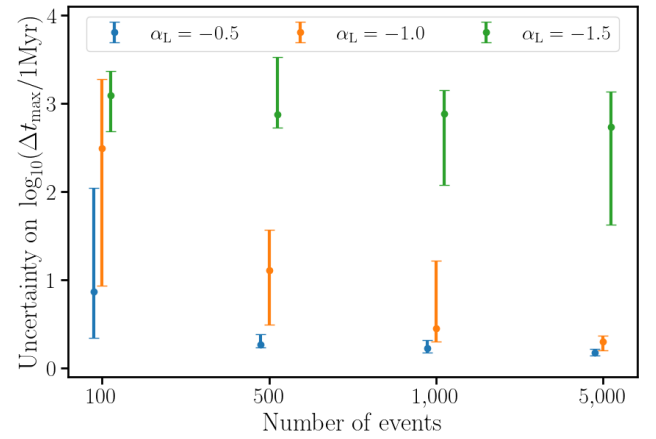


Fig. 4. Credible intervals on the width of the 90% credible interval for the maximum time delay, as a function of the number of events and for different values of α .

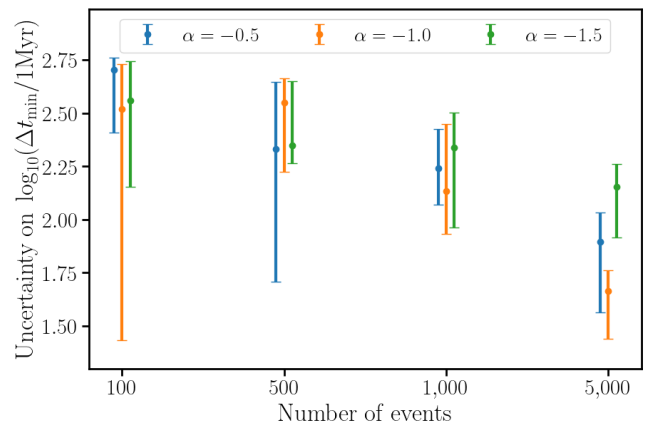


Fig. 5. Credible intervals on the width of the 90% credible interval for the minimum time delay, as a function of the number of events and for different values of α .

² The limit $\Delta t_{\max} = \Delta t_{\min}$ is not included in our prior; however, intervals shorter than 1 Myr are in practice hardly distinguishable from a delta function.

We find that, with 1000 observations, α can already be constrained within ~ 0.5 for an injected value of $\alpha = -0.5$, while a similar accuracy is achieved with ~ 5000 events for $\alpha = -1.0$.

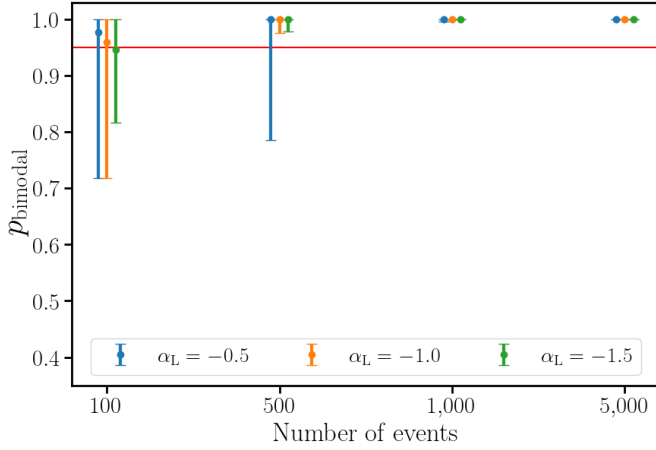


Fig. 6. Credible intervals on the confidence that the mass distribution is bimodal as a function of the number of events and for the different values of α_L . The horizontal red line corresponds to a probability of 0.95.

For $\alpha = -1.5$, however, the power-law index remains essentially unconstrained even after 5000 events. In this case, the redshift distribution closely resembles that obtained for a fixed delay, as shown in the right panel of Fig. 1. This behaviour is reflected in the large uncertainty on the maximum time delay, shown in Fig. 4, which remains uncertain at the level of several Gyr even after 5000 observations.

In contrast, for the other scenarios, the maximum delay time is very well constrained, within a few Myr, as it strongly affects the number of sources at $z \sim 0$, where the observational constraints are tightest. The minimum time delay is constrained to within a few hundred Myr in all cases, as shown in Fig. 5, consistent with our observation in Sect. 2 that varying $\Delta t_{\min}$ by a few Myr has little impact on the results.

4.2. Sub-populations

We start by discussing the recovery of the mass distribution. With 100 events, we can already confidently measure that $\lambda \neq 0$ and $\lambda \neq 1$, which indicates that the population cannot be described by a single component. In Appendix D we show examples of how the total mass and mass ratio distributions are recovered as the number of events increases. Figure 6 displays the 90% confidence interval of p_{bimodal} (centred on the median) as a function of the number of events. With 500 detections, we can already establish in most cases that the total mass distribution is bimodal, and with 1000 detections this can be determined unambiguously, regardless of α_L .

