

What to do when things get crowded? Scalable joint analysis of overlapping gravitational wave signals

James Alvey^{†, 1,*} Uddipta Bhardwaj^{‡, 2, †}
Samaya Nissanke,^{2, 3} and Christoph Weniger¹

¹GRAPPA Institute, Institute for Theoretical Physics Amsterdam,

University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

²GRAPPA Institute, Anton Pannekoek Institute for Astronomy and Institute of High-Energy Physics,

University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

³Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands

The gravitational wave sky is starting to become very crowded, with the fourth science run (O4) at LIGO expected to detect $\mathcal{O}(100)$ compact object coalescence signals. Data analysis issues start to arise as we look further forwards, however. In particular, as the event rate increases in *e.g.* next generation detectors, it will become increasingly likely that signals arrive in the detector coincidentally, eventually becoming the dominant source class. It is known that current analysis pipelines will struggle to deal with this scenario, predominantly due to the scaling of traditional methods such as Monte Carlo Markov Chains and nested sampling, where the time difference between analysing a single signal and multiple can be as significant as days to months. In this work, we argue that sequential simulation-based inference methods can solve this problem by breaking the scaling behaviour. Specifically, we apply an algorithm known as (truncated marginal) neural ratio estimation (TMNRE), implemented in the code `peregrine` and based on `swyft`. To demonstrate its applicability, we consider three case studies comprising two overlapping, spinning, and precessing binary black hole systems with merger times separated by 0.05 s, 0.2 s, and 0.5 s. We show for the first time that we can recover, with full precision (as quantified by a comparison to the analysis of each signal independently), the posterior distributions of all 30 model parameters in a full joint analysis. Crucially, we achieve this with only $\sim 15\%$ of the waveform evaluations that would be needed to analyse even a single signal with traditional methods.

GitHub: The `peregrine` analysis and inference library is available [here](#) (`peregrine-gw/peregrine`). In addition, the TMNRE implementation `swyft` is available [here](#) (`undark-lab/swyft`).

[†]These two authors contributed equally to this work.

Introduction.— The advent of gravitational wave (GW) detection [1] has already revolutionized our understanding of the Universe, providing a powerful tool for studying theories of gravity [2], astrophysics [3, 4], and cosmology [5–7]. Up to now, over 90 compact binary coalescences (CBCs) have been detected globally, and live observations by the LIGO-Virgo-Kagra Collaboration (LVKC) in its fourth observing run currently add new observations on an almost daily basis [8–13]. Looking forwards, future enhancements to the detector network include the introduction of KAGRA [14, 15], LIGO-India [16, 17], and the proposed LIGO-voyager [18], alongside the next-generation detectors such as the Einstein Telescope (ET) [19], Cosmic Explorer (CE) [20], and the space-based Laser Interferometer Space Antenna (LISA) [21]. Combined, these promise to significantly expand our coverage of the gravitational wave landscape, both in terms of number and distinct class of event, as we probe new cosmic distances and frequency ranges.

In this context, however, the improved sensitivity and expanded coverage of these advanced detectors brings forth a number of concrete data analysis challenges. For example, these include the analysis of long signals from *e.g.* large mass ratio inspirals, the impact of non-stationary or unknown detector noise characterisation on parameter inference, or the global analysis of a (set of)

stochastic background and transient signals (see *e.g.* [22–30]). In this work, we will focus specifically on an additional problem: the analysis of overlapping binary coalescence signals within a single detection window [31–35]. Although this issue is not necessarily directly relevant to the current LVKC observing run (although it is possible given the detector upgrades that such an event could be observed in O4/5), the presence of large numbers ($\mathcal{O}(1000)s$) of overlapping signals in future detector networks is overwhelmingly likely [36]. More precisely, current estimates of merger rates [3] have been used to predict the probability of detecting concurrent CBCs in multiple current and next generation configurations, see *e.g.* [36, 37]. These references have reported that observing overlapping CBCs in a 3G detector network is near certain. Another way to state the problem is the following: given current merger rate estimates, suppose an event is above the detectable threshold in a next generation detector network, the probability that there is an additional overlapping source in the same detection window is effectively 100%.

The presence of this dominant class of overlapping signals presents a significant challenge for the applicability of current parameter inference pipelines (*e.g.* [38–41]) in terms of both computational feasibility, as well as statistical robustness. There are a number of current strategies

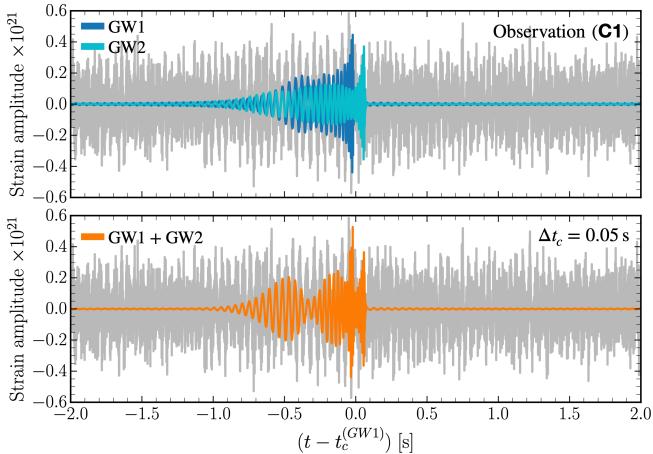


FIG. 1. Time domain representation of two spinning, precessing BBH gravitational wave signals overlapping with $\Delta t_c = 0.05$ s. Depicting one of our case studies (**C1**), the top panel shows the individual component signals **GW1** (with $M_1 = 42.2 M_\odot$, $M_2 = 32.5 M_\odot$) and **GW2** (with $M_1 = 34.2 M_\odot$, $M_2 = 27.7 M_\odot$), both at a network signal-to-noise ratio of 20. The signals are chronologically ordered according to merger times in LIGO-Hanford detector noise (grey). The bottom panel shows the overlapping signal (orange) resulting from the interference of **GW1** and **GW2**.

proposed in the literature as to how to approach this challenge. One possible method is to assume that exactly one of the signals is dominant, and then apply current parameter inference techniques such as Monte Carlo Markov Chains (MCMC) [42, 43] or nested sampling [44–47] with an individual waveform template. However, this has been shown to induce significant biases in the parameter reconstruction due to the effective mismodeling of the underlying noise realisation,¹ especially when the signals have similar merger times and/or frequency evolution.² This motivated additional efforts to develop joint parameter estimation techniques through *e.g.* modifications to the single waveform likelihood function to account for multiple overlapping CBCs (see *e.g.* Ref. [34]), or working directly with the joint likelihood.

At a fundamental level, however, since traditional sampling algorithms such as MCMC and nested sampling generate samples directly from the full joint posterior, these algorithms tend to scale poorly as the dimensionality of the problem increases. For example, whilst an analysis of a single signal LVKC event may take $\mathcal{O}(1)$ day with state of the art waveform templates, it has been shown that analysing even two waveforms could take

$\mathcal{O}(1)$ month [34]³. This renders the implementation of such techniques for the potential $\mathcal{O}(10^5)$ events [36] in a single observing run with 3G detectors computationally infeasible. It may be possible to improve this outlook with efficient high performance computing techniques or via the implementation of rapid waveform modelling techniques [48], fast stochastic samplers [49] and relative binning [50, 51].

Regardless, these challenges motivate the exploration of alternative data analysis strategies. In this work, we will argue for the benefits of modern simulation-based inference (SBI) techniques (see *e.g.* Ref. [52] for a review) in analysing overlapping gravitational wave signals. In general, SBI methods solve the Bayesian inference problem of performing parameter inference on a set of model parameters θ given some observation \mathbf{x}_0 by forward modelling the data \mathbf{x} as a function of θ . For many applications (see *e.g.* Refs. [53–61]), this crucially does not require the evaluation (or specification) of an explicit likelihood, although there are other benefits that are also relevant to GW studies such as simulation efficiency, amortized inference, and the ability to do marginal inference on subsets of a model. Particularly for GW parameter inference, an implementation of SBI called neural posterior estimation (NPE) has been shown to impressively perform fully amortized inference on compact binary mergers [62–64]. In the context of overlapping gravitational waves analysis, a normalizing flow-based approach has shown promising results [35] by obtaining wide, but (crucially) more accurate posteriors as compared to traditional likelihood-based approaches. Currently, this represents the state-of-the-art for parameter inference of overlapping gravitational wave signals.

In the current work, we will make use of another SBI algorithm, known as (truncated marginal) neural ratio estimation, TMNRE [53] to analyse overlapping GW signals. NRE functions by estimating the posterior-to-prior ratio via a binary classification task, see Refs. [53, 65] for more information. Recently, this was implemented for the first time for the analysis of individual compact binary mergers in the code **peregrine** [66], built on top of **swyft** [53, 65]. There are a number of key results from Ref. [66] that suggest that this approach could solve the challenges facing overlapping analysis. Specifically, it was shown that **peregrine** could be used to analyse individual signals with only 2% of the waveform evaluations required in a traditional approach such as nested sampling. The fundamental reason for why our approach is more simulation efficient is the fact that TMNRE focuses sequentially on lower dimensional marginal posteriors. In more detail, directly estimating (typi-

¹ This is particularly pertinent for next generation detectors where there will be very few periods without signal in-band [33].

² Previous studies using Fisher forecasting have additionally highlighted that these biases are more pronounced for binary black hole (BBH) systems [31], owing to there being fewer in-band cycles compared to binary neutron star (BNS) and neutron star-black hole (NSBH) systems.

