

This is the accepted manuscript made available via CHORUS. The article has been published as:

# Unbiased Hubble constant estimation from binary neutron star mergers

Daniel J. Mortlock, Stephen M. Feeney, Hiranya V. Peiris, Andrew R. Williamson, and Samaya M. Nissanke

Phys. Rev. D **100**, 103523 — Published 18 November 2019

DOI: [10.1103/PhysRevD.100.103523](https://doi.org/10.1103/PhysRevD.100.103523)

# Unbiased Hubble constant estimation from binary neutron star mergers

Daniel J. Mortlock,<sup>1,2,3,\*</sup> Stephen M. Feeney,<sup>4</sup> Hiranya V. Peiris,<sup>5,6</sup>

Andrew R. Williamson,<sup>7,8</sup> and Samaya M. Nissanke<sup>7,8</sup>

<sup>1</sup>*Astrophysics Group, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*

<sup>2</sup>*Department of Mathematics, Imperial College London, London SW7 2AZ, UK*

<sup>3</sup>*Department of Astronomy, Stockholm University, AlbaNova, SE-10691 Stockholm, Sweden*

<sup>4</sup>*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA*

<sup>5</sup>*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*

<sup>6</sup>*Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, AlbaNova, Stockholm SE-106 91, Sweden*

<sup>7</sup>*GRAPPA, Anton Pannekoek Institute for Astronomy and Institute of High-Energy Physics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*

<sup>8</sup>*Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands*

Gravitational wave (GW) observations of binary neutron star (BNS) mergers can be used to measure luminosity distances and hence, when coupled with estimates for the mergers' host redshifts, infer the Hubble constant,  $H_0$ . These observations are, however, affected by GW measurement noise, uncertainties in host redshifts and peculiar velocities, and are potentially biased by selection effects and the mis-specification of the cosmological model or the BNS population. The estimation of  $H_0$  from samples of BNS mergers with optical counterparts is tested here by using a phenomenological model for the GW strains that captures both the data-driven event selection and the distance-inclination degeneracy, while being simple enough to facilitate large numbers of simulations. A rigorous Bayesian approach to analyzing the data from such simulated BNS merger samples is shown to yield results that are unbiased, have the appropriate uncertainties, and are robust to model mis-specification. Applying such methods to a sample of  $N \simeq 50$  BNS merger events, as LIGO+Virgo could produce in the next  $\sim 5$  years, should yield robust and accurate Hubble constant estimates that are precise to a level of  $\lesssim 2 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , sufficient to reliably resolve the current tension between local and cosmological measurements of  $H_0$ .

## I. INTRODUCTION

The current expansion rate of the Universe is characterized by the Hubble constant,  $H_0$ . While there is an empirical consensus that  $H_0 \simeq 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , there is a  $4.4\text{-}\sigma$  tension between the most recent local and cosmological measurements: the anchor-Cepheid-supernova distance ladder gives  $H_0 = 74.2 \pm 1.8 \text{ km s}^{-1} \text{ Mpc}^{-1}$  [1]; while the *Planck* cosmic microwave background (CMB) data, combined with the assumption of a standard flat cold dark matter ( $\Lambda$ CDM) cosmological model, imply that  $H_0 = 67.3 \pm 0.6 \text{ km s}^{-1} \text{ Mpc}^{-1}$  [2]. It is tempting to use this tension as motivation for rejecting or extending  $\Lambda$ CDM [3–25], but before settling on such an exciting possibility it is necessary to ensure that the discrepancy is not due to limitations in the analysis of one or both of the datasets [26–39].

The most direct way of resolving this ‘Hubble trouble’ would be a local measurement of  $H_0$  that is completely independent of the above distance ladder. One of the most promising options is gravitational wave (GW) observations of mergers between compact/relativistic objects such as black holes (BHs) and neutron stars (NSs). The GW waveform from any such merger provides information about the distance to the system which can then be combined with a redshift measurement/estimate

to constrain  $H_0$  [40]. The advent of the Advanced Laser Interferometer Gravitational-Wave Observatory and Advanced Virgo (LIGO+Virgo) has resulted in the detection of several such mergers [41, 42], although the utility for measuring  $H_0$  depends strongly on the type of system:

- Binary neutron star (BNS) mergers are particularly promising, as a possible electromagnetic (EM) counterpart could be used to identify a host galaxy from which a spectroscopic redshift measurement could be made (*e.g.*, [43–45]). For some fraction of BNS mergers a counterpart will not be identified, in which case it is plausible to take a statistical approach, averaging over the host galaxies that are consistent with the GW localization [46–48]. It is also possible that GW data alone could be used to obtain redshift constraints by exploiting either the narrowness of the NS mass distribution [49] or NS tidal deformability [50]. The first BNS merger event detected in GWs, GW 170817/GRB 170817A [41], provided the constraint that  $H_0 = 70.0^{+12.0}_{-8.0} \text{ km s}^{-1} \text{ Mpc}^{-1}$  [45]. More precise, if necessarily model-dependent, constraints can be obtained by using EM data to estimate the inclination of the system [51], with different analyses of the data from GRB 170817A yielding  $H_0 = 74.0^{+5.3}_{-13.7} \text{ km s}^{-1} \text{ Mpc}^{-1}$  [52] and  $H_0 = 68.9^{+4.7}_{-4.6} \text{ km s}^{-1} \text{ Mpc}^{-1}$  [53]. Conversely, if it had not been possible to identify an EM counterpart to GW 170817, the GW measurements alone would

---

\* mortlock@ic.ac.uk

have yielded  $H_0 = 70.0^{+48.0}_{-23.0} \text{ km s}^{-1} \text{ Mpc}^{-1}$  [48].

- Binary black hole (BBH) mergers are the strongest GW sources and have dominated current merger samples [54]. But the expected lack of any EM emissions, along with their broad range of masses, makes it difficult to obtain useful redshift estimates. The best hope is to take the same sort of statistical approach as considered above for the BNS mergers [46, 55, 56], although the uncertainties in  $H_0$  are then dominated by this limitation (*e.g.*, [57, 58]).
- NS-BH mergers might provide tighter distance constraints than BNS mergers due to both their higher system mass [44, 59] and the possibility that the BH spin is not aligned with the orbital angular momentum, which would induce precession of the orbital plane [60]. It is less clear, however, whether there would be detectable EM emission from such mergers; if there is not then the constraints on  $H_0$  would be limited in the same way as BBH mergers and BNS mergers without counterparts.

As compact mergers with EM counterparts provide the cleanest  $H_0$  constraints and no NS-BH merger has yet been decisively confirmed in either GW or EM observations, the focus here is on BNS mergers with EM counterparts and confirmed host galaxies, although the overall approach taken here applies to all types of compact binary mergers. The third LIGO+Virgo observing run (O3) should detect several more BNS mergers which, if EM counterparts could be identified in all cases, would already give an uncertainty in  $H_0$  smaller than the difference between the local and CMB values [61, 62]. Looking further ahead, detector advances over the next five years could yield a sample of  $\sim 50$  BNS merger events, sufficient to measure  $H_0$  with a precision of  $\lesssim 2 \text{ km s}^{-1} \text{ Mpc}^{-1}$  [62, 63].

Such a sample of BNS mergers with EM counterparts would be sufficient to resolve the current  $H_0$  tension [63], but only if the data analysis produces  $H_0$  estimates that have correct uncertainties and that are free of systematic biases (*i.e.*, accurate, as well as precise). One potential source of bias is selection effects, as detection on the basis of the observed GW signal-to-noise ratio (SNR) will preferentially include events for which the measurement noise has augmented the signal, making such mergers appear closer than they are. Unless accounted for, this selection effect would result in a systematic overestimate of  $H_0$  (*e.g.*, [64, 65]). Another potential source of bias is mis-specification of the cosmological model or the BNS population, as the data-generation process links  $H_0$  to these other global parameters (*e.g.*, [45, 62, 64]). Eventually, overall calibration uncertainties are likely to provide the absolute systematic floor to the precision in  $H_0$  of a BNS merger sample; these are currently at a level of few per cent, although there are clear prospects for further improvements [66, 67].

The main aim of this paper is to test whether a Bayesian population analysis of the sort described by

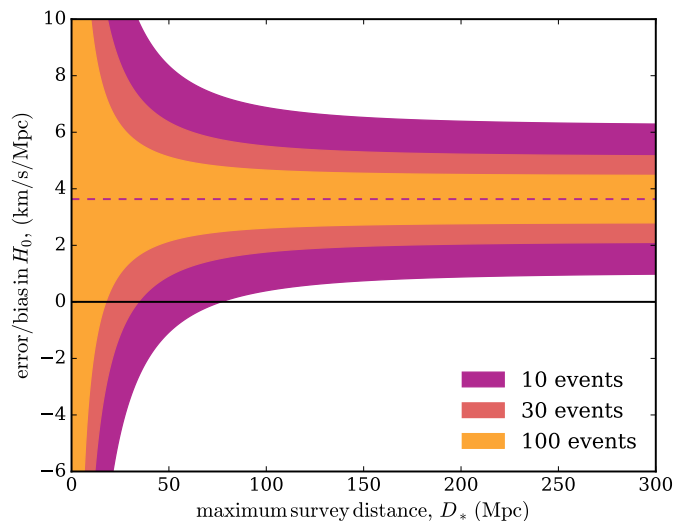


FIG. 1. The expected range of  $H_0$  estimates from sample of  $N = 10$ ,  $N = 30$  and  $N = 100$  BNS merger events, as a function of the maximum distance,  $D_*$ , to which sources can be detected. The offset from zero comes from the potential bias due to selection effects, which would dominate over sample variance for  $N \gtrsim 10$  and  $D_* \gtrsim 70 \text{ Mpc}$ .

Refs [45, 48, 62–64, 68, 69] gives unbiased  $H_0$  estimates with valid uncertainties (*i.e.*, is both accurate and precise) when applied to realistic BNS merger samples. A secondary aim is to provide a derivation from first principles of the full posterior distribution appropriate to a sample of BNS mergers with EM counterparts. Simple predictions for the uncertainty and selection bias in  $H_0$  are given in Secs II and III. The sample simulations are described in Sec. IV and the Bayesian analysis approach summarized in Sec. V. The large-sample properties of this approach are then explored in Sec. VI, with the conclusions and possibilities for future development summarized in Sec. VII. The general model and inference formalism is presented in Appendix A and the simplified BNS inspiral model is described in Appendix B.

## II. PREDICTED UNCERTAINTIES

In order to assess the performance of any data analysis method it is useful to have a prediction for the expected uncertainties in the idealized case in which there are no complicating factors like selection biases or systematic errors. This gives a target for the uncertainties obtained from real data and also a guide for assessing whether any potential systematic effects are likely to be significant.

For a local sample of BNS mergers with counterparts, the predicted uncertainty in  $H_0$  can be estimated by considering an expanding Euclidean geometry in which distance,  $D$ , is related to redshift,  $z$ , by  $D = cz/H_0$ , where  $c$  is the speed of light. Given a measured (spectroscopic) redshift,  $\hat{z}$ , an estimated peculiar velocity,  $\hat{v}$  and a GW

distance,  $\hat{D}$ , for a BNS merger event (or any other low-redshift extragalactic object), the natural estimator for the Hubble constant is

$$\hat{H}_0 = \frac{c \hat{z} - \hat{v}}{\hat{D}}, \quad (1)$$

where the peculiar velocity estimate effectively corrects the spectroscopic redshift to (hopefully) bring it closer to the cosmological value. The resultant uncertainty (really the standard deviation of the estimator) is approximated to leading order as (*cf.* [61, 62])

$$\sigma_H \simeq \frac{1}{D} (c^2 \sigma_z^2 + \sigma_v^2 + H_0^2 \sigma_D^2)^{1/2}, \quad (2)$$

where  $\sigma_D$  is the uncertainty in the distance from the GW data,  $\sigma_z$  is the observational redshift uncertainty, and  $\sigma_v$  is the uncertainty in the peculiar velocity (which should be no larger than the observed peculiar velocity dispersion of  $\sim 200 \text{ km s}^{-1}$  [70], but could be smaller if object-specific information is available).

Using representative numerical values for BNS mergers, and exploiting the fact that it is already known that  $H_0$  is close to  $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , Eq. 2 can be re-written as

$$\begin{aligned} \sigma_H &\simeq 7.8 \text{ km s}^{-1} \text{ Mpc}^{-1} \left( \frac{D}{43.8 \text{ Mpc}} \right)^{-1} \\ &\times \left[ \left( \frac{0.21 \sigma_z}{0.00024} \right)^2 + \left( \frac{0.20 \sigma_v}{70.0 \text{ km s}^{-1}} \right)^2 + \left( \frac{\sigma_D}{4.9 \text{ Mpc}} \right)^2 \right]^{1/2} \\ &\simeq 7.0 \text{ km s}^{-1} \text{ Mpc}^{-1} \left( \frac{D}{100 \text{ Mpc}} \right)^{-1} \\ &\times \left[ \left( \frac{0.43 \sigma_z}{0.001} \right)^2 + \left( \frac{0.29 \sigma_v}{200 \text{ km s}^{-1}} \right)^2 + \left( \frac{\sigma_D}{10 \text{ Mpc}} \right)^2 \right]^{1/2}, \end{aligned} \quad (3)$$

where the first case uses GW 170817 as a reference and the second uses more generic values appropriate to the simulations described below. Both cases illustrate that the error in the reconstruction of the source distance from the GW data will be the dominant uncertainty unless the SNR is sufficiently high that  $\sigma_D$  can be reduced to  $\lesssim 3 \text{ Mpc}$ , which might be possible for very nearby sources (*e.g.*, the ‘golden’ events invoked in Ref. [62]). Even for such sources, however, there would be little value in obtaining a redshift measurement more precise than the fiducial value of  $\sigma_z \simeq 0.001$  unless the peculiar velocity uncertainty of the host could be reduced significantly below the cosmological prior value.

A single BNS merger clearly cannot place interesting constraints on  $H_0$ , but if the uncertainty from a sample of  $N$  events scales as  $N^{-1/2}$  then Eq. 3 broadly confirms the numerical predictions [62, 63] that  $\sim 50$  events should yield  $\sigma_H \lesssim 2 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and hence be able to resolve the current Hubble constant tension. For the purposes of making quantitative predictions, it is useful to take a

more realistic approach that incorporates both the dependence of the uncertainty on source distance and the relative numbers of sources at different distances.