Next, we turn to the distinction between the two redshift distributions. The upper panel of Fig. 7 shows representative examples of the reconstructed redshift distributions for $\alpha_L = -1$ as the number of events increases. The lower panel shows the distribution of KS statistics, both between-population and within-population. The value above the bottom panel indicates p_{dist} . This figure provides a visual interpretation of our method for estimating the probability that the two redshift distributions differ: as the confidence bands overlap less, the between-population KS distribution is moved towards higher values than the within-population ones, and the computed p_{dist} increases. We observe that the redshift distribution of the heavy population is

typically better constrained than that of the light population, as a consequence of their larger S/N.

Our results for p_{dist} are shown in Fig. 8. For $\alpha_L = -0.5$, we can already establish that the two redshift distributions differ with more than 95% confidence using 500 events. For $\alpha_L = -1$, this level of significance is reached in over 95% of realisations with 5000 events. For $\alpha_L = -1.5$, the distinction becomes more challenging, and it appears that many more events are needed in order to distinguish the distributions, in agreement with the results found in the single population case.

5. Conclusions

Despite major progress over the last decade, the formation pathways of BNSs remain uncertain. In particular, the number of mass transfer episodes and the criterion for their stability have a strong influence on the properties of the resulting population yet are still not fully understood. The redshift distribution of BNS mergers provides an observational handle on BNS formation processes, since it encodes the distribution of time delays between binary formation and coalescence and therefore the initial BNS orbital separation. Moreover, BNS masses could also be correlated with the formation channel. The detection of GW190425 by LVK highlighted the possibility that a significant fraction of BNSs may form through unstable ‘case BB’ mass transfer, resulting in much shorter delay times than the rest of the population (Romero-Shaw et al. 2020; Galaudage et al. 2021). In this scenario, two BNS sub-populations with distinct redshift evolutions could coexist. In this paper, we explore the ability of future ground-based facilities such as the ET to characterise the time-delay distribution of BNSs and to probe the existence of distinct sub-populations.

Using the latest estimate of the BNS merger rate by the LVK (The LIGO Scientific Collaboration 2025), we generated mock BNS populations consisting of an equal mixture of light and heavy sub-populations, inspired by the parametric models of Farrow et al. (2019) and Galaudage et al. (2021). Assuming that the redshift distribution of each sub-population is given by the convolution of the SFR with a time-delay distribution, we considered two scenarios. In the first, both light and heavy BNSs share a common time-delay distribution described by a power law, with index $\alpha_L = -0.5, -1$, and -1.5 . In the second, heavy BNSs are assigned a fixed, short time delay, while light BNSs follow a power-law distribution with the same set of indices. Independently of this choice, we find that ET should observe at least a few thousand BNS mergers over its lifetime, with estimated detection rates ranging from 1710 to 56280 events per year in the fiducial $\alpha_L = -1$ case for a common redshift distribution.

We then considered multiple realisations of catalogues ranging from 100 up to 5000 events and performed hierarchical Bayesian analyses to reconstruct the astrophysical distributions. Crucially, when considering the possibility of having two distinct sub-populations, we did not constrain either sub-population to a fixed delay time in the inference, allowing for a more general and flexible model. We find that for $\alpha = -0.5$ and -1 , the power-law index can be constrained within ~ 0.5 after 1000 and 5000 observations, respectively. The minimum delay time is constrained within a few hundred megayears, while the maximum delay time can be determined to within a few megayears. For $\alpha = -1.5$, however, the parameters remain poorly constrained, as the resulting redshift distribution closely resembles that obtained for a fixed delay of ~ 100 Myr.

