

# Real-time inference for binary neutron star mergers using machine learning

<https://doi.org/10.1038/s41586-025-08593-z>

Received: 2 August 2024

Accepted: 3 January 2025

Published online: 5 March 2025

Open access

 Check for updates

Maximilian Dax<sup>1,2,3,✉</sup>, Stephen R. Green<sup>4</sup>, Jonathan Gair<sup>5</sup>, Nihar Gupte<sup>5,6</sup>, Michael Pürrer<sup>7,8</sup>, Vivien Raymond<sup>9</sup>, Jonas Wildberger<sup>3</sup>, Jakob H. Macke<sup>1,10</sup>, Alessandra Buonanno<sup>5,6</sup> & Bernhard Schölkopf<sup>1,2,3</sup>

Mergers of binary neutron stars emit signals in both the gravitational-wave (GW) and electromagnetic spectra. Famously, the 2017 multi-messenger observation of GW170817 (refs. 1,2) led to scientific discoveries across cosmology<sup>3</sup>, nuclear physics<sup>4–6</sup> and gravity<sup>7</sup>. Central to these results were the sky localization and distance obtained from the GW data, which, in the case of GW170817, helped to identify the associated electromagnetic transient, AT 2017gfo (ref. 8), 11 h after the GW signal. Fast analysis of GW data is critical for directing time-sensitive electromagnetic observations. However, owing to challenges arising from the length and complexity of signals, it is often necessary to make approximations that sacrifice accuracy. Here we present a machine-learning framework that performs complete binary neutron star inference in just 1 s without making any such approximations. Our approach enhances multi-messenger observations by providing: (1) accurate localization even before the merger; (2) improved localization precision by around 30% compared to approximate low-latency methods; and (3) detailed information on luminosity distance, inclination and masses, which can be used to prioritize expensive telescope time. Additionally, the flexibility and reduced cost of our method open new opportunities for equation-of-state studies. Finally, we demonstrate that our method scales to long signals, up to an hour in length, thus serving as a blueprint for data analysis for next-generation ground- and space-based detectors.

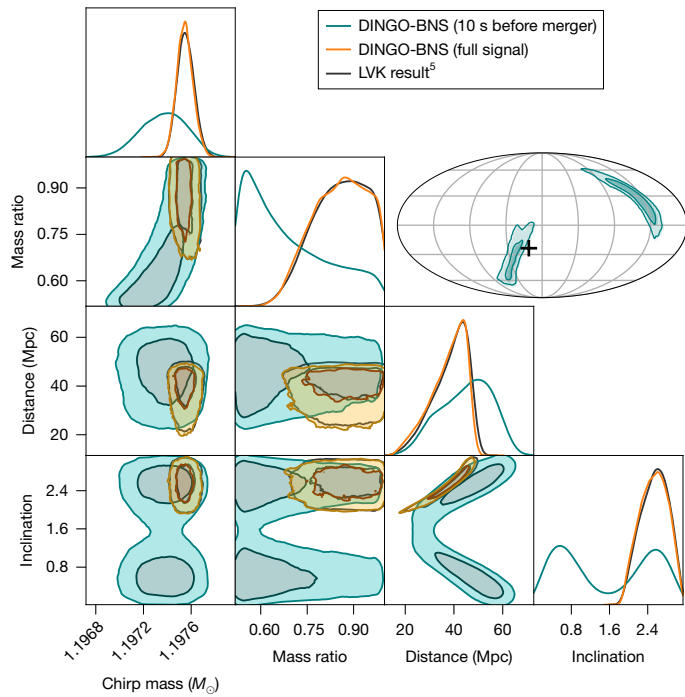
The fast and accurate inference of binary neutron stars (BNSs) from gravitational-wave (GW) data is a critical challenge facing multi-messenger astronomy. For a BNS, the GW signal is visible by the Laser Interferometer GW Observatory (LIGO)–Virgo GW Interferometer (Virgo)–Kamioka GW Detector (KAGRA) (collectively, LVK)<sup>9–11</sup> observatories minutes before any electromagnetic counterpart. The GW encodes information on the source characterization, distance, sky location and orientation necessary for pointing and prioritizing optical telescopes. However, the length of BNS signals makes conventional Bayesian inference techniques<sup>12,13</sup> too slow to be useful in low-latency applications. Instead, once a GW signal is identified by detection pipelines<sup>14,15</sup>, approximate algorithms are used for providing initial alerts (for example, BAYESTAR<sup>16</sup>, which uses the signal-to-noise ratio (SNR) time series rather than the complete strain data and gives localization in seconds). Other methods focus on accelerating likelihood evaluations without incurring loss of precision (for example, using reduced-order quadratures), with the state-of-the-art delivering localization in 6 min and full inference in 2 h (ref. 17).

Simulation-based machine learning offers a powerful alternative for GW inference (see Methods for related work). With simulation-based inference (SBI)<sup>18</sup>, neural networks are trained to encode probabilistic estimates of astrophysical source parameters conditional on data.

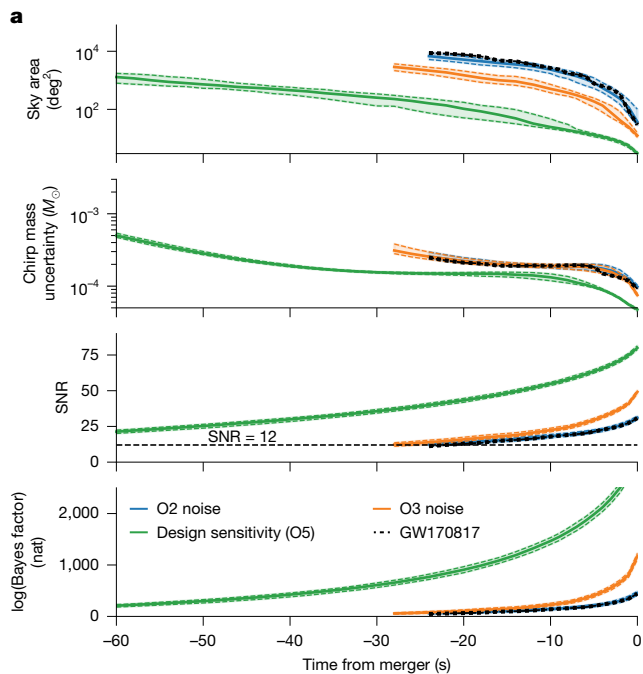
Trained networks then enable extremely fast analysis for new datasets, amortizing upfront training costs across observations. In past work, we developed the deep inference for GW observations (DINGO) framework for binary black holes (BBHs)<sup>19,20</sup>, which performs accurate inference in seconds, including strong accuracy guarantees when coupled with importance sampling. However, when applied to BNSs, machine-learning approaches, such as DINGO, are beset by the same challenges facing traditional methods because of long signal durations. Indeed, DINGO becomes unreliable even for low-mass BBHs (chirp masses  $\lesssim 15 M_{\odot}$ ) with signals longer than roughly 16 s. A BNS lasts for hundreds of seconds for the LVK and will reach hours for next-generation detectors (for example, Cosmic Explorer<sup>21</sup> and Einstein Telescope<sup>22</sup>). From the neural-network perspective, this corresponds to time or frequency series input with up to tens of millions of dimensions—a thousand-fold increase over BBH.

In this study, we overcome these challenges by leveraging perturbative BNS physics information to simplify and compress the data. However, this simplification requires approximate knowledge of the source itself and is hence valid only over a small portion of the parameter space. We solve this problem using a new algorithm, called prior conditioning, which enables us to construct networks that can be adapted at inference time to subsets of the prior volume. Our new framework, called

<sup>1</sup>Max Planck Institute for Intelligent Systems, Tübingen, Germany. <sup>2</sup>ETH Zurich, Zurich, Switzerland. <sup>3</sup>ELLIS Institute Tübingen, Tübingen, Germany. <sup>4</sup>School of Mathematical Sciences, University of Nottingham, Nottingham, UK. <sup>5</sup>Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Potsdam, Germany. <sup>6</sup>Department of Physics, University of Maryland, College Park, MD, USA. <sup>7</sup>Department of Physics, University of Rhode Island, Kingston, RI, USA. <sup>8</sup>Center for Computational Research, University of Rhode Island, Kingston, RI, USA. <sup>9</sup>Gravity Exploration Institute, Cardiff University, Cardiff, UK. <sup>10</sup>Machine Learning in Science, University of Tübingen & Tübingen AI Center, Tübingen, Germany. ✉e-mail: maximilian.dax@tuebingen.mpg.de



**Fig. 1 | Real-time inference of GW170817.** DINGO-BNS estimates all BNS parameters in just 1 s (orange, 10.8% sample efficiency), reproducing LVK results<sup>5</sup> (black, 0.1% typical efficiency) three orders of magnitude faster than existing methods<sup>17,33,39</sup>. DINGO-BNS can also analyse partial data before the merger occurs (teal, 78.9% efficiency). Fast analysis results are crucial for directing electromagnetic searches for prompt, or even precursor, signals. Note that GW170817 overlapped with a loud glitch, which could explain why the true sky position lies in the tail of the pre-merger distribution.



**Fig. 2 | Pre-merger inference with DINGO-BNS.** **a**, Evolution of pre-merger estimates for GW170817 (black) and GW170817-like simulations injected into different noise levels (colours). We display the 90% credible sky area, the standard deviation of the chirp mass, the accumulated signal-to-noise and the log(Bayes factor) in the natural unit of information (nat) comparing the signal and noise models. All of these quantities are inferred with a latency of around 1 s. Dotted lines represent the 10th/90th percentiles. We impose a

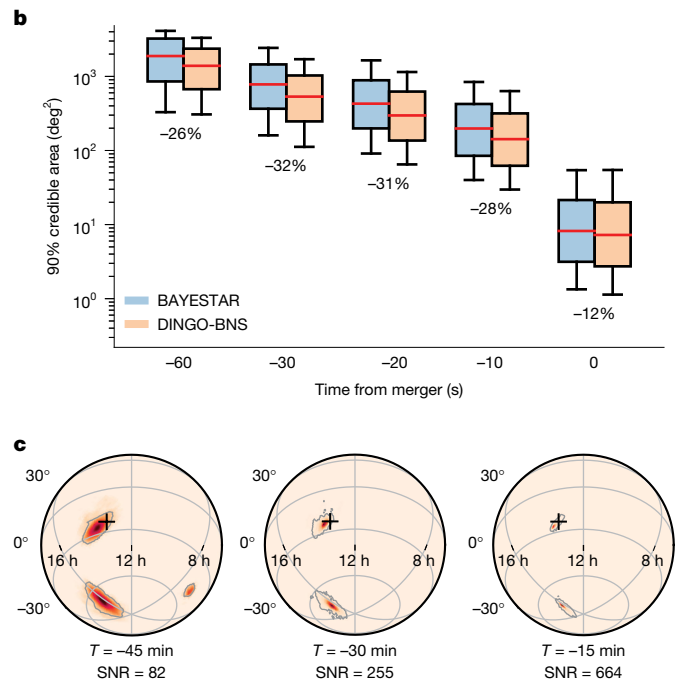
DINGO-BNS, makes no (practically relevant) approximation and takes just 1 s for accurate inference of all 17 BNS parameters (Fig. 1). Using DINGO-BNS, we can also infer all of these parameters minutes before the merger based on partial inspiral-only information—estimates that can be continuously updated as more data become available (Fig. 2a). Near-real-time or pre-merger alerts can then be provided to astronomers, facilitating potential discoveries of precursor and prompt electromagnetic counterparts<sup>23–25</sup>.

Our results are faster and more complete than any existing low-latency algorithm, with the accuracy of offline parameter estimation codes. Compared to BAYESTAR, we achieve median reductions in the size of the 90% credible sky region of 30% (Fig. 2b). Finally, DINGO-BNS exhibits excellent scaling to longer signals (Methods), and we demonstrate next-generation detector pre-merger inference for signals up to an hour in length (Fig. 2c).

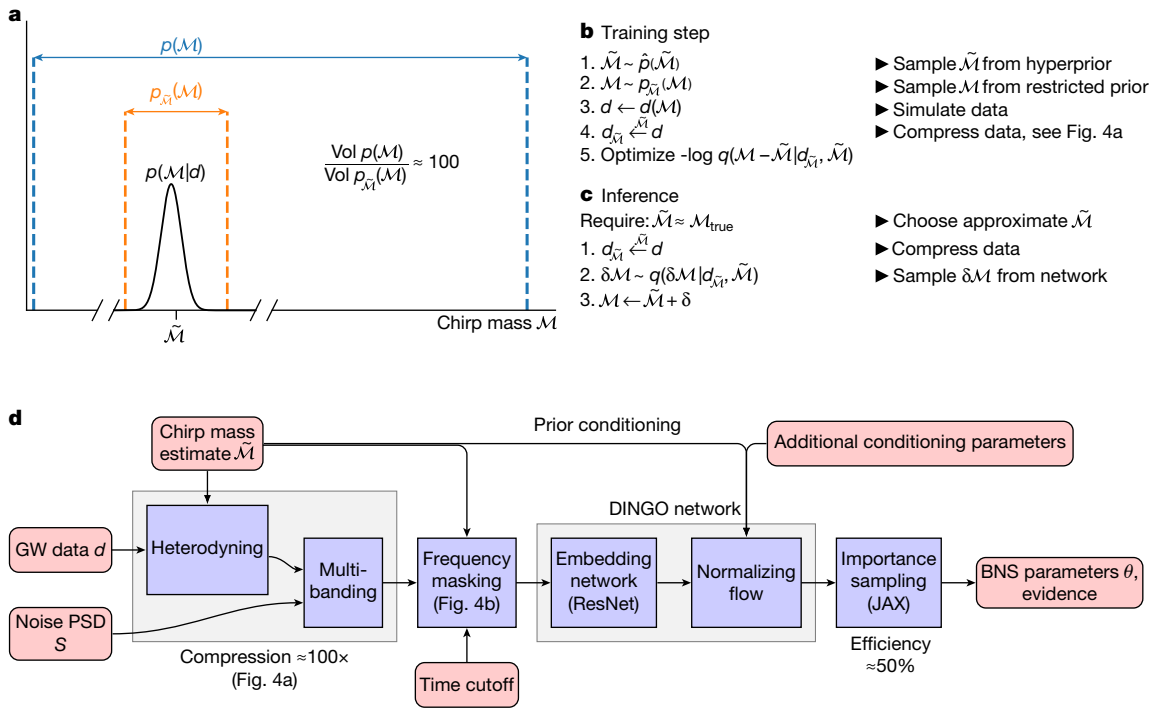
### DINGO-BNS

For given GW data,  $\mathbf{d}$ , we characterize the source in terms of the posterior probability distribution,  $p(\boldsymbol{\theta}|\mathbf{d})$ , over BNS parameters,  $\boldsymbol{\theta}$ . Parameters include component masses (two), spins (six), orientation, sky position (two), luminosity distance, polarization, time and phase of coalescence and (in contrast to black holes) tidal deformabilities (two). Following our past work<sup>19</sup>, we use simulated GW datasets to train a density estimation neural network,  $q(\boldsymbol{\theta}|\mathbf{d})$  (a normalizing flow), to approximate  $p(\boldsymbol{\theta}|\mathbf{d})$ . Once trained, the inference for new  $\mathbf{d}$  simply requires sampling  $\boldsymbol{\theta} \sim q(\boldsymbol{\theta}|\mathbf{d})$ . We obtain asymptotically exact results by augmenting samples with importance weights using the GW likelihood function<sup>26</sup>. This framework, called DINGO-IS<sup>20</sup>, has been successfully applied to black hole mergers. However, the length of BNS signals renders the naïve transfer of machine-learning methods impossible.