For a given source the GW strain signal is inversely proportional to its distance, while the strain noise is additive and a characteristic of the detector (*e.g.*, [71]). Assuming the sources are all intrinsically identical (*i.e.*, truly ‘standard’ sirens), the SNR can then be written as  $\rho \simeq \rho_* D_*/D$ , where  $\rho_*$  is the minimum SNR required for selection and  $D_*$  is the maximum distance out to which the survey can detect such sources. Assuming further that the relative uncertainty in the distance reconstruction is related to the SNR by  $\sigma_D/D \simeq 1/\rho$ , the absolute distance uncertainty is (*cf.* [61, 71, 72])

$$\sigma_D \simeq \frac{D^2}{D_*^2} \sigma_*, \quad (4)$$

where  $\sigma_*$  is the distance uncertainty for a source at  $D_*$ . Combining this with Eq. 2 gives

$$\sigma_H \simeq H_0 \frac{\sigma_*}{D_*} \left( \frac{D_0^2}{D^2} + \frac{D^2}{D_*^2} \right)^{1/2}, \quad (5)$$

where  $D_0 = (c^2 \sigma_z^2 + \sigma_v^2)^{1/2} / (H_0 \sigma_*/D_*) \simeq 30 \text{ Mpc}$  is the distance beyond which redshift/velocity uncertainties become unimportant. This relationship also identifies a distance of  $\sim (D_0 D_*)^{1/2} \simeq 30 (D_*/D_0)^{1/2} \text{ Mpc}$  as that for which a single source in such a survey would yield the tightest constraint on  $H_0$  (*cf.* the ‘sweet spot’ distance of Ref. [62]).

The number of detected sources will, out to  $\sim D_*$ , be proportional to the volume element, which for the purposes of this calculation can be approximated as  $dV/dD \simeq 4\pi D^2$ . Ignoring the effect of noise on completeness (which is discussed below in Sec. III and Sec. IV D), the probability of selection,  $S$ , is  $P(S|D, D_*) = \Theta(D_* - D)$  and so the distance distribution of detected events is then

$$P(D|S, D_*) \simeq \Theta(D) \Theta(D_* - D) \frac{3 D^2}{D_*^3}, \quad (6)$$

where  $\Theta(\cdot)$  denotes the Heaviside step function.

The optimal estimate from a sample of BNS events would be to take an inverse variance-weighted average of  $\hat{H}_0$  from Eq. 1 for each source, with the weighting proportional to  $1/\sigma_H^2$  from Eq. 5. The resultant uncertainty in  $H_0$  for a given sample of  $N$  sources with individual uncertainties  $\sigma_{H,1}, \sigma_{H,2}, \dots, \sigma_{H,N}$  would then be  $\sigma_H = [(\sum_{i=1}^N 1/\sigma_{H,i}^2)]^{-1/2}$ . However, because the number of sources increases as  $D^2$  and for  $D \gtrsim (D_0 D_*)^{1/2}$  the inverse variance weighting decreases as  $D^{-2}$ , the uncertainty produced by weighting all sources equally is within  $\sim 10\%$  of that given by the optimal scheme, so uniform weighting is used here for simplicity. The expected uncertainty from a sample of  $N$  events is hence approximated

as the average of the expected variance per event,

$$\begin{aligned}\sigma_H &\simeq \left[ \frac{1}{N} \int_0^\infty dD P(D|S, D_*) \sigma_D^2 \right]^{1/2} \\ &= \frac{1}{N^{1/2}} \left[ \int_0^{D_*} dD \frac{3D^2}{D_*^3} \left( H_0 \frac{\sigma_*}{D_*} \right)^2 \left( \frac{D_0^2}{D^2} + \frac{D^2}{D_*^2} \right) \right]^{1/2} \\ &= \frac{1}{N^{1/2}} \left( \frac{3}{5} \right)^{1/2} \frac{H_0 \sigma_*}{D_*} \left( 5 \frac{D_0^2}{D_*^2} + 1 \right)^{1/2},\end{aligned}\quad (7)$$

where Eqs. 4 and 5 have been used. These uncertainties are shown for different sample sizes as the colored bands in Fig. 1, from which it is clear that redshift and velocity uncertainties become irrelevant for sufficiently deep surveys with  $D_* \gg D_0 \simeq 30$  Mpc. In this regime Eq. 7 simplifies to be

$$\sigma_H \simeq 1.2 \text{ km s}^{-1} \text{ Mpc}^{-1} \left( \frac{N}{50} \right)^{-1/2} \left( \frac{\sigma_*/D_*}{0.15} \right), \quad (8)$$

where the fiducial value chosen for  $\sigma_*/D_*$  is, from simulations, found to be approximately  $2/\rho_*$ .

Equation 7 (or, if appropriate, Eq. 8) represents a target for any analysis of real or simulated BNS merger data, and also provides a guide to the level at which potential biases could have a significant impact.

### III. SELECTION BIAS

Given that it is likely that GW observations of BNS mergers will soon produce uncertainties in  $H_0$  sufficiently small to resolve the current Hubble constant tension, it is important to ensure that the resultant estimates are not significantly affected by systematic biases. There can be systematic errors from model mis-specification, but the most pernicious potential bias comes from the fact that any real BNS merger sample will consist of sources selected on the basis of the measured SNR of the same GW data that is then used to obtain distance constraints. This selection cut will preferentially include cases where the measurement noise has added to the signal, and which hence are inferred to be closer than they in fact are; applying any sort of simple average (*e.g.*, using Eq. 1) to such a BNS merger sample would overestimate  $H_0$ .

The magnitude of this potential bias is somewhat ambiguous, as it is only defined in the context of a specific method for obtaining a distance estimate,  $\hat{D}$ , from the GW data, but a reasonable approach is to extend the simple survey model described in Sec. II by including selection effects. The error in  $H_0$  from a single object at distance  $D$  for which the data give a distance estimate that is off by  $\Delta D = \hat{D} - D$  is

$$\begin{aligned}\hat{H}_0 - H_0 &\simeq -\frac{\Delta D}{D} H_0 \\ &\simeq -10 \text{ km s}^{-1} \text{ Mpc}^{-1} \frac{\Delta D/D}{0.15},\end{aligned}\quad (9)$$

where the (cosmological) redshift is assumed to be known perfectly and the final expression is reasonable for an object that is close to the detection threshold.

The implied bias from a sample would then be given by averaging  $\hat{H}_0 - H_0$  over all distances and possible noise realizations, which would require an explicit model for the measurement process. A more generic approximation is possible by considering the distinct behavior in three different distance ranges: objects with  $D \lesssim D_* - \sigma_*$  are all well detected, irrespective of the noise realization, and so dilute any overall bias; objects close to the survey horizon with  $|D - D_*| \lesssim \sigma_*$  are detected only if the noise augments the signal, in which case their distance is underestimated by  $|\Delta D| \simeq \sigma_*$ ; and objects with  $D \gtrsim D_* + \sigma_*$  are never detected and so do not affect the Hubble constant estimate at all. Taking the distance distribution of detected objects as

$$P(D|S) \propto \begin{cases} D^2 & \text{if } 0 \leq D \leq D_* - \sigma_* \\ (D_* + \sigma_* - D)(D_* - \sigma_*)^2 / (2\sigma_*) & \text{if } D_* - \sigma_* \leq D \leq D_* + \sigma_* \\ 0 & \text{if } D > D_* + \sigma_*, \end{cases} \quad (10)$$

the bias from selection effects would be

$$\begin{aligned}\Delta H_{\text{sel}} &\simeq \frac{3}{1 + 2\sigma_*/D_*} \left( \frac{\sigma_*}{D_*} \right)^2 H_0 \\ &\simeq 3.6 \text{ km s}^{-1} \text{ Mpc}^{-1} \left( \frac{\sigma_*/D_*}{0.15} \right)^2.\end{aligned}\quad (11)$$

The  $(\sigma_*/D_*)^2$  scaling comes about as this ratio determines both the range of distances for which selection effects are important and the magnitude of the distance underestimate for the fraction of these objects selected. This bias is shown as the offset in Fig. 1 and, aside from being comparable to the current difference between the local and cosmological  $H_0$  measurements (Sec. I), is larger than the predicted uncertainties (Sec. II) from a sample of even  $N \simeq 10$  events if  $D_* \gtrsim 70$  Mpc. While the exact value of the predicted  $H_0$  bias is somewhat model-dependent, it is clear that it must be taken into account if BNS mergers are to be useful in resolving the current Hubble tension.

### IV. SIMULATED BNS MERGER SAMPLES

In order to test both the precision and accuracy of  $H_0$  estimates from BNS mergers it is necessary to generate simulated samples of events that, in particular, are subject to appropriate selection effects. The population model and observations described here represent a specific version of the general structure described in Appendix A. The primary focus is on the self-consistency of the model, with the sample selection being performed on the same measured quantities which are subsequently

used to drive the inference calculation. As such, the model is reasonably simple, and includes neither realistic detector and noise models (*cf.* [44, 62, 63]) nor full inference of the individual BNS merger parameters (*cf.* [45, 63, 73–75]).

### A. Cosmology

For the low redshifts of  $z \lesssim 0.2$  out to which BNS mergers (and NS-BH mergers) are likely to be detected in the next decade it is sufficient to adopt the standard Taylor series approximation to the cosmological expansion history, in which the dynamics are characterized to leading order by the deceleration parameter  $q_0$  (*e.g.*, [76]) In a  $\Lambda$ CDM model  $q_0 = \Omega_m/2 - \Omega_\Lambda$ , where  $\Omega_m$  is the normalized matter density and  $\Omega_\Lambda$  is the normalized cosmological constant. The *Planck* CMB data imply that  $\Omega_m \simeq 0.31$  and  $\Omega_\Lambda \simeq 0.69$  [2] and hence that  $q_0 \simeq -0.53$ , although these values would come into question if the *Planck* value of  $H_0$  were determined to be incorrect.

In this cosmological model the luminosity distance is

$$D(z, H_0, q_0) \simeq \frac{cz}{H_0} \left[ 1 + \frac{1}{2} (1 - q_0) z \right] \quad (12)$$

and the co-moving volume element is

$$\frac{dV}{dz}(H_0, q_0) \simeq 4\pi \frac{c^3 z^2}{H_0^3} [1 - 2(1 + q_0)z]. \quad (13)$$

### B. BNS population

The local BNS population is taken to be defined by the rate of mergers per unit proper time per unit co-moving volume,  $\Gamma$ , which has been measured as  $\Gamma = 1540^{+3200}_{-1220} \text{ Gpc}^{-3} \text{ yr}^{-1}$  [41].

An individual BNS merger can, as discussed further in Appendix B, effectively be described by just two parameters: its chirp mass,  $\mathcal{M} = (M_1 M_2)^{3/5} / (M_1 + M_2)^{1/5}$ , where  $M_1$  and  $M_2$  are the masses of the two NSs; and the inclination of the system to the line-of-sight,  $\iota$ . The population prior distribution in these parameters is taken to have the redshift-independent form

$$P(\mathcal{M}, \iota | \bar{\mathcal{M}}, \sigma_{\mathcal{M}}) \propto N(\mathcal{M}; \bar{\mathcal{M}}, \sigma_{\mathcal{M}}^2) \Theta(\iota) \Theta(\pi - \iota) \frac{\sin(\iota)}{2}, \quad (14)$$

where<sup>1</sup> the sinusoidal distribution in  $\iota$  encodes the assumption that these systems have random (and independent) orientations.

The distribution of host (line-of-sight) peculiar motions is taken to be

$$P(v|\sigma_{\parallel}) = N(v; 0, \sigma_{\parallel}^2), \quad (15)$$

<sup>1</sup> Here  $N(x; \mu, \sigma^2) = \exp[-(x - \mu)^2 / (2\sigma^2)] / [(2\pi)^{1/2} \sigma]$  denotes a normal density of mean  $\mu$  and variance  $\sigma^2$ .

where  $\sigma_{\parallel} \simeq 500 \text{ km s}^{-1}$  from observations of local galaxy motions [70]. The additional motion of the BNS system relative to the host is unimportant in this context (Appendix B).

### C. Observations

Any BNS merger that has been detected from its GW emission is subject to intensive follow-up observations, the primary aim of which is to identify an EM counterpart and hence a host galaxy. While there is no guarantee of this process being successful [40, 46–48, 55, 57], it is reasonable to assume that hosts will be identified for a significant fraction of BNS mergers [77]. And, as it is these systems which will provide the best constraints on  $H_0$  (*e.g.*, [62]), they are the focus here. Three distinct types of measurement can provide information about a BNS merger with a confirmed host: the GW data from the BNS merger; the host galaxy redshift measurement; and, potentially, an estimate of the (line-of-sight) peculiar velocity of the host galaxy. Following the general formalism in Appendix A 2, the associated likelihoods depend on the details of the relevant observations and measurements, which are denoted by  $O$  for brevity.

The GW data from a merger produced by a single detector consists of the discretized and noisy linear combination of the two GW polarizations weighted by the detector response function. A full simulation of such data would require using a numerical general relativity code (*e.g.*, LALSimulation [78]) to generate such waveforms, multiplying them by the response function, and adding noise as appropriate for the instrument. The resultant time stream(s) could then be analysed to obtain parameter constraints using software such as *LALInference* [73], *BILBY* [74] or *PyCBC* [75]. Repeating this procedure for large numbers of mergers, as required here to examine the bias of population-level inference approaches (Sec. VI), would be prohibitive.

Fortunately, it is possible to make use of a more streamlined approach in which the time series data is never actually simulated (and hence, even more importantly, the full parameter estimation procedure is not required). In order to assess the inference of the population-level parameters, and  $H_0$  in particular, the only requirement on the model is that the dependence of the likelihood on the model parameters is accurately represented. For a BNS merger, almost all the constraining information about  $\mathcal{M}$ ,  $\iota$  and  $D$  is encoded in the relatively simple inspiral phase, the data from which can be summarized by three statistics that can be calculated from the full data (as described in, *e.g.*, Refs. [44, 71] and defined in Appendix B): a measured value of the redshifted chirp mass,  $\hat{\mathcal{M}}_z$ , which is obtained from the time-dependence of the chirp signal; and the measured

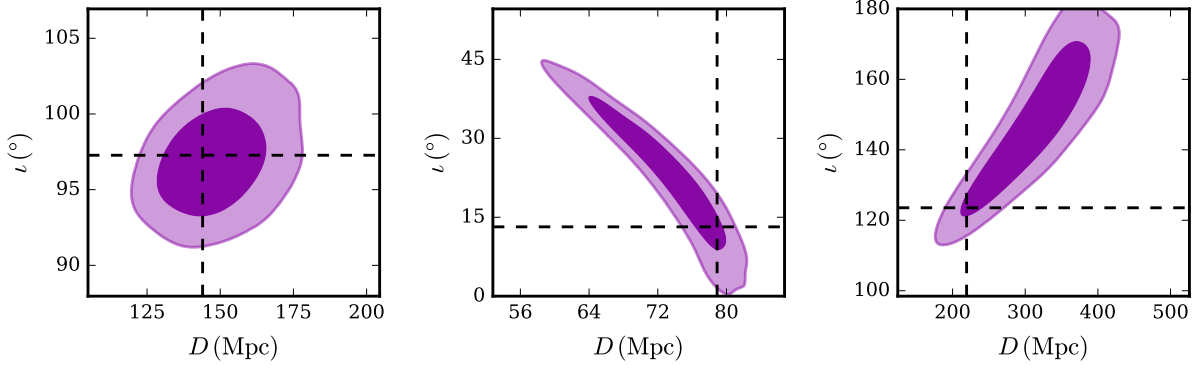


FIG. 2. Constraints on the distance,  $D$ , and inclination,  $\iota$ , for three representative BNS mergers from the GW data only, using the simplified model described in Sec. IV. The purple and mauve indicate 68% and 95% highest posterior density credible regions and the dashed lines show the input parameters.