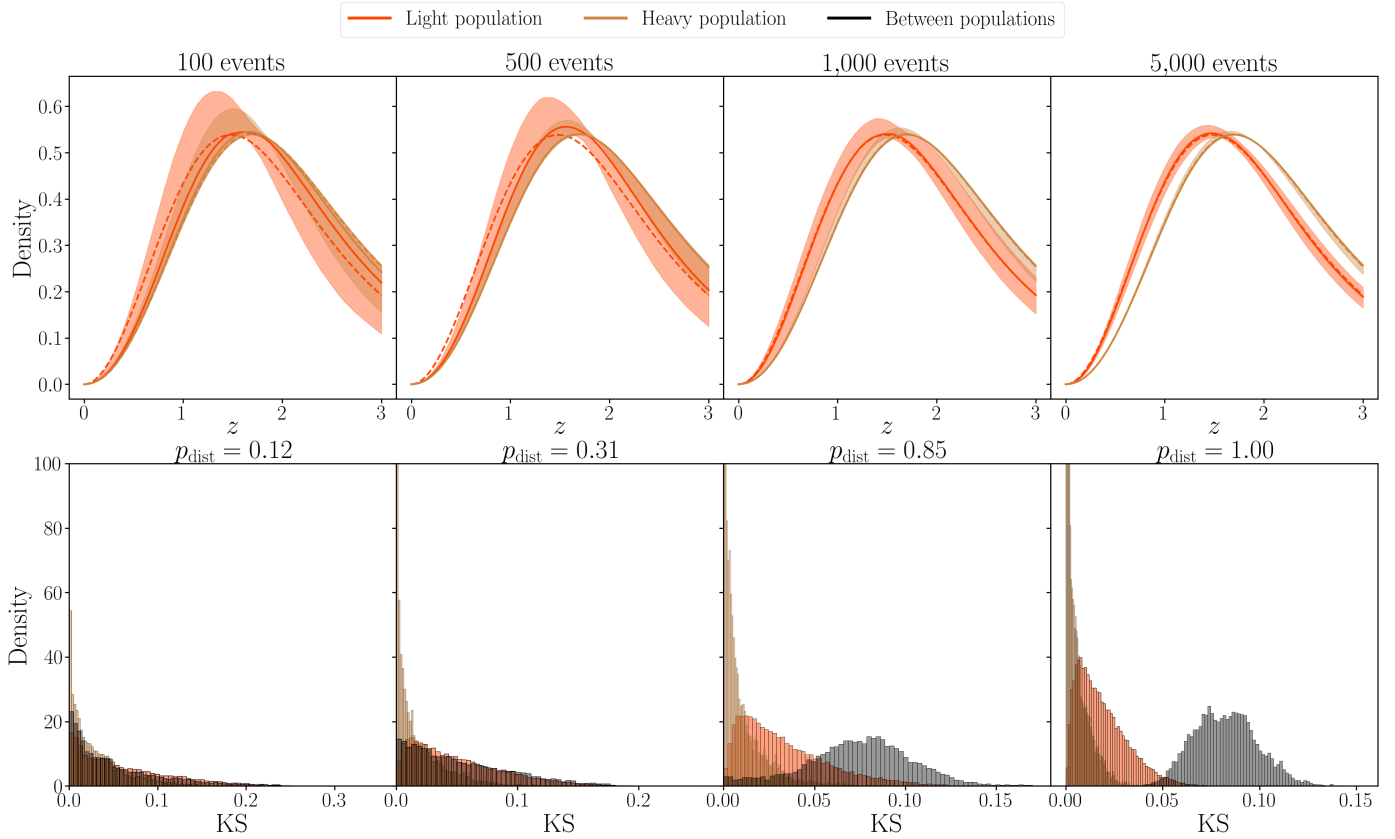


Fig. 7. Upper panel: reconstructions of the redshift distributions for increasing catalogue sizes. Coloured bands show the 90% credible intervals for the light (red) and heavy (golden) populations. Solid lines indicate the median reconstructions, while dashed lines mark the true distributions. For the heavy population, the dashed and full lines superimpose almost perfectly. Lower panel: distribution of within-population and between-population KS statistics. The value on top shows the probability that the two heavy and light population have different redshift distributions.

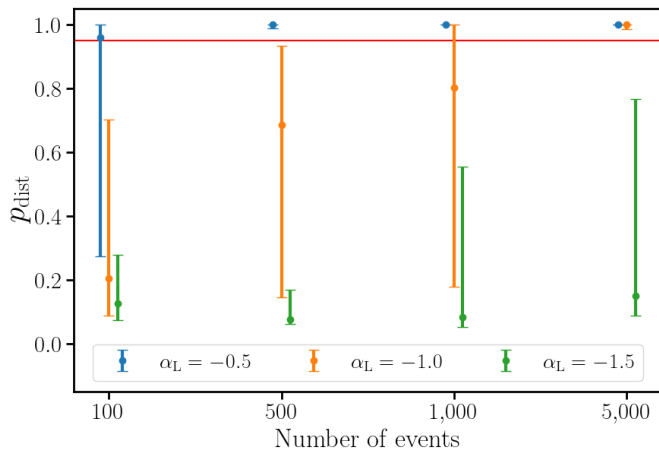


Fig. 8. Credible intervals on the confidence with which we can determine that the redshift distribution of the light and of the heavy population are different as a function of the number of events and for the different values of α_L . The horizontal red line corresponds to a probability of 0.95.

In the case of two sub-populations, for all α_L values, we find that with 500 events we can identify the bimodality of the BNS total mass distribution in most cases, and with 1000 events we can do so in practically all cases. Moreover, a few thousand events are sufficient to disentangle the redshift distributions of

the two sub-populations if $\alpha_L = -0.5$ or -1 . For $\alpha_L = -1.5$, the difference between the two distributions is very small, and more observations would be required to distinguish them, if at all. We could not explore this further due to limited computational resources. This highlights a crucial point: current population analysis frameworks will face significant challenges when the number of observations reaches several thousand.