³ In Ref. [34], the reported time taken for a joint parameter estimation using **dynesty** [47] with a 30D likelihood function is 23.8 days on 16 Intel(R) Xeon(R) Gold 6152 CPUs.

cally 1D and 2D) marginal posteriors has been shown to be extremely efficient in a number of contexts, see e.g. [56, 58]. The sequential aspect of the algorithm is achieved through a prior truncation scheme that composes well with marginalization [53] carried out over a series of inference rounds, and allows us to quickly zoom in to the relevant region of parameter space for a given observation. For example, the geocentric merger time below (see bottom panel of Fig. 2) is constrained to within $\sim 1\%$ of the initial prior range, so truncating to this region can significantly reduce the data variance. For more details on both of these aspects, see Refs. [53, 65, 66]. In the present context, assuming that the sort of simulation efficiency illustrated in Ref. [66] carries through to the overlapping case – which we will demonstrate below – the inference time could be brought down from $\mathcal{O}(1 \text{ month})$ [34] to $\mathcal{O}(1 \text{ day})$.

In the rest of the work, we will begin by setting up the case study that we will analyse – two overlapping spinning, precessing binary black hole signals that have similar signal-to-noise ratios in a detector with LIGO and Virgo O4 (current) design sensitivity. Then, we will demonstrate that we can achieve full precision and accuracy in the parameter inference across the full 30-dimensional parameter space using the inference results for the individual signals as an optimality test. Finally, we provide some discussions and conclusions regarding our approach and results. Additional technical details regarding the **peregrine** implementation can be found in either the *Supplemental Material* below, or in Ref. [66].

An Overlapping Case Study.— To explore the challenges posed by overlapping signals, we have designed three case studies with varying degrees of overlap as specified by the merger times, set at $\Delta t_c = 0.05 \text{ s}$ (**C1**), 0.2 s (**C2**) and 0.5 s (**C3**). All of these choices fall within the high-bias regime, as defined in *e.g.* [31]. In each case study, we select two spinning, precessing binary black hole (BBH) signals with similar mass ratios and chirp masses. This choice ensures that the frequency evolution of the sources is comparable, allowing for extended periods of overlap in both time and frequency domains. Moreover, we order the mergers such that the lighter mass binary merges after the heavier mass binary, aiming at maximum overlap of their corresponding signals. Furthermore, in order to eliminate the dominance of a single signal, we set the network signal-to-noise ratio (SNR) of the individual component signals to be the same at a value of 20. This is achieved by appropriately scaling the luminosity distances of the signals once we have fixed the underlying noise realization. By ensuring similar SNRs in addition to the aforementioned setup, we simulate the most challenging scenario for parameter estimation of this specific class of signals. We provide an example observation in Fig. 1, which corresponds to one of our case studies (**C1**). The figure shows overlapping GW signals in the time do-

main with a merger time separation of $\Delta t_c = 0.05 \text{ s}$. More specifically, the signals are generated using source parameters detailed in Tab. S1 (*Supplemental Material*) and injected onto a three-detector network comprising of the two LIGO detectors and Virgo.⁴

As validation of our results, we perform a separate analysis of the individual component signals using **peregrine** and the approach detailed in Ref. [66], where it was shown that the TMNRE approach [53, 65] is highly efficient in achieving excellent agreement with *e.g.* MCMC. This allows us to refer to the parameter estimation performance for non-overlapping signals and establish a baseline for comparison. This is an important step forward in terms of the state-of-the-art, where we look to move towards full precision as opposed to simply unbiased coverage, as was impressively demonstrated in Ref. [34]. We emphasize that the comparison to the full overlapping analysis should not be interpreted in terms of “superimposed” posteriors but rather as a means to understand the loss of precision introduced by signal overlap and the corresponding reduction in SNR.

For data generation, we consider a network comprising three detectors: the two LIGO detectors (Hanford and Livingston) and Virgo⁵, all operating at the sensitivity level of the O4 observing run with a minimum frequency of 20 Hz and a sampling rate of 2048 Hz. To simulate detector noise, we use the LIGO and Virgo O4 power spectral density (PSD) curve [67, 68]. The GW signals in our simulations are evaluated using the state-of-the-art **IMRPhenomXPHM** [69] waveform approximant. This waveform model is chosen due to its ability to accurately account for higher-order modes and precession effects⁶. It is worth noting that alternative waveform approximants, such as ones from the **SEOBNR** (*e.g.* [72–74]) and **IMRPhenom** (*e.g.* [69, 75, 76]) families, can also be readily employed in our setup depending on the specific requirements and objectives of the analysis.

Key Results with peregrine.— All of the key results in this work are in the context of the case studies described above. To summarise them briefly, our main discussion points are highlighted in Fig. 2 and Fig. 3. They illustrate two general points: our ability to achieve high precision and accuracy in our reconstruction of the posteriors for

⁴ We note that for next generation detectors, the substantial increase in bandwidth will lead to a significantly higher number of cycles in-band for such signals [19, 20], which may improve the outlook.

⁵ In the current analysis, we do not consider the KAGRA detector, since the sources are at the edge of its projected sensitivity range [14].

⁶ Studies involving harmonic decomposition of precessing waveforms have demonstrated how the effects of precession on the gravitational wave signal can mimic the beating of two waveforms overlapping and lead to degeneracies in parameter inference [70, 71].

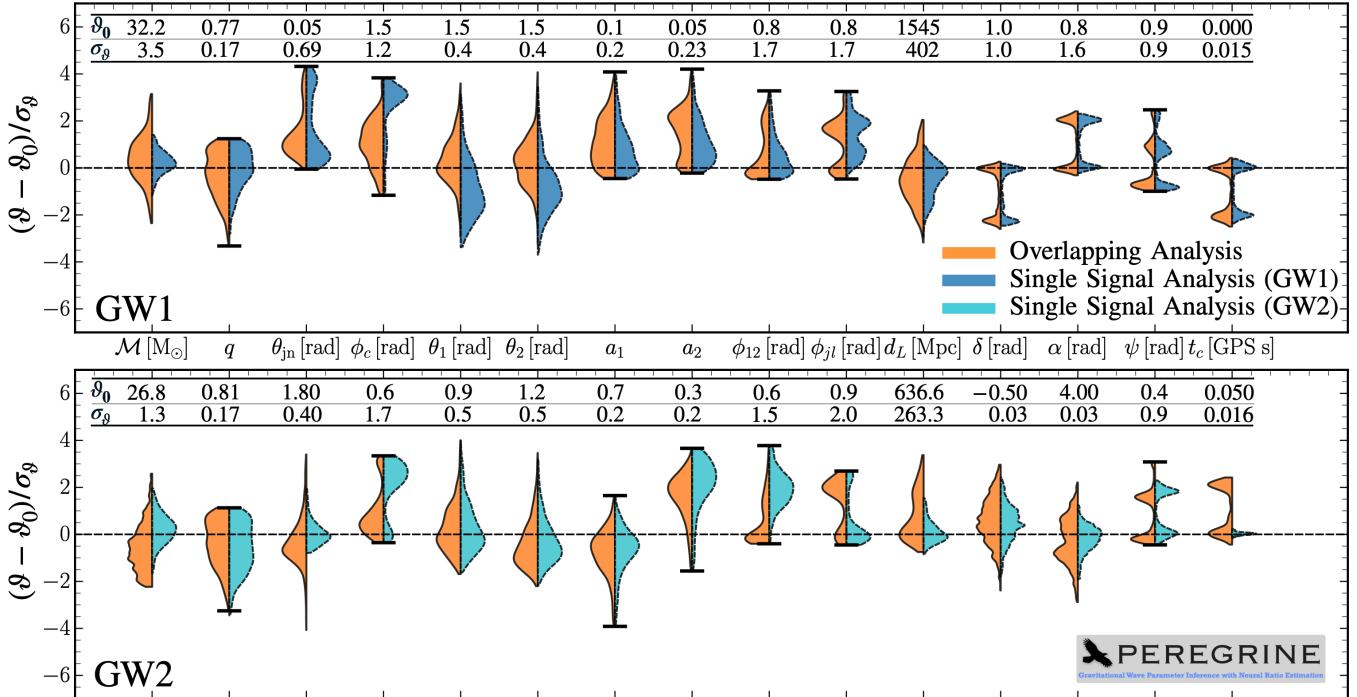


FIG. 2. Violin plots showing 1D marginal posterior distributions of all 30 parameters characterising an overlapping signal comprising two concurrent BBH signals with $\Delta t_c = 0.05$ s (as shown in Fig. 1). The left of each violin shows the 1D marginal posterior obtained from our overlapping signal analysis (C1). The corresponding distributions on the right of each violin represent the posterior distributions obtained via a single signal analysis of each component signal in the absence of the other waveform. The top and bottom panels indicate parameters of the first (**GW1**) and second (**GW2**) signal respectively. To aid comparison across parameters, all 1D posterior samples are scaled such that the injected value, ϑ_0 is at 0 (indicated by the dashed horizontal line), and normalized by the standard deviations of the overlapping posterior samples, σ_ϑ . When shown, black horizontal lines at the ends of the posteriors indicate the (physical) edges of the prior boundary, *e.g.* for the mass ratio q , the limits are at $q = 0, 1$. When not shown, this means that the posterior lies well within the initial prior range.

all 30 model parameters (15 for each signal), as well as our ability to investigate correlations between the two signals and the implications regarding the separation of the two waveforms in terms of precision and coverage. We discuss these points in more detail below, as well as the additional validation tests that we have carried out to support our results, and the computational efficiency of our algorithm in the present context. Full sets of detailed posteriors and coverage tests can be found in the *Supplemental Material* (see Figs. S1–S8).