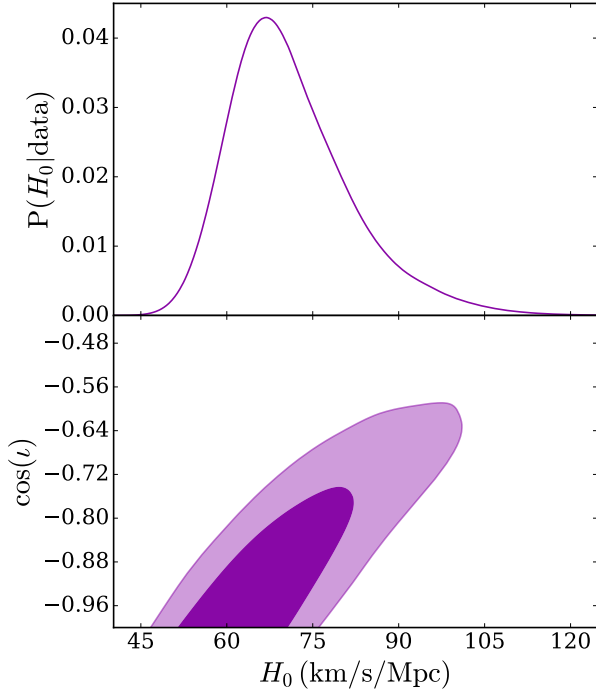


FIG. 3. Constraints on the Hubble constant,  $H_0$ , and the cosine of the system inclination,  $\cos(\iota)$ , from a BNS merger designed to mimic GW 170817. These distributions are hence directly comparable to Figures 1 and 2 of Ref. [45]. The purple and mauve indicate 68% and 95% highest posterior density credible regions.

amplitudes<sup>2</sup> of the two orthogonal strain components,  $\hat{A}_+$  and  $\hat{A}_\times$ .

The uncertainties in these three quantities,  $\sigma_{\mathcal{M}_z}$ ,  $\sigma_{A_+}$  and  $\sigma_{A_\times}$ , are determined primarily by the properties of

the observations (noise level, correlations, time resolution, *etc.*), and here are taken to effectively characterize the survey. The utility of BNS mergers as a means of measuring distances rests on the fact that the redshifted chirp mass can be measured to high precision (*e.g.*, [40, 41]), and the small error in this quantity does not contribute significantly to the overall uncertainty on the merger distance. As such, it is a reasonable approximation to ignore the uncertainty in  $\mathcal{M}_z$  completely, treating the measured redshifted chirp mass as exact. Both amplitude measurements, however, have appreciable uncertainties of up to  $\sim 10\%$  for a realistic sample of detected mergers (*cf.* Sec. IV F). The errors in these measured amplitudes not only are the dominant contributions to the distance uncertainty, but are also linked to the event selection (Sec. IV D), so must be included self-consistently to assess any potential biases. In general the uncertainties of the amplitudes of the two components will differ, but for the purpose of the bias analysis undertaken here it is sufficient to assume they are the same, so that  $\sigma_A = \sigma_{A_+}$  and  $\sigma_{A_\times}$  is the primary quantity used to characterise the GW observations here.

The main result of applying the above approximations and simplifying assumptions is, as detailed in Appendix B, that the likelihood for a single BNS merger can be written as

$$\begin{aligned} & \mathcal{P}(\hat{\mathcal{M}}_z, \hat{A}_+, \hat{A}_\times | \mathcal{M}, \iota, z, H_0, q_0, \sigma_A) \\ &= \delta \left[ \hat{\mathcal{M}}_z - (1+z) \mathcal{M} \right] \\ & \times \mathcal{N} \left[ \hat{A}_+; \frac{G(1+z) \mathcal{M}/c^2}{D(z, H_0, q_0)} \frac{1 + \cos^2(\iota)}{2}, \sigma_A^2 \right] \\ & \times \mathcal{N} \left[ \hat{A}_\times; -\frac{G(1+z) \mathcal{M}/c^2}{D(z, H_0, q_0)} \cos(\iota), \sigma_A^2 \right], \end{aligned} \quad (16)$$

where<sup>3</sup>  $D(z, H_0, q_0)$  is the luminosity distance as defined

<sup>2</sup> The term ‘amplitude’ is something of a misnomer, as these quantities can be positive or negative.

<sup>3</sup> Here  $\delta(x)$  denotes a Dirac delta function.



in Sec. IV A. It is likelihoods of this form that are incorporated into the inference formalism described in Sec. V, although in the numerical implementation a small uncertainty in the redshifted chirp mass is included.

Despite the mathematical simplicity of the model summarized in Eq. 16, it captures all the behavior that is relevant to obtaining constraints on the source distance and inclination. Examples of the resultant posterior distributions are shown in Fig. 2 for fiducial BNS mergers with  $\mathcal{M} \simeq M_\odot$  and  $D \simeq 100$  Mpc, which hence have amplitudes (*i.e.*,  $|A_+|$  and  $|A_\times|$ ) of up to  $\sim 5 \times 10^{-22}$ . The uncertainties of  $\sigma_A \simeq 10^{-23}$  were chosen to give constraints consistent with LIGO+Virgo in its current configuration. The posterior distributions have a range of morphologies and, most importantly, exhibit the strong distance-inclination degeneracy that is the main source of distance uncertainty (*cf.* [45, 51, 53]).

The reported spectroscopic redshift of a BNS host,  $\hat{z}$ , is taken to have a measurement uncertainty  $\sigma_z$ , with the associated likelihood

$$P(\hat{z}|z, v, \sigma_z) = N\left[\hat{z}; z + (1+z)\frac{v}{c}, \sigma_z^2\right]. \quad (17)$$

Any BNS merger event with a confirmed host galaxy is likely to be of considerable scientific interest, implying that high-quality spectroscopic data will be obtained, so a typical uncertainty would be  $\sigma_z \lesssim 0.001$  (*cf.* the measurement of  $\sigma_z = 0.00024$  for the host of GW 170807 [79]).

It is possible that the (line-of-sight) peculiar velocity of a BNS host can be estimated from the positions and/or motions of nearby galaxies, yielding an estimate  $\hat{v}$ , with uncertainty  $\sigma_v$ . The associated likelihood is hence taken to be

$$P(\hat{v}|v, \sigma_v) = N(\hat{v}; v, \sigma_v^2). \quad (18)$$

The case in which there is no useful peculiar velocity information can be encoded by taking  $\sigma_v \rightarrow \infty$ , (and, optionally,  $\hat{v} = 0$ ), in which case the uncertainty on the peculiar velocity is given by the value of  $\sigma_{||}$  assumed from Sec. IV B.

Combining the simple GW likelihood (Eq. 16) with the above models of the redshift measurement (Eq. 17) and peculiar velocity data (Eq. 18), it is possible to produce constraints on  $H_0$  from mock data. This is illustrated in Fig. 3, which shows results for  $\hat{A}_+ = 1.24 \times 10^{-21}$ ,  $\hat{A}_\times = 1.14 \times 10^{-21}$ ,  $\sigma_A = 4.8 \times 10^{-23}$ ,  $\hat{z} = 0.01$ ,  $\sigma_z = 0$ ,  $\hat{v} = 0$ , and  $\sigma_v = 200 \text{ km s}^{-1}$ . These values were chosen to mimic the joint  $H_0$  and  $\cos(\iota)$  posterior from GW170817 as presented in Ref. [45]. The essential phenomenology is reproduced correctly, with the distance-inclination degeneracy still present but less distinct due to the extra uncorrelated uncertainty from the lack of knowledge about the (cosmological) redshift.

## D. Selection function

The selection of a BNS merger event into the survey is assumed to be determined by the GW data alone, and to take the form of a hard cut on the observed SNR,  $\rho_*$ . For the simple model described in Appendix B, with measured strain amplitudes of  $\hat{A}_+$  and  $\hat{A}_\times$  and uncertainty  $\sigma_A$ , the observed SNR is

$$\hat{\rho} = \frac{(\hat{A}_+^2 + \hat{A}_\times^2)^{1/2}}{\sigma_A}. \quad (19)$$

The selection probability is then

$$P(S|\hat{A}_+, \hat{A}_\times, \sigma_A) = \Theta\left[\frac{(\hat{A}_+^2 + \hat{A}_\times^2)^{1/2}}{\sigma_A} - \rho_*\right]. \quad (20)$$

The implied selection function is given in terms of the chirp mass,  $\mathcal{M}$ , inclination  $\iota$  and source distance  $D$  as

$$P(S|\mathcal{M}, \iota, z, H_0, q_0, \sigma_A, \rho_*) \quad (21)$$

$$\begin{aligned} &= \int d\hat{A}_+ \int d\hat{A}_\times P(\hat{A}_+, \hat{A}_\times | \mathcal{M}, \iota, z, H_0, q_0, \sigma_A) P(S|\hat{A}_+, \hat{A}_\times, \sigma_A) \\ &= 1 - \frac{1}{\pi} \int_0^{\phi_{\max}} d\phi \left\{ \exp[-r_{\min}^2(\phi)/2] - \exp[-r_{\max}^2(\phi)/2] \right\}, \end{aligned}$$

where  $\phi_{\max} = \min[\arcsin(\rho/\rho_*), \pi]$ ,

$$r_{\min}(\phi) = \max\left\{0, \rho \cos(\phi) - [\rho_*^2 - \rho^2 \sin^2(\phi)]^{1/2}\right\}, \quad (22)$$

$$r_{\max}(\phi) = \rho \cos(\phi) + [\rho_*^2 - \rho^2 \sin^2(\phi)]^{1/2},$$

and

$$\rho = \frac{G(1+z)\mathcal{M}/c^2 [\cos^4(\iota) + 6 \cos^2(\iota) + 1]^{1/2}}{D(z, H_0, q_0)} \frac{1}{2 \sigma_A} \quad (23)$$

is, from Eq. 16, the mean SNR (*cf.* [71]). The combination of additive noise and the inverse distance dependence of the signal strength effectively defines a maximum distance out to which a survey can detect sources. For the simple model defined by Eq. 16, and assuming redshifts of  $z \ll 1$ , this is given by

$$D_* \simeq \frac{G\mathcal{M}}{c^2 \rho_* \sigma_A}, \quad (24)$$

but in many cases (*e.g.*, Sec. II) it is most useful to think of  $D_*$  itself as a characteristic of a survey.

In reality, the selection of a sample of BNS merger events with confirmed hosts also depends on whether it is possible to confirm a host galaxy from EM follow-up observations (*e.g.*, [48]). For the moment all GW events are going to be the subject of intense follow-up observing campaigns, and so it is assumed here both that every BNS merger that occurs in a galaxy will have its host identified and, further, that a spectroscopic redshift measurement will subsequently be made. While some fraction of events might be host-less (*e.g.* [80, 81]) this will just reduce the useful sample size produced by any given survey; the inference of  $H_0$  should not be affected.



### E. Expected number of events

Assuming BNS mergers are independent of each other, and the parent population of BNS systems is large, it is reasonable to model the sample size,  $N$ , as a draw from a Poisson distribution. This is characterized purely by the expected number of events, which (as described further in Appendix A 3 b) is given by integrating over the BNS parameters and the observing period  $T$  to obtain

$$\begin{aligned} \bar{N}(\Gamma, \bar{\mathcal{M}}, \sigma_{\mathcal{M}}, H_0, q_0, \sigma_A, \rho_*, T) \\ = T \int_0^\infty dz \frac{\Gamma}{1+z} \frac{dV}{dz}(H_0, q_0) \\ \int_0^\infty d\mathcal{M} \int_0^\pi d\iota P(\mathcal{M}, \iota | \bar{\mathcal{M}}, \sigma_{\mathcal{M}}) P(S | \mathcal{M}, \iota, z, H_0, q_0, \sigma_A, \rho_*), \end{aligned} \quad (25)$$

where  $P(\mathcal{M}, \iota | \bar{\mathcal{M}}, \sigma_{\mathcal{M}})$  is given in Eq. 14 and  $P(S | \mathcal{M}, \iota, z, H_0, q_0, \sigma_A, \rho_*)$  is given in Eq. 21. In general, these integrals must be evaluated numerically, *e.g.*, using a Monte Carlo approach such as that described below in Sec. IV F.

Particular care needs to be taken here as the Gaussian GW noise model defined in Sec. IV C results in a non-zero selection probability even for sources at an infinite distance with a true SNR of  $\rho = 0$ . Given that the volume element as specified in Sec. IV B potentially increases to infinite redshifts, the combination is an unnormalizable distribution with an infinite number of expected sources. This is not a problem in practice as the integration in Eq. 25 can be truncated at a finite redshift beyond the detection horizon; this model is hence best understood as a numerical approximation.

Another important aspect of Eq. 25 is that  $\bar{N}$  is nearly independent of  $H_0$ , and can hence potentially be ignored in the data analysis step. For low redshifts the comoving volume element scales as  $H_0^{-3}$ , but the effective maximum redshift of the survey is  $z_* \simeq H_0 D_*/c \propto H_0$ , so the implied volume in redshift space scales as  $H_0^3$ ; the two effects cancel out to leading order (*cf.* [45, 62]). This result can also be understood from a purely physical point of view: the detectability of a GW event is not significantly affected by its recession velocity; only its (luminosity) distance is important, and so changing the radial integration variable to  $D$  would (largely) remove  $H_0$  from the calculation of  $\bar{N}$ .

### F. Simulation algorithm

The algorithm used here to generate a self-consistently selected sample of low-redshift BNS merger events (as used in Sec. VI) is:

1. Choose values of the cosmological model parameters,  $H_0$  and  $q_0$  (Sec. IV A), and BNS population properties,  $\Gamma$ ,  $\bar{\mathcal{M}}$  and  $\sigma_{\mathcal{M}}$  (Sec. IV B), along with the observational characteristics of the GW and EM observations,  $T$ ,  $\sigma_A$ ,  $\rho_*$ ,  $\sigma_z$  and  $\sigma_v$  (Sec. IV C).