Our study shows that next-generation ground-based detectors will provide information on the processes leading to BNS coalescence, and could reveal a sub-population of heavy BNSs merging on much shorter timescales than the Galactic population. The most likely scenario behind it would be unstable ‘case BB’ mass transfer, which tightens the binary orbit, allowing the system to remain bound after the second supernova and resulting in shorter merger delays. Detecting such a population would provide insight into BNS formation channels, accessible only through GW observations, since rapidly merging heavy BNSs would be largely absent in the Milky Way, explaining the lack of Galactic binaries as massive as GW190425. We emphasise once again that, although this scenario provides a viable explanation for the discrepancy between GW190425 and the Galactic population, it is not the only one capable of doing so. Alternatively, if heavy BNSs exist in our Galaxy but are radio-invisible (Safarzadeh et al. 2020; Chu et al. 2025), the Laser Interferometer Space Antenna (LISA) should enable them to be found in the Milky Way (Korol & Safarzadeh 2021).

An important caveat of our analysis is that it was limited to two sub-populations. For instance, allowing for a light and

heavy coupling would introduce a third sub-population, although the (currently limited) data tend to favour mostly heavy and heavy and light and light coupling. Moreover, several formation pathways may contribute to the BNS population (Tauris et al. 2017; Vigna-Gómez et al. 2018), potentially resulting in a more complex distribution. Neglecting this could severely bias population reconstruction when using parametric or astrophysical models for the population prior (Zevin et al. 2021; Toubiana et al. 2021; Cheng et al. 2023). Future work will explore more comprehensive analyses that combine population synthesis models (Pellouin et al. 2025) with multi-dimensional non-parametric approaches, such as that of Tenorio et al. (2025).

Acknowledgements. We are thankful to S. Borhanian, T. Bruel and M. Quartin for fruitful discussions. A.T. is supported by MUR Young Researchers Grant No. SOE2024-0000125, ERC Starting Grant No. 945155–GWmining, Cariplo Foundation Grant No. 2021-0555, MUR PRIN Grant No. 2022-Z9X4XS, Italian-French University (UIF/UFJ) Grant No. 2025-C3-386, MUR Grant “Progetto Dipartimenti di Eccellenza 2023-2027” (BiCoQ), and the ICSC National Research Centre funded by NextGenerationEU.

