

Sequential simulation-based inference for extreme mass ratio inspirals

Philippa S. Cole^{1,2,*} James Alvey^{3,4} Lorenzo Speri⁵,

Christoph Weniger⁶ Uddipta Bhardwaj⁶ Davide Gerosa^{1,2} and Gianfranco Bertone⁶

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

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

³*Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*

⁴*Institute of Astronomy, University of Cambridge,*

Madingley Road, Cambridge CB3 0HA, United Kingdom

⁵*European Space Agency (ESA), European Space Research and Technology*

Centre (ESTEC), Keplerlaan 1, 2201 AZ Noordwijk, the Netherlands

⁶*Gravitation Astroparticle Physics Amsterdam (GRAPPA),*

Institute for Theoretical Physics Amsterdam and Delta Institute for Theoretical Physics,

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

Extreme mass-ratio inspirals pose a difficult challenge in terms of both search and parameter estimation for upcoming space-based gravitational-wave detectors such as LISA. Their signals are long and of complex morphology, meaning they carry a large amount of information about their source, but are also difficult to search for and analyse. We explore how sequential simulation-based inference methods, specifically truncated marginal neural ratio estimation, could offer solutions to some of the challenges surrounding extreme-mass-ratio inspiral data analysis. We show that this method can efficiently narrow down the volume of the complex 11-dimensional search parameter space by a factor of $10^6 - 10^7$ and provide 1-dimensional marginal proposal distributions for non-spinning extreme-mass-ratio inspirals. We discuss the current limitations of this approach and place it in the broader context of a global strategy for future space-based gravitational-wave data analysis.

I. INTRODUCTION

Extreme-mass-ratio inspirals (EMRIs) are a key target for future space-based gravitational wave detectors such as the Laser Interferometer Space Antenna (LISA) [1], as well as TianQin [2] and Taiji [3]. Defined as a binary black-hole system where the primary mass is much larger than the secondary, their interesting dynamics have a rich phenomenology that could offer new astrophysics and fundamental physics insights [4]. For example, EMRIs are ideal testbeds for beyond-General-Relativity effects [5–9], environmental effects due to dark matter [10–14], scalar fields [15–18], or accretion disks [19–22], as well as probing intermediate and supermassive black-hole formation mechanisms [23].

However, with great richness comes great complication. EMRIs are expected to have eccentric and inclined orbits and many thousands of harmonics contributing to the gravitational waveform, which makes computing the waveforms challenging. The validity of post-Newtonian methods for long duration waveforms do not extend down to such small mass ratios and numerical-relativity simulations of systems with such large differences in scales is computationally unfeasible, see Ref. [24] for attempts to resolve this. Furthermore, these systems are expected to remain in the LISA band for months or years, meaning that many thousands or millions of orbits must be computed. Currently, adiabatic approximations [25], effective-one-body approaches [26–28] and self-force theory [29–35] are relied upon, and it is an active area of

research to beat down waveform systematics from these methods so that they are accurate enough to be used in gravitational-wave data analysis.

On the detector side, the challenges of extracting signals from realistic space-based gravitational-wave detector data are numerous [36]. The noise is expected to be non-stationary and non-Gaussian, due to glitches, drifting noise levels in the components of the instruments, as well as having gaps due to both planned and unplanned downtime [37–46]. Furthermore, there will be many overlapping signals in the data, including thousands of quasi-monochromatic galactic sources that will remain in band for the duration of the LISA mission [47–50].

Searches and parameter-estimation algorithms for ground-based gravitational-wave detector data traditionally rely upon match-filtering and likelihood-based methods [51]. However, given the computational expense of computing a waveform template bank large enough for the search portion – $\sim \mathcal{O}(10^{40})$ templates will be required to perform a fully coherent matched filter search for year-long inspiral waveforms [52, 53] – and the highly complex likelihood surface in the context of parameter estimation, these methods may not be currently optimised for space-based gravitational-wave data.