2. Identify a maximum redshift (or, equivalently, distance) such that there is a negligible probability of any merger beyond this being detected in the survey. If a fully realistic population model were used this redshift could be arbitrarily large, as evolution would ensure a finite sample size; but, for the simple non-evolving model used here, care needs to be taken to avoid including the spurious secondary peak discussed in Sec. IV E. Less fundamentally, using a more carefully chosen maximum redshift is also important for computational efficiency, to avoid the need for simulating large numbers of undetectable sources. The approach used for setting  $z_{\max}$  here is to calculate the distance at which a merger with a chirp mass 3 standard deviations above the population mean would require a  $3\sigma_A$  positive noise deviation to be detected. Further adopting a linear redshift-distance relationship, this gives

$$z_{\max} = \frac{G H_0}{c^3} \frac{\bar{\mathcal{M}} + 3\sigma_{\mathcal{M}}}{(\rho_* - 3)\sigma_A}. \quad (26)$$

In the simulations shown in Fig. 4  $z_{\max}$  corresponds to a maximum distance of  $D_{\max} \simeq 900$  Mpc, which is clearly more than sufficient to include all detectable sources.

3. Calculate the expected number of BNS mergers with  $z \leq z_{\max}$  during the observing period  $T$  as

$$\bar{N}_{\max} = T \int_0^{z_{\max}} dz \frac{\Gamma}{1+z} \frac{dV}{dz}(H_0, q_0). \quad (27)$$

4. Draw the actual number of mergers in this volume,  $N_{\max}$ , from a Poisson distribution of mean  $\bar{N}_{\max}$ .
5. For each of these  $N_{\max}$  BNS mergers:

- (a) Draw a redshift from the distribution

$$\begin{aligned} P(z | z_{\max}, H_0, q_0) \\ \propto \Theta(z) \Theta(z_{\max} - z) \frac{1}{1+z} \frac{dV}{dz}(H_0, q_0), \end{aligned} \quad (28)$$

and both a chirp mass and inclination from  $P(\mathcal{M}, \iota | \bar{\mathcal{M}}, \sigma_{\mathcal{M}})$  as given in Eq. 14.

- (b) Draw measured amplitudes  $\hat{A}_+$  and  $\hat{A}_\times$  (and, optionally, a measured chirp mass,  $\hat{\mathcal{M}}_z$ ) from Eq. 16, and hence calculate the observed SNR,  $\hat{\rho} = \rho(\hat{A}_+, \hat{A}_\times, \sigma_A)$ , according to Eq. 19.
- (c) Select the merger into the sample of detected events if  $\hat{\rho} \geq \rho_*$ .
- (d) If the merger is selected then draw a host peculiar velocity  $v$  from Eq. 15, a measured redshift,  $\hat{z}$ , from Eq. 17 and, optionally, a measured peculiar velocity,  $\hat{v}$ , from Eq. 18.

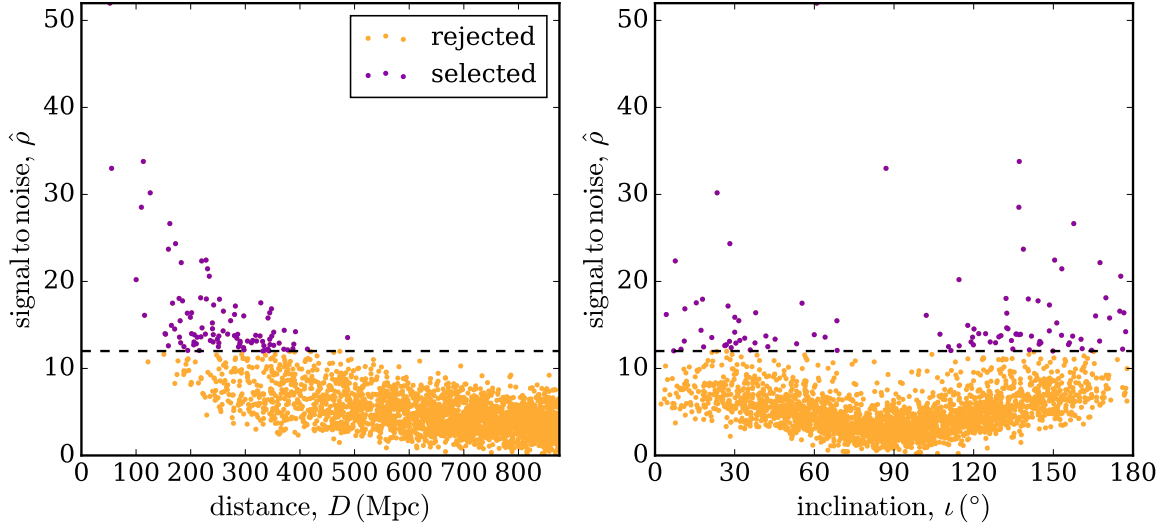


FIG. 4. Distributions of the measured SNR,  $\hat{\rho}$ , and distance (left) and inclination (right) for a simulated sample of 100 selected BNS merger events (purple) with  $\rho \geq \rho_* = 12$ . The much larger number of mergers which were not selected are also shown in orange.

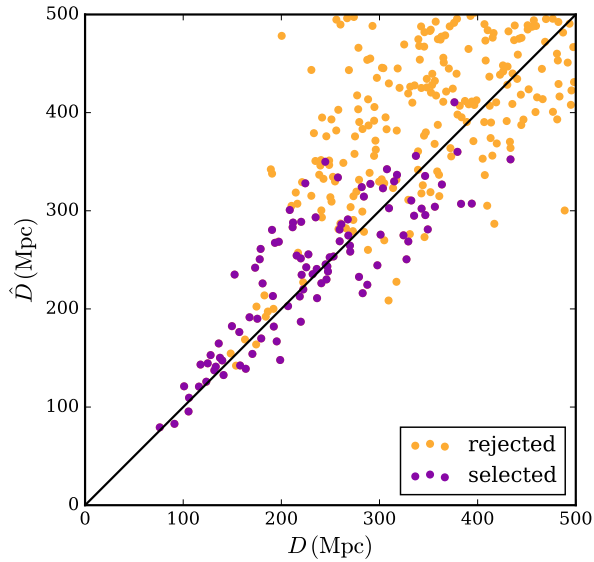


FIG. 5. Comparison of the fiducial best-fit distance,  $\hat{D}$ , and the true distance,  $D$ , for both selected (purple) and rejected (orange) sources. The over-abundance of sources close to the maximum survey distance with  $\hat{D} < D$  would potentially give a biased estimate of  $H_0$ , as described in Sec. III.

The output of the above algorithm is a sample of  $N$  detected BNS mergers along with their redshifted chirp masses,  $\hat{\mathcal{M}}_z = (\hat{\mathcal{M}}_{z,1}, \hat{\mathcal{M}}_{z,2}, \dots, \hat{\mathcal{M}}_{z,N})$ , measured GW amplitudes,  $\hat{\mathbf{A}} = (\hat{A}_{+,1}, \hat{A}_{\times,1}, \hat{A}_{+,2}, \hat{A}_{\times,2}, \dots, \hat{A}_{+,N}, \hat{A}_{\times,N})$ , their hosts' spectroscopic redshifts,  $\hat{\mathbf{z}} = (\hat{z}_1, \hat{z}_2, \dots, \hat{z}_N)$ , and (possibly) their estimated peculiar velocities,  $\hat{\mathbf{v}} = (\hat{v}_1, \hat{v}_2, \dots, \hat{v}_N)$ , along with the associated

measurement uncertainties. The algorithm also produces the true redshifts,  $\mathbf{z} = (z_1, z_2, \dots, z_N)$ , chirp masses,  $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_N)$ , inclinations,  $\boldsymbol{\iota} = (\iota_1, \iota_2, \dots, \iota_N)$ , and host peculiar velocities,  $\mathbf{v} = (v_1, v_2, \dots, v_N)$ , for these selected mergers, although these quantities are not actually part of a simulated catalog, and cannot be used in the parameter inference described in Sec. V.

One important feature of the above approach is that, while  $N$  is a draw from the appropriate Poisson distribution of mean  $\bar{N}$ , the integral in Eq. 25 is never actually evaluated explicitly. This is not particularly important for the fairly simple model used here, but is potentially critical in the more general cosmological case described in Appendix A 1. An example of a sample generated in this way is shown in Fig. 4, in particular illustrating that the most distant detected sources in the sample are those for which the noise has conspired to increase the SNR. These are also typically face-on or face-off sources, despite these configurations having a low prior probability (Eq. 14). If such data were treated naively, it would produce systematic underestimates for the distances to mergers close to the maximum survey distance. For the sample shown in Fig. 5 this would induce a bias of  $\sim 2 \text{ km s}^{-1} \text{ Mpc}^{-1}$  in the inferred value of  $H_0$  (*cf.* Sec. III).

## V. BAYESIAN PARAMETER INFERENCE

A sample of BNS mergers with measured GW data, spectroscopic host redshifts and (possibly) estimated peculiar velocities places constraints on the cosmological model, the BNS population and the properties of the mergers. This is encoded in the joint posterior distribution of all the parameters. For the population and

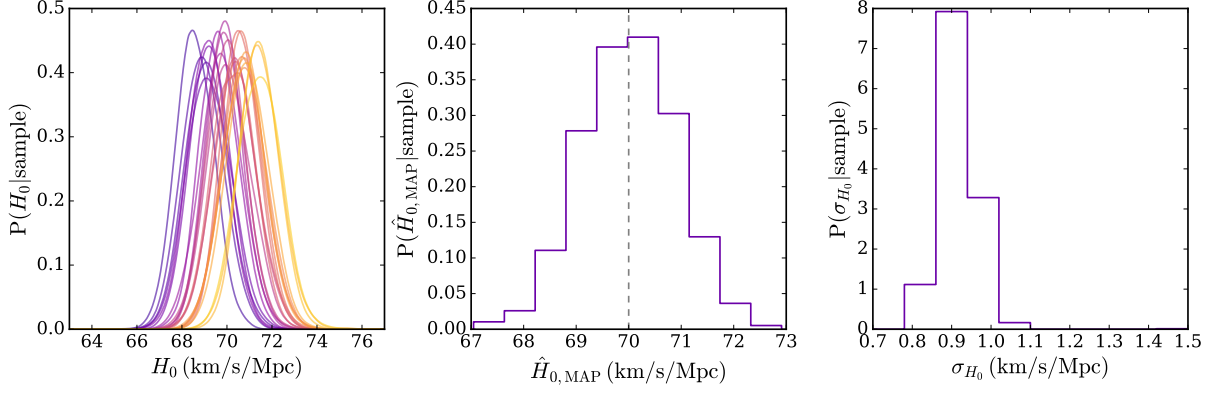


FIG. 6. Left:  $H_0$  posteriors for 25 independent 100-BNS samples, generated assuming a linear Hubble relation. The posteriors are colored dark purple to light orange by their MAP  $H_0$  values. Center/right: distribution of MAP  $H_0$  values and posterior standard deviations for the full set of 1000 100-BNS samples.

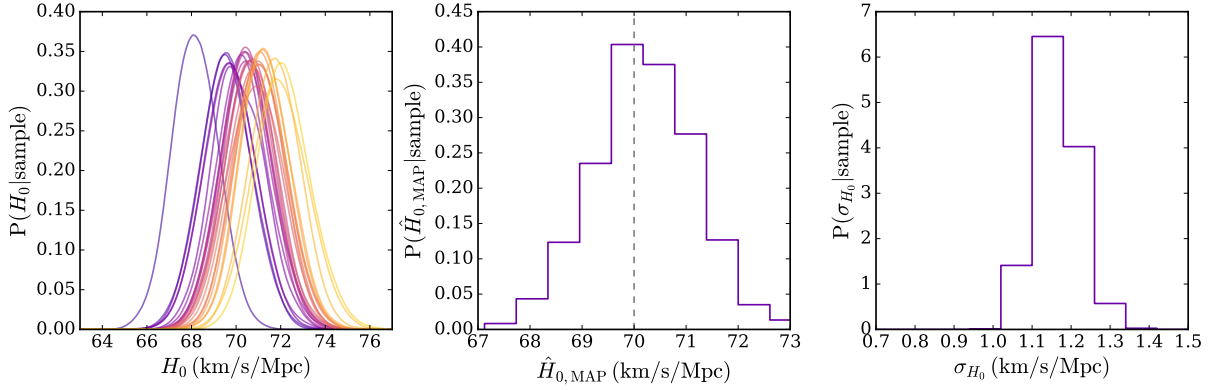


FIG. 7. As Fig. 6, but generated assuming a quadratic Hubble relation. From left to right: example  $H_0$  posteriors, colored from low to high MAP value; distributions of MAP  $H_0$  values and posterior standard deviations for 1000 100-BNS samples.

data model described above in Sec. IV this has the form  $P(\mathcal{M}, \boldsymbol{\iota}, \mathbf{z}, \mathbf{v}, \Gamma, \bar{\mathcal{M}}, \sigma_{\mathcal{M}}, H_0, q_0 | N, \hat{\mathcal{M}}_z, \hat{\mathbf{A}}, \hat{\mathbf{z}}, \hat{\mathbf{v}}, \sigma_A, \sigma_z, \sigma_v, \rho_*, T, I)$ , where  $I$  represents the prior information assumed about the cosmological and BNS population parameters. A derivation from first principles of the general form of this posterior distribution, valid for arbitrary cosmologies

and BNS population models, is given in Sec. A 4. This has the same structure and dependencies as the simpler model described in Sec. IV, so the above posterior can be obtained from Eq. A9 by making the identifications  $\boldsymbol{\theta} \rightarrow \mathcal{M}$ ,  $\beta \rightarrow (\Gamma, \bar{\mathcal{M}}, \sigma_{\mathcal{M}})$ ,  $\Omega \rightarrow (H_0, q_0)$  and  $O \rightarrow (\sigma_A, \sigma_z, \sigma_v, \rho_*, T)$ . This gives

$$\begin{aligned}
 & P(\mathcal{M}, \boldsymbol{\iota}, \mathbf{z}, \mathbf{v}, \Gamma, \bar{\mathcal{M}}, \sigma_{\mathcal{M}}, H_0, q_0 | N, \hat{\mathcal{M}}_z, \hat{\mathbf{A}}, \hat{\mathbf{z}}, \hat{\mathbf{v}}, \sigma_A, \sigma_z, \sigma_v, \rho_*, T, I) \\
 & \propto P(\Gamma, \bar{\mathcal{M}}, \sigma_{\mathcal{M}} | I) P(H_0, q_0 | I) \exp[-\bar{N}(\Gamma, \bar{\mathcal{M}}, \sigma_{\mathcal{M}}, H_0, q_0, \sigma_A, \rho_*, T)] \\
 & \times \prod_{i=1}^N \frac{\Gamma}{1+z_i} \frac{dV}{dz}(H_0, q_0) P(\mathcal{M}_i, \boldsymbol{\iota}_i; \bar{\mathcal{M}}, \sigma_{\mathcal{M}}) P(\hat{\mathcal{M}}_{z,i}, \hat{A}_{+,i}, \hat{A}_{\times,i} | \mathcal{M}_i, \boldsymbol{\iota}_i, z_i, H_0, q_0, \sigma_A) P(v_i | \sigma_{||}) P(\hat{z}_i | z_i, v_i, \sigma_z) P(\hat{v}_i | v_i, \sigma_v),
 \end{aligned} \tag{29}$$

where the expected number of detected mergers  $\bar{N}(\Gamma, \bar{\mathcal{M}}, \sigma_{\mathcal{M}}, H_0, q_0, \sigma_A, \rho_*, T)$  is defined in Eq. 25, the volume element  $dV/dz(H_0, q_0)$  is given in Eq. 13, the BNS demographic  $P(\mathcal{M}, \boldsymbol{\iota}; \bar{\mathcal{M}}, \sigma_{\mathcal{M}})$  is given in Eq. 14,

the GW likelihood  $P(\hat{\mathcal{M}}_z, \hat{A}_+, \hat{A}_{\times} | \mathcal{M}, \boldsymbol{\iota}, z, H_0, q_0, \sigma_A)$  is given in Eq. 16, the peculiar velocity prior  $P(v | \sigma_{||})$  is given in Eq. 15, the redshift likelihood  $P(\hat{z} | z, v, \sigma_z)$  is given in Eq. 17, and the peculiar velocity likelihood