References

- Abac, A., Abramo, R., Albanesi, S., et al. 2026, *J. Cosmology Astropart. Phys.*, **2026**, 081
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, *Phys. Rev. Lett.*, **119**, 161101
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2020, *ApJ*, **892**, L3
- Alsing, J., Silva, H. O., & Berti, E. 2018, *MNRAS*, **478**, 1377
- Antoniadis, J., Tauris, T. M., Özel, F., et al. 2016, arXiv e-prints [arXiv:1605.01665]
- Belczynski, K., Kalogera, V., & Bulik, T. 2002, *ApJ*, **572**, 407
- Belczynski, K., Askar, A., Arca-Sedda, M., et al. 2018, *A&A*, **615**, A91
- Borhanian, S. 2021, *Class. Quantum Gravity*, **38**, 175014
- Branchesi, M., Maggiore, M., Alonso, D., et al. 2023, *J. Cosmology Astropart. Phys.*, **2023**, 068
- Cheng, A. Q., Zevin, M., & Vitale, S. 2023, *ApJ*, **955**, 127
- Chruslinska, M., Belczynski, K., Klencki, J., & Benacquista, M. 2018, *MNRAS*, **474**, 2937
- Chu, Q., Lu, Y., & Yu, S. 2025, *ApJ*, **980**, 181
- Danilishin, S., & Zhang, T. 2023, ET sensitivity curves used for CoBA Science Study, Technical Report ET-0304B-22, Einstein Telescope Collaboration
- de Sá, L. M., Rocha, L. S., Bernardo, A., Bachega, R. R. A., & Horvath, J. E. 2024, *MNRAS*, **535**, 2041
- Dewi, J. D. M., & Pols, O. R. 2003, *MNRAS*, **344**, 629
- Essick, R. 2023, *Phys. Rev. D*, **108**, 043011
- Evans, M., Adhikari, R. X., Afle, C., et al. 2021, arXiv e-prints [arXiv:2109.09882]
- Evans, M., Corsi, A., Afle, C., et al. 2023, arXiv e-prints [arXiv:2306.13745]
- Farah, A. 2022, GWMockCat: A lightweight code to generate mock catalogs of gravitational-wave sources, <https://git.ligo.org/amanda.farah/GWMockCat>, gitLab repository, release v1 (May 03, 2022)
- Farah, A. M., Edelman, B., Zevin, M., et al. 2023, *ApJ*, **955**, 107
- Farr, W. M., & Chatziioannou, K. 2020, *Res. Notes Am. Astron. Soc.*, **4**, 65
- Farrow, N., Zhu, X.-J., & Thrane, E. 2019, *ApJ*, **876**, 18
- Finn, L. S., & Chernoff, D. F. 1993, *Phys. Rev. D*, **47**, 2198
- Fishbach, M., Farr, W. M., & Holz, D. E. 2020, *ApJ*, **891**, L31
- Galaudage, S., Adamcewicz, C., Zhu, X.-J., Stevenson, S., & Thrane, E. 2021, *ApJ*, **909**, L19
- García-Quirós, C., Colleoni, M., Husa, S., et al. 2020, *Phys. Rev. D*, **102**, 064002
- Gerosa, D., & Bellotti, M. 2024, *Class. Quantum Gravity*, **41**, 125002
- Giacobbo, N., & Mapelli, M. 2018, *MNRAS*, **480**, 2011
- Hartigan, J., & Hartigan, P. 1985, *Ann. Stat.*, **13**, 70
- Iorio, G., Mapelli, M., Costa, G., et al. 2023, *MNRAS*, **524**, 426
- Karnesis, N., Katz, M. L., Korsakova, N., Gair, J. R., & Stergioulas, N. 2023, *MNRAS*, **526**, 4814
- Kopparapu, R. K., Hanna, C., Kalogera, V., et al. 2008, *ApJ*, **675**, 1459
- Korol, V., & Safarzadeh, M. 2021, *MNRAS*, **502**, 5576
- Kruckow, M. U. 2020, *A&A*, **639**, A123
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Mandel, I., Farr, W. M., & Gair, J. R. 2019, *MNRAS*, **486**, 1086
- Maoz, D., & Nakar, E. 2025, *ApJ*, **982**, 179
- Nair, A., & Stevenson, S. 2025, *MNRAS*, **543**, 233
- Neijssel, C. J., Vigna-Gómez, A., Stevenson, S., et al. 2019, *MNRAS*, **490**, 3740
- Özel, F., & Freire, P. 2016, *ARA&A*, **54**, 401
- Özel, F., Psaltis, D., Narayan, R., & Santos Villarreal, A. 2012, *ApJ*, **757**, 55
- Pellouin, C., Dvorkin, I., & Lehoucq, L. 2025, *A&A*, **693**, A283
- Qin, Y., Zhu, J.-P., Meynet, G., et al. 2024, *A&A*, **691**, A214
- Regimbau, T., Dent, T., Del Pozzo, W., et al. 2012, *Phys. Rev. D*, **86**, 122001
- Romero-Shaw, I. M., Farrow, N., Stevenson, S., Thrane, E., & Zhu, X.-J. 2020, *MNRAS*, **496**, L64
- Safarzadeh, M., Ramirez-Ruiz, E., & Berger, E. 2020, *ApJ*, **900**, 13
- Santoliquido, F., Mapelli, M., Giacobbo, N., Bouffanais, Y., & Artale, M. C. 2021, *MNRAS*, **502**, 4877
- Shao, D.-S., Tang, S.-P., Jiang, J.-L., & Fan, Y.-Z. 2020, *Phys. Rev. D*, **102**, 063006
- Talbot, C., & Golomb, J. 2023, *MNRAS*, **526**, 3495
- Tauris, T. M., Kramer, M., Freire, P. C. C., et al. 2017, *ApJ*, **846**, 170
- Tenorio, R., Toubiana, A., Bruel, T., Gerosa, D., & Gair, J. R. 2025, *ApJ*, **994**, L52
- The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al. 2025, arXiv e-prints [arXiv:2508.18083]
- Toubiana, A., Wong, K. W. K., Babak, S., et al. 2021, *Phys. Rev. D*, **104**, 083027
- Urlus, R. 2025, diptest: Hartigan’s dip test for unimodality (Python/C++), python package implementing Hartigan and Hartigan’s dip test for unimodality
- Vigna-Gómez, A., Neijssel, C. J., Stevenson, S., et al. 2018, *MNRAS*, **481**, 4009
- Vitale, S., Gerosa, D., Farr, W. M., & Taylor, S. R. 2022, in *Handbook of Gravitational Wave Astronomy*, eds. C. Bambi, S. Katsanevas, & K. D. Kokkotas (Berlin: Springer), 45
- Ye, C. S., Fong, W.-f., Kremer, K., et al. 2020, *ApJ*, **888**, L10
- Zevin, M., Bavera, S. S., Berry, C. P. L., et al. 2021, *ApJ*, **910**, 152
- Zevin, M., Nugent, A. E., Adhikari, S., et al. 2022, *ApJ*, **940**, L18

Appendix A: Comparison between population parametrisations

Figures A.1 and A.2 compare our population model with that of Galaudage et al. (2021), with the extra condition that only light/light and heavy/heavy pairings are allowed. The total mass distributions are identical, as both models describe them as the sum of two Gaussians. The corresponding mass-ratio distributions are qualitatively similar. The discontinuity visible in the (m_t, q) correlation for the light population originates from the fact that in Galaudage et al. (2021), the component masses are not ordered; enforcing $q \leq 1$ thus introduces this artificial feature. Still, our model has a similar support to the original distribution.