To tackle this challenge, DINGO-BNS makes several innovations (Fig. 3d), including using knowledge of specific BNS signal morphology



minimum SNR<sup>17</sup> of 12. **b**, Sky localization area at 90% credible level for various pre-merger times compared against BAYESTAR. The boxplots display the medians (percentage changes indicated), quartiles and 10th/90th percentiles. DINGO-BNS localization is consistently more precise. **c**, Pre-merger sky localization for a GW170817-like event injected into Cosmic Explorer noise, using a minimum frequency of 6 Hz. The black marker indicates the injection coordinates and the grey outline is the 90% credible area.



**Fig. 3 | Prior conditioning and method flowchart.** **a**, For a typical event, the chirp mass posterior (black) is tightly constrained compared to the prior (blue), so a tighter prior (orange) that still covers the posterior is sufficient for inference. In addition, the narrow prior can simplify the analysis. Our prior-conditioning technique therefore trains a single neural network that can be tuned at inference time to an event-specific prior. **b**, Training is accomplished by simulating data from randomly chosen narrow priors,  $p_{\tilde{\mathcal{M}}}(\mathcal{M})$ , each parametrized by a reference chirp mass,  $\tilde{\mathcal{M}}$ , on which the network is also

conditioned. Prior conditioning enables prior-specific heterodyning based on  $\tilde{\mathcal{M}}$ , followed by multibanding compression (Fig. 4a), effectively simplifying the data distribution that the model must learn and reducing its dimensionality. **c**, At inference time, an initial chirp mass estimate,  $\tilde{\mathcal{M}}$ , determines the event-specific prior and compression. **d**, Prior conditioning and the other technical innovations are integrated into a single neural network that can be trained end-to-end and produce  $10^5$  weighted samples per second, with typical sampling efficiencies of 50%.

to compress data in a non-lossy way, conditioning the network on the compressor using prior conditioning, frequency masking based on the pre-merger time and chirp mass, and conditioning on parameter subsets for incorporating multi-messenger information or expectations from nuclear models. The philosophy underlying our approach is that the full BNS problem is too hard for existing neural architectures, so we divide the parameter and data spaces into manageable portions based on known physical information. We then combine all of these variable design choices into a single network. By passing relevant control parameters at inference time, the network can be tuned to the context at hand.

### Data compression and prior conditioning

We adapted two GW analysis techniques to the SBI context—heterodyning<sup>27</sup> to simplify the data and multibanding<sup>28,29</sup> to reduce the data dimension without loss of information. During the long inspiral period, a BNS signal exhibits a ‘chirp’, with phase evolution (to leading order in the post-Newtonian expansion<sup>30</sup>),

$$\varphi(f; \mathcal{M}) = \frac{3}{128} \left( \frac{\pi G \mathcal{M} f}{c^3} \right)^{-5/3}, \quad (1)$$

where  $f$  is the frequency,  $c$  is the speed of light in vacuum and  $\mathcal{M} = (m_1 m_2)^{3/5} / (m_1 + m_2)^{1/5}$  is the chirp mass of the system, with  $m_1$  and  $m_2$  being the component masses. Given an approximation,  $\tilde{\mathcal{M}}$ , to the chirp mass, we heterodyne the (frequency-domain) data by multiplying by  $e^{i\varphi(f; \tilde{\mathcal{M}})}$ , reducing the number of oscillations in the signal by several orders of magnitude (Fig. 4a). Given the heterodyned data, we apply multibanding by partitioning the domain into (empirically determined)

frequency bands and coarsening the resolution in higher bands, such that the (heterodyned) signal is preserved.

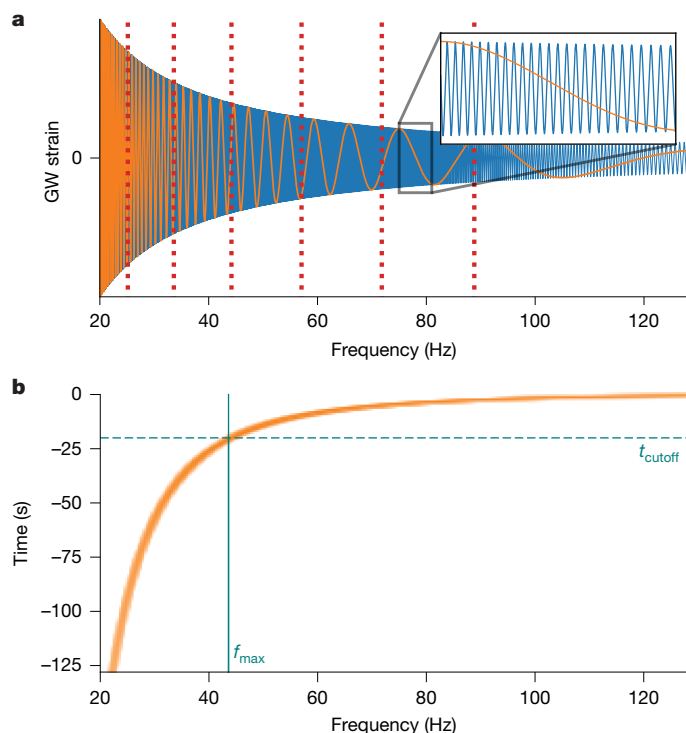
Because the compression described requires  $\tilde{\mathcal{M}}$  to approximate the chirp mass, it cannot be done across the entire BNS prior volume using a single  $\tilde{\mathcal{M}}$  value. Therefore, DINGO-BNS uses prior conditioning to restrict to an event-specific prior over which data are compressed. The restricted volume additionally simplifies the density estimation task. By conditioning on the choice of restriction, prior conditioning trains a network that is tunable to this choice, but otherwise applicable over the whole volume (Fig. 3a). Inference requires an estimate,  $\tilde{\mathcal{M}}$ , of the chirp mass,  $\mathcal{M}$ , which can be determined quickly by sweeping across the prior (Methods).

### Frequency masking

In contrast to past work, DINGO-BNS also allows strain frequency series with varying minimum and maximum frequencies ( $f_{\min}$  and  $f_{\max}$ , respectively). For a given analysis,  $f_{\min}$  is chosen based on  $\tilde{\mathcal{M}}$  and the segment duration as the minimum frequency present in the signal in a given GW-detector network. This masking is necessary for consistency with frequency-domain waveform models, which assume infinite duration. Choosing  $f_{\max}$ , by contrast, determines the end time of the data stream analysed to enable pre-merger inference (Fig. 4b and Methods).

### Conditioning on parameter subsets

The DINGO-BNS framework (and SBI in general) allows considerable flexibility in terms of quickly marginalizing over, and conditioning on, parameters. Conditioning on a parameter allows us to set it to a fixed



**Fig. 4 | Compression and frequency masking.** **a**, We compress data by a factor of around 100 by first factoring out (that is, heterodyning) the predominant phase evolution of the signal (blue). The resulting simplified signal (orange) is down-sampled in resolution, reducing data dimensionality (coarser resolution at high frequencies; bands indicated by dotted red lines). **b**, To enable pre-merger inference, we mask out the strain frequency series according to the cutoff time.

value—for example, to incorporate knowledge of that parameter from other sources. In our study, we trained DINGO-BNS networks conditioned on the sky position—that is, we learned  $p(\boldsymbol{\theta}|\{\alpha, \delta\}|\mathbf{d}, \alpha, \delta)$ , where  $\alpha$  and  $\delta$  denote the right ascension and declination, respectively. Such a network allows us to incorporate precise multi-messenger localization to obtain tighter constraints on the remaining parameters, potentially enabling real-time feedback on whether optical candidates should be prioritized for detailed spectroscopy<sup>31</sup>. In this way, DINGO-BNS can enable new modes of interaction between GW and electromagnetic observers, potentially transforming how we prioritize and respond to multi-messenger events. We have also explored parameter-conditioning to accelerate offline nuclear equation-of-state (EOS) analyses (Methods).

## Experiments

We generated training data using simulated BNS waveforms (including spin-precession and tidal contributions, but without higher angular multipoles<sup>32</sup>) with additive stationary Gaussian detector noise. When relevant, networks are also trained with power spectral density (PSD)-conditioning to enable instant tuning to noise levels at the time of an event. At inference time, we validate and correct results using importance sampling, thus guaranteeing their accuracy, provided a sufficient effective sample size is obtained<sup>20</sup>. We accelerate the importance sampling step using JAX waveform and likelihood implementations<sup>33–35</sup>.

We performed four studies using DINGO-BNS: (1) a pre-merger analysis of the first BNS detected—GW170817—as well as equivalent injections (simulated datasets) at varying noise levels; (2) a pre-merger analysis of a range of injections in LVK design sensitivity noise; (3) an after-merger analysis of the two detected BNS events—GW170817 and

GW190425—reproducing published LVK results; and (4) a pre-merger analysis of injections in Cosmic Explorer noise (with a minimum frequency of 6 Hz, corresponding to an hour-long signal). We use the importance sampling efficiency as a primary performance metric, finding average values of 63.3%, 47.0%, 31.0% and 35.6% in experiments (1), (2), (3) and (4), respectively. With these high efficiencies, inference for  $10^4$  effective samples takes roughly 1 s on an H100 GPU (Methods). Efficiencies are generally higher for pre-merger, probably because the waveform morphology is simplest in the early inspiral.

## Discussion

Prior conditioning works well for BNS inference, and it could be extended to address further challenges in GW astronomy (for example, the isolation of events from overlapping background signals in next-generation detectors) and other scientific domains. In the future, we would like to explore our prior-conditioning approach to data compression for black hole–neutron star systems and low-mass BBHs. This is non-trivial because such systems can emit GWs in higher angular radiation multipoles (that is, beyond the  $(l, m) = (2, 2)$  mode that we assume here), which evolve according to integer multiples of (1), and so would require an improved heterodyning algorithm to factor out the chirp. Higher modes are not present in BNS signals, because the stars are very nearly equal in mass.

Another exciting prospect for SBI is a more realistic treatment of detector noise. Indeed, because BNS inspirals have long durations, noise non-stationarities and non-Gaussianities are more likely to manifest. Currently, DINGO-BNS assumes stationary Gaussian noise and is supplied with an estimate of the PSD (possibly resulting in additional latency for data preparation, see Methods). However, by training on realistic detector noise, our approach can, in principle, learn to fully characterize the noise jointly with the signal, including any deviations from stationarity and Gaussianity. This approach is akin to on-source PSD and glitch modelling<sup>36</sup>, but allows more general noise and automatically marginalizes over uncertainties. Initial steps in this direction have already been taken for intermediate-mass BBHs<sup>37</sup>. Improved noise treatments, such as those afforded by SBI, will become crucial for reducing systematic error as detectors become more sensitive<sup>38</sup>.

Finally, although DINGO-BNS is intended to be used for parameter estimation following a trigger by dedicated search pipelines, its speed opens up the possibility of being run continuously on all data as they are taken. Either the SNR or Bayesian evidence time series generated by DINGO-BNS could then be used as a detection statistic, forming an end-to-end detection and parameter estimation pipeline. To implement this would require calibrating these statistics to determine false alarm rates, as well as careful comparisons against existing algorithms to establish efficacy.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-025-08593-z>.

- Abbott, B. P. et al. GW170817: observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.* **119**, 161101 (2017).
- Abbott, B. P. et al. Multi-messenger observations of a binary neutron star merger. *Astrophys. J. Lett.* **848**, L12 (2017).
- Abbott, B. P. et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature* **551**, 85–88 (2017).
- Abbott, B. P. et al. Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A. *Astrophys. J. Lett.* **848**, L13 (2017).
- Abbott, B. P. et al. Properties of the binary neutron star merger GW170817. *Phys. Rev. X* **9**, 011001 (2019).
- Abbott, B. P. et al. GW170817: measurements of neutron star radii and equation of state. *Phys. Rev. Lett.* **121**, 161101 (2018).