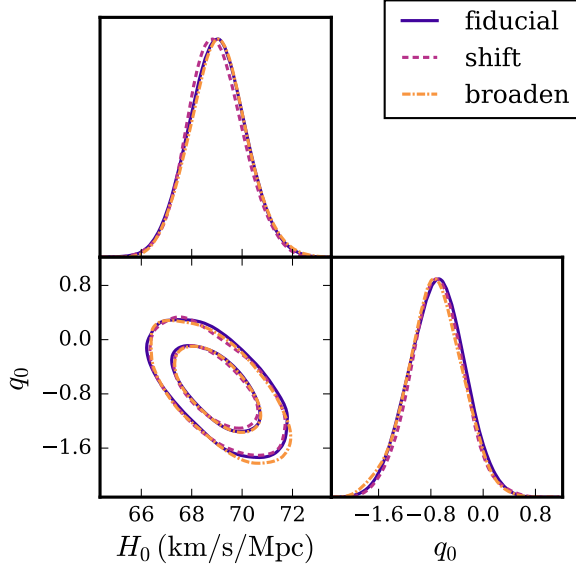


FIG. 8. Constraints on  $H_0$  and  $q_0$  for a single 100-event sample, inferred using our fiducial  $\mathcal{M}$  prior (purple solid); a prior shifted upwards by one sigma (pink dashed); and a prior with four times the variance (orange dot-dashed). The inference is robust to reasonable variations in the chirp mass prior.

$P(\hat{v}|v, \sigma_v)$  is given in Eq. 18.

The prior in the cosmological parameters is taken to be

$$P(H_0, q_0|I) \propto N(H_0; \hat{H}_0, \sigma_H^2) \Theta(q_0 + 2) \Theta(1 - q_0) N(q_0; -0.5, 0.5^2), \quad (30)$$

where the  $H_0$  prior is centred on  $\hat{H}_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and has a width of  $\sigma_H = 20 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . The truncation in  $q_0$  avoids the region of parameter space ( $q_0 > 1/2 z_{\text{max}} - 1 \simeq 1.1$ ) in which  $dV/dz$  drops to zero inside the simulation volume described in Sec. IV F. For samples of more than a few BNS mergers these priors have a minimal effect on the inference of  $H_0$  (Sec. VI).

The BNS population parameters  $\mathcal{M}$ ,  $\sigma_{\mathcal{M}}$  and  $\Gamma$  could be fit along with the cosmological parameters, but doing so is prohibitive numerically (as explained below), given the large number of simulations needed for the tests described in Sec. VI. The prior in the population parameters is hence taken to be

$$P(\Gamma, \bar{\mathcal{M}}, \sigma_{\mathcal{M}}) = \delta(\Gamma - \hat{\Gamma}) \delta(\bar{\mathcal{M}} - \hat{\bar{\mathcal{M}}}) \delta(\sigma_{\mathcal{M}} - \hat{\sigma}_{\mathcal{M}}), \quad (31)$$

where the assumed values are  $\hat{\Gamma}_0 = 1540 \text{ Gpc}^{-3} \text{ yr}^{-1}$ ,  $\hat{\bar{\mathcal{M}}} = 1.2 M_{\odot}$  and  $\hat{\sigma}_{\mathcal{M}} = 0.12 M_{\odot}$ .

Evaluating the posterior (Eq. 29) requires calculating the expected number of detectable events,  $\bar{N}$ , for different values of the unspecified model parameters: in this case,  $H_0$  and, if relevant,  $q_0$ . Rather than evaluate Eq. 25 at every location in parameter space, it is more efficient to estimate this integral using a Monte Carlo approach

order	posterior bias	posterior width
linear	$0.01 \pm 0.89$	$0.92 \pm 0.30$
quadratic	$0.20 \pm 0.97$	$1.17 \pm 0.36$

TABLE I. Average bias and width of  $H_0$  posteriors calculated from our 1000 simulations of 100-event samples.

for a grid of models and then fit a smooth function to capture its parameter dependence. A 25-point grid of  $H_0$  (and  $q_0$ ) values and a fourth-order polynomial were used, although the results are insensitive to these specific choices.

With this fit in hand, the 402-parameter<sup>4</sup> joint posterior for a given catalog was explored using Hamiltonian Monte Carlo [82] as implemented in Stan [83, 84]. This yields  $\sim 700$  independent samples from each marginal  $H_0$  posterior (from 2000 total samples) in 25 seconds on four Intel Xeon (2.4 GHz) CPUs, sufficient to analyze large numbers of simulations. The Stan model code and Python driver used in this analysis are publicly available at <https://github.com/sfeeney/hh0>.

## VI. RESULTS

The inference formalism described in Sec. V can be tested on simulated BNS merger samples generated self-consistently as described in Sec. IV F. The datasets were generated with these parameter values:  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ;  $q_0 = -0.5$  or a linear Hubble relationship;  $\Gamma = 1540 \text{ Gpc}^{-3} \text{ yr}^{-1}$ ;  $\bar{\mathcal{M}} = 1.2 M_{\odot}$ ; and  $\sigma_{\mathcal{M}} = 0.12 M_{\odot}$ . The observations are defined by:  $\sigma_A = 2 \times 10^{-23}$ ;  $\rho_* = 12$ ;  $\sigma_{||} = 500 \text{ km s}^{-1}$ ;  $\sigma_v = 200 \text{ km s}^{-1}$ ; and  $\sigma_z = 0.001$ . This set-up gives a maximum survey distance of  $D_* \simeq 250 \text{ Mpc}$  and typical Hubble-flow velocity uncertainties of  $\sim 360 \text{ km s}^{-1}$ .

1000 independent merger catalogs of  $N = 100$  selected events were generated for each of two cosmological scenarios: a linear Hubble relationship described by  $H_0$  alone; and the quadratic relationship described by  $H_0$  and  $q_0$ , as set out in Sec. IV A. The resultant posterior distributions in  $H_0$  are shown for 25 independent simulated samples of  $N = 100$  mergers in the left panels of Fig. 6 (for the linear Hubble relation) and Fig. 7 (for the quadratic Hubble relation). Samples of this size are clearly approaching the asymptotic Gaussian regime, with the quadratic Hubble relation posteriors slightly broader (and more skewed) than in the linear setting due to the degeneracy between  $H_0$  and  $q_0$ . The right panels of Figs 6 and 7 show the distributions of posterior widths for the full sets of 1000 catalogs. These distributions, whose

<sup>4</sup> The parameters are  $H_0$ ,  $q_0$  and a true redshift, peculiar velocity, chirp mass and inclination for each of the 100 BNS mergers.

characteristics are summarized in Table I, show that posteriors formed from samples of 100 BNS mergers have similar uncertainties (with  $\sim 33\%$  variations in width), indicating that there is minimal inter-sample variance. However, the peaks of the  $H_0$  posterior from individual samples can be scattered significantly high or low (as was the case for the single sample used in Ref. [63]).

The distribution of maximum-posterior ( $\hat{H}_{0,\text{MAP}}$ ) values from samples of  $N = 100$  events are shown in the center panels of Figs 6 and 7. For both Hubble relations, the distribution of  $\hat{H}_{0,\text{MAP}}$  values is centered on the true underlying value, indicating that the method is unbiased in the presence of selection effects. The average bias for the two Hubble relations, defined as the difference between  $\hat{H}_{0,\text{MAP}}$  and  $H_0$ , is listed in Table I and is within a standard deviation of zero in both cases.

The robustness of this approach, and particularly the sensitivity of the  $H_0$  constraints to mis-specification of  $\bar{\mathcal{M}}$  and  $\sigma_{\mathcal{M}}$ , can be assessed by conditioning the inference on values for these parameters that are different from those used in the simulations. Figure 8 shows posteriors on  $H_0$  and  $q_0$  inferred using three different BNS population priors for a single 100-event sample with chirp masses generated from the fiducial model (Eq. 14). The joint posteriors obtained for an offset prior with  $\hat{\mathcal{M}} = \bar{\mathcal{M}} + \sigma_{\mathcal{M}}$  and  $\hat{\sigma}_{\mathcal{M}} = \sigma_{\mathcal{M}}$  are shown by the pink dashed curves; results for a broadened prior with  $\hat{\mathcal{M}} = \bar{\mathcal{M}}$  and  $\hat{\sigma}_{\mathcal{M}} = 4\sigma_{\mathcal{M}}$  are shown by the orange dot-dashed curves. The constraints on the cosmological parameters in both cases differ insignificantly from those obtained using the correct model ( $\hat{\mathcal{M}} = \bar{\mathcal{M}}$  and  $\hat{\sigma}_{\mathcal{M}} = \sigma_{\mathcal{M}}$ ), as shown by the solid purple curves.

## VII. CONCLUSIONS

Using a rigorous Bayesian procedure to analyze realistic samples of BNS mergers gives estimates of the Hubble constant that are unbiased in the presence of selection effects, and robust to mis-specification of the other cosmological parameters or the BNS population model. The resultant uncertainties match expectations, confirming that this method is both accurate and precise. Applying such methods to the sample of  $\sim 50$  BNS mergers that could be produced by LIGO+Virgo in the next  $\sim 5$  years should yield a robust  $H_0$  estimate accurate to less than  $2 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . This would be sufficient to resolve the current tension between local and cosmological measurements of  $H_0$ .

While the formalism derived and implemented here is very general and applicable in an arbitrary cosmological setting, the primary focus of the simulations was on self-consistency, in particular with the same selection rule(s) used both in the sample generation and the posterior calculations. In order to take such an approach for suf-

ficiently large numbers of simulations, it was necessary to adopt a simplified model of the GW waveform and fix some population-level parameters. One obvious extension of this work would be to use full numerical simulations of BNS and NS-BH inspirals and their resultant time series, using a simplified GW waveform and likelihood analysis such as BAYESTAR [85] or, more ambitiously, full inference using LALInference [73], BILBY [74] or PyCBC [75]. A single simulated sample of 51 BNS mergers was analyzed using LALInference in Ref. [63], suggesting that even with realistic individual object constraints the regime of asymptotic normality has been reached, but this remains to be demonstrated fully.

While the near-complete decoupling of the BNS mass distribution from the cosmological parameters due to the tight chirp mass constraints is central to this entire method of measuring the Hubble constant [40], it would also be preferable to extend the Bayesian inference formalism to include the population parameters as well, something which will be more important for NS-BH merger samples for which there is less prior knowledge. A related extension would be to include a more realistic model of the EM counterpart detection process, in particular the correlation between GW selection and counterpart identification due to viewing angle dependencies. This will become particularly important in the context of GW-based constraints on  $H_0$  and other cosmological parameters from potential combined samples of hundreds to thousands of BNS, NS-BH and BBH mergers expected in the next decade with upgraded and new GW detectors (*e.g.*, [86, 87]).

## VIII. ACKNOWLEDGMENTS

The authors acknowledge useful discussions with Will Farr, Jon Gair, Archisman Ghosh and Leo Singer, along with the detailed comments made by the (anonymous) referee. This work was performed in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. The Flatiron Institute is supported by the Simons Foundation. This work was also partially supported by a grant from the Simons Foundation. HVP was partially supported by the European Research Council (ERC) under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement number 306478-CosmicDawn, and the research environment grant 'Gravitational Radiation and Electromagnetic Astrophysical Transients (GREAT)' funded by the Swedish Research Council (VR) under Dnr 2016-06012. This work was partially enabled by funding from the UCL Cosmoparticle Initiative. ARW and SMN acknowledge the generous financial support of the Netherlands Organization for Scientific Research through the NWO VIDI and NWO TOP Grants (PI: Nissanke).