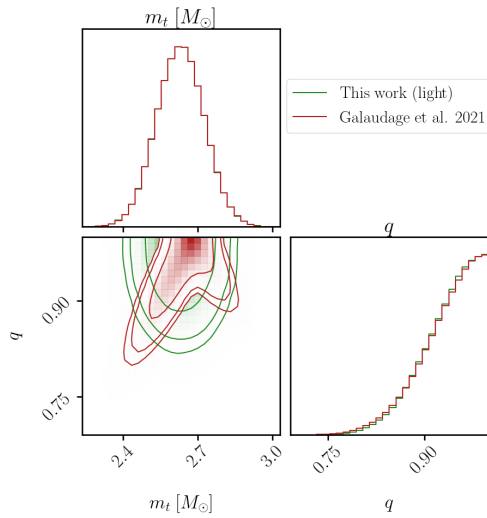


Fig. A.1. Comparison between the parametrisation of Galaudage et al. (2021) for the light population and ours in terms of total mass and mass ratio.

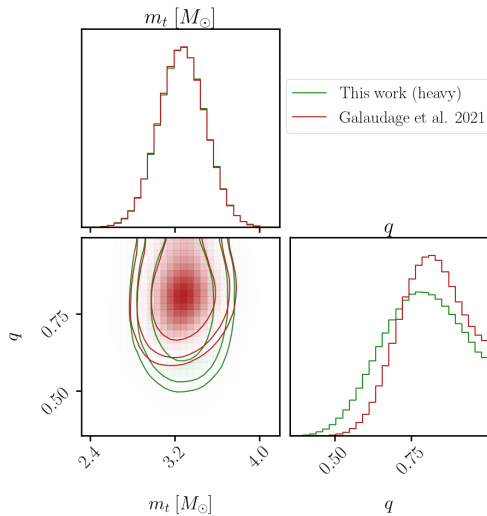


Fig. A.2. Comparison between the parametrisation of Galaudage et al. (2021) for the heavy population of ours in terms of total mass and mass ratio.

Appendix B: Mock detection and parameter estimation

In their appendices, Fishbach et al. (2020) and Farah et al. (2023) outline a prescription for generating mock gravitational-wave (GW) catalogues that self-consistently incorporate parameter estimation uncertainty and selection effects, while also reproducing the expected scaling of measurement errors with the inverse of the S/N. Their approach leverages the fact that the S/N, ρ , can be written as

$$\rho = \frac{F(\theta')}{D_L}, \quad (\text{B.1})$$

where θ' denotes all binary parameters entering the S/N computation except for the luminosity distance D_L (detector frame masses, spins, sky location...). For fixed θ' , this expression allows one to transform straightforwardly between D_L and ρ .

The key idea is to treat the S/N itself as a sampling parameter during mock parameter estimation, rather than the luminosity distance. This makes it possible to enforce that the measured S/N, $\tilde{\rho}$, is normally distributed around the true S/N, with unit variance: $\tilde{\rho} \sim \mathcal{N}(\rho, 1)$, which reflects the expected behaviour of a matched-filter S/N in the case of a single detector (Finn & Chernoff 1993). Selection effects are then included as follows: for each source, the true S/N is computed and scattered according to the above distribution to obtain $\tilde{\rho}$; the event is retained only if $\tilde{\rho}$ exceeds a chosen S/N threshold, ρ_{th} .

For retained events, posterior samples of the S/N, $\hat{\rho}$, are then drawn from a normal distribution centred at $\tilde{\rho}$ with unit variance. To model measurement uncertainty on the remaining parameters θ' , “noisy” values $\tilde{\theta}'$ are generated by sampling from a normal distribution with variance proportional to $\tilde{\rho}^{-1}$, mimicking the typical scaling of parameter errors with inverse S/N. Posterior samples $\hat{\theta}'$ are obtained by further drawing from normal distributions centred at $\tilde{\theta}'$ with variances also proportional to $\tilde{\rho}^{-1}$. The $\{\hat{\theta}', \hat{\rho}\}$ samples are then transformed back into $\{\hat{\theta}', \hat{D}_L\}$ samples via Eq. (B.1). The proportionality constants for the measurement errors on θ' were calibrated against mock parameter estimation runs.

The problem can be simplified by noting that the S/N may be expressed as

$$\rho = \rho_{\text{opt}} \Theta, \quad (\text{B.2})$$

where ρ_{opt} is the optimal S/N for a source located directly overhead the detector with face-on inclination, and Θ encodes the dependence on sky location, polarisation, and inclination. For a single detector, Θ is bounded by $0 \leq \Theta \leq 1$, while for ET the range becomes $0 \leq \Theta_{\text{ET}} \leq \sqrt{3}$. Denoting by $F_{+,i}$ and $F_{\times,i}$ the antenna pattern functions of the three equivalent ET detectors, as given in Regimbau et al. (2012), the factor Θ_{ET} can be written as

$$\Theta_{\text{ET}} = \sqrt{\sum_{i=1}^3 (F_{+,i}^2 + F_{\times,i}^2)}. \quad (\text{B.3})$$

This expression generalises the single-detector relation derived in Finn & Chernoff (1993), but is strictly valid only for the case of a triangular ET configuration, where all three detectors are co-located. We note that in a network of multiple detectors, the distribution of the measured S/N, $\tilde{\rho}$, deviates slightly from the simple normal form assumed in the single-detector case (Essick 2023; Gerosa & Bellotti 2024). However, for the purposes of this

study, we neglect this correction and adopt the single-detector approximation for simplicity.