7. Abbott, B. P. et al. Tests of general relativity with GW170817. *Phys. Rev. Lett.* **123**, 011102 (2019).
8. Coulter, D. A. et al. Swope supernova survey 2017a (SSS17a), the optical counterpart to a gravitational wave source. *Science* **358**, 1556 (2017).
9. Aasi, J. et al. Advanced LIGO. *Class. Quant. Grav.* **32**, 074001 (2015).
10. Acernese, F. et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.* **32**, 024001 (2015).
11. Aso, Y. et al. Interferometer design of the KAGRA gravitational wave detector. *Phys. Rev. D* **88**, 043007 (2013).
12. Veitch, J. et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D* **91**, 042003 (2015).
13. Ashton, G. et al. BILBY: a user-friendly Bayesian inference library for gravitational-wave astronomy. *Astrophys. J. Suppl.* **241**, 27 (2019).
14. Nitz, A. H., Dal Canton, T., Davis, D. & Reyes, S. Rapid detection of gravitational waves from compact binary mergers with PyCBC Live. *Phys. Rev. D* **98**, 024050 (2018).
15. Cannon, K. et al. GstLAL: a software framework for gravitational wave discovery. *SoftwareX* **14**, 100680 (2021).
16. Singer, L. P. & Price, L. R. Rapid Bayesian position reconstruction for gravitational-wave transients. *Phys. Rev. D* **93**, 024013 (2016).
17. Morisaki, S. et al. Rapid localization and inference on compact binary coalescences with the advanced LIGO-Virgo-KAGRA gravitational-wave detector network. *Phys. Rev. D* **108**, 123040 (2023).
18. Cranmer, K., Brehmer, J. & Louppe, G. The frontier of simulation-based inference. *Proc. Natl Acad. Sci. USA* **117**, 30055–30062 (2020).
19. Dax, M. et al. Real-time gravitational wave science with neural posterior estimation. *Phys. Rev. Lett.* **127**, 241103 (2021).
20. Dax, M. et al. Neural importance sampling for rapid and reliable gravitational-wave inference. *Phys. Rev. Lett.* **130**, 171403 (2023).
21. Reitze, D. et al. Cosmic Explorer: The U.S. contribution to gravitational-wave astronomy beyond LIGO. *Bull. Am. Astron. Soc.* **51**, 035 (2019).
22. Punturo, M. et al. The third generation of gravitational wave observatories and their science reach. *Class. Quant. Grav.* **27**, 084007 (2010).
23. Chaudhary, S. S. et al. Low-latency gravitational wave alert products and their performance at the time of the fourth LIGO-Virgo-KAGRA observing run. *Proc. Natl Acad. Sci. USA* **121**, e2316474121 (2024).
24. Sridhar, N., Zrake, J., Metzger, B. D., Sironi, L. & Giannios, D. Shock-powered radio precursors of neutron star mergers from accelerating relativistic binary winds. *Mon. Not. Roy. Astron. Soc.* **501**, 3184–3202 (2021).
25. Most, E. R. & Philippov, A. A. Electromagnetic precursor flares from the late inspiral of neutron star binaries. *Mon. Not. Roy. Astron. Soc.* **515**, 2710–2724 (2022).
26. Abbott, B. P. et al. A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals. *Class. Quant. Grav.* **37**, 055002 (2020).
27. Cornish, N. J. Fast Fisher matrices and lazy likelihoods. Preprint at <https://arxiv.org/abs/1007.4820> (2010).
28. Vinciguerra, S., Veitch, J. & Mandel, I. Accelerating gravitational wave parameter estimation with multi-band template interpolation. *Class. Quant. Grav.* **34**, 115006 (2017).
29. Morisaki, S. Accelerating parameter estimation of gravitational waves from compact binary coalescence using adaptive frequency resolutions. *Phys. Rev. D* **104**, 044062 (2021).
30. Blanchet, L. Gravitational radiation from post-Newtonian sources and inspiralling compact binaries. *Living Rev. Rel.* **17**, 2 (2014).
31. Abbott, B. P. et al. Low-latency gravitational-wave alerts for multimessenger astronomy during the second advanced LIGO and Virgo observing run. *Astrophys. J.* **875**, 161 (2019).
32. Dietrich, T. et al. Matter imprints in waveform models for neutron star binaries: Tidal and self-spin effects. *Phys. Rev. D* **99**, 024029 (2019).
33. Wong, K. W. K., Isi, M. & Edwards, T. D. P. Fast gravitational-wave parameter estimation without compromises. *Astrophys. J.* **958**, 129 (2023).
34. Edwards, T. D. P. et al. Differentiable and hardware-accelerated waveforms for gravitational wave data analysis. *Phys. Rev. D* **110**, 064028 (2024).
35. Wouters, T., Pang, P. T. H., Dietrich, T. & Van Den Broeck, C. Robust parameter estimation within minutes on gravitational wave signals from binary neutron star inspirals. *Phys. Rev. D* **110**, 083033 (2024).
36. Cornish, N. J. & Littenberg, T. B. BayesWave: Bayesian inference for gravitational wave bursts and instrument glitches. *Class. Quant. Grav.* **32**, 135012 (2015).
37. Raymond, V., Al-Shammari, S. & Göttel, A. Simulation-based inference for gravitational-waves from intermediate-mass binary black holes in real noise. Preprint at <https://arxiv.org/abs/2406.03935> (2024).
38. Davis, D. et al. LIGO detector characterization in the second and third observing runs. *Class. Quant. Grav.* **38**, 135014 (2021).
39. Williams, M. J., Veitch, J. & Messenger, C. Importance nested sampling with normalising flows. *Mach. Learn. Sci. Tech.* **4**, 035011 (2023).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

### Machine-learning framework

The Bayesian posterior,  $p(\boldsymbol{\theta}|\mathbf{d}) = p(\mathbf{d}|\boldsymbol{\theta})p(\boldsymbol{\theta})/p(\mathbf{d})$ , is defined in terms of a prior  $p(\boldsymbol{\theta})$  and a likelihood  $p(\mathbf{d}|\boldsymbol{\theta})$ . For GW inference, the likelihood is constructed by combining models for waveforms and detector noise. The Bayesian evidence,  $p(\mathbf{d})$ , corresponds to the normalization of the posterior and can be used for model comparison.

Our framework is based on neural posterior estimation (NPE)<sup>40–42</sup>, which trains a density estimation neural network,  $q(\boldsymbol{\theta}|\mathbf{d})$ , to estimate  $p(\boldsymbol{\theta}|\mathbf{d})$ . We parameterize  $q(\boldsymbol{\theta}|\mathbf{d})$  with a conditional normalizing flow<sup>43,44</sup>. Training minimizes the loss  $L = \mathbb{E}_{p(\boldsymbol{\theta}, \mathbf{d})}[-\log q(\boldsymbol{\theta}|\mathbf{d})]$ , where the expectation value is computed across a dataset  $(\boldsymbol{\theta}_i, \mathbf{d}_i)$  of parameters,  $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta})$ , paired with corresponding likelihood simulations,  $\mathbf{d}_i \sim p(\mathbf{d}|\boldsymbol{\theta}_i)$ . After training,  $q(\boldsymbol{\theta}|\mathbf{d})$  serves as a surrogate for  $p(\boldsymbol{\theta}|\mathbf{d})$  and inference for any observed data,  $\mathbf{d}_o$ , can be performed by sampling  $\boldsymbol{\theta} \sim q(\boldsymbol{\theta}|\mathbf{d}_o)$ . With DINGO<sup>19,20</sup> using a group-equivariant formulation of NPE (GNPE<sup>19,45</sup>), the GW data are simplified by aligning coalescence times in the different detectors. However, this comes at the cost of longer inference times, so we do not use GNPE for DINGO-BNS.

At inference, we correct for potential inaccuracies of  $q(\boldsymbol{\theta}|\mathbf{d})$  using importance sampling<sup>20</sup>, by assigning weight,  $w_i = p(\mathbf{d}|\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)/q(\boldsymbol{\theta}_i|\mathbf{d})$ , to each sample,  $\boldsymbol{\theta}_i \sim q(\boldsymbol{\theta}_i|\mathbf{d})$ . A set of  $n$  weighted samples  $(w_i, \boldsymbol{\theta}_i)$  corresponds to  $n_{\text{eff}} = (\sum_i w_i)^2 / (\sum_i w_i^2)$  effective samples from the posterior,  $p(\boldsymbol{\theta}|\mathbf{d})$ . This reweighting enables asymptotically exact results and the sample efficiency,  $\epsilon = n_{\text{eff}}/n$ , serves as a performance metric. The normalization of the weights further provides an unbiased estimate of the Bayesian evidence,  $p(\mathbf{d}) = (\sum_i w_i)/n$ .

Below, we describe in more detail the technical innovations of DINGO-BNS that enable scaling of this framework to BNS signals.

**Prior conditioning.** An NPE model,  $q(\boldsymbol{\theta}|\mathbf{d})$ , estimates the posterior,  $p(\boldsymbol{\theta}|\mathbf{d})$ , for a fixed prior,  $p(\boldsymbol{\theta})$ . Choosing a broad prior enhances the general applicability of the NPE model, but it also implies worse tuning to specific events (for which smaller priors may be sufficient). This is a general trade-off in NPE, but it is particularly notable for BNS inference, where typical events constrain the chirp mass to around  $10^{-3}$  of the prior volume. Thus, for an individual BNS event, a tight chirp mass prior would have been sufficient (Extended Data Fig. 1b) and moreover would have enabled effective heterodyning<sup>27,46,47</sup>. However, to cover generic BNS events, we need to train the NPE network with a large prior (Extended Data Table 2).

We resolve this trade-off with a new technique called prior conditioning. The key idea is to train an NPE model with multiple different (restricted) priors simultaneously. Training a prior-conditioned model requires hierarchical sampling:

$$\boldsymbol{\theta} \sim p_{\boldsymbol{\theta}}(\boldsymbol{\theta}), \boldsymbol{\rho} \sim \hat{p}(\boldsymbol{\rho}), \quad (2)$$

where  $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$  is a prior family parameterized by  $\boldsymbol{\rho}$  and  $\hat{p}(\boldsymbol{\rho})$  is a corresponding hyperprior. We additionally condition the NPE model,  $q(\boldsymbol{\theta}|\mathbf{d}, \boldsymbol{\rho})$ , on  $\boldsymbol{\rho}$ . This model can then perform inference for any desired prior,  $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ , by simply providing the corresponding  $\boldsymbol{\rho}$ . This effectively amortizes the training cost over different choices of the prior. On each of the restricted priors, we are furthermore allowed to transform the data in a  $\boldsymbol{\rho}$ -dependent way. This is because, having been conditioned on the prior choice, the network has all the information necessary to properly interpret the transformed data. We use this freedom to heterodyne the GW strain with respect to the approximate chirp mass.

To apply prior conditioning for the chirp mass,  $\mathcal{M}$ , we use a set of priors,  $p_{\widetilde{\mathcal{M}}}(\mathcal{M}) = U_{m_1, m_2}(\widetilde{\mathcal{M}} - \Delta\mathcal{M}, \widetilde{\mathcal{M}} + \Delta\mathcal{M})$ . Here,  $U_{m_1, m_2}(\mathcal{M}_{\min}, \mathcal{M}_{\max})$  denotes a distribution over  $\mathcal{M}$  with support,  $[\mathcal{M}_{\min}, \mathcal{M}_{\max}]$ , within which component masses,  $m_1, m_2$ , are uniformly distributed. We use a fixed  $\Delta\mathcal{M} = 0.005 M_{\odot}$  and choose a hyperprior,  $\hat{p}(\widetilde{\mathcal{M}})$ , covering the expected range of  $\mathcal{M}$  for LVK detections of BNS (Extended Data Table 2).

Because  $\Delta\mathcal{M}$  is small,  $\widetilde{\mathcal{M}}$  is a good approximation for any  $\mathcal{M}$  within the restricted prior,  $p_{\widetilde{\mathcal{M}}}(\mathcal{M})$ , and we can thus use  $\widetilde{\mathcal{M}}$  for heterodyning. The resulting model,  $q(\boldsymbol{\theta}|\mathbf{d}, \widetilde{\mathcal{M}})$ , can then perform inference with event-optimized heterodyning and priors (via the choice of appropriate  $\widetilde{\mathcal{M}}$ ), but is nevertheless applicable to the entire range of the hyperprior.

Inference results are independent of  $\widetilde{\mathcal{M}}$  as long as the posterior,  $p(\mathcal{M}|\mathbf{d})$ , is fully covered by  $[\widetilde{\mathcal{M}} - \Delta\mathcal{M}, \widetilde{\mathcal{M}} + \Delta\mathcal{M}]$ . For BNS,  $p(\mathcal{M}|\mathbf{d})$  is typically tightly constrained and we can use a coarse estimate of  $\mathcal{M}$  for  $\widetilde{\mathcal{M}}$ . This can either be taken from a GW search pipeline or rapidly computed from  $q(\boldsymbol{\theta}|\mathbf{d}, \widetilde{\mathcal{M}})$  itself by sweeping the hyperprior (see below). Note that, for shorter GW signals from black hole mergers,  $p(\mathcal{M}|\mathbf{d})$  is generally less well constrained. The transfer of prior conditioning would thus require larger (and potentially flexible) values of  $\Delta\mathcal{M}$ . Alternatively, the prior range can be extended at inference time by iterative Gibbs sampling of  $\mathcal{M}$  and  $\widetilde{\mathcal{M}}$ , similar to the GNPE algorithm<sup>19,45</sup>.

Prior conditioning is a general SBI technique that enables a choice of prior at inference time. This can also be achieved with sequential NPE<sup>40–42,48</sup>. However, in contrast to prior conditioning, these techniques require simulations and retraining for each observation, resulting in more expensive and slower inference. We here use prior conditioning with priors of fixed width for the chirp mass, and optional additional conditioning on fixed values for other parameters (corresponding to Dirac delta priors). Extension to more complicated priors and hyperpriors is straightforward.

**Independent estimation of chirp mass and merger times.** Running DINGO-BNS requires an initial estimate of the chirp mass,  $\mathcal{M}$  (to determine  $\widetilde{\mathcal{M}}$  for the network), and the merger time,  $t_c$  (to trigger the analysis). Matched filter searches can identify the presence of a compact binary signal and its chirp mass and merger time in low latency<sup>14,15,49–51</sup>. Specialized early warning searches are designed to produce output before the coalescence can further provide a rough indication of sky position and distance<sup>52–54</sup>. When available, the output of such pipelines can be used to trigger a DINGO analysis and provide estimates for  $\mathcal{M}$  and  $t_c$ .

We here describe an alternative independent approach of obtaining these parameters, using only the trained DINGO-BNS model. We compute  $\widetilde{\mathcal{M}}$  by sweeping the entire hyperprior,  $\hat{p}(\widetilde{\mathcal{M}}) = U_{m_1, m_2}(\widetilde{\mathcal{M}}_{\min}, \widetilde{\mathcal{M}}_{\max})$ . Specifically, we run DINGO-BNS with a set of prior centres,

$$\widetilde{\mathcal{M}}_i = \widetilde{\mathcal{M}}_{\min} + i\Delta\mathcal{M}, \quad (3)$$

where  $i$  takes integer values between 0 and  $(\widetilde{\mathcal{M}}_{\max} - \widetilde{\mathcal{M}}_{\min})/\Delta\mathcal{M}$ . The inference models in this study were trained with hyperprior ranges of up to  $[1.0, 2.2] M_{\odot}$ . For  $\Delta\mathcal{M} = 0.005 M_{\odot}$ , we can thus cover the entire global chirp mass range using 241 (overlapping) local priors. We run DINGO-BNS for all local priors,  $\widetilde{\mathcal{M}}_i$ , in parallel, with 10 samples per  $\widetilde{\mathcal{M}}_i$ . This requires a DINGO-BNS inference of only a few thousand samples, which takes less than 1 s. We use the chirp mass,  $\mathcal{M}$ , of the maximum likelihood sample as the prior centre,  $\widetilde{\mathcal{M}}$ , for the analysis (Extended Data Fig. 1a). Note that the exact choice of  $\widetilde{\mathcal{M}}$  does not matter, as long as the inferred posterior is fully covered by  $[\widetilde{\mathcal{M}} - \Delta\mathcal{M}, \widetilde{\mathcal{M}} + \Delta\mathcal{M}]$  (Extended Data Fig. 1b).