- 
- [1] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic, *ApJ* **876**, 85 (2019).
- [2] Planck Collaboration, arXiv e-prints (2018), arXiv:1807.06209 [astro-ph.CO].
- [3] M. Wyman, D. H. Rudd, R. A. Vanderveld, and W. Hu, *PRL* **112**, 051302 (2014).
- [4] A. Pourtsidou and T. Tram, *PRD* **94**, 043518 (2016).
- [5] E. Di Valentino, A. Melchiorri, and J. Silk, *Physics Letters B* **761**, 242 (2016).
- [6] Q.-G. Huang and K. Wang, *EPJC* **76**, 506 (2016).
- [7] J. L. Bernal, L. Verde, and A. G. Riess, *JCAP* **10**, 019 (2016).
- [8] P. Ko and Y. Tang, *Physics Letters B* **762**, 462 (2016).
- [9] T. Karwal and M. Kamionkowski, *PRD* **94**, 103523 (2016).
- [10] S. Kumar and R. C. Nunes, *PRD* **94**, 123511 (2016).
- [11] B. Santos, A. A. Coley, N. Chandrachani Devi, and J. S. Alcaniz, *JCAP* **2**, 047 (2017).
- [12] V. Prilepina and Y. Tsai, *JHEP* **9**, 33 (2017).
- [13] G.-B. Zhao *et al.*, *Nature Astronomy* **1**, 627 (2017).
- [14] M.-M. Zhao, D.-Z. He, J.-F. Zhang, and X. Zhang, *PRD* **96**, 043520 (2017).
- [15] J. Solà, A. Gómez-Valent, and J. de Cruz Pérez, *Physics Letters B* **774**, 317 (2017).
- [16] E. Di Valentino, E. V. Linder, and A. Melchiorri, *PRD* **97**, 043528 (2018).
- [17] E. Di Valentino, C. Boehm, E. Hivon, and F. R. Bouchet, *PRD* **97**, 043513 (2018).
- [18] L. L. Graef, M. Benetti, and J. S. Alcaniz, *PRD* **99**, 043519 (2019).
- [19] W. Yang, A. Mukherjee, E. Di Valentino, and S. Pan, *PRD* **98**, 123527 (2018).
- [20] A. El-Zant, W. El Hanafy, and S. Elgammal, *ApJ* **871**, 210 (2019).
- [21] G. Benevento, M. Raveri, A. Lazanu, N. Bartolo, M. Liguori, P. Brax, and P. Valageas, *JCAP* **2019**, 027 (2019).
- [22] W. Yang, S. Pan, E. Di Valentino, E. N. Saridakis, and S. Chakraborty, *PRD* **99**, 043543 (2019).
- [23] K. Aylor, M. Joy, L. Knox, M. Millea, S. Raghunathan, and W. L. Kimmy Wu, *ApJ* **874**, 4 (2019).
- [24] C.-T. Chiang and A. Slosar, ArXiv e-prints (2018), arXiv:1811.03624.
- [25] V. Poulin, T. L. Smith, T. Karwal, and M. Kamionkowski, *PRL* **122**, 221301 (2019).
- [26] G. Efstathiou, *MNRAS* **440**, 1138 (2014).
- [27] D. N. Spergel, R. Flauger, and R. Hložek, *PRD* **91**, 023518 (2015).
- [28] M. Rigault *et al.*, *ApJ* **802**, 20 (2015).
- [29] D. O. Jones, A. G. Riess, and D. M. Scolnic, *ApJ* **812**, 31 (2015).
- [30] G. E. Addison, Y. Huang, D. J. Watts, C. L. Bennett, M. Halpern, G. Hinshaw, and J. L. Weiland, *ApJ* **818**, 132 (2016).
- [31] Planck Collaboration, *A&A* **596**, A107 (2016).
- [32] W. Cardona, M. Kunz, and V. Pettorino, *JCAP* **3**, 056 (2017).
- [33] B. R. Zhang, M. J. Childress, T. M. Davis, N. V. Karpenka, C. Lidman, B. P. Schmidt, and M. Smith, *MNRAS* **471**, 2254 (2017).
- [34] H.-Y. Wu and D. Huterer, *MNRAS* **471**, 4946 (2017).
- [35] S. M. Feeney, D. J. Mortlock, and N. Dalmasso, *MNRAS* **476**, 3861 (2018).
- [36] B. Follin and L. Knox, *MNRAS* **477**, 4534 (2018).
- [37] S. Dhawan, S. W. Jha, and B. Leibundgut, *A&A* **609**, A72 (2018).
- [38] M. Rigault *et al.*, ArXiv e-prints (2018), arXiv:1806.03849.
- [39] C. A. P. Bengaly, U. Andrade, and J. S. Alcaniz, ArXiv e-prints (2018), arXiv:1810.04966.
- [40] B. F. Schutz, *Nature* **323**, 310 (1986).
- [41] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, and et al., *PRL* **119**, 161101 (2017).
- [42] The LIGO Scientific Collaboration, the Virgo Collaboration, *et al.*, arXiv e-prints (2018), arXiv:1811.12907 [astro-ph.HE].
- [43] N. Dalal, D. E. Holz, S. A. Hughes, and B. Jain, *PRD* **74**, 063006 (2006).
- [44] S. Nissanke, D. E. Holz, S. A. Hughes, N. Dalal, and J. L. Sievers, *ApJ* **725**, 496 (2010).
- [45] B. P. Abbott *et al.*, *Nature* **551**, 85 (2017).
- [46] W. Del Pozzo, *PRD* **86**, 043011 (2012).
- [47] M. Oguri, *PRD* **93**, 083511 (2016).
- [48] M. Fishbach, R. Gray, I. Magaña Hernandez, H. Qi, A. Sur, F. Acernese, L. Aiello, A. Allocca, M. A. Aloy, and A. Amato, *ApJL* **871**, L13 (2019).
- [49] S. R. Taylor and J. R. Gair, *PRD* **86**, 023502 (2012).
- [50] C. Messenger and J. Read, *PRL* **108**, 091101 (2012).
- [51] K. P. Mooley, A. T. Deller, O. Gottlieb, E. Nakar, G. Hallinan, S. Bourke, D. A. Frail, A. Horeh, A. Corsi, and K. Hotokezaka, *Nature* **561**, 355 (2018).
- [52] C. Guidorzi, R. Margutti, D. Brout, D. Scolnic, W. Fong, K. D. Alexander, P. S. Cowperthwaite, J. Annis, E. Berger, P. K. Blanchard, R. Chornock, D. L. Coppejans, T. Eftekhari, J. A. Frieman, D. Huterer, M. Nicholl, M. Soares-Santos, G. Terreran, V. A. Villar, and P. K. G. Williams, *ApJL* **851**, L36 (2017).
- [53] K. Hotokezaka, E. Nakar, O. Gottlieb, S. Nissanke, K. Masuda, G. Hallinan, K. P. Mooley, and A. T. Deller, *Nature Astronomy*, 385 (2019).
- [54] The LIGO Scientific Collaboration, the Virgo Collaboration, *et al.*, arXiv e-prints (2018), arXiv:1811.12940 [astro-ph.HE].
- [55] R. Nair, S. Bose, and T. D. Saini, *PRD* **98**, 023502 (2018).
- [56] R. Gray, I. Magaña Hernandez, H. Qi, A. Sur, P. R. Brady, H.-Y. Chen, W. M. Farr, M. Fishbach, J. R. Gair, A. Ghosh, D. E. Holz, S. Mastrogiiovanni, C. Messenger, D. A. Steer, and J. Veitch, arXiv e-prints (2019), arXiv:1908.06050 [gr-qc].
- [57] M. Soares-Santos, A. Palmese, W. Hartley, J. Annis, J. Garcia-Bellido, O. Lahav, Z. Doctor, M. Fishbach, D. E. Holz, and H. Lin, *ApJL* **876**, L7 (2019).
- [58] The LIGO Scientific Collaboration, the Virgo Collaboration, *et al.*, arXiv e-prints (2019), arXiv:1908.06060 [astro-ph.CO].
- [59] S. Nissanke, D. E. Holz, N. Dalal, S. A. Hughes, J. L. Sievers, and C. M. Hirata, ArXiv e-prints (2013), arXiv:1307.2638.
- [60] S. Vitale and H.-Y. Chen, *PRL* **121**, 021303 (2018).
- [61] N. Seto and K. Kyutoku, *MNRAS* **475**, 4133 (2018).

- [62] H.-Y. Chen, M. Fishbach, and D. Holz, *Nature* **562**, 545 (2018).
- [63] S. M. Feeney, H. V. Peiris, A. R. Williamson, S. M. Nissanke, D. J. Mortlock, J. Alsing, and D. Scolnic, *PRL* **122**, 061105 (2019).
- [64] S. R. Taylor, J. R. Gair, and I. Mandel, *PRD* **85**, 023535 (2012).
- [65] C. Messenger and J. Veitch, *New Journal of Physics* **15**, 053027 (2013).
- [66] S. Karki, D. Tuyenbayev, S. Kandhasamy, B. P. Abbott, T. D. Abbott, E. H. Anders, J. Berliner, J. Betzwieser, C. Cahillane, L. Canete, C. Conley, H. P. Daveloza, N. De Lillo, J. R. Gleason, E. Goetz, K. Izumi, J. S. Kissel, G. Mendell, V. Quetschke, M. Rodruck, S. Sachdev, T. Sadecki, P. B. Schwinberg, A. Sottile, M. Wade, A. J. Weinstein, M. West, and R. L. Savage, *Review of Scientific Instruments* **87**, 114503 (2016).
- [67] C. Cahillane, J. Betzwieser, D. A. Brown, E. Goetz, E. D. Hall, K. Izumi, S. Kandhasamy, S. Karki, J. S. Kissel, G. Mendell, R. L. Savage, D. Tuyenbayev, A. Urban, A. Viets, M. Wade, and A. J. Weinstein, *PRD* **96**, 102001 (2017).
- [68] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al., *Physical Review X* **6**, 041015 (2016).
- [69] I. Mandel, W. M. Farr, and J. R. Gair, *MNRAS* **486**, 1086 (2019).
- [70] D. Burstein, *Reports on Progress in Physics* **53**, 421 (1990).
- [71] C. Cutler and É. E. Flanagan, *PRD* **49**, 2658 (1994).
- [72] D. Marković, *PRD* **48**, 4738 (1993).
- [73] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C. J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O’Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. van Der Sluys, A. Vecchio, W. Voudsen, and L. Wade, *PRD* **91**, 042003 (2015).
- [74] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov, F. Hernandez Vivanco, J. Harms, M. E. Lower, G. D. Meadors, D. Melchor, E. Payne, M. D. Pitkin, J. Powell, N. Sarin, R. J. E. Smith, and E. Thrane, *ApJS* **241**, 27 (2019).
- [75] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, *PASP* **131**, 024503 (2019).
- [76] S. Weinberg, *Cosmology* (Oxford University Press, 2008).
- [77] S. Nissanke, M. Kasliwal, and A. Georgieva, *ApJ* **767**, 124 (2013), arXiv:1210.6362 [astro-ph.HE].
- [78] LIGO Scientific Collaboration, “LIGO Algorithm Library - LALSuite,” free software (GPL) (2018).
- [79] A. C. Crook, J. P. Huchra, N. Martimbeau, K. L. Masters, T. Jarrett, and L. M. Macri, *ApJ* **655**, 790 (2007).
- [80] E. Berger, *ApJ* **722**, 1946 (2010).
- [81] W. Fong and E. Berger, *ApJ* **776**, 18 (2013).
- [82] R. M. Neal, *ArXiv e-prints* (2012), arXiv:1206.1901 [stat.CO].
- [83] B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, *Journal of Statistical Software, Articles* **76**, 1 (2017).
- [84] Stan Development Team, “PyStan: the python interface to stan, Version 2.14.0.0,” (2016).
- [85] L. P. Singer and L. R. Price, *PRD* **93**, 024013 (2016).
- [86] B. S. Sathyaprakash, B. F. Schutz, and C. Van Den Broeck, *Classical and Quantum Gravity* **27**, 215006 (2010).
- [87] A. Nishizawa, K. Yagi, A. Taruya, and T. Tanaka, in *Journal of Physics Conference Series*, Journal of Physics Conference Series, Vol. 363 (2012) p. 012052.
- [88] W. Del Pozzo, T. G. F. Li, and C. Messenger, *PRD* **95**, 043502 (2017).
- [89] E. Belgacem, Y. Dirian, S. Foffa, and M. Maggiore, *PRD* **98**, 023510 (2018).
- [90] S. M. Carroll, W. H. Press, and E. L. Turner, *ARA&A* **30**, 499 (1992).
- [91] E. L. Wright, *PASP* **118**, 1711 (2006).
- [92] T. J. Loredo, in *American Institute of Physics Conference Series*, American Institute of Physics Conference Series, Vol. 735, edited by R. Fischer, R. Preuss, and U. V. Toussaint (2004) pp. 195–206.
- [93] R. L. Streit, *Poisson Point Processes* (Springer, 2010).
- [94] L. Blanchet, *Living Rev. Rel.* **17**, 2 (2014).
- [95] M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments*, Oxford Masters Series in Physics (Oxford University Press, 2007).

## Appendix A: Bayesian inference of cosmological parameters from BNS mergers

The task of inferring the Hubble constant from BNS merger events can be thought of as a special case of the more general problem of cosmological parameter inference from such data (*e.g.*, [46, 72, 88, 89]). Combined with the fact that a Euclidean approximation is insufficient even for current low-redshift BNS merger samples, taking a fully general approach and then making low-redshift approximations ensures that any such simplifications are done rigorously. The BNS merger population model (Appendix A1), data (Appendix A2) and sample selection (Appendix A3) have the same structure as the more specific simulation described in Sec. IV. These ingredients are then combined self-consistently to obtain both the full posterior distribution and the marginalized posterior in the cosmological parameters (Appendix A4).



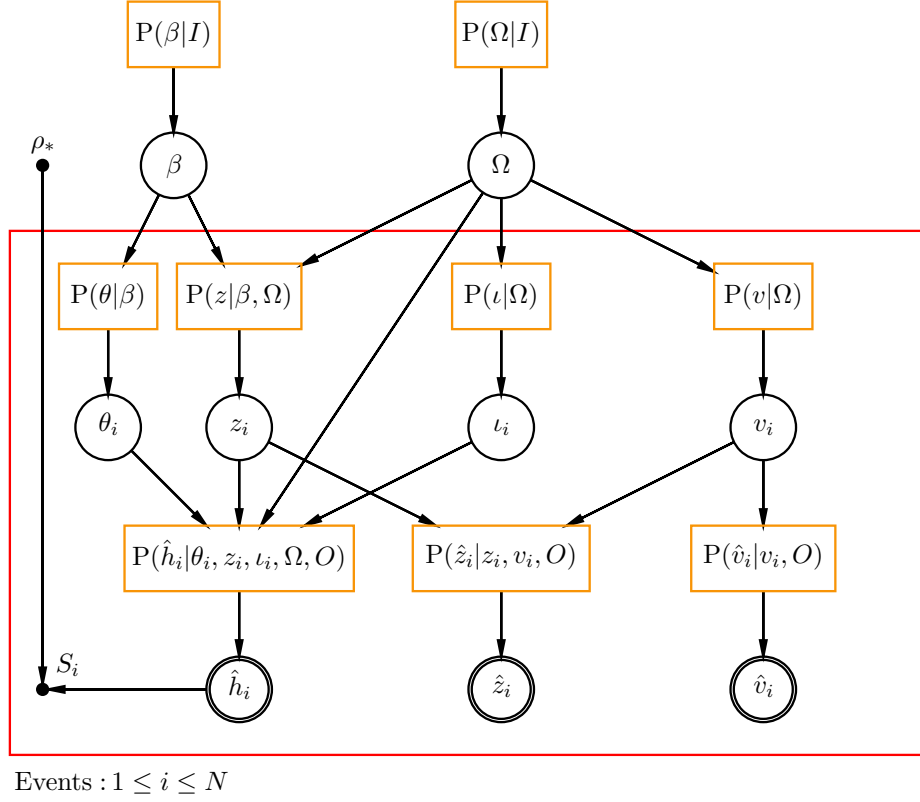


FIG. 9. Network diagram for the hierarchical model describing the BNS population and data. Quantities in single circles are model parameters; quantities in double circles are measured values. These quantities are linked by the probability distributions in orange rectangles: the top row are parameter priors; the middle row describe the population model; the bottom row define the (object-level) likelihood. The arrows linking the quantities and distributions define the forward/generative model that could be used to simulate samples and measurements. The quantities inside the red rectangular plate are specific to a single BNS merger event; the detected events are indexed by  $i \in \{1, 2, \dots, N\}$ .

## 1. Physical model

A necessary ingredient for any Bayesian inference formalism is a generative model which could be used to simulate a mock dataset. For the case of BNS mergers this includes both a cosmological model (Appendix A 1 a) and the BNS population (Appendix A 1 b), although it is primarily the dependence structure of the model, rather than specific functional forms, which are required at this stage. This physical model includes neither measurements (Appendix A 2) or sample selection (Appendix A 3 a), which are kept distinct as they play different roles in the simulation and inference. The overall structure of the model is summarized in Fig. 9.