In this paper, we explore how simulation-based inference methods, which have been shown to be successful in other gravitational-wave applications [54–60], can help tackle the challenges of EMRI searches and parameter estimation. We build a pipeline to extract the parameters of Schwarzschild EMRIs in vacuum using Truncated Marginal Neural Ratio Estimation (TMNRE) [61, 62], assuming Gaussian noise and including the LISA antenna response. We show that this sequential method can be

* philippa.cole@unimib.it

a very powerful tool for narrowing down the parameter space from initially wide priors, emulating a search strategy, as well as producing proposal distributions that encompass the true injected values. We note here that by “search,” we refer to the reduction in prior volume for realistic values of the parameters of the source, rather than the identification of specific data segments to analyse. These notions are somewhat combined in a LIGO-Virgo-KAGRA-style matched filtering analysis, which is likely to be infeasible for EMRIs in the mHz band.

This paper is organised as follows: in Sec. II, we describe how we produce EMRI waveforms for this study; in Sec. III, we briefly describe the noise and detector response that we employ; in Sec. IV, we review the state of the art for EMRI searches and parameter estimation; in Sec. V, we introduce our TMNRE approach; in Sec. VI, we present our results; in Sec. VII, we discuss future prospects.

II. EMRI WAVEFORMS

We model EMRI signals using the time domain implementation of the Fast EMRI Waveform (FEW) package [63–72], which is GPU-accelerated. The six parameters that describe a Schwarzschild EMRI system in the source frame¹ are the primary black-hole mass m_1 , the secondary black-hole mass m_2 , the initial semi-latus rectum p_0 , the initial eccentricity of the system e , and the initial radial and azimuthal phases $\Phi_{r,0}$ and $\Phi_{\phi,0}$. Five extrinsic parameters are then required to describe the source in the detector frame (see Sec. III), namely the polar angle of the direction of the orbital angular momentum unit vector θ_k , the azimuthal angle of the direction of the orbital angular momentum unit vector ϕ_k , the luminosity distance d_L , the polar sky location angle θ_S , and the azimuthal sky location angle ϕ_S .

FEW computes the phase and the amplitude of the gravitational waveform up to first order in gravitational self-force theory, i.e. the adiabatic approximation, which relies on expanding the metric of the binary in powers of the mass ratio $q = m_2/m_1$. The amplitudes of the modes are computed to lower accuracy, and are pre-computed for a subset of geodesic parameter sets, $\{p, e\}$ in the Schwarzschild case, and interpolated to cover the wider parameter space using a combination of reduced-order-model [73] surrogates and deep-learning methods. The amplitude and the phase are combined for each mode, and the modes are summed over to produce an adiabatic waveform $h(t) = h_+(t) - ih_\times(t)$ in the time domain.

While in this paper we only consider Schwarzschild adiabatic (1st-order self-force) systems, we stress that the simulation module can be easily adapted to include more parameters and different waveform models, so we expect

to be able to extend our results to the Kerr case (requiring three extra parameters: the dimensionless spin of the massive black hole a , the inclination angle between the orbital angular momentum unit vector and the spin vector x_I , and the initial polar phase $\Phi_{\theta,0}$), and to include 2nd-order self-force implementations in the near future, see e.g. Ref. [35].

III. LISA RESPONSE AND NOISE

LISA will consist of three satellites in heliocentric orbits, forming an equilateral triangular constellation with arm lengths of approximately 2.5×10^6 km [1]. Each spacecraft will exchange laser beams with the other two, and measuring light propagation times in both directions results in six interferometric links. Incoming gravitational waves induce perturbations in these light-travel times, which can be extracted via time-delay interferometry (TDI), a technique that cancels otherwise overwhelming laser phase noise by constructing specific linear combinations of the delayed signals [74–78].

The Michelson variables, X, Y, Z have correlated noise properties, whereas linear combinations of these, the so-called A, E and T channels, are orthogonal and uncorrelated in the approximation that the arm lengths are fixed and equal. In reality, both the “breathing” motion of the constellation, as well as orbital adjustments to combat drift, break this orthogonality. For simplicity and computational efficiency, we do not use the T channel in our analysis, as it suppresses gravitational-wave power with respect to the other two, although see e.g. Ref. [79] for more details. We feed our simulated time-domain signals into the `fastlisaresponse` [80] code, which is GPU-accelerated, to compute the projection onto the A and E TDI variables, making use of the European Space Agency’s simulations of LISA orbits [81].

In this work we assume stationary, Gaussian noise according to the power spectral density (PSD) as given in Eq. (58) of Ref. [82] and implemented in `pycbc` v.2.4.0 [83]. We consider TDI 1.0 variables and account for the removal of confusion noise due to galactic white-dwarf binaries using the analytic fit implemented in Ref. [83]. This is a function of the duration that LISA has been taking data for, which we take to be 1 year, and decreases the amplitude of the effective PSD around the frequencies $\lesssim 10^{-3}$ Hz.