Our approach differs from that of Fishbach et al. (2020); Farah et al. (2023) in two main respects. First, although we enforce hard prior boundaries when drawing the posterior samples $\hat{\theta}'$, we do not impose such constraints when generating the “noisy” parameter values $\tilde{\theta}'$. This choice reflects the fact that noise can make it appear as the likelihood peaks outside of the prior limits, which manifests as railing of the posterior against the prior boundary. This is also simpler, as it means that the likelihood is truly a Gaussian and not a truncated Gaussian, avoiding the need to adjust the sampling procedure to account for a modified functional form.

Second, Fishbach et al. (2020); Farah et al. (2023) adopt $\theta' = (\mathcal{M}_{c,d}, \eta, \Theta)$ as sampling parameters, where $\mathcal{M}_{c,d}$ is the detector-frame chirp mass and η is the symmetric mass ratio. However, we find that sampling in the symmetric mass ratio leads to large variances in the Monte Carlo estimators used when evaluating the population likelihood (see App. D). This issue arises because our underlying population model favours nearly equal-mass systems, while sampling in η tends to underpopulate the region near $q \sim 1$, as also discussed in Farah (2022). To mitigate this, we sample directly in the mass ratio q instead of η . We assume that uncertainties in the inferred mass ratio scale as $0.15\rho_{\text{th}}/\rho$, which is a conservative estimate based on the measurement uncertainties reported for GW170817 (Abbott et al. 2017) and GW190425 (Abbott et al. 2020), particularly when tight spin priors are used. Although the mapping from detector-frame chirp mass to total mass is less problematic, we likewise choose to sample directly in the detector-frame total mass, adopting a similarly conservative uncertainty scaling of $0.1\rho_{\text{th}}/\rho$. For Θ we use the same as Fishbach et al. (2020) and Farah et al. (2023): $0.21\rho_{\text{th}}/\rho$.

We compute S/Ns using the GWBENCH package (Borhanian 2021), setting the spins to 0 and using the IMRPhenomXHM waveform (García-Quirós et al. 2020). For the S/N threshold, we take $\rho_{\text{th}} = 8$.

Appendix C: Hierarchical Bayesian analysis

We denote with θ the set of parameters describing a GW event, $p(d|\theta)$ the single event likelihood and $p(\theta|\Lambda)$ the population prior on θ , which depends on hyperparameters Λ that we wish to infer. The population likelihood for observing N_{obs} events $\{d\} = \{d_1, \dots, d_{N_{\text{obs}}}\}$, marginalised over the event rate is Mandel et al. (2019); Vitale et al. (2022)

$$p(\{d\}|\Lambda) = \prod_i^{N_{\text{obs}}} \int \frac{p(d_i|\theta)p(\theta|\Lambda)}{p_{\text{det}}(\Lambda)} d\theta. \quad (\text{C.1})$$

We have introduced the selection function defined via

$$p_{\text{det}}(\Lambda) = \int p_{\text{det}}(\theta)p(\theta|\Lambda) d\theta, \quad (\text{C.2})$$

$$p_{\text{det}}(\theta) = \int_{d > \text{threshold}} p(d|\theta) dd, \quad (\text{C.3})$$

where the second integral is restricted to realisations d that exceed the detection threshold (defined via the chosen ranking statistic), here the measured matched filter S/N $\hat{\rho}$.

Alternatively, if we do not wish to marginalise over the rate, we can write

$$p(\{d\}|\Lambda) = e^{-N(\Lambda)p_{\text{det}}(\Lambda)} \prod_i^{N_{\text{obs}}} \int p(d_i|\theta) \frac{dN}{d\theta}(\Lambda) d\theta. \quad (\text{C.4})$$

where $\frac{dN}{d\theta}$ is the differential number of events, $N(\Lambda) = \int \frac{dN}{d\theta}(\Lambda) d\theta$ is the total expected number of events, and we have the relation $\frac{dN}{d\theta} = N(\Lambda)p(\theta|\Lambda)$.

In practice, the single-event likelihood can be written in terms of the posterior and the parameter estimation prior via Bayes’ theorem. The integral over θ is evaluated using Monte Carlo integration, based on posterior samples obtained as described in App. B.

Similarly, the selection function is estimated via an injection campaign in which events are drawn from an injection distribution $\pi_{\text{inj}}(\theta)$, and only those satisfying the detection criterion are retained, forming the set θ_{det} . The detection probability is then computed using importance sampling:

$$p_{\text{det}}(\Lambda) = \sum_{\theta_{\text{det}}} \frac{p(\theta_{\text{det}}|\Lambda)}{\pi_{\text{inj}}(\theta_{\text{det}})}. \quad (\text{C.5})$$

As discussed in Talbot & Golomb (2023), the finite number of samples used to evaluate the integrals in population analyses can introduce significant variance in the Monte Carlo estimators, which may in turn lead to biases. To mitigate this issue in the estimation of the selection function, we choose the injection prior to match the true population, thereby reducing the variance in the importance sampling weights. Additionally, in line with the recommendations of Talbot & Golomb (2023), we impose a threshold on the total variance of the log-likelihood and discard hyperparameters exceeding this limit.