### a. Cosmology

GW signals from BNS events can, in principle, be seen to cosmological distances and so must in general be analyzed in the context of a full cosmological model. This is described by a set of parameters  $\Omega = (H_0, \Omega_m, \Omega_\Lambda, \omega, \dots)$  which between them specify the contents of the universe and its expansion history. None of the results presented here depend strongly on the particular type of cosmological model adopted, so  $\Omega$  is used in order to preserve generality, although it is implicit that  $H_0$  is always included in this list.

The radial coordinate used to specify position along the line-of-sight is the (cosmological) redshift,  $z$ , defined strictly in terms of the ratio of the cosmological scale factor to the current value. This is distinct from the observable redshift,  $\hat{z}$ , which in general differs from  $z$  due to the source's peculiar velocity relative to the Hubble-Lemaître flow. Assuming this motion is non-relativistic, the observable redshift is given in terms of the line-of-sight component of the source's

peculiar velocity,  $v$ , as

$$\tilde{z} = (1 + z) \left( 1 + \frac{v}{c} \right) - 1 = z + (1 + z) \frac{v}{c}, \quad (\text{A1})$$

where positive  $v$  hence corresponds to motion away from the observer.

The redshift and cosmological model combine to specify both the luminosity distance,  $D(z, \Omega)$  and the co-moving volume element  $dV/dz(\Omega)$ . Expressions for  $D(z, \Omega)$  and  $dV/dz(\Omega)$  are available for standard cosmological models [90, 91], although the model is kept general here. The numerical marginalization scheme described in Sec. IV A requires the function  $z(D, \Omega)$ , defined such that  $z[D(z, \Omega), \Omega] = D[z(D, \Omega), \Omega] = 1$  for any cosmological model (and any non-negative value of  $z$  or  $D$ ). These inverses exist in standard cosmological models as the luminosity distance increases monotonically with  $z$  [76].

### b. The BNS merger population

The BNS population is defined primarily by the rate of mergers per unit proper time per unit co-moving volume,  $\Gamma(z, \beta)$ , where  $\beta$  are the parameters which describe the population model. The expected number of events in the redshift range  $z$  to  $z + dz$  that would be registered by a perfect all-sky detector in time interval  $dt$  is hence

$$d\bar{N} = \frac{\Gamma(z, \beta)}{1 + z} \frac{dV}{dz}(\Omega) dz dt, \quad (\text{A2})$$

where the reduction by a factor of  $(1 + z)$  comes about due to time dilation from the source frame, and  $dV/dz(\Omega)$  is the co-moving volume element defined in Appendix A 1 a.

The properties of a single BNS merger are separated into intrinsic system parameters,  $\theta$  (the NS masses, spins, *etc.*), and observer-dependent quantities, taken here to be the inclination of the system with respect to the line-of-sight,  $\iota$ , and the line-of-sight peculiar velocity of the host galaxy,  $v$  (and  $z$ , although this is already incorporated in the overall rate). For a given redshift, the population model is a (normalized) probability distribution of the form

$$P(\theta, v, \iota | z, \beta, \Omega) = P(\theta | z, \beta) P(v | z, \Omega) \Theta(\iota) \Theta(\pi - \iota) \frac{\sin(\iota)}{2}. \quad (\text{A3})$$

This encodes the assumption that BNS systems are oriented randomly, leading to the sinusoidal distribution in  $\iota$ .

Some care is required in treating the peculiar velocity of the BNS merger, as it has two distinct contributions: the motion of the host galaxy relative to the Hubble-Lemaître flow; and the orbital motion of the merger system within the host galaxy. (Only BNS mergers with counterparts are considered here, *cf.* [77], and all such systems have, by definition, a host galaxy.) Both these motions are expected to have typical speeds of  $\sim 10^2 \text{ km s}^{-1}$ , although they enter the inference formalism in distinct ways. In physical terms the peculiar velocity of the host galaxy,  $v$ , is determined by the local distribution of matter, but even in the absence of source-specific data (Appendix A 2 c), knowledge of cosmological structure formation implies a distribution of the form  $P(v | z, \Omega)$ . While there is a formal link between the peculiar velocity and the cosmological parameters, this is negligible compared to the link through the GW data, and so the dependence of  $P(v | z, \Omega)$  on  $\Omega$  can be ignored. This reflects the status of the host's peculiar velocity as a nuisance parameter in this context.

## 2. BNS merger data

Three distinct types of measurement provide information about the properties of a BNS merger event: GW strain time-series data (Appendix A 2 a); a spectroscopic redshift for the host galaxy (Appendix A 2 b); and, potentially, an estimated (line-of-sight) peculiar velocity for the host galaxy (Appendix A 2 c). All of these measured quantities have associated uncertainties (and other details associated with the measurement process); for brevity, all such quantities are combined into a single parameter,  $O$ , that characterizes the observations.

### a. GW data

The GW signal from a merger event comprises two orthogonal polarizations,  $h_p(t; \theta, \iota, v, z, \Omega)$ , with  $p \in \{+, \times\}$ , where the form of the  $t$  dependence is determined by the intrinsic merger properties,  $\theta$ , the observer-dependent quantities  $\iota$ ,  $v$  and  $z$ , and the cosmological parameters,  $\Omega$ . The full waveforms depend on the complicated non-linear

physics of the merger, but the dependence on  $v$  and  $z$  is determined purely by the physics of GW propagation in an expanding universe. The GW signal in the far field regime is subject to a time dilation by a factor of  $(1+z)(1+v/c)$  and the amplitude scales as  $1/D_P(z, \Omega)$ , where  $D_P(z, \Omega)$  is the proper distance to redshift  $z$ , so the expression  $D_P(z, \Omega) h_p[(t - t_0)/(1+z)(1+v/c); \theta, \iota, v, z, \Omega]$  is independent of redshift and peculiar velocity. The dependence of the strain on the peculiar velocity is sufficiently small that it can be ignored (Appendix B), so  $h_p(t; \theta, \iota, v, z, \Omega) \rightarrow h_p(t; \theta, \iota, z, \Omega)$  is assumed.

The GW data for a merger comprise a discretized time-series of measured strains, denoted  $\hat{h}$  for simplicity, along with associated uncertainties, implicitly included in  $O$ . The likelihood for the strain data hence has the form  $P(\hat{h}|\theta, \iota, z, \Omega, O)$ , which is kept general in this derivation; the specific form used in the simulations described in Sec. IV is detailed in Appendix B.

### *b. Redshift measurement*

The measured spectroscopic redshift of a BNS host,  $\hat{z}$ , is linked to both its cosmological redshift,  $z$ , and its line-of-sight peculiar velocity,  $v$ , leading to a likelihood of the form  $P(\hat{z}|z, v, O)$ , with the observable redshift given in Eq. A1.

### *c. Peculiar velocity estimates*

It is possible that the (line-of-sight) peculiar velocity of a BNS host can be estimated from the positions and/or motions of nearby galaxies, yielding an estimate  $\hat{v}$ . In the absence of further information, this contribution to the likelihood has the form  $P(\hat{v}|v, O)$ , where the information on the positions and/or velocities of nearby galaxies is left implicit. It is unrealistic to consider the case that the uncertainty is small or negligible, as there is no credible way to get precise peculiar velocity information; but it is useful to consider the limit that the uncertainty is infinite, which is equivalent to there being no useful peculiar velocity data at all for a given system. This situation could also be recovered by removing  $\hat{v}$  and the associated likelihood from the calculation altogether.

It is also possible in principle that the line-of-sight component of the orbital motion of the merger relative to its host galaxy could be estimated (*e.g.*, through the combination of a precise location and rotation curve), although the impact of this on the GW data is minor anyway (Appendix B) so this is not explored further here.

## **3. Sample of BNS events**

The full BNS merger dataset from a GW survey (and follow-up observations) is a catalog of  $N$  selected merger events, each with associated GW data,  $\hat{\mathbf{h}} = (\hat{h}_1, \hat{h}_2, \dots, \hat{h}_N)$  redshift measurements,  $\hat{\mathbf{z}} = (\hat{z}_1, \hat{z}_2, \dots, \hat{z}_N)$  and (possibly) host peculiar velocity information,  $\hat{\mathbf{v}} = (\hat{v}_1, \hat{v}_2, \dots, \hat{v}_N)$ , as described in Appendix A 2. The additional model ingredients needed to define the sample generation process is the selection rule (Appendix A 3 a), which then determines the the number of selected events (Appendix A 3 b).

### *a. Selection*

The selection of a BNS merger event into a sample is assumed to be determined by the GW data alone, and to take the form of a hard cut on some statistic calculated from the GW data,  $\hat{h}$ . This function is denoted  $\rho(\hat{h}, O)$  and can be thought of as the observed SNR (or a proxy for this). The form of  $\rho(\hat{h}, O)$  is not as important as the fact that it is deterministic, meaning the selection probability can be written as

$$P(S|\hat{h}, O) = \Theta[\rho(\hat{h}, O) - \rho_*]. \quad (\text{A4})$$

The fact that the selection is a deterministic function of the data means that for any event in a selected sample,  $\hat{\rho} = \rho(\hat{h}, O) \geq \rho_*$  and so  $P(S|\hat{h}, O) = 1$ , a critical fact [92] which simplifies the parameter inference calculation (Sec. V).

### b. Number of events

Modeling the BNS merger population as a realization of a Poisson point process [93], the number of detected events in a sample is drawn from the Poisson distribution

$$P(N|\Omega, \beta, O) = \Theta(N) \frac{[\bar{N}(\beta, \Omega, O)]^N \exp[-\bar{N}(\beta, \Omega, O)]}{N!}, \quad (\text{A5})$$

which is characterized purely by the expected number of events,  $\bar{N}(\Omega, \beta, O)$ . This is obtained by integrating the product of the event rate and the detection probability (assumed to depend only on the GW data) over the BNS merger properties, which gives

$$\bar{N}(\beta, \Omega, O) = T \int_0^\infty dz \frac{\Gamma(z, \beta)}{1+z} \frac{dV}{dz}(\Omega) \int d\theta P(\theta|z, \beta) \int_0^\pi d\iota \frac{\sin(\iota)}{2} \int d\hat{h} P(\hat{h}|\theta, \iota, z, \Omega, O) \Theta[\rho(\hat{h}, O) - \rho_*], \quad (\text{A6})$$

where the integral with respect to  $t$  yields the observing time,  $T$ , which is included in  $O$  along with  $\rho_*$ . In general, this integral must be evaluated numerically, *e.g.*, using a Monte Carlo approach such as that described in Sec. IV F, although in some applications they can be avoided (*e.g.*, [64]).

## 4. Parameter inference

The general inference task here is to obtain constraints on the full set of model parameters given all the data on a sample of  $N$  BNS merger events, indexed by  $i \in \{1, 2, \dots, N\}$ . The model, defined in Sec. IV, is characterized by the cosmological parameters,  $\Omega$ , the BNS population parameters,  $\beta$ , and the BNS mergers' physical properties,  $\theta = (\theta_1, \theta_2, \dots, \theta_N)$ , host redshifts,  $z = (z_1, z_2, \dots, z_N)$ , and peculiar velocities,  $v = (v_1, v_2, \dots, v_N)$ ; the data for the  $N$  events are, as described in Sec. IV C, the GW measurements  $\hat{h} = (\hat{h}_1, \hat{h}_2, \dots, \hat{h}_N)$ , redshift measurements  $\hat{z} = (\hat{z}_1, \hat{z}_2, \dots, \hat{z}_N)$  and (possibly) peculiar velocity information,  $\hat{v} = (\hat{v}_1, \hat{v}_2, \dots, \hat{v}_N)$ . Combined with a description the sample selection (Appendix A 3), this is sufficient to calculate the joint posterior distribution in all the parameters (Appendix A 4 a), and the marginalized distribution in the parameters of interest (Appendix A 4 b).

### a. Joint posterior distribution

The constraints on the cosmological parameters, BNS population and the individual events implied by a sample of  $N$  detected BNS mergers are fully described by the joint posterior distribution in all the model parameters,  $P(\theta, \iota, z, v, \beta, \Omega|N, \hat{h}, \hat{z}, \hat{v}, O, I)$ , where  $I$  is the prior information assumed about  $\Omega$  and  $\beta$ . The knowledge about the cosmological model and the BNS population are assumed to be independent, so that the prior distribution factorizes as  $P(\Omega, \beta|I) = P(\Omega|I) P(\beta|I)$ ; the prior information on all the other parameters is specified by  $\Omega$  and  $\beta$ , and so is conditionally-independent of  $I$ . The posterior is conditioned not only on the obvious data for each of the merger events (*i.e.*,  $\hat{h}$ ,  $\hat{z}$  and  $\hat{v}$ ), but also on the size of the sample (*i.e.*, the value of  $N$ ) and hence implicitly on the fact that the GW data from each detected event must satisfy the selection criterion outlined in Appendix A 3 a.

The task now is to write the joint posterior distribution in terms of the functions/distributions defined in Appendix A 1, Appendix A 2 and Appendix A 3. It is useful to introduce explicitly the fact that each of the merger events was selected, denoted  $\mathbf{S} = (S_1, S_2, \dots, S_N)$  following Appendix A 3 a. This can be added to the list of quantities being conditioned on because the selection process defined in Eq. 21 is deterministic – whether  $S$  is true can be determined from  $\hat{h}$  – which means that including  $\mathbf{S}$  does not add any extra information:  $P(\theta, \iota, z, v, \beta, \Omega|N, \hat{h}, \hat{z}, \hat{v}, O, I) = P(\theta, \iota, z, v, \beta, \Omega|N, \mathbf{S}, \hat{h}, \hat{z}, \hat{v}, O, I)$ . The reason for including  $\mathbf{S}$  is that it gives the freedom to condition on selection alone, which allows the the unnormalized posterior to be written as

$$P(\theta, \iota, z, v, \beta, \Omega|N, \hat{h}, \hat{z}, \hat{v}, O, I) \propto P(\Omega|I) P(\beta|I) P(N|\Omega, \beta, O) \prod_{i=1}^N P(\theta_i, \iota_i, z_i, v_i, \hat{h}_i, \hat{z}_i, \hat{v}_i|S_i, \beta, \Omega, O). \quad (\text{A7})$$

The first two terms are the prior distributions on the cosmological and BNS population parameters; the third term encodes the constraints provided by the number of detected events (Sec. IV E); but the terms in the product still require some manipulation to be written in terms of the distributions which define the model. Taking any one such

term and successively applying Bayes's theorem, the chain rule, and the law of total probability then yields the joint distribution in the intrinsic and observed properties of a selected merger event as

$$\begin{aligned} & P(\theta, \iota, z, v, \hat{h}, \hat{z}, \hat{v} | S, \beta, \Omega, O) \\ &= \frac{1}{\bar{N}(\Omega, \beta, O)} \frac{\Gamma(z, \beta)}{1+z} \frac{dV}{dz}(\Omega) P(\theta | z, \beta) \Theta(\iota) \Theta(\pi - \iota) \frac{\sin(\iota)}{2} P(v | z, \Omega) P(\hat{h} | \theta, \iota, z, \Omega, O) P(\hat{z} | z, v, O) P(\hat{v} | v, O), \end{aligned} \quad (\text{A8})$$

where the fact that  $P(S | \hat{h}, O) = 1$  for selected events has been used to omit this term, and the normalizing constant is equal to the expected number of events in the sample, given in Eq. A6.