IV. STATE OF THE ART

Various strategies have been explored for EMRI searches and parameter estimation. On the search side, EMRIs span such a vast parameter space that fully coherent template bank methods become impractical, since their posterior distributions occupy only a tiny fraction of this space [52]. As a result, conventional grid-based approaches are infeasible, leading to the adoption of

¹ Note that FEW waveform generator takes detector masses as input.

Parameter	Prior Range
$m_1/10^5 [M_\odot]$	$\mathcal{U}[4, 9]$
$m_2 [M_\odot]$	$\mathcal{U}[20, 60]$
$p_0/10 [m_1]$	$\mathcal{U}[8.5, 14.0]$
e_0	$\mathcal{U}[0.05, 0.60]$
$\cos \theta_K$	$\mathcal{U}[-1, 1]$
ϕ_K	$\mathcal{U}[0, 2\pi]$
$d_L/100 [\text{Gpc}]$	$\mathcal{U}[8.75, 50.0]$
$\Phi_{\phi,0}$	$\mathcal{U}[-\pi, \pi]$
$\Phi_{r,0}$	$\mathcal{U}[-\pi, \pi]$
$\cos \theta_S$	$\mathcal{U}[-1, 1]$
ϕ_S	$\mathcal{U}[0, 2\pi]$

TABLE I. Uniform prior ranges used for all parameter estimation results shown in this paper.

stochastic search techniques. However, the presence of numerous local maxima within the search landscape due to intricate, non-local parameter degeneracies in the signal space [84] implies that many combinations of parameters can produce waveforms that closely match the data. Distinguishing the true global maximum is a complex puzzle, as it can be located several σ away from the local maxima, see [84] for more details.

Previous works have adopted several strategies to address these problems. Stochastic searches that use hybrid methods between genetic algorithms and MCMC [85], as well as methods built around the Metropolis–Hastings algorithm [86, 87] have been explored as a means for using information about local maxima to try and identify the global maximum.

An alternative approach was proposed in Ref. [88] which used a veto likelihood to remove the local maxima, allowing only the global maximum to remain. Other approaches used phenomenological waveforms to first identify the EMRI harmonics before carrying out a parameter search [89]. More recently, Ref. [90] performed a three-stage semi-coherent search beginning with physical harmonic waveforms followed by phenomenological waveforms. Ref. [91] used deep-learning techniques on time-frequency domain waveforms obtained in both the analytic-kludge and effective-one-body-Teukolsky frameworks to search for the mass, spin, and eccentricity of the injected systems. Furthermore, Ref. [92] used convolutional neural networks to search for analytic-kludge and augmented analytic-kludge EMRI waveforms with noise according to the expected sensitivities of TianQin [2], assuming the detector behaves like two Michelson interferometers (i.e. no TDI). Finally, Ref. [93] used sparse dictionary learning to search for and reconstruct analytic-kludge EMRI waveforms with primary mass and eccentricity parameters allowed to vary over broad ranges but with all other parameters fixed.

On the parameter estimation side, likelihood-based approaches have been the most utilised. For example, Ref. [94] recovered EMRI signals from the mock LISA

data challenge 1.B [95] employing a parallel tempering MCMC algorithm on two-week-long signals with the low-frequency approximation of the LISA response. The GPU-accelerated MCMC pipeline presented in Ref. [96] performs parameter estimation on noise-free analytic-kludge waveforms and a simplified response function. Works that utilise the `eryn` MCMC sampler [97], which we use for comparison in Sec. VI, include Refs. [63, 64], which reconstruct a subset of parameters of noise-free frequency and time domain generated FEW signals, without including the LISA response; Ref. [20, 22, 98], which includes environmental effects; Ref. [7] which investigates EMRI parameter estimation in beyond general relativity theories; Ref [35] which includes post-adiabatic effects, and; Ref. [99] which explores strong lensing.

More recently, simulation-based inference (SBI) methods have been investigated as an alternative to likelihood-based approaches. Normalizing flows are used in Ref. [100] to reconstruct the parameters of very similar signals to those we use in this work, namely Schwarzschild FEW waveforms with the LISA response implemented as in Ref. [80], however without noise.