The merger time,  $t_c$ , can be inferred by continuously running this  $\widetilde{\mathcal{M}}$  scan on the input data stream, sliding the  $t_c$  prior in real time over the incoming data. With inference times of 1 s, continuous analysis can be achieved on just a few parallel computational nodes (or even a single node when also parallelizing over the  $t_c$  grid), constantly running on the input data stream. Event candidates can then be identified by analysing the SNR, triggering upon exceeding some defined threshold (Extended Data Fig. 1c). This scan can be performed at an arbitrary (but fixed) time before the merger.

This scan successfully estimates  $\mathcal{M}$  and  $t_c$  for both real BNS events (Extended Data Fig. 1). However, we have not tested this at a large scale

on detector noise to compute false alarm rates because DINGO-BNS is primarily intended for parameter estimation. Existing search and early warning pipelines are probably more robust for event identification, particularly in the presence of non-stationary detector noise.

**Frequency multibanding.** Although the native resolution of a frequency series is determined by the duration,  $T$ , of the corresponding time series, ( $\Delta f = 1/T$ ), we can average adjacent frequency bins whenever the signal is roughly constant. This enables data compression with only negligible loss of information. Here we use frequency multibanding, which divides the frequency range,  $[f_{\min}, f_{\max}]$ , into  $N$  bands of decreasing resolution. Frequency band  $i$  covers the range  $[\hat{f}_i, \hat{f}_{i+1}]$  with  $\Delta f_i = 2^i \Delta f_0$ , where  $\hat{f}_0 = f_{\min}$ ,  $\hat{f}_N = f_{\max}$  and  $\Delta f_0$  is the native resolution of the frequency series. Within band  $i$ , the multibanded domain thus compresses the data by a factor of  $2^i$  (Extended Data Fig. 2), which is achieved by averaging  $2^i$  sequential bins from the original frequency series (called decimation). To achieve optimal compression, we empirically choose the smallest possible nodes,  $\hat{f}_i$ , for which GW signals are still fully resolved. Specifically, we simulate a set of  $10^3$  heterodyned GW signals and demand that every period of these signals is covered by at least 32 bins in the resulting multibanded frequency domain. This is done before generating the training dataset, and the multibanded domain then remains fixed during dataset generation and training. The optimized resolution achieves compression factors between 60 and 650 (Extended Data Fig. 2c).

Traditionally, multibanding has been used without additional heterodyning—an approach we could also apply to DINGO-BNS. However, this would lead to lower compression factors, ranging from 14 to 56. More importantly, using multibanding alone would result in substantially more complicated input data to the DINGO-BNS network. Indeed, heterodyning enables additional truncation of the waveform singular value decomposition used in initializing the first network layer<sup>19</sup>. For LVK data, this results in a reduction from roughly 1,800 to 200 basis elements, thereby simplifying the learning task.

Care needs to be taken that the approximations are valid in the presence of detector noise. We now investigate how multibanding affects data simulation (for training) and the likelihood (for importance sampling).

**Data simulation.** The GW data are simulated as the sum of a signal and detector noise,  $\mathbf{d} = \mathbf{h}(\boldsymbol{\theta}) + \mathbf{n}$ . The detector noise in frequency bin  $j$  is given by

$$n_j \sim \mathcal{N}(0, \sigma \sqrt{S_j}), \sigma = \sqrt{\frac{w}{4\Delta f}}, \quad (4)$$

where  $\mathbf{S}$  denotes the detector noise PSD, and  $\sigma$  takes into account the frequency resolution and the Tukey window factor,  $w$ . Note that  $\mathbf{n}$  is a complex frequency series, which we ignore in our notation, as the considerations here hold for real and imaginary parts individually. It is conventional to work with whitened data,

$$d_j^w = h_j^w(\boldsymbol{\theta}) + n_j^w = \frac{h_j(\boldsymbol{\theta}) + n_j}{\sqrt{S_j}}, \quad (5)$$

in which case  $n_j^w \sim \mathcal{N}(0, \sigma)$ .

We convert to the multibanded frequency domain by averaging sets of  $N_i = 2^i$  bins,

$$\bar{d}_j^w = \frac{1}{N_i} \sum_{k=m_j}^{m_j+N_i-1} (h_k^w + n_k^w) = \bar{h}_j^w + \bar{n}_j^w, \quad (6)$$

where  $j$  denotes the bin in the multibanded domain,  $m_j$  denotes the starting index of the decimation window for  $j$  in the native domain and  $i$  indexes the frequency band associated with  $j$ . Because  $\bar{n}_j^w$  is an average of  $N_i$  Gaussian random variables with standard deviation  $\sigma$ , it follows that  $\bar{n}_j^w$  is also Gaussian with standard deviation,

$$\sigma_i = \sigma / \sqrt{N_i} = \sqrt{\frac{w}{4\Delta f N_i}} = \sqrt{\frac{w}{4\Delta f_i}}. \quad (7)$$

We can thus simulate the detector noise directly in the multibanded domain by updating  $\sigma \rightarrow \sigma_i$ , corresponding to  $\Delta f \rightarrow \Delta f_i$ . For the whitened signal we find

$$\bar{h}_j^w = \frac{1}{N_i} \sum_{k=m_j}^{m_j+N_i-1} \frac{h_k}{\sqrt{S_k}} \approx \bar{h}_j \sum_{k=m_j}^{m_j+N_i-1} \frac{1}{\sqrt{S_k}}, \quad (8)$$

assuming an approximately constant signal,  $\mathbf{h}$ , within the decimation window,  $\bar{h}_j \approx h_k, \forall k \in [m_j, m_j + N_i - 1]$ . For frequency-domain waveform models, we can thus directly compute the signal  $\bar{h}_j$  in the multibanded domain by simply evaluating the model at frequencies  $\bar{f}_j$ .

In summary, we can directly generate BNS data in the multibanded frequency domain by: (1) updating the noise standard deviation according to the multibanded resolution; (2) appropriately decimating noise PSDs; and (3) computing signals and noise realizations in the compressed domain. These operations are carefully designed to be consistent with the data processing of real BNS observations, which for DINGO-BNS are first whitened in the native domain and then decimated to the multibanded domain. This process relies on the assumption that signals are constant within decimation windows, and we ensure that this is (approximately) fulfilled when determining the multibanded resolution. Indeed, for signals generated directly in the multibanded domain, we find mismatches of at most around  $10^{-7}$  when comparing to signals that are properly decimated from the native domain.

**Likelihood evaluations.** We also use frequency multibanding to evaluate the likelihood for importance sampling. The standard Whittle likelihood used in GW astronomy<sup>26</sup> reads:

$$\log p(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} \sum_k \frac{|d_k^w - h_k^w(\boldsymbol{\theta})|^2}{\sigma^2}, \quad (9)$$

up to a normalization constant. The sum extends over all bins,  $k$ , in the native frequency domain. Assuming a constant signal (as above) and PSD within each decimation window, we can directly compute the likelihood in the multibanded domain,

$$\log p(\mathbf{d}|\boldsymbol{\theta}) \approx -\frac{1}{2} \sum_j \frac{|\bar{d}_j^w - \bar{h}_j^w(\boldsymbol{\theta})|^2}{\sigma_{(j)}^2}. \quad (10)$$

The assumptions are not exactly fulfilled in practice—for additional corrections, see ref. 29. For importance sampling, we can always evaluate the exact likelihood in the native frequency domain instead. In this case, the result is no longer subject to any approximations, even if the DINGO-BNS proposal is generated with a network using multibanded data. With the full likelihood for GW170817, we found a sample efficiency of 11.0% with an inference time of 13 s for 50,000 samples. The deviation from the result obtained with the multibanded likelihood is negligible (Jensen–Shannon divergence of less than  $5 \times 10^{-4}$  nat for all parameters). This demonstrates that use of the multibanded resolution has no practically relevant impact on the results.

**Frequency masking.** Because the GW likelihood (and our framework) uses the frequency domain, but data are taken in the time domain, it is necessary to convert data by windowing and Fourier transforming. However, frequency domain waveform models assume infinite time duration, leading to inconsistencies with finite time segments,  $[t_{\min}, t_{\max}]$ . Because the frequency evolution of the inspiral is tightly constrained by the chirp mass,  $\mathcal{M}$ , we can compute boundaries,  $f_{\min}(t_{\min}, \mathcal{M})$  and  $f_{\max}(t_{\max}, \mathcal{M})$ , such that the signals are not corrupted by the finite-duration effects within  $[f_{\min}, f_{\max}]$  and are negligibly small outside of that range (Extended Data Fig. 3).

We approximate the lower bound,  $f_{\min}(t_{\min}, \mathcal{M})$ , using the leading order in the post-Newtonian relationship between time and frequency,

$$f_{\text{OPN}}(t, \mathcal{M}) = \frac{1}{8\pi} \left(\frac{-t}{5}\right)^{-3/8} \left(\frac{G\mathcal{M}}{c^3}\right)^{-5/8}. \quad (11)$$

For a network designed for fixed data duration,  $T$ , we set  $f_{\min}(T, \mathcal{M}) = f_{\text{OPN}}(-T, \mathcal{M}) + f_{\text{buffer}}$  (we use  $f_{\text{buffer}} = 1$  Hz for LVK and  $f_{\text{buffer}} = 0.5$  Hz for next-generation detector setups).

For the upper bound, we found that  $f_{\text{OPN}}(t, \mathcal{M})$  is not sufficiently accurate. Instead, we determined  $f_{\max}(t, \mathcal{M})$  empirically by simulating a set of signals (with parameters  $\theta - p(\theta)$ ) and computing mismatches between signals with and without truncation at  $t > t_{\max}$ . For a given set of simulations, we choose  $f_{\max}(t, \mathcal{M})$  as the highest frequency at which all mismatches are at the most  $10^{-3}$ . To avoid additional computation at inference time, we cache the results in a lookup table for  $f_{\max}(t, \mathcal{M})$ . The lookup table used here was generated with 20 waveforms per element. We verified for random elements that this matches the result obtained using 1,000 waveforms with an accuracy of approximately 0.1 Hz. For production use, the lookup table may need to be generated with more waveforms.

Both bounds depend on the chirp mass,  $\mathcal{M}$ , with the upper bound additionally depending on the pre-merger time. To enable inference for arbitrary configurations, we trained a single network with variable frequency bounds. During training, we computed  $f_{\min}(T, \tilde{\mathcal{M}})$  with the centre  $\tilde{\mathcal{M}}$  of the local chirp mass prior. The upper frequency bound,  $f_{\max}$ , is sampled randomly (uniform in frequency bins of the multibanded frequency domain) to allow arbitrary pre-merger times. Data outside of  $[f_{\min}, f_{\max}]$  are zero-masked.

Such masked networks can perform pre-merger inference. Given an alert (for example, from a detection pipeline) predicting a merger at time  $\tilde{t}_c$  (in the future) with chirp mass  $\tilde{\mathcal{M}}$ , DINGO-BNS uses the frequency range  $[f_{\min}(T, \tilde{\mathcal{M}}), f_{\max}(-\tilde{t}_c, \tilde{\mathcal{M}})]$ . The frequency domain data are further shifted by  $\tilde{t}_c$ , such that the prior  $p(t_c)$  is centred around  $\tilde{t}_c$ . The resulting pre-merger posterior can then be used to update the alert predictions  $(\tilde{\mathcal{M}}, \tilde{t}_c)$  (for example, as the centre of the inferred posterior marginals) and to trigger a new DINGO-BNS analysis with the new strain data that became available in the meantime. Such iterative posterior updates allow continuous optimization of the prior centres  $(\tilde{\mathcal{M}}, \tilde{t}_c)$ . As a result, the  $t_c$  prior for the after-merger analysis (Extended Data Table 2) does not need to account for large trigger uncertainties and only needs to be large enough to capture expected posteriors,  $p(t_c | \mathbf{d})$ .

In the absence of external alerts, the independent DINGO-BNS search described earlier (Extended Data Fig. 1) can also be run as a pre-merger scan, searching for mergers around a freely chosen (but fixed) time,  $\tilde{t}_c$ , in the future. By continuously running this scan over the incoming data stream, alerts are triggered at roughly time  $\tilde{t}_c$  before potential BNS mergers. Note that, in the absence of chirp mass predictions, we need to apply a conservative frequency range,  $[f_{\min}(T, \mathcal{M}_{\min}), f_{\max}(-\tilde{t}_c, \mathcal{M}_{\min})]$ , where  $\mathcal{M}_{\min/\max}$  are the prior bounds for  $\mathcal{M}$ .

**EOS likelihood.** A nuclear EOS implies a functional relationship between neutron star masses,  $m_i$ , and tidal deformabilities,  $\lambda_i$ . The likelihood,  $p(\mathbf{d} | \mathcal{E})$ , for a given EOS,  $\mathcal{E}$ , and data,  $\mathbf{d}$ , can be computed by integrating the GW likelihood along the hyperplane defined by the EOS constraint,  $\lambda_i = \lambda_i^\mathcal{E}(m_i)$ ,

$$p(\mathbf{d} | \mathcal{E}) = \int p(\mathbf{d} | \theta) p(\theta) \delta(\lambda_i - \lambda_i^\mathcal{E}(m_i)) d\theta \\ = \int \int p(\mathbf{d} | m_1, m_2, \lambda_1^\mathcal{E}(m_1), \lambda_2^\mathcal{E}(m_2)) dm_1 dm_2 \quad (12)$$

Here  $p(\mathbf{d} | m_1, m_2, \lambda_1, \lambda_2)$  is the Bayesian evidence of  $\mathbf{d}$  conditional on  $(m_1, m_2, \lambda_1, \lambda_2)$ . To calculate equation (12) using Monte Carlo integration, it is necessary to repeatedly evaluate the integrand, which is extremely expensive using traditional methods (for example, nested sampling).