The first result to discuss in detail is summarised in Fig. 2. This highlights the comparison⁷ between the 1D posteriors for all model parameters from our joint analysis of the overlapping signal in case study (C1, $\Delta t_c = 0.05$ s) as compared to the result for the corresponding individual signal. There are a number of things to comment on: firstly, we see that we achieve excellent coverage of the true injection value, represented by the

horizontal dashed line. Secondly, and arguably the most important result of this work, we achieve remarkably similar precision compared to the corresponding analysis of the individual signal.⁸ The posteriors are slightly broader in the overlapping analysis (see *e.g.* the detector-frame chirp mass M for an illustrative example), which is to be expected since the presence of an additional signal must act to slightly reduce the overall SNR of the waveform under consideration. Nonetheless, we show for the first time that it is possible to obtain accurate parameter constraints with little to no loss of information. This should be compared to *e.g.* the results in Ref. [34] which achieved impressive unbiased coverage, but at the cost of significantly broader posteriors. For completeness, we comment on the one posterior that shows a distinct difference in sensitivity compared to our individual signal analysis baseline. Specifically, if one looks at the posteriors for the merger time of the second signal in the bottom panel of Fig. 2, we see that the overlapping analysis

⁷ The figure shows each 1D marginal posterior rescaled with respect to its injection value ϑ_0 and normalized by the standard deviation of the overlapping posterior samples σ_ϑ to ease comparison across parameters.

⁸ The single signal analysis here really acts as a sensitivity baseline, *i.e.* it should not be possible to exceed the constraining power of the analysis in the absence of the second signal.

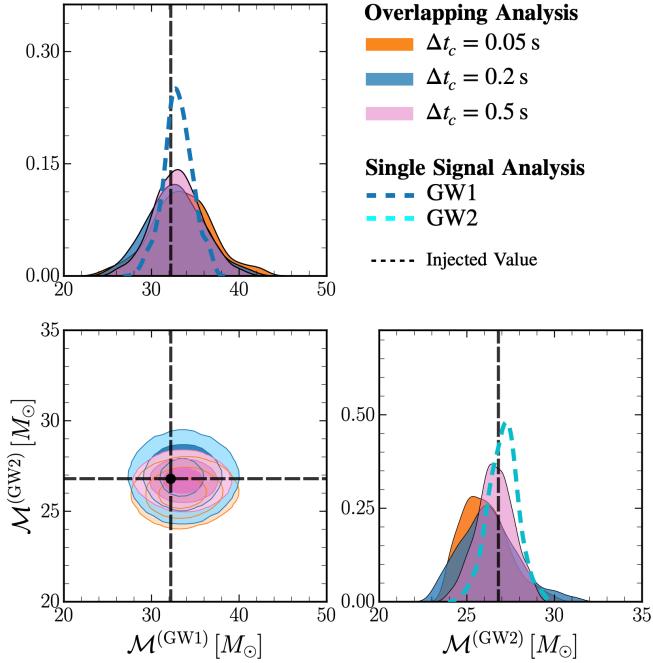


FIG. 3. Corner plot showing the 1D marginal posteriors for the chirp masses of the two BBH signals overlapping with three different degrees of time-overlap ($\Delta t_c = 0.05 \text{ s}, 0.2 \text{ s}, 0.5 \text{ s}$ as shown in the figure legend). The 2D marginal posteriors are shown with $1\sigma, 2\sigma$ and 3σ contours indicated by subsequently lighter shading. The dashed 1D marginal posteriors indicate the maximum precision and accuracy with which one can expect to constrain the parameters, as determined by analysing the individual signals.

results in a bimodal, and broader posterior on t_c . We believe this is fully explained by looking at the corresponding merger time of the first signal (**GW1**). The presence of two distinct modes (also evident in the individual-signal analysis) leads to a direct degeneracy between the merger time of the first and second signal that is fundamentally not present when analysing only the latter (**GW2**). Regardless, it is still extremely well measured as compared to the initial prior range (see Tab. S1).

We now focus attention on Fig. 3 which presents the 1D and 2D posteriors for the chirp masses of each source (**GW1** and **GW2**) for all three case studies. We see again the level of precision that we achieve compared to the individual signal analysis, now compared across the different merger time differences. The general trend is clear that there is a slight, but noticeable, improvement in sensitivity on the chirp masses as the signals are separated. This is more prevalent in the case of **GW2**, where indeed for $\Delta t_c = 0.5 \text{ s}$ it actually matches that of the individual signal. This brings up an interesting discussion regarding the impact of the underlying noise realisation on the precise position and shape of the posterior. Specifically, one might wonder why we do not exactly recover the posterior of the individual signal. We believe that the

slight relative shift is a result of the (sub-dominant) presence of the ringdown of the first signal close to the merger of the second signal which effectively modifies slightly the underlying noise realisation. The result of this should exactly be a small observed shift in the posterior. The second distinct point to make about these results is the behaviour of the 2D contours. We see that, as expected, they shrink as the merger times move further apart, but also that the two chirp masses are broadly uncorrelated. This is expected physically since the signals are totally distinct, however, it is also highlighting the fact that this measurement is not noise-dominated which could lead to induced degeneracies between the parameters. We believe this is the first time this has been explicitly shown for this type of overlapping analysis, and paints a hopeful picture for our ability to obtain high quality reconstructions of coincident signals in future detectors.

Although the individual signal analysis is a strong indication that our method is performing well, it is important to carry out additional validation tests. Coverage tests, see *e.g.* [77, 78] are now standard practice in simulation-based inference applications. The full outcome of these tests are shown in the *Supplemental Material* in Figs. S6–S8. The takeaway message from our analysis is that for all parameters, across all case studies, our posteriors are extremely well calibrated (lie along the diagonal as plotted in the (empirical coverage)-(expected coverage) plane).

The final important discussion regarding our analysis concerns computational efficiency. It has been explicitly shown that using traditional methods to carry out this type of joint inference is extremely costly, taking on average over 3 weeks to run (see Footnote 3 in Ref. [34]). Taking our analysis of the first case study (**C1**) as a concrete example, we will make a direct comparison with this state-of-the-art joint inference benchmark. Across 7 sequential rounds of inference using **peregrine** (for details regarding our TMNRE implementation see Ref. [66]), we required only 7×10^6 overlapping waveform simulations to achieve our analysis results. This is only a factor of 10 more than we required to perform analysis on a single signal in Ref. [66] for twice as many parameters. This therefore convincingly breaks the expected scaling of traditional methods solving this joint inference problem, and is an order of magnitude fewer waveform evaluations than is typically required to analyse even a single signal with traditional methods such as MCMC. In terms of time, we used an 18-core CPU cluster node to generate the 7×10^6 waveform evaluations, which took on average around 2-3 hours per round of 1 million examples. Training the network each round is currently the main time cost of the analysis, taking between 8 and 12 hours per round for 10^6 simulations. Overall, this means our analysis takes around 48 hours to complete, depending on the specific

hardware choice.⁹ For more details on the network specification and training procedure, please see the section in the *Supplemental Material*.

Conclusions and Outlook.— In this work, we have presented the application of the sequential simulation-based inference method in Ref. [66] to the analysis of overlapping gravitational wave signals. We showed for the first time that we can obtain full precision and accuracy over all spinning, precessing binary black hole parameters in a joint inference analysis. Our main results are presented in Figs. 2 and 3 for the three case studies described above. Crucially, to do so, we required only 7×10^6 waveform evaluations, which is an order of magnitude fewer than traditional methods typically require to analyse *even a single signal*.

As far as outlook is concerned, there are a number of important points to make regarding our method. The first is scalability: we have highlighted that we can perform full inference on two overlapping signals with a computational cost and simulation efficiency that is at least an order of magnitude lower than the current state-of-the-art joint inference methods (see *e.g.* Refs. [31–35]). There is significant room for improvement in this direction, for example, we could use pre-trained networks from earlier inference rounds and fine-tune them to significantly accelerate our training, utilise accelerated waveform models [48], or implement a more efficient truncation scheme that accounts for parameter correlations. Secondly, if we look towards the science case of future experiments such as the Einstein Telescope [19], Cosmic Explorer [20], and LISA [21], overlapping signals (of various types) will be a dominant source class. In order to use our joint analysis framework for an entirely realistic next generation gravitational wave scenario, natural development is required in the sense of using 3G interferometer setups, considering multiple (*i.e.* more than 2) signal overlaps, carrying out the initial triggering of these overlapping events, and dealing with very long signals (for *e.g.* next generation detection of binary neutron star signals). Another class of analysis problems in the future will be the characterisation of detector noise, since almost all of the data is expected to contain some signal component. This is in addition to the fact that the noise may potentially be auto-correlated. It will be interesting to investigate the simulation-efficiency achieved in such a next generation detector (ET/CE) setting. Partly as a result of the implicit likelihood nature of `peregrine` [66] as a method, combined with its scalability, we believe it offers a flexible starting point for this development towards a data analysis pipeline suitable for the next generation of gravitational wave detectors.

⁹ Specifically, we ran our waveform simulations on 18 (shared) *Intel(R) Xeon(R) Platinum 8360Y* CPUs and network training on a single, shared *NVIDIA A100* graphics card.

ACKNOWLEDGEMENTS

This work is part of the project CORTEX (NWA.1160.18.316) of the research programme NWA-ORC which is (partly) financed by the Dutch Research Council (NWO). Additionally, CW acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 864035). UB is supported through the CORTEX project of the NWA-ORC with project number NWA.1160.18.316 which is partly financed by the Dutch Research Council (NWO). JA is supported through the research program “The Hidden Universe of Weakly Interacting Particles” with project number 680.92.18.03 (NWO Vrije Programma), which is partly financed by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Dutch Research Council). SN acknowledges support from the NWO Projectruimte grant (Samaya Nissanke). Additionally, CW acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 864035). We thank Thomas Edwards and Pablo Bosch for useful conversations and comments on a draft of this manuscript. The main analysis for this work was carried out on the Snellius Computing Cluster at SURFsara.