The posterior on Λ is then obtained through Bayes’ theorem: $p(\Lambda|\{d\}) \propto p(\{d\}|\Lambda)\pi(\Lambda)$, with the priors described in Sec. 2 and in the next one. The sampling is performed with the Eryn sampler³ (Karnesis et al. 2023).

Appendix D: Reconstruction of the mass distribution

Figure D.1 shows representative examples of how the measurement of the total mass and the mass ratio distribution improves as the number of events increases. These plots correspond to realisations of the $\alpha_L = -1$ case; however, the choice of α_L has little impact on this result.

³ Publicly available at <https://github.com/mikekatz04/Eryn>.

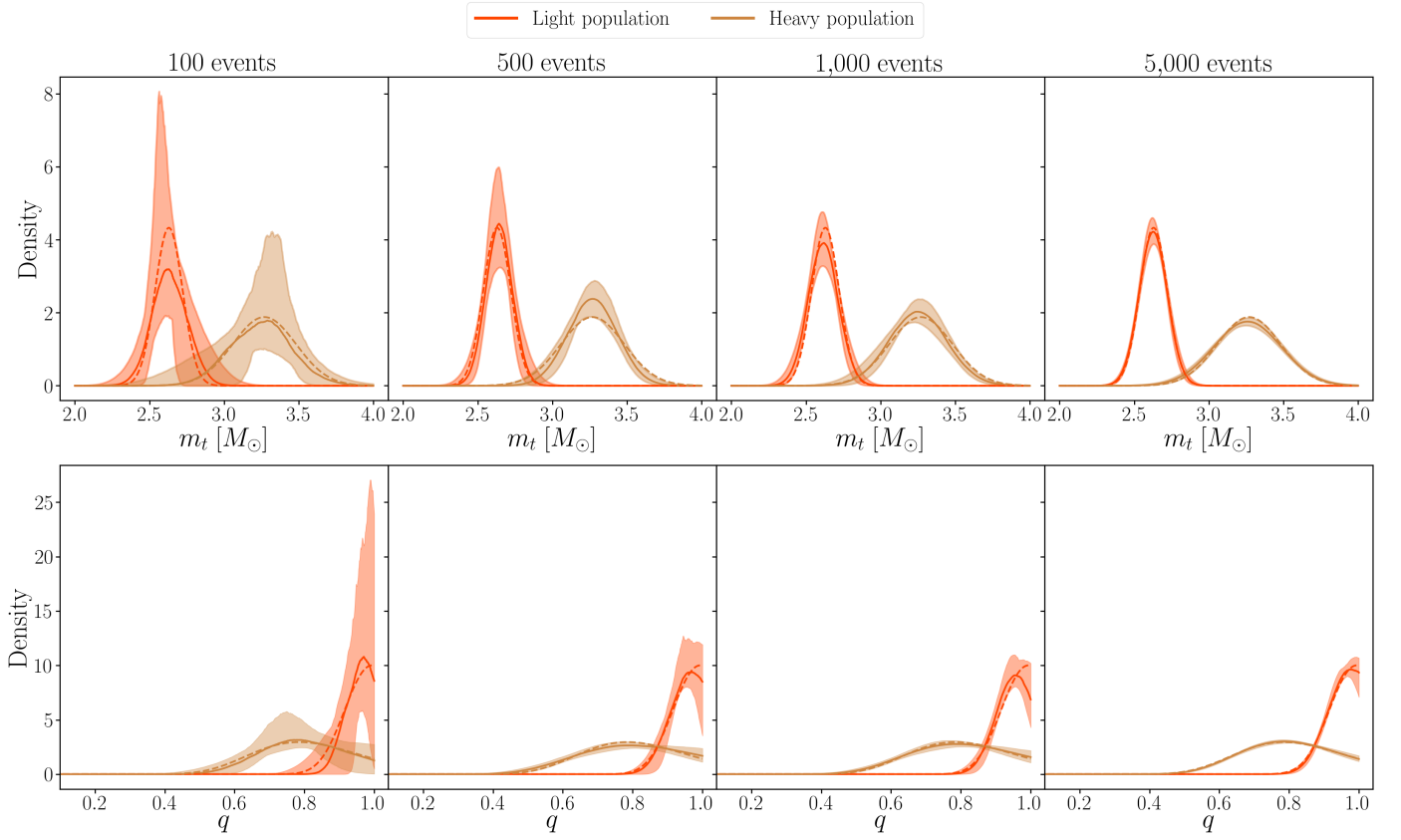


Fig. D.1. Reconstructions of the total mass and mass ratio distributions for increasing catalogue sizes. Coloured bands show the 90% credible intervals for the light (red) and heavy (golden) populations. Solid lines indicate the median reconstructions, while dashed lines mark the true distributions.