With DINGO-BNS, there are two fast ways to evaluate the integrand, using either a conditional or a marginal network: (1) a marginal network,  $q(m_1, m_2, \lambda_1, \lambda_2 | \mathbf{d})$ , directly provides an unnormalized estimate of the conditional evidence,  $p(\mathbf{d} | m_1, m_2, \lambda_1, \lambda_2)$  (sufficient for model comparison, but not subject to our usual accuracy guarantees); or (2) a conditional network,  $q(\theta | \mathbf{d}; m_1, m_2, \lambda_1, \lambda_2)$ , provides the normalized conditional evidence via importance sampling (including accuracy guarantees). Option (1) allows  $10^5$  evaluations per second, whereas option (2) only allows  $10^3$ , assuming  $10^2$  weighted samples per evaluation.

By combining (1) and (2), we can achieve speed and accuracy, using the marginal network (1) to define a proposal distribution for Monte Carlo integration with the integrand from (2). Specifically, the density of the marginal network,  $q(m_1, m_2, \lambda_1, \lambda_2 | \mathbf{d})$ , is evaluated on an  $(m_1, m_2)$  grid with  $\lambda_i = \lambda_i^\mathcal{E}(m_i)$ . This provides a discretized estimate of the integrand,  $p(\mathbf{d} | m_1, m_2, \lambda_1^\mathcal{E}(m_1), \lambda_2^\mathcal{E}(m_2))$ , which we use as a proposal distribution for the integration in equation (12) when computing the integrand with the more accurate method (2). We tested this on GW170817 data using two polynomial EOS constraints,  $\lambda = \lambda^\mathcal{E}(m)$  (Extended Data Fig. 4), finding good sample efficiencies of around 50%, small uncertainties,  $\sigma_{\log p(\mathbf{d} | \mathcal{E})} \approx 0.01$ , and computation times of 1–3 s for the integral equation (12). Alternatively, the proposal could also be generated using a network,  $q(m_1, m_2 | \mathbf{d})$ , which additionally marginalizes over  $\lambda_i$ . Finally, for a parametric EOS, a DINGO-BNS network could be conditioned on EOS parameters, allowing for direct EOS inference. This variety of approaches emphasizes the flexibility of SBI for EOS inference.

**Related work.** Machine learning for GW astronomy is an active area of research<sup>55</sup>. Several studies have explored machine-learning inference for black hole mergers<sup>19,20,56–65</sup>. There have also been applications specific to BNS inference (Extended Data Table 1). The GW-SkyLocator algorithm<sup>66</sup> estimates the sky position using the SNR time series (similar to BAYESTAR), whereas Jim<sup>33,35</sup> uses hardware acceleration and machine learning to speed up conventional samplers and achieve full inference in 21–33 min. The i-nessai framework<sup>39</sup> achieves BNS inference in 24 min by combining normalizing flows using importance nested sampling. Ref. 67 explores pre-merger BNS detection and parameter estimation with normalizing flows, also reporting 1-s analysis times. However, it has not demonstrated accurate results on real data, and is subject to several other limitations. SBI has also been used for neutron star EOS inference from GWs<sup>68</sup> and electromagnetic data<sup>69</sup>. Pre-merger localization with conventional techniques has been explored for ground-based third-generation detectors<sup>70–72</sup> and for space-based detectors<sup>73,74</sup>.

## Experimental details

For our experiments, we trained DINGO-BNS networks using the hyperparameters and neural architecture<sup>44,75</sup> from ref. 19, with a few modifications. The embedding network consisted of a sequence of 34 two-layer, fully-connected residual blocks with hidden dimensions of 2,048 ( $\times 8$ ), 1,024 ( $\times 8$ ), 512 ( $\times 6$ ), 256 ( $\times 6$ ) and 128 ( $\times 6$ ) after the initial projection layer. Compared with ref. 19, this added ten new blocks, increasing the number of trainable parameters in this part of the embedding network from 17 million to 91 million. For the LVK experiments, we used a dataset with  $3 \times 10^7$  training samples and trained it for 200 epochs. For the Cosmic Explorer experiments, we used  $6 \times 10^7$  training samples and trained it for 100 epochs. Training took between 5 and 7 days on one H100 GPU. We used three detectors for LVK (LIGO-Hanford, LIGO-Livingston and Virgo) and two detectors for Cosmic Explorer (primary detector at the location of LIGO-Hanford, secondary detector at the location of LIGO-Livingston). The networks were trained with the priors displayed in Extended Data Table 2. The DINGO-BNS network marginalized over the phase of coalescence,  $\phi_c$ . During importance sampling, we reconstructed  $\phi_c$  (ref. 20) (Extended Data Figs. 6a and 8) or used a phase-marginalized likelihood<sup>76,77</sup> (in all other experiments).



The phase reconstruction used here made the same assumptions as conventional phase marginalization<sup>76,77</sup>.

In the first experiment, we evaluated DINGO-BNS models on 200 simulated GW datasets, generated using a fixed GW signal with GW170817-like parameters and simulated LVK detector noise. We used noise PSDs from the second (O2) and third (O3) LVK, observing runs as well as LVK design sensitivity. For each noise level, we trained one pre-merger network ( $f \in [23, 200]$  Hz) and one network for inference with the full signal, including the merger ( $f \in [23, 1024]$ ). The latter network was only used for after-merger inference because we found that separation into two networks improved the performance. The pre-merger network was trained with frequency masking, the masking bound,  $f_{\max}$ , being sampled in the range  $[28, 200]$  Hz, enabling inference up to 60 s before the merger.

In the second experiment, we analysed  $10^4$  simulated GW datasets, with GW signal parameters randomly sampled from the prior (Extended Data Table 2), the  $\mathcal{M}$  prior reduced to the range  $[1.0, 1.5] M_{\odot}$  and the  $d_l$  prior reweighted to a uniform distribution in the comoving volume, with design sensitivity noise PSDs. We again trained one pre-merger network ( $f \in [19.4, 200]$  Hz) and one after-merger network ( $f \in [19.4, 1, 024]$  Hz). The pre-merger network was trained with frequency masking with the masking bound,  $f_{\max}$ , sampled in the range  $[25, 200]$  Hz, enabling inference up to 60 s before the merger for  $\mathcal{M} \leq 1.5 M_{\odot}$ . Both networks were additionally trained with lower-frequency masking, with  $f_{\min}(\mathcal{M})$  determined as explained above, ensuring an optimal frequency range for any chirp mass. Following ref. 17, we only considered events with  $\text{SNR} \geq 12$ . Before each analysis, we performed a  $t_c$  scan by generating 2,500 samples for four time-shifted copies of the strain, followed by joint importance sampling of the combined results. Specifically, for a network with a  $t_c$  prior  $U(-\tau, \tau)$ , we applied the time shifts  $(-3\tau, -\tau, \tau, 3\tau)$ , effectively increasing the prior to  $U(-4\tau, 4\tau)$ . For the subsequent analysis, we time shifted the data such that the  $t_c$  posterior was fully covered by the prior.

For each DINGO-BNS result, we generated a skymap using a kernel density estimator implemented by `ligo.skymap`<sup>78</sup>. For the sky localization comparison between DINGO-BNS and BAYESTAR, we ran BAYESTAR based on the GW signal template generated with the maximum likelihood parameters from the DINGO-BNS analysis. We noted that BAYESTAR was designed as a low-latency pipeline and typically run with (coarser) parameter estimates from search templates. Therefore, the reported BAYESTAR runs may deviate slightly from the realistic LVK setup. However, our results are consistent with those of ref. 17, which also had an approximately 30% precision improvement over BAYESTAR localization (using LVK search triggers). Both DINGO-BNS and ref. 17 performed full Bayesian BNS inference and should therefore have had identical localization improvements over BAYESTAR (assuming ideal accuracy, which for DINGO-BNS was validated with consistently high importance sampling efficiency). Differences to the localization comparison in ref. 17 are thus primarily attributed to different configurations for BAYESTAR and slightly different injection priors. Additional results for the localization comparison are shown in Extended Data Fig. 5. A probability–probability (P–P) plot for the after-merger analysis is shown in Extended Data Fig. 6a, which shows no significant bias for any parameter.

In the third experiment, we reproduced the public LVK results for GW170817 (refs. 1,5) and GW190425 (ref. 79) with DINGO-BNS. We used the same priors and data settings as the LVK, but we did not marginalize over calibration uncertainty. The GW170817 after-merger analysis (see also Extended Data Fig. 7) was performed with a DINGO-BNS model conditioned on the sky position,  $\{\alpha, \beta\}$ . Following the LVK analysis<sup>5</sup>, we used the localization  $\alpha = 3.44616$  rad and  $\delta = -0.408084$  rad from the electromagnetic counterpart AT 2017gfo (ref. 8) at inference. Note that the localization uncertainty ( $\sigma_{\alpha} = 10^{-6}$  rad,  $\sigma_{\delta} = 10^{-6}$  rad (ref. 8)) was negligible for GW parameter estimation, but such effects could, in principle, be integrated by convolving the conditional DINGO-BNS

network with a distribution over  $\{\alpha, \beta\}$ . We found good sample efficiencies for both events (10.8% for GW170817 and 51.3% for GW190425) and good agreement with the LVK results (Extended Data Fig. 6b). The LVK results used detector noise PSDs generated with BayesWave<sup>36</sup>, which were not available before the merger. For our pre-merger analysis of GW170817 in the main part (sample efficiency 78.9%), we thus used a PSD generated using the Welch method. The GW170817 signal overlapped with a loud glitch in the LIGO-Livingston detector<sup>1</sup>, and we used the glitch-subtracted data provided by the LVK in our analyses. Because such data would not be available before the merger, the pre-merger inference of BNS events overlapping with glitches would, in practice, also require fast glitch mitigation methods.

In the fourth experiment, we analysed simulated Cosmic Explorer data using the anticipated noise PSDs for the primary and secondary detectors. We trained a DINGO-BNS network for pre-merger inference using  $f \in [6, 11]$  Hz, with the upper frequency masking bound,  $f_{\max}$ , sampled in the range  $[7, 11]$  Hz. This supported a signal length of 4,096 s, with a pre-merger inference between 45 and 15 min before the merger. We injected signals with GW170817-like parameters for distance, masses and inclination to investigate how well a GW170817-like event could be localized in the Cosmic Explorer detector. We also trained a network on the full frequency range,  $[6, 1, 024]$  Hz, for after-merger inference, with a reduced distance prior to control the SNR (Extended Data Table 2).

**Sample efficiencies.** We report sample efficiencies for all injection studies in Extended Data Fig. 6. The importance-sampled DINGO-BNS results are accurate, even with low efficiency, provided that a sufficient absolute number of effective samples can be generated. The efficiency nevertheless is a valuable diagnostic for assessing the performance of the trained inference networks.

In the LVK experiments, we found consistently high efficiencies, comparable to or greater than those reported for BBHs<sup>20</sup>. As a general trend, we observed that higher noise levels (Extended Data Fig. 6c) and earlier pre-merger times (Extended Data Fig. 6d) led to higher efficiencies. This is because low SNR events generally have broader posteriors, which are simpler to model for DINGO-BNS density estimators. Furthermore, the GW signal morphology is most complicated around the merger, making pre-merger inference much simpler than inference based on the full signal.

For Cosmic Explorer injections with GW170817-like parameters (Extended Data Fig. 6e), DINGO-BNS achieved extremely high efficiency for early pre-merger analyses, but the performance decreased substantially for later analysis times. This effect can again be attributed to the increase in SNR, which was  $O(10^3)$  15 min before the merger. Improving DINGO-BNS for such high SNR events will probably require improved density estimators<sup>64</sup> that can better deal with tighter posteriors. When limiting the SNR by increasing the distance prior (Extended Data Table 2), we found good sample efficiencies for an after-merger Cosmic Explorer analysis that used the full 4,096-s-long signal (Extended Data Fig. 6e).

**Inference times.** The computational cost of inference with DINGO-BNS is dominated by: (1) neural-network forward passes to sample from the approximate posterior,  $\theta \sim q(\theta | \mathbf{d}_{\mathcal{N}}, \mathcal{M})$ ; and (2) likelihood evaluations,  $p(\theta | \mathbf{d})$ , used for importance sampling. For 50,000 samples on an H100 GPU, (1) takes around 0.370 s and (2) takes around 0.190 s, resulting in an inference time of less than 0.6 s. The speed of the likelihood evaluations is enabled by using JAX waveform and likelihood implementations<sup>33–35</sup>, combined with the heterodyning and multiband-ing step that we also used to compress the data for the DINGO-BNS network. We extend the open-source implementations<sup>34,35</sup> by combining NRTidalv1 (refs. 32,80) with IMRPhenomPv2 (refs. 81–83), as well as re-implementing the DINGO likelihood functions in JAX. The JAX functions are usually just-in-time compiled to run efficiently. The input dimension of the likelihood (determined by likelihood batch size and

# Article

number of frequency bins) is fixed, enabling compilation ahead of time on random data of the correct input dimension. After compilation, a running DINGO-BNS script can perform any number of analyses without recompiling. Thus, we can leave the compilation time (18 s) out of the timing estimate for importance sampling. Compilation could further be transferred between separate DINGO-BNS runs using the persistent compilation cache in JAX, although we have not implemented this option. Likelihood evaluations can also be done without JAX, which takes less than 10 s on a single node with 64 CPUs for 50,000 samples. For the vast majority of DINGO-BNS analyses in this study, the sample efficiency was sufficiently high such that 50,000 samples corresponded to several thousands of effective samples after importance sampling, enabling full importance sampling inference in less than 1 s.

**Additional sources of latency.** The inference times quoted above assume that the data have already been provided to DINGO-BNS. In practice, there are various additional sources of latency.