* j.b.g.alvey@uva.nl; ORCID: 0000-0003-2020-0803

† u.bhardwaj@uva.nl; ORCID: 0000-0003-1233-4174

- [1] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **116**, 061102 (2016), [arXiv:1602.03837 \[gr-qc\]](https://arxiv.org/abs/1602.03837).
- [2] R. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. D* **103**, 122002 (2021), [arXiv:2010.14529 \[gr-qc\]](https://arxiv.org/abs/2010.14529).
- [3] R. Abbott *et al.* (LIGO Scientific, Virgo), *Astrophys. J. Lett.* **913**, L7 (2021), [arXiv:2010.14533 \[astro-ph.HE\]](https://arxiv.org/abs/2010.14533).
- [4] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **121**, 161101 (2018), [arXiv:1805.11581 \[gr-qc\]](https://arxiv.org/abs/1805.11581).
- [5] B. P. Abbott *et al.* (LIGO Scientific, Virgo, VIRGO), *Astrophys. J.* **909**, 218 (2021), [arXiv:1908.06060 \[astro-ph.CO\]](https://arxiv.org/abs/1908.06060).
- [6] S. Nissanke, D. E. Holz, S. A. Hughes, N. Dalal, and J. L. Sievers, *Astrophys. J.* **725**, 496 (2010), [arXiv:0904.1017 \[astro-ph.CO\]](https://arxiv.org/abs/0904.1017).
- [7] S. Nissanke, J. Sievers, N. Dalal, and D. Holz, *Astrophys. J.* **739**, 99 (2011), [arXiv:1105.3184 \[astro-ph.CO\]](https://arxiv.org/abs/1105.3184).
- [8] R. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. X* **11**, 021053 (2021), [arXiv:2010.14527 \[gr-qc\]](https://arxiv.org/abs/2010.14527).
- [9] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), [arXiv:2111.03606 \[gr-qc\]](https://arxiv.org/abs/2111.03606).
- [10] R. Abbott *et al.* (LIGO Scientific, VIRGO), (2021), [arXiv:2108.01045 \[gr-qc\]](https://arxiv.org/abs/2108.01045).
- [11] A. H. Nitz, C. D. Capano, S. Kumar, Y.-F. Wang, S. Kastha, M. Schäfer, R. Dhurkunde, and M. Cabero, *Astrophys. J.* **922**, 76 (2021), [arXiv:2105.09151 \[astro-ph.HE\]](https://arxiv.org/abs/2105.09151).

- [12] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, *Phys. Rev. D* **100**, 023011 (2019).
- [13] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, *Phys. Rev. D* **101**, 083030 (2020).
- [14] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, Virgo, VIRGO), *Living Rev. Rel.* **21**, 3 (2018), arXiv:1304.0670 [gr-qc].
- [15] T. Akutsu *et al.* (KAGRA), *PTEP* **2021**, 05A101 (2021), arXiv:2005.05574 [physics.ins-det].
- [16] B. Iyer *et al.*, LIGO-India: Proposal for an interferometric gravitational wave observatory. (2021).
- [17] C. S. Unnikrishnan, *Int. J. Mod. Phys. D* **22**, 1341010 (2013), arXiv:1510.06059 [physics.ins-det].
- [18] R. X. Adhikari *et al.* (LIGO), *Class. Quant. Grav.* **37**, 165003 (2020), arXiv:2001.11173 [astro-ph.IM].
- [19] M. Maggiore *et al.*, *JCAP* **03**, 050 (2020), arXiv:1912.02622 [astro-ph.CO].
- [20] D. Reitze *et al.*, *Bull. Am. Astron. Soc.* **51**, 035 (2019), arXiv:1907.04833 [astro-ph.IM].
- [21] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, *et al.*, (2017), arXiv:1702.00786 [astro-ph.IM].
- [22] T. B. Littenberg and N. J. Cornish, *Phys. Rev. D* **107**, 063004 (2023), arXiv:2301.03673 [gr-qc].
- [23] S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. L. Berry, E. Berti, P. Amaro-Seoane, A. Pettiteau, and A. Klein, *Phys. Rev. D* **95**, 103012 (2017), arXiv:1703.09722 [gr-qc].
- [24] A. Spadaro, R. Buscicchio, D. Vetrugno, A. Klein, D. Gerosa, S. Vitale, R. Dolesi, W. J. Weber, and M. Colpi, (2023), arXiv:2306.03923 [gr-qc].
- [25] C. R. Weaving, L. K. Nuttall, I. W. Harry, S. Wu, and A. Nitz, (2023), arXiv:2306.16429 [astro-ph.IM].
- [26] M. R. Sah and S. Mukherjee, (2023), arXiv:2307.06405 [gr-qc].
- [27] F. Pozzoli, S. Babak, A. Sesana, M. Bonetti, and N. Karnesis, (2023), arXiv:2302.07043 [astro-ph.GA].
- [28] R. Flauger, N. Karnesis, G. Nardini, M. Pieroni, A. Ricciardone, and J. Torrado, *JCAP* **01**, 059 (2021), arXiv:2009.11845 [astro-ph.CO].
- [29] D. Bandopadhyay and C. J. Moore, (2023), arXiv:2305.18048 [gr-qc].
- [30] M. L. Katz, A. J. K. Chua, L. Speri, N. Warburton, and S. A. Hughes, *Phys. Rev. D* **104**, 064047 (2021), arXiv:2104.04582 [gr-qc].
- [31] E. Pizzati, S. Sachdev, A. Gupta, and B. Sathyaprakash, *Phys. Rev. D* **105**, 104016 (2022), arXiv:2102.07692 [gr-qc].
- [32] A. Antonelli, O. Burke, and J. R. Gair, *Mon. Not. Roy. Astron. Soc.* **507**, 5069 (2021), arXiv:2104.01897 [gr-qc].
- [33] A. Samajdar, J. Janquart, C. Van Den Broeck, and T. Dietrich, *Phys. Rev. D* **104**, 044003 (2021), arXiv:2102.07544 [gr-qc].
- [34] J. Janquart, T. Baka, A. Samajdar, T. Dietrich, and C. Van Den Broeck, (2022), arXiv:2211.01304 [gr-qc].
- [35] J. Langendorff, A. Kohlmuß, J. Janquart, and C. Van Den Broeck, *Phys. Rev. Lett.* **130**, 171402 (2023), arXiv:2211.15097 [gr-qc].
- [36] P. Relton and V. Raymond, *Phys. Rev. D* **104**, 084039 (2021), arXiv:2103.16225 [gr-qc].
- [37] P. Relton, A. Virtuoso, S. Bini, V. Raymond, I. Harry, M. Drago, C. Lazzaro, A. Miani, and S. Tiwari, *Phys. Rev. D* **106**, 104045 (2022), arXiv:2208.00261 [gr-qc].
- [38] G. Ashton *et al.*, *Astrophys. J. Suppl.* **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].
- [39] I. M. Romero-Shaw *et al.*, *Mon. Not. Roy. Astron. Soc.* **499**, 3295 (2020), arXiv:2006.00714 [astro-ph.IM].
- [40] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, *Publ. Astron. Soc. Pac.* **131**, 024503 (2019), arXiv:1807.10312 [astro-ph.IM].
- [41] J. Veitch *et al.*, *Phys. Rev. D* **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].
- [42] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *Publ. Astron. Soc. Pac.* **125**, 306 (2013), arXiv:1202.3665 [astro-ph.IM].
- [43] W. D. Vousden, W. M. Farr, and I. Mandel, *Mon. Not. Roy. Astron. Soc.* **455**, 1919 (2016), arXiv:1501.05823.
- [44] J. Skilling, *Bayesian Analysis* **1**, 833 (2006).
- [45] W. J. Handley, M. P. Hobson, and A. N. Lasenby, *Mon. Not. Roy. Astron. Soc.* **453**, 4384 (2015), arXiv:1506.00171 [astro-ph.IM].
- [46] G. Ashton *et al.*, *Nature* **2** (2022), 10.1038/s43586-022-00121-x, arXiv:2205.15570 [stat.CO].
- [47] J. S. Speagle, *Mon. Not. Roy. Astron. Soc.* **493**, 3132 (2020), arXiv:1904.02180 [astro-ph.IM].
- [48] T. D. P. Edwards, K. W. K. Wong, K. K. H. Lam, A. Coogan, D. Foreman-Mackey, M. Isi, and A. Zimmerman, (2023), arXiv:2302.05329 [astro-ph.IM].
- [49] K. W. K. Wong, M. Isi, and T. D. P. Edwards, (2023), arXiv:2302.05333 [astro-ph.IM].
- [50] B. Zackay, L. Dai, and T. Venumadhav, (2018), arXiv:1806.08792 [astro-ph.IM].
- [51] N. Leslie, L. Dai, and G. Pratten, *Phys. Rev. D* **104**, 123030 (2021), arXiv:2109.09872 [astro-ph.IM].
- [52] K. Cranmer, J. Brehmer, and G. Louppe, *Proc. Nat. Acad. Sci.* **117**, 30055 (2020), arXiv:1911.01429 [stat.ML].
- [53] B. K. Miller, A. Cole, P. Forré, G. Louppe, and C. Weniger, in *35th Conference on Neural Information Processing Systems* (2021) arXiv:2107.01214 [stat.ML].
- [54] A. Tejero-Cantero, J. Boelts, M. Deistler, J.-M. Lueckmann, C. Durkan, P. J. Gonçalves, D. S. Greenberg, and J. H. Macke, *Journal of Open Source Software* **5**, 2505 (2020).
- [55] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, *Mon. Not. Roy. Astron. Soc.* **488**, 4440 (2019), arXiv:1903.00007 [astro-ph.CO].
- [56] A. Cole, B. K. Miller, S. J. Witte, M. X. Cai, M. W. Grootes, F. Nattino, and C. Weniger, *JCAP* **09**, 004 (2022), arXiv:2111.08030 [astro-ph.CO].
- [57] N. A. Montel, A. Coogan, C. Correa, K. Karchev, and C. Weniger, *Mon. Not. Roy. Astron. Soc.* **518**, 2746 (2022), arXiv:2205.09126 [astro-ph.CO].
- [58] N. Anau Montel and C. Weniger, in *36th Conference on Neural Information Processing Systems* (2022) arXiv:2211.04291 [astro-ph.IM].
- [59] A. Dimitriou, C. Weniger, and C. A. Correa, (2022), arXiv:2206.11312 [astro-ph.CO].
- [60] K. Karchev, R. Trotta, and C. Weniger, *Mon. Not. Roy. Astron. Soc.* **520**, 1056 (2022), arXiv:2209.06733 [astro-ph.CO].
- [61] J. Alvey, M. Gerdes, and C. Weniger, (2023), arXiv:2304.02032 [astro-ph.GA].
- [62] M. Dax, S. R. Green, J. Gair, M. Deistler, B. Schölkopf, and J. H. Macke, (2021), arXiv:2111.13139 [cs.LG].
- [63] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, *Phys. Rev. Lett.* **127**, 241103 (2021),