We will now describe our investigation of using a sequential simulation-based inference method for EMRI search and parameter estimation. To the best of our knowledge, this is the first time that sequential simulation-based methods have been used to reconstruct all 11 parameters of a Schwarzschild EMRI waveform computed with FEW, including the LISA response function and Gaussian, stationary noise according to the LISA PSD. Crucially, we also start our analysis from broad priors, see Table I, and in no sense do we initialise our pipeline close to the injected values that we want to infer. For example, Ref. [87] searched the Mock LISA Data Challenge 1.B dataset [95] for low-mass EMRIs using the following priors (we only report parameters common to our analysis): $m_1 \in \mathcal{U}[0.95, 1.05] \times 10^6 M_\odot$, $m_2 \in \mathcal{U}[9.5, 10.5] M_\odot$, and eccentricity at plunge (note we use instead initial eccentricity) $e_{pl} \in \mathcal{U}[0.15, 0.25]$. Ref. [85] use $m_1 \in \mathcal{U}[4.8, 5.4] \times 10^6 M_\odot$, $m_2 \in \mathcal{U}[9.4, 11.4] M_\odot$, and eccentricity $e \in \mathcal{U}[0.195, 0.23]$. The most similar work to ours that uses simulation-based inference [101] used $m_1 \in \mathcal{U}[0.95, 1.1] \times 10^6 M_\odot$, $m_2 \in \mathcal{U}[9, 11] M_\odot$, $e_0 \in \mathcal{U}[0.1, 0.3]$, $p_0/m_1 \in \mathcal{U}[9.1, 9.3]$ and the full domains for the four angular parameters. It is unclear how they treat the initial phases. The broadest priors implemented in the literature that we refer to in this section are those of Ref. [91], $m_1 \in \mathcal{U}[10^5, 10^8] M_\odot$, $m_2 \in \mathcal{U}[10, 100] M_\odot$, $e \in \mathcal{U}[0.005, 0.6]$, $p_0/m_1 \in \mathcal{U}[10, 12]$ and Ref. [90]: $m_1 \in \mathcal{U} \ln[10^5, 10^7] M_\odot$, $m_2 \in \mathcal{U}[5, 15] M_\odot$, $e \in \mathcal{U}[0, 0.7]$, with a fixed luminosity distance of 1 Gpc. Both of these references use full domains for the angular parameters.

V. TRUNCATED MARGINAL NEURAL RATIO ESTIMATION

We use TMNRE [61, 62] in order to run parameter estimation with wide priors on EMRI systems. Specifically, we developed a version of the `peregrine` software [56] suitable for EMRIs. The `peregrine` code was initially developed to be an optimised implementation of the TMNRE scheme for gravitational-wave signals at ground-based interferometers and was shown to perform well in the case of overlapping events [102]. Other use cases of the same scheme in the context of gravitational waves include `saqqara` [60, 103] for inferring the noise properties of LISA from stochastic background observations. See also e.g. Refs. [104–107] for other astrophysical and cosmological use cases.

At the basic level, TMNRE trains a neural network to approximate the likelihood-to-evidence ratio, which, when re-weighted by prior samples, yields an estimate of the posterior distributions over the parameters of interest. Unlike direct joint posterior estimation (via likelihood- or simulation-based approaches), TMNRE learns lower (e.g. 1-d) dimensional marginal ratios directly. This has been shown to have the potential to significantly reduce the number of required waveform evaluations compared to full joint posterior estimation followed by marginalisation [56].

From a machine-learning perspective, the training objective for TMNRE is to minimise the binary cross-entropy loss [61, 62]:

$$\begin{aligned} \mathcal{L}[\hat{\rho}_{k,\phi}] = & - \int \{ p(x, \theta_k) \ln \sigma(\hat{\rho}_{k,\phi}(x, \theta_k)) \\ & + p(x)p(\theta_k) \ln [1 - \sigma(\hat{\rho}_{k,\phi}(x, \theta_k))] \} dx d\theta_k. \end{aligned} \quad (1)$$