Inserting this object-level distribution and the Poisson distribution of the number of events in Eq. A5 into Eq. A7 then gives the unnormalized joint posterior distribution in all the model parameters as

$$\begin{aligned} & P(\theta, \iota, z, v, \beta, \Omega | N, \hat{h}, \hat{z}, \hat{v}, O, I) \propto P(\Omega | I) P(\beta | I) \exp[-\bar{N}(\beta, \Omega, O)] \\ & \times \prod_{i=1}^N \frac{\Gamma(z_i, \beta)}{1+z_i} \frac{dV}{dz_i}(\Omega) \Theta(\iota_i) \Theta(\pi - \iota_i) \sin(\iota_i) P(\theta_i | z_i, \beta) P(\hat{h}_i | \theta_i, \iota_i, z_i, \Omega, O) P(v_i | z_i, \Omega) P(\hat{z}_i | z_i, v_i, O) P(\hat{v}_i | v_i, O), \end{aligned} \quad (\text{A9})$$

where the selection indicators  $\mathbf{S}$  have been omitted because (as argued above) they do not contain any extra information beyond that already encoded in  $\hat{\mathbf{h}}$ . This has the standard structure for the full likelihood of a Poisson point process [93], in particular with the expected number of detected sources appearing only in the exponential term, although the number of different component distributions in the product somewhat obscures the link to this standard statistical model. This posterior has the characteristic Poisson point process structure that also appears in *e.g.*, Refs. [64, 69].

Equation A9 represents the main result of the above model formulation as it encodes all the links between the measured data and the quantities of interest. Other than the assumption of random orientations encoded in the sinusoidal inclination distribution, it is deliberately kept completely general, with only the structure of the conditional dependencies from Fig. 9 enforced. Importantly, this result is valid in a fully cosmological context, which means it also provides a rigorous route to obtaining low-redshift approximations without the need for any heuristic arguments about the relationship between distance, redshift and line-of-sight velocity.

### *b. Marginalization over nuisance parameters*

The main scientific aim here is inference of the cosmological parameters, and specifically  $H_0$ , which casts all the individual and population-level BNS quantities as nuisance parameters to be integrated out. For the simple model described in Sec. IV this is done by sampling the joint posterior distribution, as described in Sec. V, and so there is no need to explicitly calculate the marginalized posterior distribution. The alternative approach [45, 62, 63], which will likely be needed when analyzing real data, is to separately marginalize over the redshift parameters constrained by the EM data and then use samples from the individual merger posteriors to marginalize over those parameters. Equation A9 can be used as the basis for this approach, with the only potential difference coming from how the pre-factor of  $\exp[-\bar{N}(\beta, \Omega, O)]$  is handled. Fortunately, this term has only a weak dependence on  $H_0$  for low-redshift samples ([45, 62] and Sec. IV E), so such choices should not strongly affect the final marginal posterior distribution in  $\Omega$ .

## **Appendix B: Likelihood for the merger inspiral**

A BNS merger is controlled by non-linear physics in dynamical and strongly curved spacetime, and is described by a large number of parameters. Within the frequency band of the current GW detectors, however, the emission relevant for distance constraints arise primarily from the comparatively simple inspiral phase preceding the merger, which can be modeled using the post-Newtonian approximation to general relativity *e.g.*, [94]. For the type of system under consideration here several additional simplifications can be made. Most importantly, it is assumed that an EM counterpart has been identified, meaning that the sky position is known. As the known Galactic BNSs that would merge within a Hubble time have low dimensionless spin parameters of  $\lesssim 0.04$  [41], the spins of the merging NSs are ignored and set to zero. Adopting these simplifications, the GW signal from a merger event takes the form of two orthogonal strain waveforms that, to leading order in the post-Newtonian expansion parameter, can be written

as (*e.g.*, [95])

$$\begin{aligned}
 h_+(t) &= \frac{G\mathcal{M}/c^2}{D_P(z, \Omega)} \frac{1 + \cos^2(\iota)}{2} \phi_+ \left[ t; \frac{5G\mathcal{M}}{c^3} (1+z) \left( 1 + \frac{u}{c} + \frac{v}{c} \right), t_c, \Phi_c \right] \\
 h_\times(t) &= \frac{G\mathcal{M}/c^2}{D_P(z, \Omega)} [-\cos(\iota)] \phi_\times \left[ t; \frac{5G\mathcal{M}}{c^3} (1+z) \left( 1 + \frac{u}{c} + \frac{v}{c} \right), t_c, \Phi_c \right],
 \end{aligned} \tag{B1}$$

with the detector-frame time dependence encoded in the functions

$$\begin{aligned}
 \phi_+(t; \tau, t_c, \Phi_c) &= \left( \frac{t_c - t}{\tau} \right)^{-1/4} \cos \left\{ 2 \left[ \Phi_c - \left( \frac{t_c - t}{\tau} \right)^{5/8} \right] \right\} \\
 \phi_\times(t; \tau, t_c, \Phi_c) &= \left( \frac{t_c - t}{\tau} \right)^{-1/4} \sin \left\{ 2 \left[ \Phi_c - \left( \frac{t_c - t}{\tau} \right)^{5/8} \right] \right\},
 \end{aligned} \tag{B2}$$

where, with  $M_1$  and  $M_2$  as the masses of the two NSs,  $\mathcal{M} = (M_1 M_2)^{3/5} / (M_1 + M_2)^{1/5}$  is the chirp mass,  $t_c$  is the detector-frame time of coalescence,  $\Phi_c$  is the orbital phase at this time,  $\iota$  is the inclination of the system relative to the line-of-sight,  $z$  is the (cosmological) redshift of the system,  $v$  is the (line-of-sight) peculiar velocity of the host galaxy,  $u$  is the (line-of-sight) velocity of the merger relative to the host, and  $D_P(z, \Omega)$  the proper distance to redshift  $z$  given the cosmological parameters  $\Omega$ . The detector-frame coalescence time-scale,  $\tau$ , is equal to the source-frame value of  $5G\mathcal{M}/c^3$  dilated by a factor of  $(1+z)(1+u/c+v/c)$ .

The GW data from a single detector is a time-series of measurements given by a linear combination of the orthogonal polarization waveforms given in Eq. B1, where the weightings depend on the merger's sky position and the detector's orientation and geometry, to which detector noise is added. If a single merger is observed by multiple detectors then the joint dataset can be used to extract information about both polarizations (*e.g.*, [95]). Taking the full GW data on a merger to be  $\mathbf{d}$ , the likelihood has the form  $P(\mathbf{d}|\mathcal{M}, t_c, \Phi_c, \iota, z, u, v, \Omega, O)$ , where the merger and cosmological parameters are defined above, and  $O$  encodes all the relevant information about the observations, such as detector geometries and noise properties. The sky position of the merger is not included in the list of parameters here as they are taken to be determined precisely by the identification of a host galaxy from EM data. The general expression for the GW likelihood is complicated, but for a well-measured event, such as GW 170807, a comparatively simple form can be adopted. The oscillatory nature of the several thousand cycles of the inspiral signal allows precise constraints to be placed on the parameters,  $\tau$ ,  $t_c$  and  $\Phi_c$ , that determine the time-dependence of the waveform according to Eq. B2. The measured amplitude(s) then place more uncertain constraints on the pre-factors of the two strain waveforms given in Eq. B1. The parameter-dependence of the likelihood can hence be well captured by calculating a few (nearly) sufficient statistics (*e.g.*, Refs. [44, 71]): the precisely measured detector-frame coalescence time-scale,  $\hat{\tau}(\mathbf{d})$ ; the precisely measured detector-frame coalescence time,  $\hat{t}_c(\mathbf{d})$ ; the precisely measured coalescence phase,  $\hat{\Phi}_c(\mathbf{d})$ ; and estimates of the two orthogonal amplitudes,  $\hat{A}_+(\mathbf{d})$  and  $\hat{A}_\times(\mathbf{d})$ . The uncertainties on these amplitudes,  $\sigma_{A_+}(O)$  and  $\sigma_{A_\times}(O)$ , are determined by a combination of the sky position of the source (assumed to be known) and the detector's geometry and noise properties (encoded in  $O$ ). These two uncertainties cannot be ignored, as they are the dominant contributions to whether a merger is detected in the first place and, for those which pass the selection criteria, the distance uncertainty. Assuming no uncertainty on the temporal parameters, and that the noise on the amplitudes is Gaussian and uncorrelated, the GW likelihood can be written in the form (*cf.* Eq. B1)

$$\begin{aligned}
 P(\mathbf{d}|\mathcal{M}, t_c, \Phi_c, \iota, z, u, v, \Omega, O) &= \delta[\hat{t}_c(\mathbf{d}) - t_c] \delta[\hat{\Phi}_c(\mathbf{d}) - \Phi_c] \\
 &\times \delta \left[ \hat{\tau}(\mathbf{d}) - \frac{5G}{c^3} (1+z) \left( 1 + \frac{u}{c} + \frac{v}{c} \right) \mathcal{M} \right] \mathcal{N} \left[ \hat{A}_+(\mathbf{d}); \frac{G\mathcal{M}/c^2}{D_P(z, \Omega)} \frac{1 + \cos^2(\iota)}{2}, \sigma_{A_+}^2(O) \right] \mathcal{N} \left[ \hat{A}_\times(\mathbf{d}); -\frac{G\mathcal{M}/c^2}{D_P(z, \Omega)} \cos(\iota), \sigma_{A_\times}^2(O) \right].
 \end{aligned} \tag{B3}$$

Calculating these statistics from real data is a challenging numerical task, although a number of approaches have been demonstrated that can achieve this (*e.g.*, [44, 71]).

Fortunately, for the purposes of the bias analysis presented here, there is no need to actually simulate full time streams and go through this complicated procedure to calculate these statistics; it is sufficient instead to include  $\hat{\tau}$ ,  $\hat{t}_c$ ,  $\hat{\Phi}_c$ ,  $\hat{A}_+$  and  $\hat{A}_\times$  in the formalism as if these quantities were measured directly. It is hence possible to replace  $\mathbf{d}$  with the above statistics and similarly give the noise amplitudes in place of  $O$ , meaning that the likelihood is given entirely in terms of explicitly defined quantities. Also defining the (accurately) measured redshifted chirp mass as

$\hat{\mathcal{M}}_z = \hat{\tau}/(5 G/c^2)$ , Eq. B3 can be rewritten as

$$\begin{aligned} & P(\hat{t}_c, \hat{\Phi}_c, \hat{\mathcal{M}}_z, \hat{A}_+, \hat{A}_\times | \mathcal{M}, t_c, \Phi_c, \iota, z, u, v, \Omega, \sigma_{A_+}, \sigma_{A_\times}) \\ &= \delta(\hat{t}_c - t_c) \delta(\hat{\Phi}_c - \Phi_c) \delta\left[\hat{\mathcal{M}}_z - (1+z)\left(1 + \frac{u}{c} + \frac{v}{c}\right)\mathcal{M}\right] \text{N}\left[\hat{A}_+; \frac{G\mathcal{M}/c^2}{D_P(z, \Omega)} \frac{1 + \cos^2(\iota)}{2}, \sigma_{A_+}^2\right] \text{N}\left[\hat{A}_\times; -\frac{G\mathcal{M}/c^2}{D_P(z, \Omega)} \cos(\iota), \sigma_{A_\times}^2\right]. \end{aligned} \quad (\text{B4})$$

This could be used either as the sampling distribution for a merger with specified properties to generate the effective data (*i.e.*,  $\hat{t}_c$ ,  $\hat{\Phi}_c$ ,  $\hat{\mathcal{M}}_z$ ,  $\hat{A}_+$  and  $\hat{A}_\times$ ) or as the basis for simulating parameter constraints, in either case without any reference to a full timestream.

In the context of cosmological parameter estimation several further simplifications can usefully be made:

- Given that both the galactic motions relative to the Hubble flow and the orbital speeds within galaxies are typically a few hundred km/s, the peculiar velocity terms  $u/c$  and  $v/c$  produce negligible relative offsets of  $\lesssim 0.1\%$ . As such, in the analysis of the GW data it is reasonable to ignore the peculiar velocities completely (*i.e.*, setting  $u = v = 0$ ). (The peculiar velocity of the host galaxy cannot be ignored as its relative contribution to the spectroscopic redshift is given by the ratio  $|v|/(cz)$ , which can be  $\gtrsim 10\%$  at  $z \lesssim 0.01$ . It is, however, the host galaxy's peculiar velocity that is relevant here, as it is assumed that no redshift is obtained for the merger/burst itself. Hence it is only  $v$  which affects the measured redshift;  $u$  can be ignored in the analysis of the EM data as well as the GW data.)
- The coalescence time,  $t_c$ , and coalescence phase,  $\Phi_c$ , are both nuisance parameters that do not provide any useful information about the properties of the merger. They would hence be marginalized over in any cosmological or population analysis, potentially increasing the uncertainty on other parameters. However, under the approximation that they are constrained perfectly by the time evolution of the GW signal,  $t_c$  and  $\Phi_c$  can simply be omitted from the likelihood completely.

Applying these simplifications leaves the the significant part of Eq. B4 as

$$\begin{aligned} & P(\hat{\mathcal{M}}_z, \hat{A}_+, \hat{A}_\times | \mathcal{M}, \iota, z, \Omega, \sigma_{A_\times}, \sigma_{A_+}) \\ & \simeq \delta\left[\hat{\mathcal{M}}_z - (1+z)\mathcal{M}\right] \text{N}\left[\hat{A}_+; \frac{G(1+z)\mathcal{M}/c^2}{D(z, \Omega)} \frac{1 + \cos^2(\iota)}{2}, \sigma_{A_+}^2\right] \text{N}\left[\hat{A}_\times; -\frac{G(1+z)\mathcal{M}/c^2}{D(z, \Omega)} \cos(\iota), \sigma_{A_\times}^2\right], \end{aligned} \quad (\text{B5})$$

where  $D(z, \Omega) = (1+z) D_P(z, \Omega)$  is the luminosity distance to redshift  $z$ . This change is made so that it is the tightly constrained redshifted chirp mass  $(1+z)\mathcal{M}$  that appears in all three terms; in an inference context the substitution  $(1+z)\mathcal{M} \rightarrow \hat{\mathcal{M}}_z$  could be made, leaving only  $D(z, \Omega)$  and  $\iota$  to be constrained by the measured amplitudes  $\hat{A}_+$  and  $\hat{A}_\times$ . Equation B5 is the form of the likelihood used in Sec. IV C and hence is the basis for all the simulations in this paper.