- [arXiv:2106.12594 \[gr-qc\]](#).
- [64] M. Dax, S. R. Green, J. Gair, M. Pürrer, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, *Phys. Rev. Lett.* **130**, 171403 (2023), [arXiv:2210.05686 \[gr-qc\]](#).
- [65] B. K. Miller, A. Cole, C. Weniger, F. Nattino, O. Ku, and M. W. Grootes, *J. Open Source Softw.* **7**, 4205 (2022).
- [66] U. Bhardwaj, J. Alvey, B. K. Miller, S. Nissanke, and C. Weniger, (2023), [arXiv:2304.02035 \[gr-qc\]](#).
- [67] G. Ashton and C. Talbot, *Mon. Not. Roy. Astron. Soc.* **507**, 2037 (2021), [arXiv:2106.08730 \[gr-qc\]](#).
- [68] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Class. Quant. Grav.* **37**, 055002 (2020), [arXiv:1908.11170 \[gr-qc\]](#).
- [69] G. Pratten *et al.*, *Phys. Rev. D* **103**, 104056 (2021), [arXiv:2004.06503 \[gr-qc\]](#).
- [70] R. D. Green, *There's more than one way to ride the wave: A multi-disciplinary approach to gravitational wave data analysis*, Ph.D. thesis, Cardiff U. (2021).
- [71] K. Kim and A. Liu, (2023), [arXiv:2301.07253 \[gr-qc\]](#).
- [72] A. Taracchini *et al.*, *Phys. Rev. D* **89**, 061502 (2014), [arXiv:1311.2544 \[gr-qc\]](#).
- [73] A. Bohé *et al.*, *Phys. Rev. D* **95**, 044028 (2017), [arXiv:1611.03703 \[gr-qc\]](#).
- [74] S. Ossokine *et al.*, *Phys. Rev. D* **102**, 044055 (2020), [arXiv:2004.09442 \[gr-qc\]](#).
- [75] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044007 (2016), [arXiv:1508.07253 \[gr-qc\]](#).
- [76] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, *Phys. Rev. Lett.* **113**, 151101 (2014), [arXiv:1308.3271 \[gr-qc\]](#).
- [77] J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, V. Begy, and G. Louppe, (2021), [arXiv:2110.06581 \[stat.ML\]](#).
- [78] P. Lemos, A. Coogan, Y. Hezaveh, and L. Perreault-Levasseur, (2023), [arXiv:2302.03026 \[stat.ML\]](#).

What to do when things get crowded?

Scalable joint analysis of overlapping gravitational wave signals

Supplemental Material

James Alvey^{†, 1,*} Uddipta Bhattacharyya^{†, 2, †}
 Samaya Nissanke^{2, 3} and Christoph Weniger¹

¹GRAPPA Institute, Institute for Theoretical Physics Amsterdam,

University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

²GRAPPA Institute, Anton Pannekoek Institute for Astronomy and Institute of High-Energy Physics,

University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

³Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands

COVERAGE TESTS

Coverage tests explicitly check the behaviour of the posterior over various realizations of the signal. The intuition is that if our posterior estimates are behaving correctly, then *e.g.* for $p\%$ of realizations, the true value of the parameter(s) should lie inside the $p\%$ confidence interval. If our posteriors are too overconfident, then this will be true for more than $p\%$ of the realizations, and vice versa for under-confident inference results. For each case study, to carry out these tests we generated an additional 2000 simulations from the truncated prior in the final round. Since our sequential SBI framework achieves local amortization, inference on these 2000 signals needed to produce the coverage plots is almost immediate, requiring only a single network evaluation for each example. The full set of these tests are shown in Figs. S6–S8.

RELEVANT DETAILS OF THE PEREGRINE IMPLEMENTATION

As discussed in the main text, the majority of the **peregrine** implementation described in Ref. [66] has remained unchanged for this application to overlapping signals. In this supplementary section, we remind the reader of the key components and indicate any places where slight changes were made compared to this reference.

The **peregrine** library is built on top of the **swyft** [53, 65] implementation of the truncated marginal neural ratio estimation (TMNRE) algorithm. For this work, we use exactly the same pytorch-based embedding network described in the Appendix and Sec. II.3 of Ref. [66]. In addition, we use the same truncation scheme detailed in Sec. II.2 of the same reference, with $\epsilon = 10^{-5}$, as defined in Refs. [53, 66]. Furthermore, we use almost identical settings for the TMNRE algorithm provided in Tab. II of [66]. The only differences in this work are the simulation schedule and the training/validation batch size. In this work, we use 10^6 simulations per round to account for the more complex parameter space (although we expect that this is overly conservative). We also take a correspondingly larger training/validation batch size of 512 instead of 256. As in the single signal analysis, we make use of the noise shuffling scheme explained in Sec. II.3

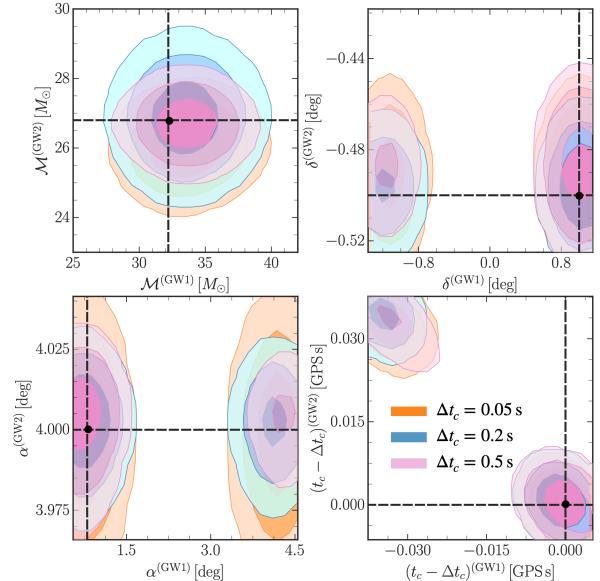


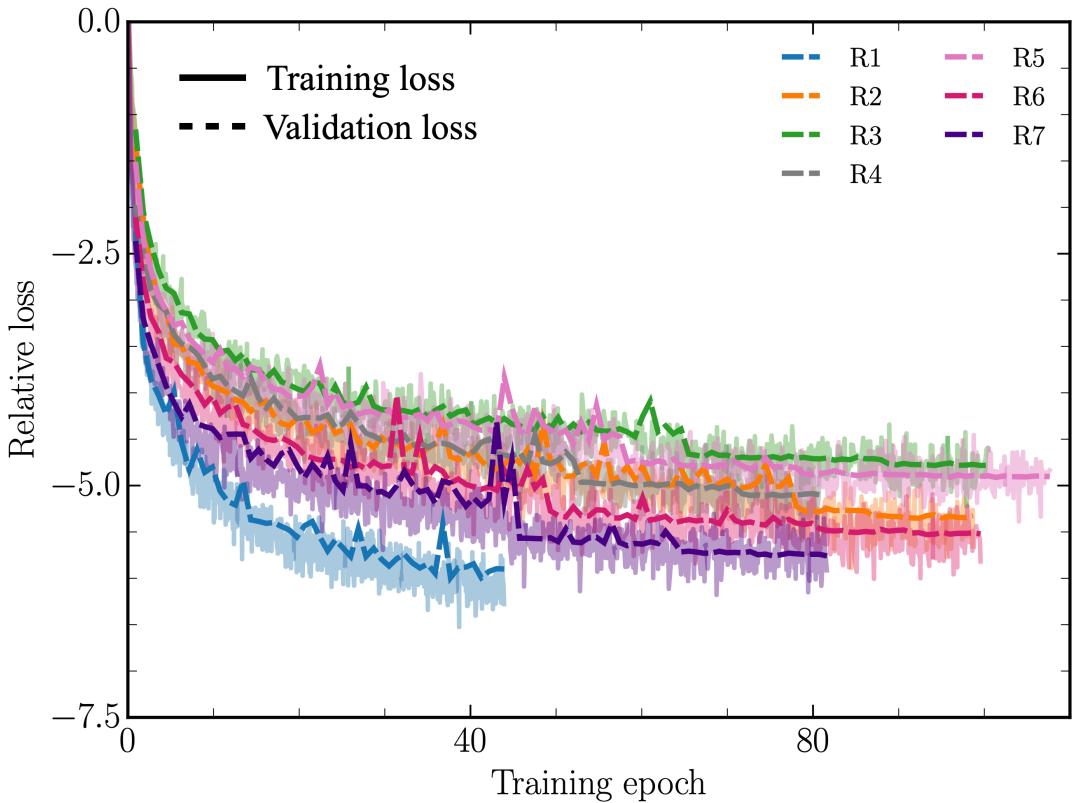
FIG. S1. Additional 2D marginals for all three case studies (**C1**, **C2**, and **C3**). From top-left to bottom-right, we show the 2D posteriors for the two chirp masses, declinations, right ascensions and merger times. Broadly, we see two features: there are few correlations between parameters from different signals, and that generally we achieve slightly better sensitivity as the merger times become more separated.

of [66], which efficiently avoids overfitting. For the application in this work, we show the training and validation curves across the 7 rounds of inference in Fig. S2. We see that across all curves, the training proceeds smoothly and the training and validation curves are well calibrated.