Here, x denotes the simulated data (including both signal and noise), θ_k represents some subset of the full parameter vector θ , σ is the sigmoid function, and ϕ denotes the parameters (weights and biases) of the neural network. In addition, $\hat{\rho}_{k,\phi}(x, \theta_k)$ represents the parameterised network trained to estimate the marginal posterior corresponding to the parameter set θ_k . In contrast to density estimation, this formulation treats the problem as a simple binary classification task, requiring the network to distinguish between samples drawn jointly $x, \theta_k \sim p(x, \theta_k)$ versus those drawn marginally $x, \theta_k \sim p(x)p(\theta)$, where here $p(x)$ is the marginal distribution of the data x , i.e. the joint distribution $p(x, \theta)$ marginalised over θ , and $p(\theta_k)$ is the prior over the marginal parameters of interest. The main motivation for this is that it can be shown analytically (by taking a functional derivative with respect to $\hat{\rho}_{k,\phi}$) that the optimal classifier $\rho_{k,\phi}^*$ is precisely $\rho_{k,\phi}^* = \ln[p(x|\theta_k)/p(x)]$ [61]. In practice, $x, \theta_k \sim p(x, \theta_k)$ is obtained by drawing a parameter sample $\theta \sim p(\theta)$, simulating the data x that corresponds to those parameter values, and then only keeping the parts of $\theta_k \subset \theta$ that one wants to perform inference

on. The samples $x, \theta_k \sim p(x)p(\theta_k)$ are obtained by the same procedure, followed by re-sampling the parameters θ_k so that the data was not generated according to the re-sampled parameter set. From a technical perspective, we minimise the loss using AdamW optimiser [108], and monitor the training and validation loss (evaluated at the end of each epoch on an independent dataset) throughout with an early stopping criteria imposed to avoid overfitting.

Simulation and training proceed in iterative rounds, with sequential truncation of the prior range [61, 62]. In each round, new simulations are generated from the current truncated prior, the network is trained, and updated posteriors (or proposal distributions) are estimated. We refer to the estimated posteriors as proposal distributions in the sense that they could inform the initial conditions of a subsequent parameter estimation pipeline, and are not to be confused with proposal distributions in the sequential SBI context. If the value of the likelihood-to-evidence ratio for a given parameter value is less than a predefined threshold (nominally 10^{-5}), the prior is truncated accordingly, and the next round begins with a refined prior. This process continues until convergence, typically defined as the point where further truncation is negligible due to significant weight persisting at the edges of the final prior. In the current setup [56], it is only possible to truncate from the outer edges of the marginal prior which, as we will see later, is sub-optimal for parameters which exhibit multi-modality. This is more of a technical hurdle than a fundamental one, however, and other options already exist to generate targeted training data including autoregressive ratio estimation algorithms [109] or active learning/truncation techniques in the context of density estimation [110, 111] which could be useful in this context.

We show the time-domain data to the network in 3 channels: the time array, the A -channel signal + noise, and the E -channel signal + noise. First it is fed through a U-Net with a similar structure to that described in [56], then the data is fed through a linear compression network which results in a vector of 16 features per parameter for each channel. Finally this feature vector is given to the logratio ($\ln[p(x|\theta_k)/p(x)]$) estimator which is a residual network [112] that performs the marginal classification as defined in `swyft` [61, 62]. For all rounds, the training and validation batch sizes are set to 128, with 90% of the simulation budget allocated for training and 10% for validation. We use a learning rate of 10^{-4} and employ an early stopping criterion of seven epochs, such that training halts if the validation loss does not decrease for seven consecutive epochs. At this point, the network weights revert to the state corresponding to the minimum validation loss, and proposal distributions are computed from this configuration. To mitigate noise-induced biases, we apply a “noise shuffling” strategy – each epoch, the noise realisations shown to the network are shuffled within each batch. This has the benefit that the network is highly unlikely to be shown the same combinations of

signal plus noise, without having to generate new noise realisations every epoch. This prevents the network from learning features of the noise as characteristics of the signals which would lead to overfitting or biased parameter inference.

VI. RESULTS

We now present parameter estimation results for EMRI waveform injections whose parameters are sampled from within the prior ranges shown in Table I. Signals are produced in the time domain, with a duration of 3.15×10^6 seconds (one tenth of a year) and a sampling cadence $dt = 10$ s. This value of dt is small enough to avoid aliasing for the values of m_1 we consider here [64]. We choose a mode content threshold of $\epsilon = 10^{-5}$ as defined in FEW, meaning that if a particular mode contributes more than a fraction of 10^{-5} of the total power emitted by all modes then it is included in the summation when the waveform is evaluated. For the fiducial system we consider here, 581 modes are included.