First, PSD estimation typically uses strain data taken during or after the merger. Obtaining low-latency PSDs represents a general challenge for low-latency analyses, encountered also in the field of GW searches<sup>14,53</sup>. Indeed, most PSDs used in this work would not be available in very low latency—the GW170817 and GW190425 PSDs from the LVK analyses<sup>5,79</sup> use on-source noise estimation<sup>36</sup> based on data during the merger, whereas LVK design sensitivity and Cosmic Explorer PSDs are not based on real detector noise, but rather reflect anticipated future configurations. These PSDs were chosen for comparability with existing studies. To test DINGO-BNS in a more realistic low-latency setting, we performed inference with PSDs estimated using Welch’s method<sup>84</sup> and only data from before the merger (Extended Data Fig. 8). The resulting posterior for GW170817 only deviated slightly from the posterior obtained from the on-source PSDs. However, we note that, in general, such pre-merger Welch PSDs may be less reliable in the presence of non-stationary detector noise, which could lead to larger biases for parameter estimation.

Strain data can also be contaminated with instrumental or environmental glitches, which need to be removed for parameter estimation<sup>85</sup>. This adds additional latency for events that overlap with glitches, as noted above in the case of GW170817. Finally, data transfer between detectors and computing facilities add some additional latency. Using DINGO-BNS to its full potential will therefore require careful integration with low-latency pipelines<sup>23</sup> and further acceleration of existing components.

**PSD tuning.** Although most of the networks used in this study were trained with only a single PSD per detector, in practice we would generally train DINGO-BNS with an entire distribution of PSDs to enable instant tuning to drifting detector noise<sup>19</sup>. (This is not relevant to tests involving, for example, design sensitivity noise.) Of the experiments in this study, only the pre-merger result of GW170817 and the result for the Welch PSD (Extended Data Fig. 8) were generated with a PSD-conditioned DINGO-BNS network. This network was trained with a PSD distribution covering the entire second LVK observing run (O2). Conditioning on the PSD makes the inference task more complicated and therefore leads to slightly reduced performance. For example, when repeating the first injection experiment (Extended Data Fig. 6c) with the PSD-conditioned DINGO-BNS network from above, the mean efficiency was reduced from 71% to 22%. Such networks can, in principle, also be trained before the start of an observing run, by training with a synthetic dataset designed to reflect the expected noise PSDs<sup>86</sup>.

## Data availability

Public LVK data is available at <https://gwosc.org/events/GW170817/> for GW170817 and at <https://dcc.ligo.org/LIGO-T1900685/public> for GW190425.

## Code availability

The code for DINGO-BNS is publicly available as an extension of the DINGO python package at <https://github.com/dingo-gw/dingo>. We provide a demo at <https://github.com/dingo-gw/binary-neutron-star-demo> and a trained network via Zenodo at <https://zenodo.org/records/13321251> (ref. 88).

- Papamakarios, G. & Murray, I. Fast  $\epsilon$ -free inference of simulation models with Bayesian conditional density estimation. In *NIPS'16: Proceedings 30th International Conference on Neural Information Processing Systems* (eds Lee, D. D. et al.) 1036–1044 (Curran Associates Inc., 2016).
- Lueckmann, J.-M. et al. Flexible statistical inference for mechanistic models of neural dynamics. In *NIPS'17: Proceedings 31st International Conference on Neural Information Processing Systems* (eds von Luxburg, U. et al.) 1289–1299 (Curran Associates Inc., 2017).
- Greenberg, D., Nonnenmacher, M. & Macke, J. Automatic posterior transformation for likelihood-free inference. *Proc. Machine Learn. Res.* **97**, 2404–2414 (2019).
- Rezende, D. & Mohamed, S. Variational inference with normalizing flows. *Proc. Machine Learn. Res.* **37**, 1530–1538 (2015).
- Durkan, C., Bekasov, A., Murray, I. & Papamakarios, G. Neural spline flows. In *NeurIPS 2019: Proceedings 33rd Annual Conference on Neural Information Processing Systems* (eds Wallach, H. et al.) 7509–7520 (Curran Associates Inc., 2019).
- Dax, M. et al. Group equivariant neural posterior estimation. In *Proceedings 10th International Conference on Learning Representations* (2022).
- Cornish, N. J. Heterodyned likelihood for rapid gravitational wave parameter inference. *Phys. Rev. D* **104**, 104054 (2021).
- Zackay, B., Dai, L. & Venumadhav, T. Relative binning and fast likelihood evaluation for gravitational wave parameter estimation. Preprint at <https://arxiv.org/abs/1806.08792> (2018).
- Deistler, M., Goncalves, P. J. & Macke, J. H. Truncated proposals for scalable and hassle-free simulation-based inference. In *NeurIPS 2022: Proceedings 36th Annual Conference on Neural Information Processing Systems* (eds Koyejo, S. et al.) 23135–23149 (Curran Associates Inc., 2022).
- Finn, L. S. Detection, measurement and gravitational radiation. *Phys. Rev. D* **46**, 5236–5249 (1992).
- Adams, T. et al. Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era. *Class. Quant. Grav.* **33**, 175012 (2016).
- Chu, Q. et al. SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences. *Phys. Rev. D* **105**, 024023 (2022).
- Sachdev, S. et al. An early-warning system for electromagnetic follow-up of gravitational-wave events. *Astrophys. J. Lett.* **905**, L25 (2020).
- Nitz, A. H., Schäfer, M. & Dal Canton, T. Gravitational-wave merger forecasting: scenarios for the early detection and localization of compact-binary mergers with ground-based observatories. *Astrophys. J. Lett.* **902**, L29 (2020).
- Kovalam, M. et al. Early warnings of binary neutron star coalescence using the SPIIR search. *Astrophys. J. Lett.* **927**, L9 (2022).
- Cuoco, E. et al. Enhancing gravitational-wave science with machine learning. *Mach. Learn. Sci. Tech.* **2**, 011002 (2020).
- Gabbard, H., Messenger, C., Heng, I. S., Tonolini, F. & Murray-Smith, R. Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy. *Nat. Phys.* **18**, 112–117 (2022).
- Green, S. R., Simpson, C. & Gair, J. Gravitational-wave parameter estimation with autoregressive neural network flows. *Phys. Rev. D* **102**, 104057 (2020).
- Delaunoy, A. et al. Lightning-fast gravitational wave parameter inference through neural amortization. In *Proceedings 3rd Workshop on Machine Learning and the Physical Sciences* (2020).
- Green, S. R. & Gair, J. Complete parameter inference for GW150914 using deep learning. *Mach. Learn. Sci. Tech.* **2**, 03LT01 (2021).
- Williams, M. J., Veitch, J. & Messenger, C. Nested sampling with normalizing flows for gravitational-wave inference. *Phys. Rev. D* **103**, 103006 (2021).
- Alvey, J., Bhardwaj, U., Nissanke, S. & Weniger, C. What to do when things get crowded? Scalable joint analysis of overlapping gravitational wave signals. Preprint at <https://arxiv.org/abs/2308.06318> (2023).
- Crisostomi, M., Dey, K., Barausse, E. & Trota, R. Neural posterior estimation with guaranteed exact coverage: the ringdown of GW150914. *Phys. Rev. D* **108**, 044029 (2023).
- Bhardwaj, U., Alvey, J., Miller, B. K., Nissanke, S. & Weniger, C. Sequential simulation-based inference for gravitational wave signals. *Phys. Rev. D* **108**, 042004 (2023).
- Wildberger, J. et al. Flow matching for scalable simulation-based inference. *NeurIPS 2023* (2023).
- Kolmus, A. et al. Tuning neural posterior estimation for gravitational wave inference. Preprint at <https://arxiv.org/abs/2403.02443> (2024).
- Chatterjee, C. & Wen, L. Premerger sky localization of gravitational waves from binary neutron star mergers using deep learning. *Astrophys. J.* **959**, 76 (2023).
- van Straalen, W., Kolmus, A., Janquart, J. & Van Den Broeck, C. Pre-merger detection and characterization of inspiraling binary neutron stars derived from neural posterior estimation. Preprint at <https://www.arxiv.org/abs/2407.10263> (2024).
- McGinn, J. et al. Rapid neutron star equation of state inference with normalising flows. Preprint at <https://arxiv.org/abs/2403.17462> (2024).
- Brandes, L. et al. Neural simulation-based inference of the neutron star equation of state directly from telescope spectra. *J. Cosmol. Astropart. Phys.* **09**, 009 (2024).
- Chan, M. L., Messenger, C., Heng, I. S. & Hendry, M. Binary neutron star mergers and third generation detectors: localization and early warning. *Phys. Rev. D* **97**, 123014 (2018).
- Nitz, A. H. & Dal Canton, T. Pre-merger localization of compact-binary mergers with third-generation observatories. *Astrophys. J. Lett.* **917**, L27 (2021).

72. Hu, Q. & Veitch, J. Rapid premerger localization of binary neutron stars in third-generation gravitational-wave detectors. *Astrophys. J. Lett.* **958**, L43 (2023).
73. Lops, G. et al. Galaxy fields of LISA massive black hole mergers in a simulated universe. *Mon. Not. Roy. Astron. Soc.* **519**, 5962–5986 (2023).
74. Piro, L. et al. Chasing super-massive black hole merging events with Athena and LISA. *Mon. Not. Roy. Astron. Soc.* **521**, 2577–2592 (2023).
75. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (IEEE, 2015).
76. Veitch, J. & Del Pozzo, W. Analytic marginalisation of phase parameter. *Ligo DCC* <https://dcc.ligo.org/LIGO-T1300326/public> (2013).
77. Thrane, E. & Talbot, C. An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models. *Publ. Astron. Soc. Austral.* **36**, e010 (2019); erratum **37**, e036 (2020).
78. Singer, L. *ligo.skymap*. *ligo.skymap:docs* <https://lscsoft.docs.ligo.org/ligo.skymap/> (2020).
79. Abbott, B. et al. GW190425: observation of a compact binary coalescence with total mass  $\sim 3.4 M_{\odot}$ . *Astrophys. J. Lett.* **892**, L3 (2020).
80. Dietrich, T., Bernuzzi, S. & Tichy, W. Closed-form tidal approximants for binary neutron star gravitational waveforms constructed from high-resolution numerical relativity simulations. *Phys. Rev. D* **96**, 121501 (2017).
81. Hannam, M. et al. Simple model of complete precessing black-hole-binary gravitational waveforms. *Phys. Rev. Lett.* **113**, 151101 (2014).
82. Khan, S. et al. Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D* **93**, 044007 (2016).
83. Bohé, A. et al. PhenomPv2 – technical notes for the LAL implementation. *LIGO Technical Document*, LIGO-T1500602-v4 (2016).
84. Welch, P. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **15**, 70–73 (1967).
85. Pankow, C. et al. Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817. *Phys. Rev. D* **98**, 084016 (2018).
86. Wildberger, J. et al. Adapting to noise distribution shifts in flow-based gravitational-wave inference. *Phys. Rev. D* **107**, 084046 (2023).
87. Hernandez Vivanco, F. et al. Measuring the neutron star equation of state with gravitational waves: the first forty binary neutron star merger observations. *Phys. Rev. D* **100**, 103009 (2019).

**Acknowledgements** We thank S. Buchholz, T. Dent, A. Kofler and S. Morisaki for useful discussions. This research made use of data and software obtained from the Gravitational Wave Open Science Center ([gwosc.org](http://gwosc.org)), a service of the LIGO Scientific Collaboration, the

Virgo Collaboration and KAGRA. This material is based upon work supported by the National Science Foundation's (NSF's) LIGO Laboratory—a major facility, fully funded by the NSF—as well as the Science and Technology Facilities Council (STFC) of the UK, the Max Planck Society (MPS) and the State of Niedersachsen, Germany, which supported the construction of the Advanced LIGO and construction and operation of the GEO600 detector. Additional support for the Advanced LIGO was provided by the Australian Research Council. Virgo is funded through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions from institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal and Spain. KAGRA is supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the Japan Society for the Promotion of Science (JSPS) in Japan; the National Research Foundation (NRF) and Ministry of Science and Information and Communications Technology (MSIT) in Korea; and Academia Sinica (AS) and the National Science and Technology Council (NSTC) in Taiwan. M.D. thanks the Hector Fellow Academy for support. This work was supported by the German Research Foundation (DFG) through Germany's Excellence Strategy (EXC-Number 2064/1, project number 390727645) and the German Federal Ministry of Education and Research (BMBF) (Tübingen AI Center, FKZ: 01IS18039A). S.R.G. was supported by a UK Research and Innovation (UKRI) Future Leaders Fellowship (grant no. MR/Y018060/1). The computational work for this manuscript was carried out on the Atlas cluster at the Max Planck Institute for Intelligent Systems in Tübingen, Germany, and the Lakshmi and Hypatia clusters at the Max Planck Institute for Gravitational Physics in Potsdam, Germany. V.R. is supported by the UK's Science and Technology Facilities Council (grant no. ST/V005618/1).

**Author contributions** M.D. led the research with input from S.R.G. and B.S. M.D. and S.R.G. conceived the main methodology. M.D. devised and carried out the implementation and experimental analysis. M.D. and S.R.G. developed the DINGO code, with contributions from N.G., M.P. and J.W. N.G. implemented the JAX waveform model. V.R. proposed the exploration of the EOS inference. M.D. and S.R.G. wrote the paper. M.D., S.R.G., J.G., N.G., M.P., V.R., J.W., J.H.M., A.B. and B.S. contributed to scientific discussions and paper editing.

**Funding** Open access funding provided by Max Planck Society.

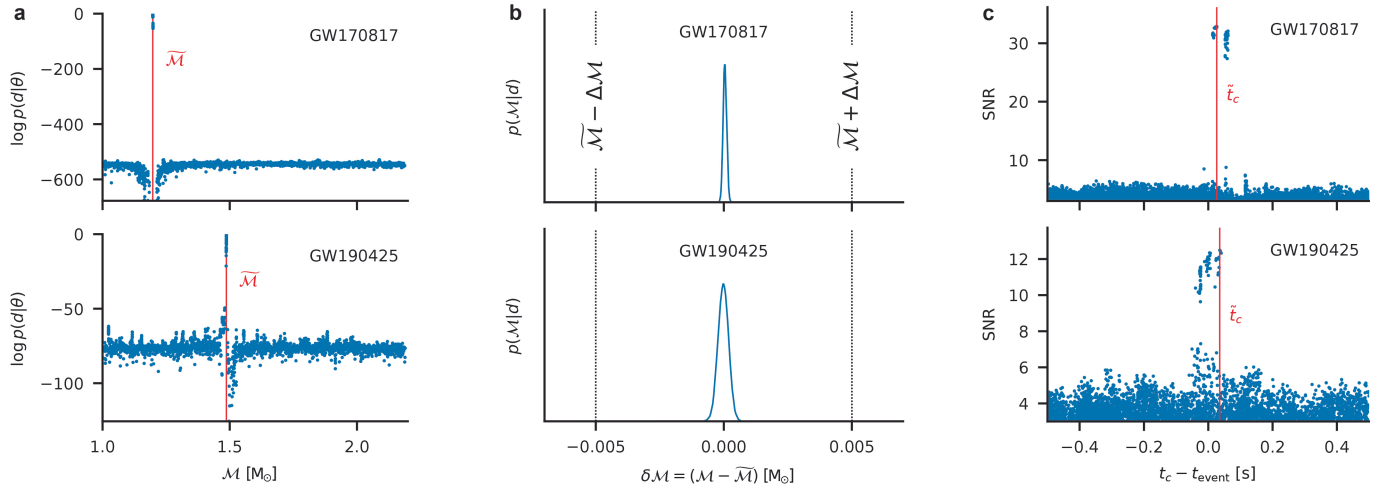
**Competing interests** The authors declare no competing interests.