In terms of simulations, we generate waveform simulations using the waveform templates provided in **bilby** [38]. In this paper, we use the **IMRPhenomXPHM** [69] waveform approximant to generate both signals, however, any waveform choice is possible within our framework. For example, in Ref. [66], we instead used the **SEOBNRv4PHM** model [74]. For the case studies described in the main text, we generated signals of total duration 4 s, which we explicitly checked was long enough (in the sense of the **bilby** functionality to compute `get_safe_signal_duration`) for all parts of the prior parameter space.

Parameter	Prior Choice	Injection (GW1)	Injection (GW2)
Chirp mass $\mathcal{M} [\text{M}_\odot]$	$\text{U}(20, 100)$	32.2	26.8
Mass ratio, q	$\text{U}(0.125, 1)$	0.77	0.81
Inclination angle θ_{jn} [rad]	$\text{sine}(0, \pi)$	0.05	1.8
Polarisation angle ψ [rad]	$\text{U}(0, \pi)$	0.9	0.4
Phase ϕ_c [rad]	$\text{U}(0, 2\pi)$	1.5	0.6
Tilt angles θ_1, θ_2 [rad]	$\text{sine}(0, \pi)$	1.5, 1.5	0.9, 1.2
Dimensionless spins a_1, a_2	$\text{U}(0, 1)$	0.1, 0.05	0.7, 0.3
Spin angles ϕ_{12}, ϕ_{jl} [rad]	$\text{U}(0, 2\pi)$	0.8, 0.8	0.6, 0.9
Right ascension α [rad]	$\text{U}(0, 2\pi)$	0.8	4.0
Declination δ [rad]	$\text{cosine}(-\pi/2, \pi/2)$	1.0	-0.5
Merger time t_c [GPS s]	$\text{U}(-0.1, 0.1)$	0.0	0.05(C1), 0.2(C2), 0.5(C3)
Luminosity Distance d_L [Mpc]	$\text{U}_{\text{vol.}}(100, 2500)^*$	1545	636.5

TABLE S1. Definitions and prior choices for all relevant gravitational wave parameters in this work.

[†] Note that these are the default priors used in bilby analyses for BBH systems (subject to calibration ranges of the waveform approximants).^{*} Specifically, the luminosity distance prior is taken to be uniform in comoving volume in the *source frame*.FIG. S2. Training and validation loss curves for the various rounds of inference performed in this work. Values of the loss on the training dataset are shown by the solid lines, whilst the values on the validation set are shown by dashed curves. The example given here is for the first case study (**C1**) described in the main text, but a similar picture holds across all cases.

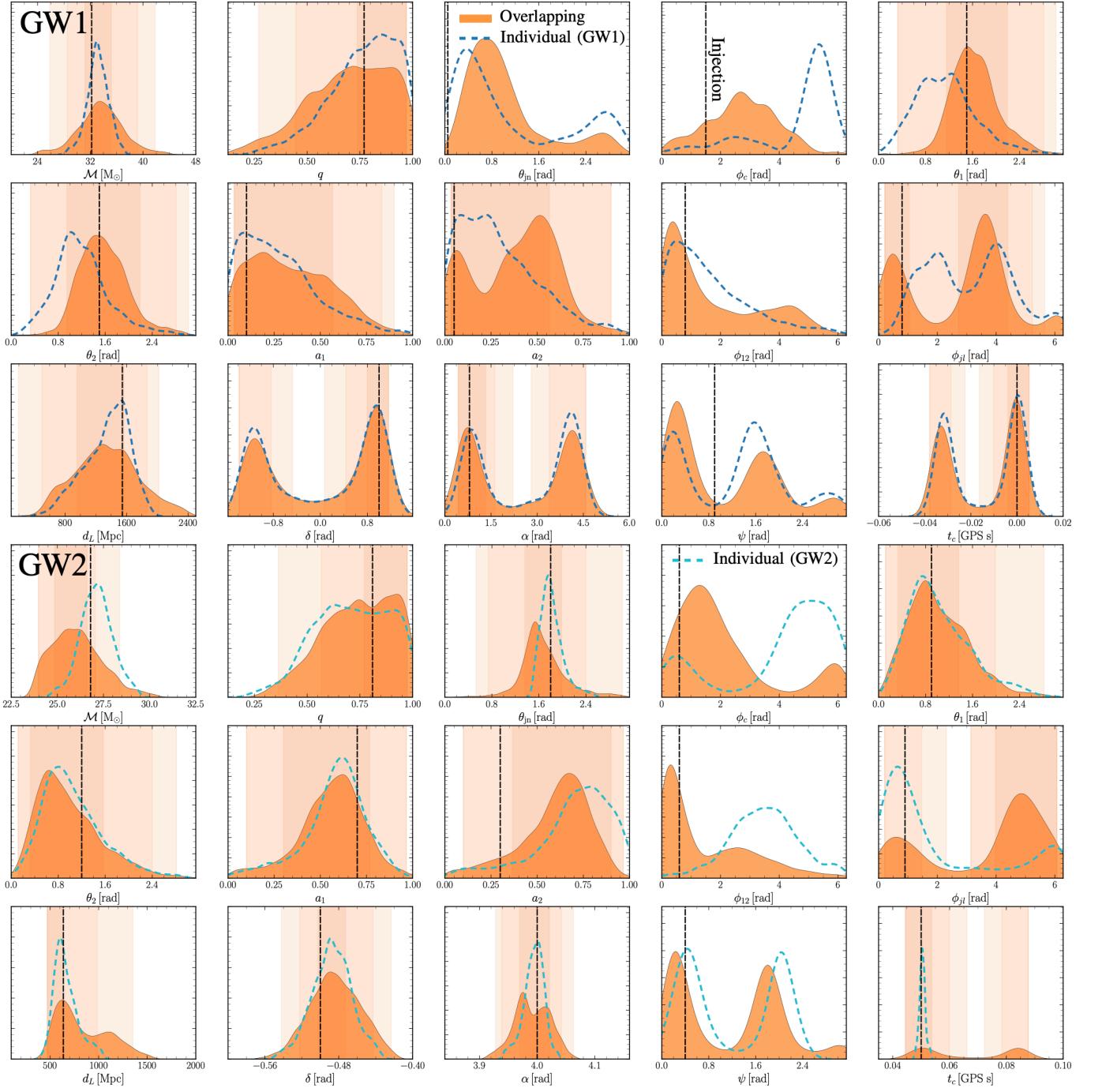


FIG. S3. Full set of 1D marginals for our case study with $\Delta t_c = 0.05$ s (**C1**, as shown in Fig. 1). Inference results from `peregrine` joint analysis are shown in orange whereas the corresponding single signal analyses are shown in blue and cyan respectively (in order of t_c), as shown in the figure legend. The black dashed lines indicate the injected parameter values. For parameters that are well measured, the 1σ , 2σ and 3σ contours are indicated in subsequently lighter shading.

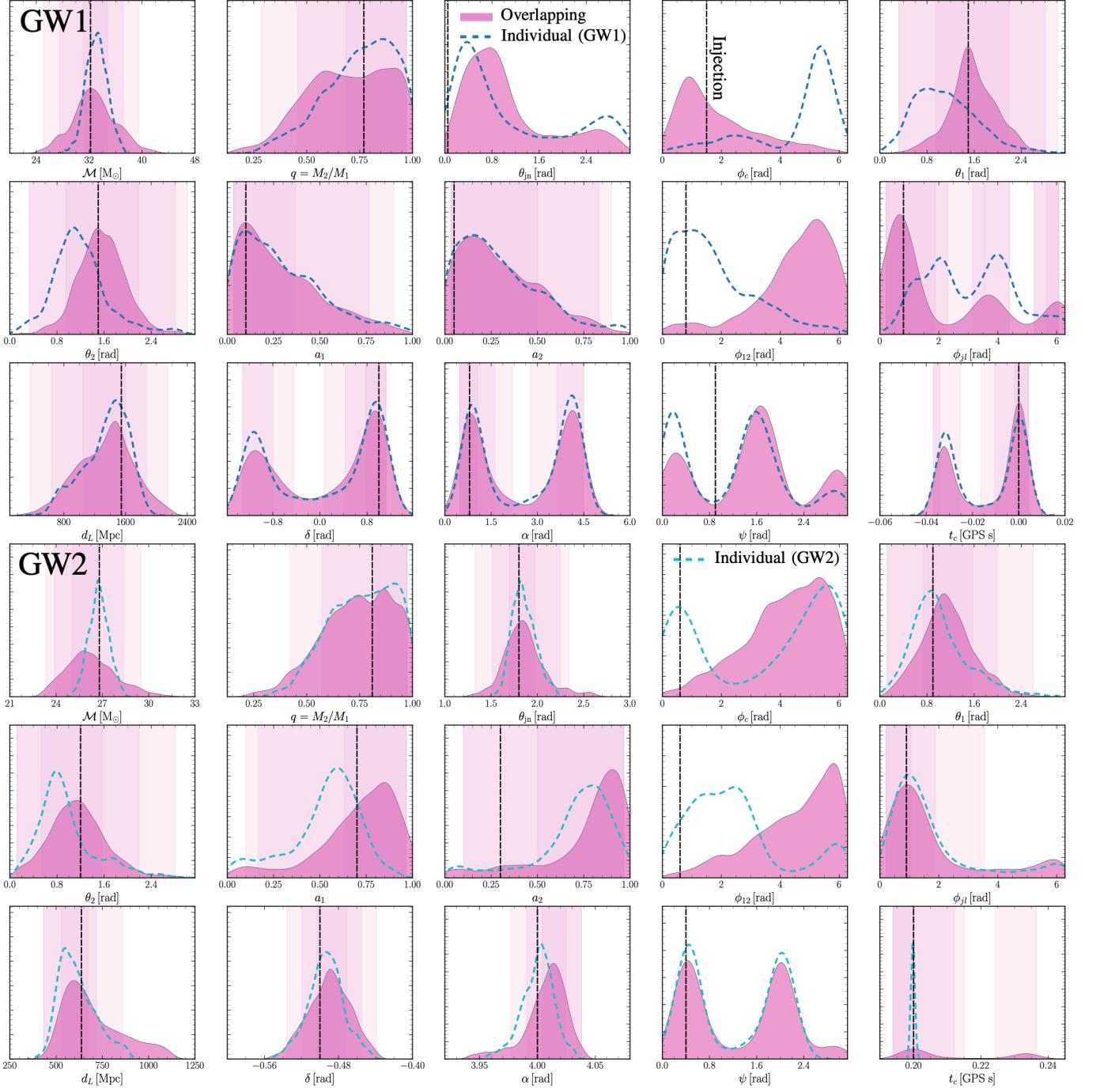


FIG. S4. Full set of 1D marginals for our case study with $\Delta t_c = 0.2$ s (C2). Inference results from **peregrine** joint analysis are shown in pink whereas the corresponding single signal analyses are shown in blue and cyan respectively (in order of t_c), as shown in the figure legend. The black dashed lines indicate the injected parameter values. For parameters that are well measured, the 1σ , 2σ and 3σ contours are indicated in subsequently lighter shading.

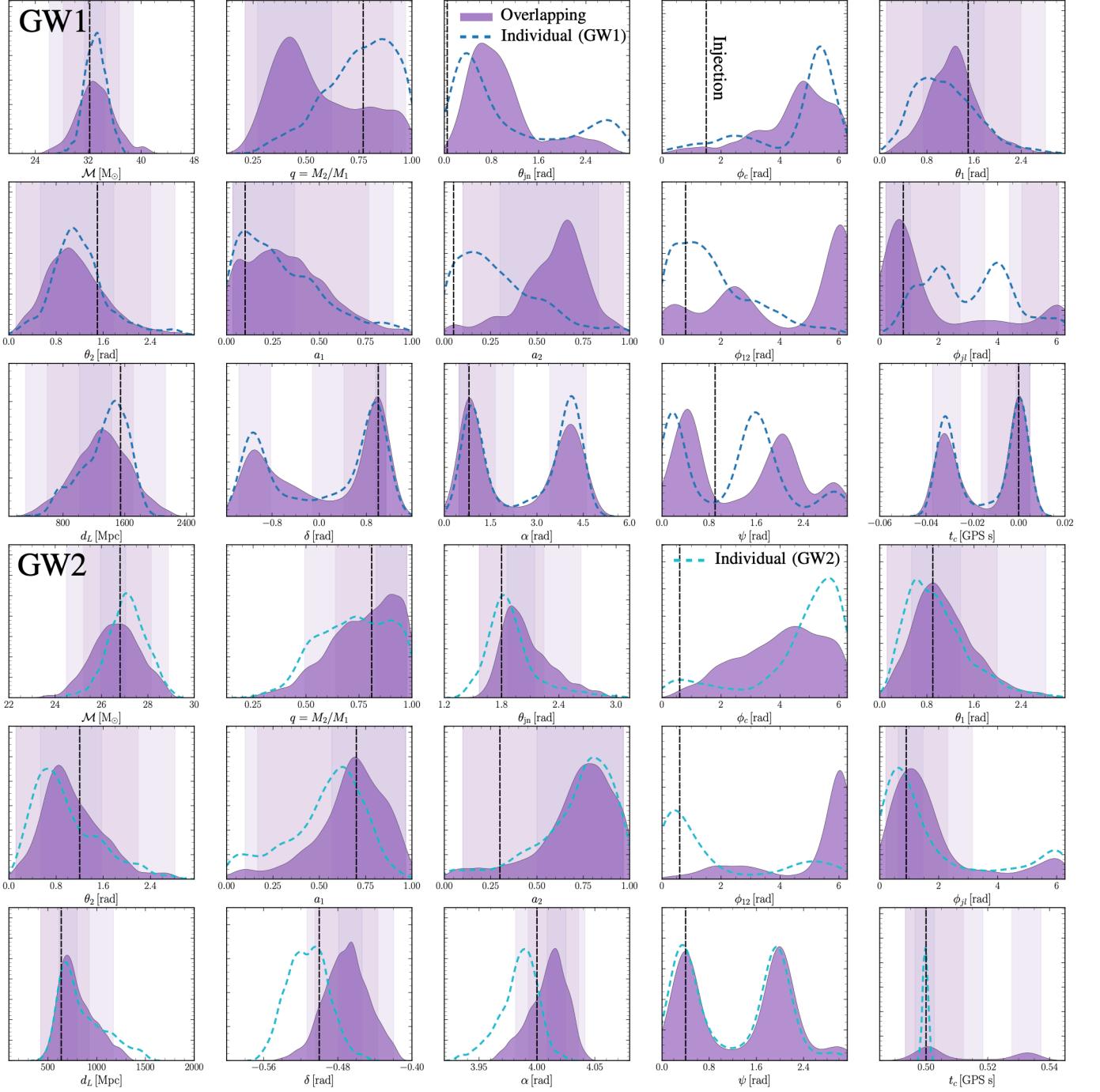


FIG. S5. Full set of 1D marginals for our case study with $\Delta t_c = 0.5$ s (**C3**). Inference results from **peregrine** joint analysis are shown in purple whereas the corresponding single signal analyses are shown in blue and cyan respectively (in order of t_c), as shown in the figure legend. The black dashed lines indicate the injected parameter values. For parameters that are well measured, the 1σ , 2σ and 3σ contours are indicated in subsequently lighter shading.

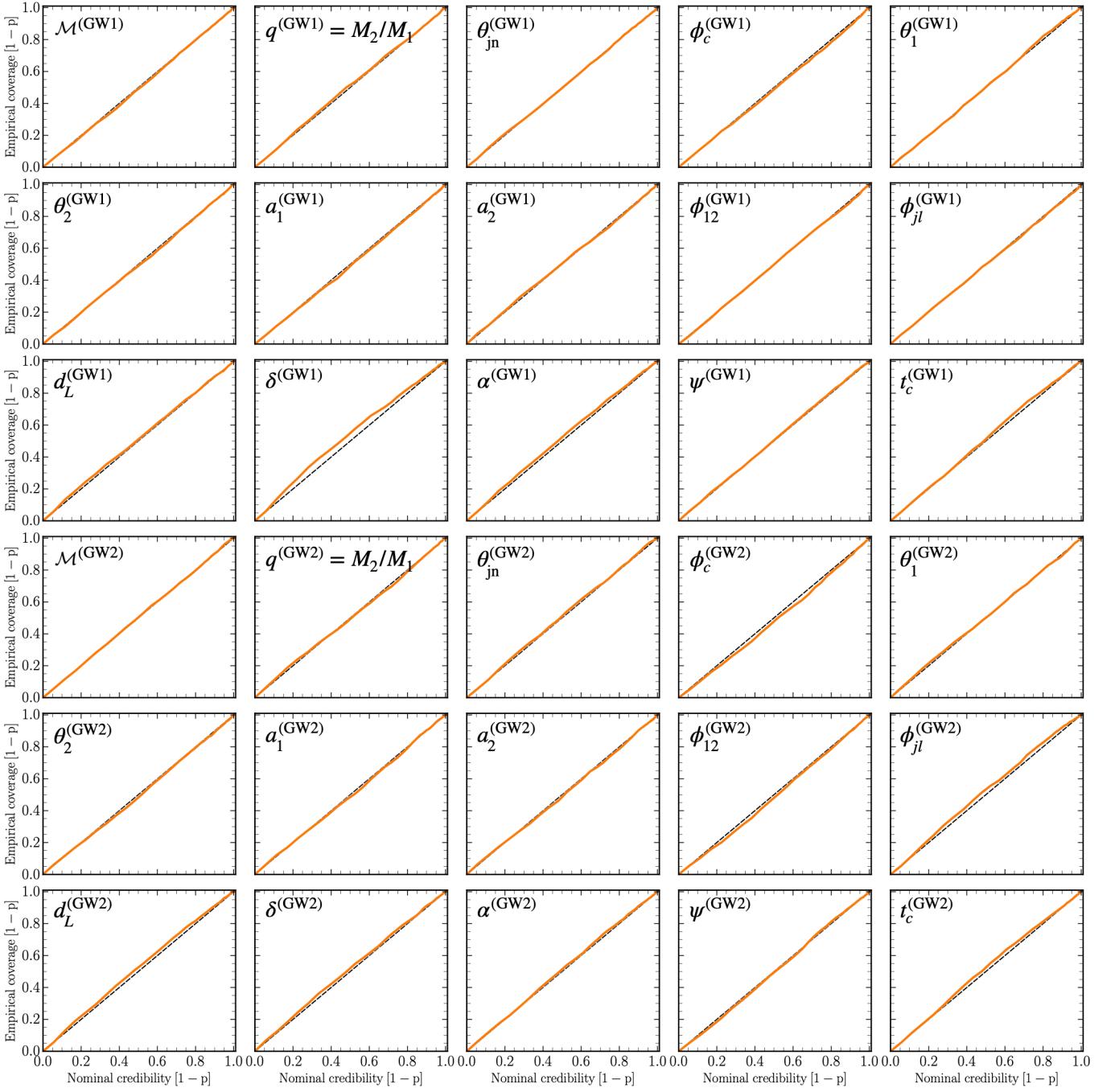


FIG. S6. Coverage results for **C1** for all 30 parameters characterising an overlapping signal comprising two overlapping binary black hole gravitational wave signals. This compares the expected coverage of the true value as a percentage on the x-axis against the actual coverage of our ratio estimator on the y-axis. The orange lines indicate the average coverage and their strong alignment along the diagonal (black-dashed) shows that our posteriors are well-calibrated.