#### Additional information

**Correspondence and requests for materials** should be addressed to Maximilian Dax.

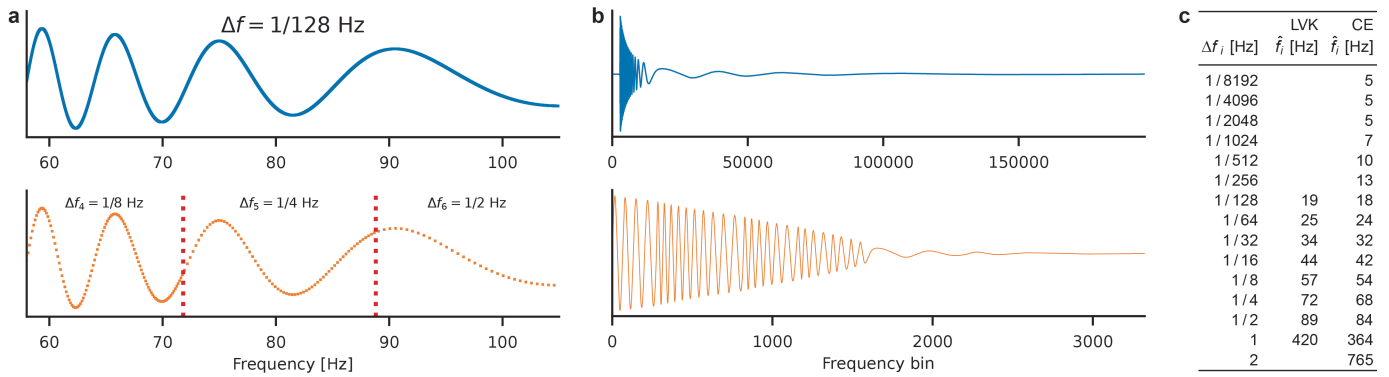
**Peer review information** *Nature* thanks Rory Smith, Michael Williams and Kaze Wong for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



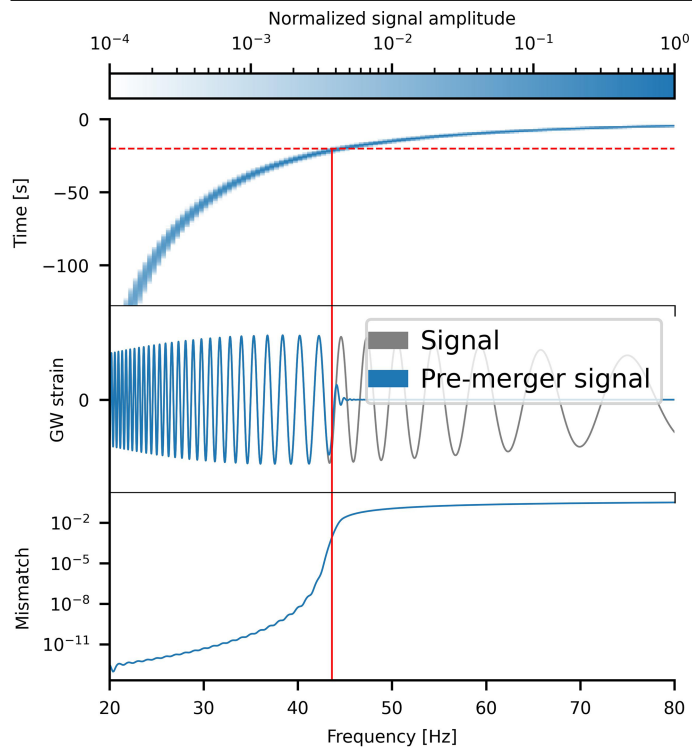
**Extended Data Fig. 1 | DINGO-BNS scan estimates the chirp mass and merger time.** (a) Log likelihoods generated from a scan over different values of  $\tilde{\mathcal{M}}$  with a DINGO-BNS network. The final  $\tilde{\mathcal{M}}$  is chosen as the maximum likelihood  $\mathcal{M}$  (red line;  $\tilde{\mathcal{M}} = 1.1975$  M<sub>⊙</sub> for GW170817,  $\tilde{\mathcal{M}} = 1.4868$  M<sub>⊙</sub> for GW190425).

(b) Posterior marginal  $p(\mathcal{M}|\mathbf{d})$ . The prior (dashed lines) determined by the scan from (a) fully covers the marginal. (c) A combined scan over  $\mathcal{M}$  and  $t_c$  successfully identifies GW170817 (with  $\hat{t}_c = 1187008882.43$ ) and GW190425 (with  $\hat{t}_c = 1240215503.04$ ).

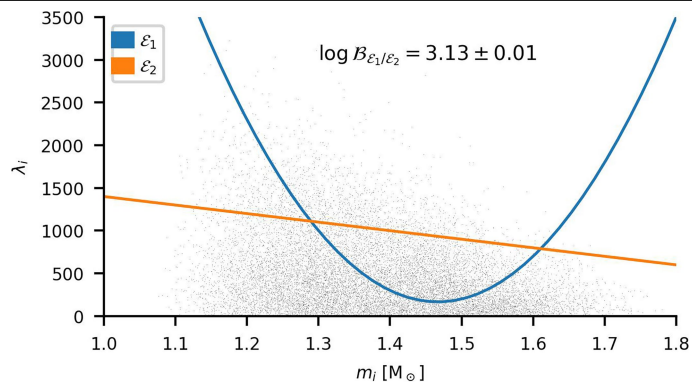


**Extended Data Fig. 2 | Frequency multibanding.** (a) The period of (heterodyned) GW signals decreases with increasing frequency. The native frequency resolution (blue) thus oversamples the signal at high frequencies. Frequency multibanding (band boundaries indicated by dotted red lines) adapts to the signal variation, decreasing the resolution at higher frequencies (orange). (b) The multibanded domain therefore requires fewer frequency bins, and the signal variation is more homogeneous across bins. (c) Multibanded

frequency domain partitions for LVK ( $f_{\min} = 19.4$  Hz, compression factor  $\approx 60$ ) and CE ( $f_{\min} = 5$  Hz, compression factor  $\approx 650$ ) experiments. We use a smaller chirp mass prior for the CE experiments (Extended Data Table 2), which allows a slightly coarser resolution compared to LVK (corresponding to lower  $\hat{f}_i$ ). The first two bands for CE are skipped entirely, which is a consequence of the reduced signal variation with heterodyning.

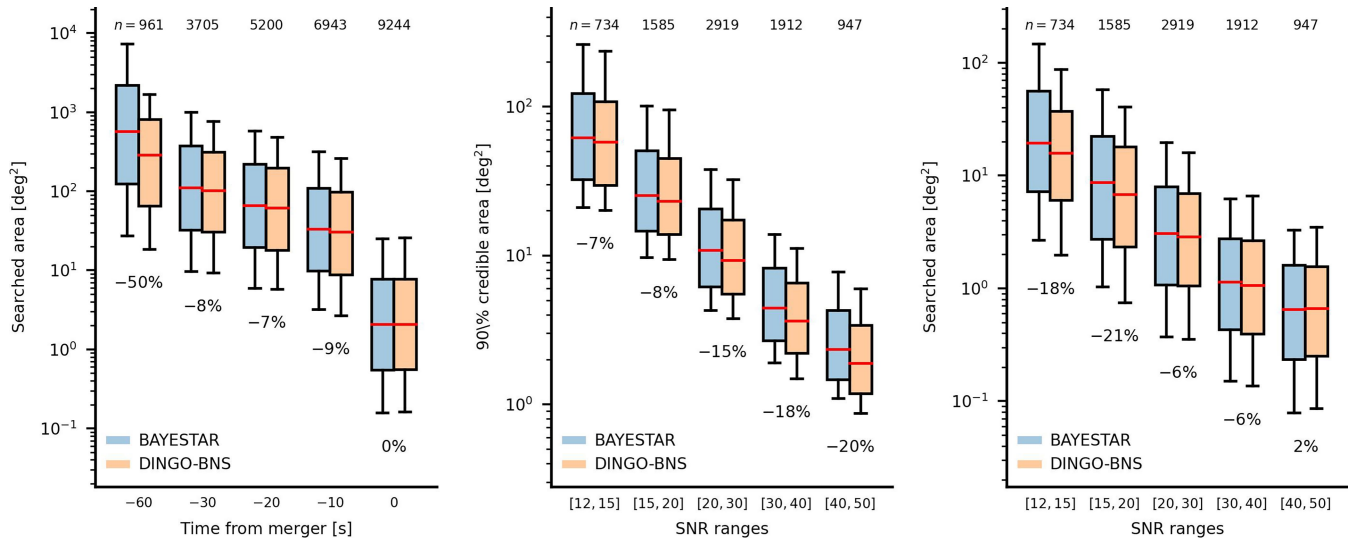


**Extended Data Fig. 3 | Time-domain truncation of BNS signals.** Truncation at time  $t_{\max}$  (red dashed line) before the merger can be approximated by truncation at a corresponding maximum frequency (red solid line) in frequency domain. Below frequency  $f_{\max}(t, \mathcal{M})$ , the truncated signal (blue in center panel) matches the original signal (gray). Above  $f_{\max}(t, \mathcal{M})$ , the amplitude of the truncated signal quickly approaches zero. We determine  $f_{\max}(t, \mathcal{M})$  empirically, by allowing mismatches between truncated and original signals of at most  $10^{-3}$  (lower panel). Analogously, truncation for  $t < t_{\min}$  can be achieved by imposing a minimum frequency cutoff  $f_{\min}(t, \mathcal{M})$ .



**Extended Data Fig. 4 | Neutron-star equation-of-state likelihood.**

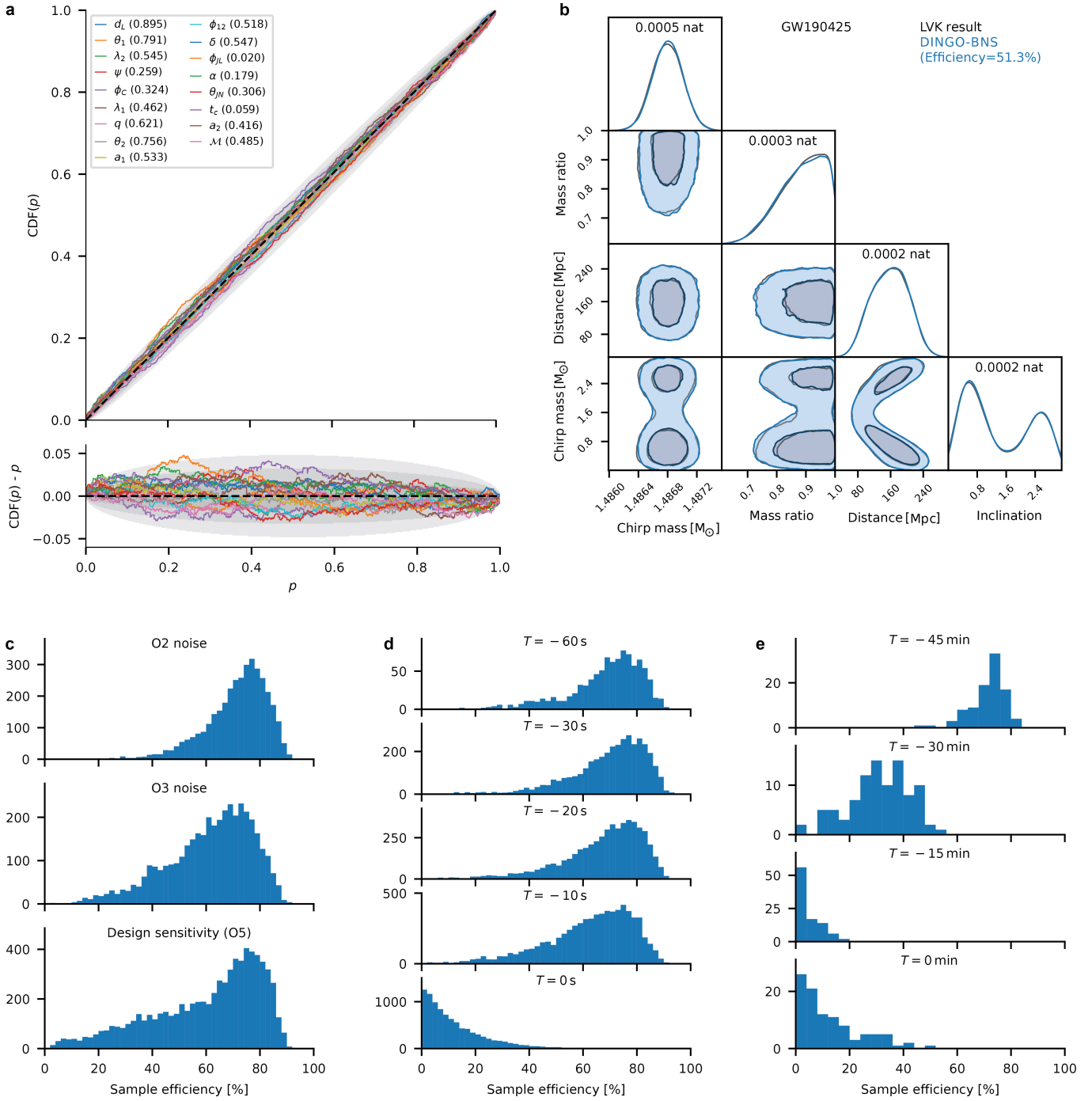
Neutron-star EOS imply a functional relation  $\lambda_i = \lambda_i(m_i)$  between tidal parameters  $\lambda_i$  and component masses  $m_i$ . The likelihood  $p(\mathbf{d}|\mathcal{E})$  for an EOS  $\mathcal{E}$  given the GW data  $\mathbf{d}$  requires integrating the posterior  $p(\boldsymbol{\theta}|\mathbf{d})$  along the corresponding hyperplane. No posterior samples (gray) will be exactly on such hyperplanes (coloured lines), hence the standard Bayesian inference techniques are not directly applicable<sup>87</sup>. DINGO-BNS provides various possibilities to directly compute this quantity, enabling comparison of different EOS in terms of the Bayes factors  $\mathcal{B}$ .