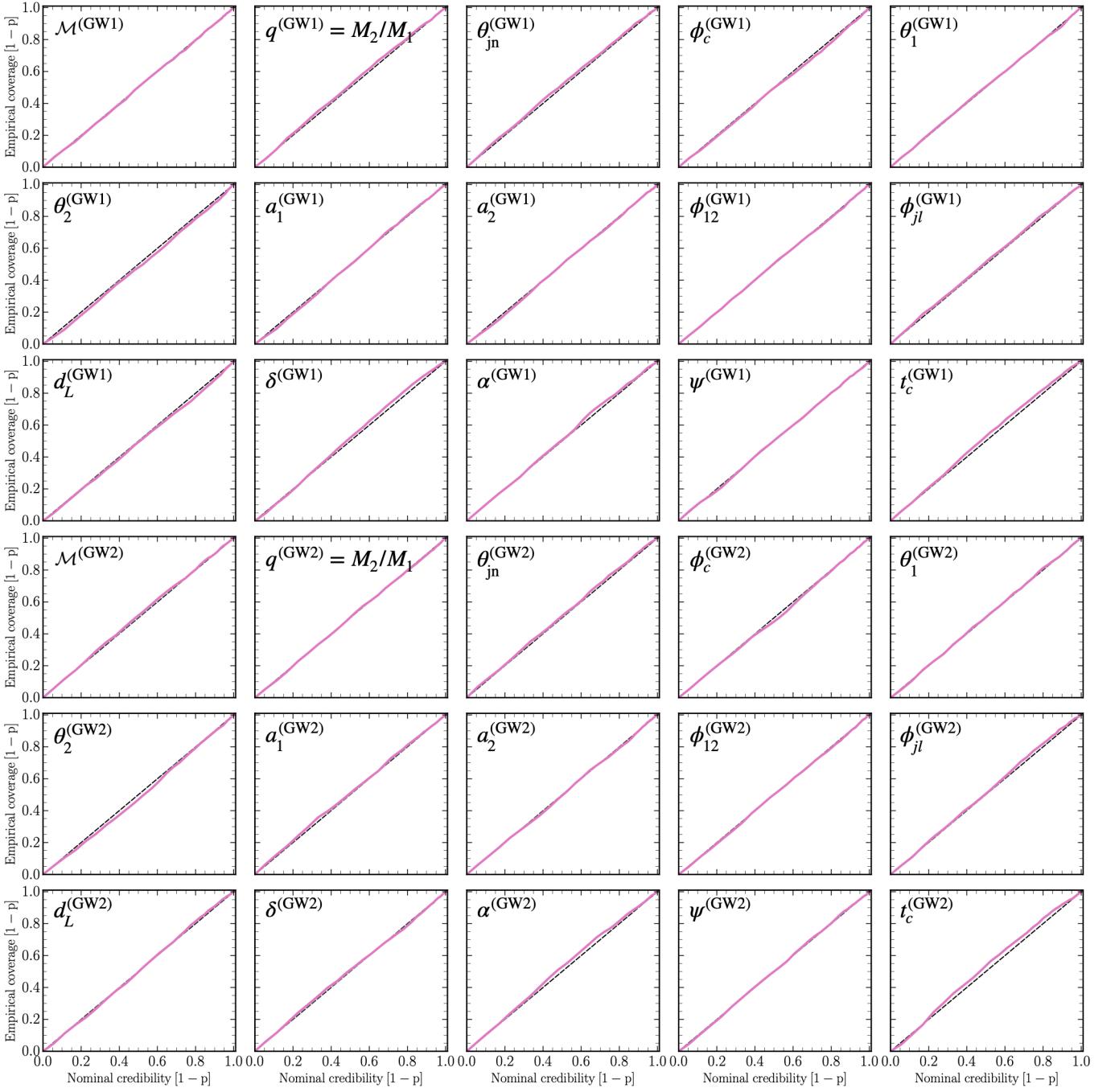


FIG. S7. Coverage results for **C2** for all 30 parameters characterising an overlapping signal comprising two overlapping binary black hole gravitational wave signals. This compares the expected coverage of the true value as a percentage on the x-axis against the actual coverage of our ratio estimator on the y-axis. The pink lines indicate the average coverage and their strong alignment along the diagonal (black-dashed) shows that our posteriors are well-calibrated.

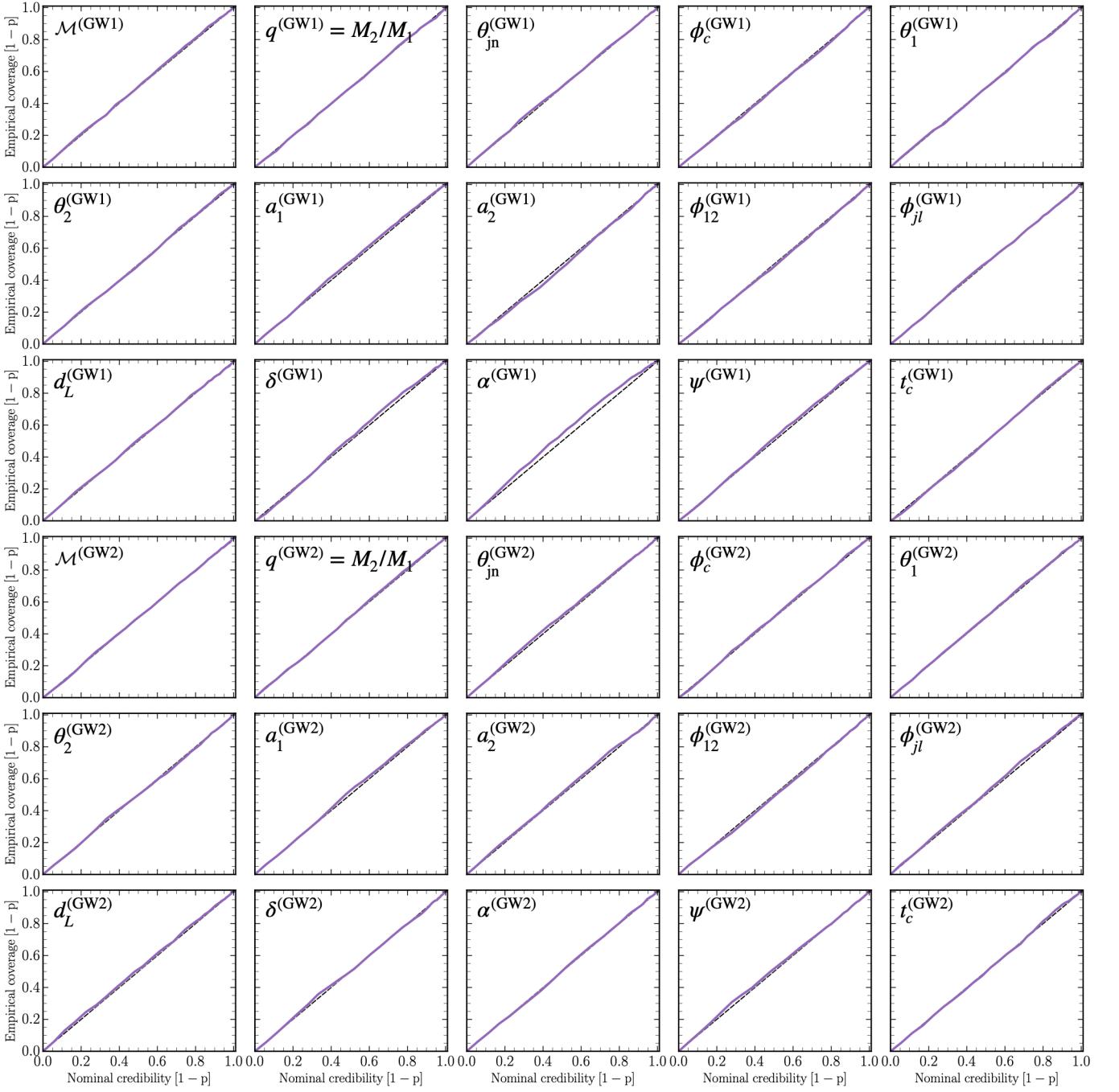


FIG. S8. Coverage results for **C3** for all 30 parameters characterising an overlapping signal comprising two overlapping binary black hole gravitational wave signals. This compares the expected coverage of the true value as a percentage on the x-axis against the actual coverage of our ratio estimator on the y-axis. The purple lines indicate the average coverage and their strong alignment along the diagonal (black-dashed) shows that our posteriors are well-calibrated.