**Extended Data Fig. 5 | Localization comparison between BAYESTAR and DINGO-BNS.** We evaluate the 90% credible area and the searched area. The boxplots display the medians (percentage changes indicated), quartiles and

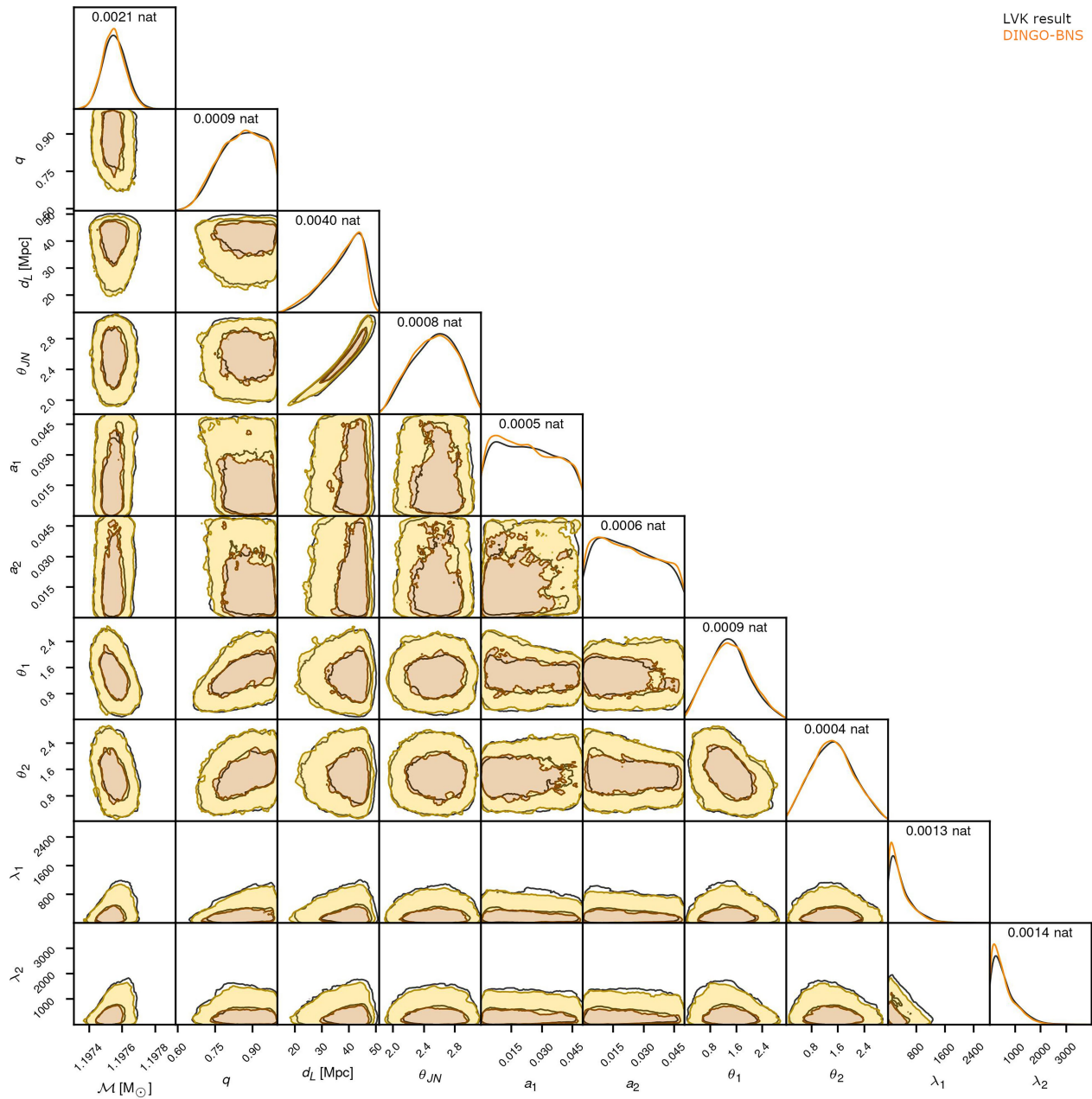
10th/90th percentiles. The comparisons according to SNR are based only on results after the merger.





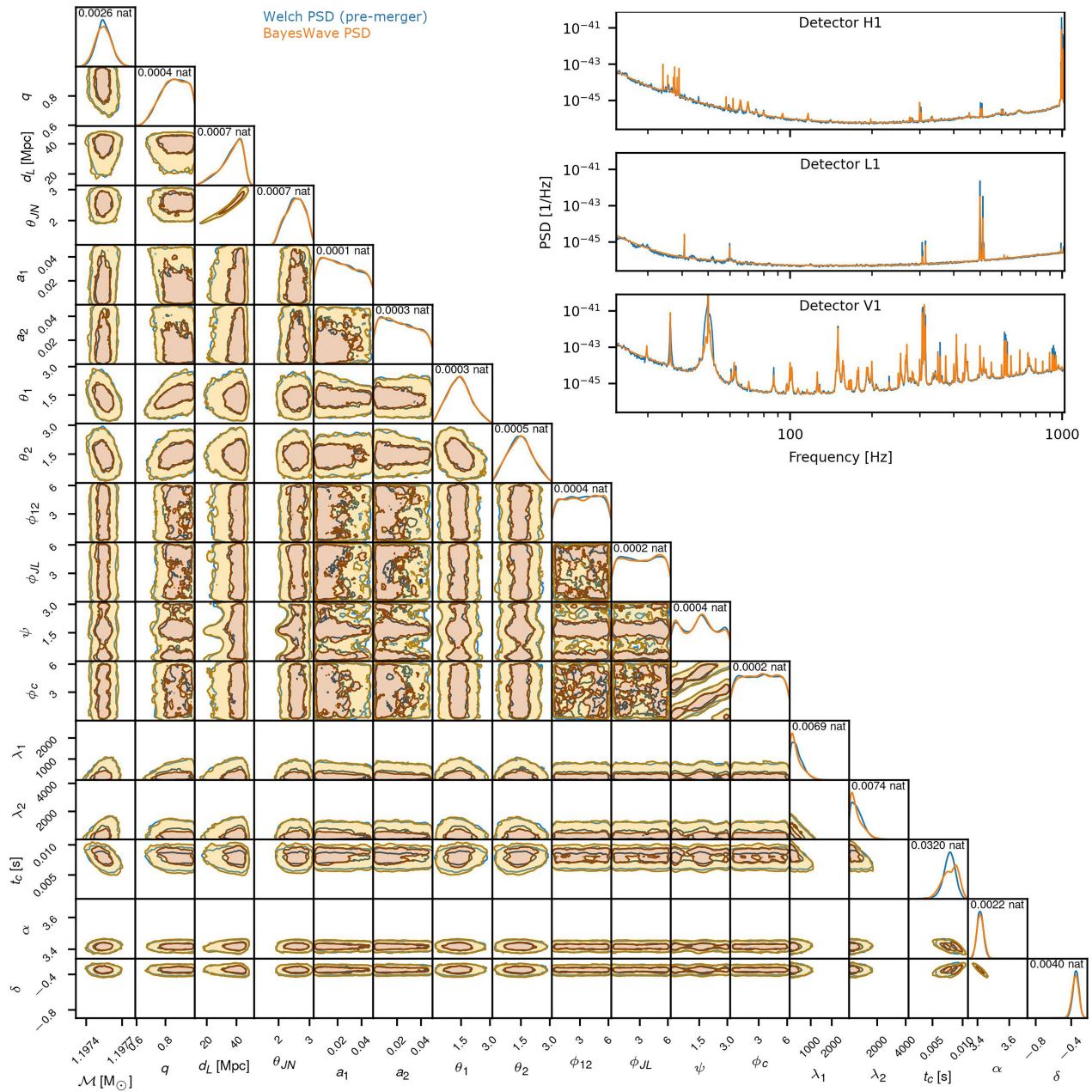
**Extended Data Fig. 6 | DINGO-BNS performance tests.** (a) P-P plot for 1000 simulated GW datasets. For each simulation and parameter, we compute the percentile  $p$  of the injected parameter under the corresponding one-dimensional marginal posterior inferred with DINGO-BNS. The plot shows the cumulative distribution function (CDF) of  $p$ . For an unbiased sampler we expect  $\text{CDF}(p) = p$  within statistical uncertainties. Gray bands indicate confidence intervals corresponding to one, two and three standard deviations. The legend displays the  $p$ -value for each parameter, the combined  $p$ -value is 0.37. The GW datasets for this plot are generated with the settings from the after-merger analysis of the second experiment, except for the luminosity distance prior,

which for this statistical test needs to be set to follow the uniform training prior (Extended Data Table 2). (b) Posterior for GW190425. 2D marginals are displayed in terms of 50% and 90% credible regions. For the 1D marginals, the deviation between the two results is quantified in terms of the Jensen-Shannon divergence (numbers on the diagonal). DINGO-BNS shows good agreement with the public LVK result<sup>79</sup>. (c)–(e) Sample efficiencies for DINGO-BNS, for (c) GW170817-like injections using different detector noise levels, (d) injections using LVK design sensitivity PSDs and (e) injections using CEPDs. Panel b adapted from ref. 79 under a Creative Commons licence CC BY 4.0.



**Extended Data Fig. 7 | Extended posterior for GW170817.** The corner plot compares the DINGO-BNS result from the main text (orange) with the LVK result (black)<sup>5</sup>. 2D marginals are displayed in terms of 50% and 90% credible regions. For the 1D marginals, the deviation between the two results is quantified in terms of the Jensen–Shannon divergence (numbers on the diagonal). For the

definition of the parameters see Extended Data Table 2. The small deviations between the results could be a consequence of the marginalization over detector calibration uncertainties in the LVK result, which is not applied with DINGO-BNS.



**Extended Data Fig. 8 | GW170817 posterior with different PSDs.** The corner plot displays DINGO-BNS inference results using the BayesWave PSD (orange) from the LVK analysis and a Welch PSD (blue). 2D marginals are displayed in terms of 50% and 90% credible regions. For the 1D marginals, the deviation between the two results is quantified in terms of the Jensen–Shannon divergence (numbers on the diagonal). For the definition of the parameters see Extended Data Table 2. The Welch PSD is estimated based on strain data

taken 1184 to 160 s before the BNS merger (using 128 segments of length eight seconds and subsequent upsampling to the desired resolution  $\Delta f = 1/128$  Hz), and is thus available in low latency for pre-merger analysis. The DINGO-BNS network for the Welch analysis is trained with a distribution of Welch PSDs covering the entire second LVK observing run, and tuned to the GW170817 PSDs via network conditioning<sup>19</sup>.

# Article

## Extended Data Table 1 | Comparison of fast methods for BNS inference

Method	Inference time	Comment
BAYESTAR <sup>16</sup>	10–60 s	Uses SNR time series. Localization only.
GW-SkyLocator <sup>66</sup>	$\ll$ 1 s	Uses SNR time series. Localization only.
Ref. 67	1 s	Only infers six parameters. Accuracy not validated by comparison to other samplers. Not tested with real data. No tuning to variable noise levels.
BILBY-DYNESTY with ROQ <sup>17</sup>	108 min	Localization results available in 6 min.
i-nessai <sup>39</sup>	24 min	
JIM <sup>33,35</sup>	21–33 min	Faster inference achieved when using simpler waveform models <sup>33</sup> .
DINGO-BNS	1 s	

The four methods in the lower block all perform full parameter estimation and their accuracy on real data has been validated by comparison against established samplers.

## Extended Data Table 2 | Training priors

	LVK	CE
$\mathcal{M}$ [ $M_{\odot}$ ]	[1.0, 2.2]	[1.15, 1.25]
$m_1$ [ $M_{\odot}$ ]	[1.0, 3.2]	[0.95, 2.4]
$m_2$ [ $M_{\odot}$ ]	[1.0, 2.0]	[0.95, 2.4]
$a_{1,2}$	[0, 0.05]	[0, 0.05]
$\lambda_1$	[0, 5000]	[0, 5000]
$\lambda_2$	[0, 10000]	[0, 10000]
$d_l$ [Mpc]	[10, 100]	[20, 50] / [1000, 2000]
$t_c$ [s]	[-0.3, 0.3] / [-0.03, 0.03]	[-1, 1] / [-0.03, 0.03]

Parameters include chirp mass  $\mathcal{M}$ , component masses  $m_{1,2}$ , spin magnitudes  $a_{1,2}$ , tidal deformabilities  $\lambda_{1,2}$ , luminosity distance  $d_l$  and merger time  $t_c$ . All priors are uniform, except for chirp mass, which is sampled uniform in component masses. At inference,  $d_l$  can be reweighted to the standard prior (uniform in comoving volume). For  $t_c$  we use a broader prior for pre-merger inference than for full inference (separated by “/” symbol) to account for higher uncertainties. We have verified that the  $t_c$  priors are sufficiently broad for the data analyzed here, although broader priors may be required for analyzing events with lower SNR or earlier pre-merger times. LVK priors are chosen to cover expected LVK BNS detections. CE priors for  $\mathcal{M}$  and  $d_l$  are reduced compared to LVK to decrease the computational cost of training. Priors for parameters not displayed here (sky position  $(\alpha, \delta)$ , spin angles  $(\theta_1, \theta_2, \Phi_{12}, \Phi_{1L})$ , inclination  $\theta_{in}$ , polarization angle  $\psi$  and reference phase  $\Phi_c$ ) are standard. The prior for mass ratio  $q$  is determined via the priors for  $\mathcal{M}$  and  $m_{1,2}$ .