We use wide priors within the domain of validity of FEW, in order to understand the potential of this algorithm for both searches and parameter estimation, cf. Table I. The exceptions to this are the initial semi-latus recti, for which we choose the prior such that the simulations capture the final part of the merger, i.e. not time segments that are much earlier in the inspiral. Depending on the parameter set, some systems will merge just within the data segment, while others will not quite reach merger. We find (as expected) that the measurement precision on a given injection degrades considerably if the simulations only capture portions of the signals that are long before merger.

We are also limited by the memory budget on the HPC cluster that we use for this analysis; specifically, we run on an exclusive node with total 494 GB of memory, utilising a single NVIDIA A100 GPU with 64 GB of memory. With the current iteration of the algorithm, increasing the signal duration we choose would require different hardware. In order to accumulate a signal-to-noise ratio that is sufficiently large for parameters to be recovered successfully, sources must be relatively closeby. The prior range on d_L varies from the “closest possible” distance of an EMRI system, thought to be at $d_L \sim 87.5$ Mpc or $z \sim 0.02$ [113] up to $d_L = 0.5$ Gpc or $z \sim 0.1$. Probing larger distances requires analysing longer-duration signals, which in turn necessitate more sophisticated down-sampling and/or data-reduction schemes. See e.g. Ref. [59] for a method for simplifying and compressing binary neutron star gravitational wave data that utilises approximate information about the source. Innovations in this direction would alleviate the hardware constraints discussed above, and enable more information about the source to be extracted from the time evolution of the signal over a larger number of orbits. We expect this to improve the precision to which this pipeline is able to

measure EMRI parameters, and we leave a full investigation to future work.

A. Marginal proposal distributions

For our first application, we consider an EMRI signal with the following injected parameters: primary mass $m_1 = 5.385 \times 10^5 M_\odot$, secondary mass $m_2 = 50.55 M_\odot$, initial semi-latus rectum $p_0 = 10.35 m_1$, initial eccentricity $e = 0.2927$, cosine of the polar angle in the detector frame $\cos \theta_K = 0.0384$, azimuthal angle in the detector frame $\phi_K = 5.212$, luminosity distance to the source $d_L = 232.8$ Mpc, initial azimuthal phase $\Phi_\phi = 2.966$, initial radial phase $\Phi_r = 2.014$, cosine of the polar sky location coordinate $\cos \theta_S = 0.4107$, and azimuthal sky location coordinate $\phi_S = 3.124$. The system has $\text{SNR} \simeq 640$ and the duration of the inspiral up to merger is 3.1×10^6 s, with respect to the 3.15×10^6 s observation window. The injected signal and noise in the A -channel are shown in Fig. 1.

We show in Fig. 2 the estimated 1D marginal proposal distributions produced with six rounds of sequential TMNRE. We note that these are the estimated posterior distributions output from the TMNRE pipeline, which we call proposal distributions as we expect them to be most useful in terms of reducing the prior volume for subsequent parameter estimation pipelines. We use 150000 simulations in each round and find that the algorithm converges after six rounds.

The reconstructed proposal distributions are informative for all of the parameters except for the initial phases $\Phi_{\phi,0}$ and $\Phi_{r,0}$, where the algorithm returns the underlying uniform priors. This is likely due to the fact that we are resolving only the frequency evolution of the harmonics and not their relative phases. This also happens when secondary likelihood peaks are found due to degeneracies, as shown in Ref. [84]. The initial phases $\Phi_{\phi,0}$ and $\Phi_{r,0}$ enter each harmonic (m, n) as $m\Phi_{\phi,0} + n\Phi_{r,0}$.

In terms of precision, the relative half-widths of the proposal distributions for the intrinsic parameters, corresponding to the 2σ credible intervals, are 0.3% for m_1 , 2% for m_2 , 0.3% for p_0 and 2% for e_0 . We comment on the comparison with the relative half-widths of typical MCMC posterior distributions in Sec. VIB. We present various diagnostics of convergence in Appendix A, including $p-p$ plots and training/validation losses. Each round in this analysis took an average of 14 hours on 1 GPU, with ~ 3 hours devoted to simulations and the remainder to training the network. Inference on a single observation (inside the truncated prior region) then takes a matter of seconds.

We note here that we also conducted this analysis in the frequency domain by taking a fast Fourier transform of the signal plus noise, and found comparable results.

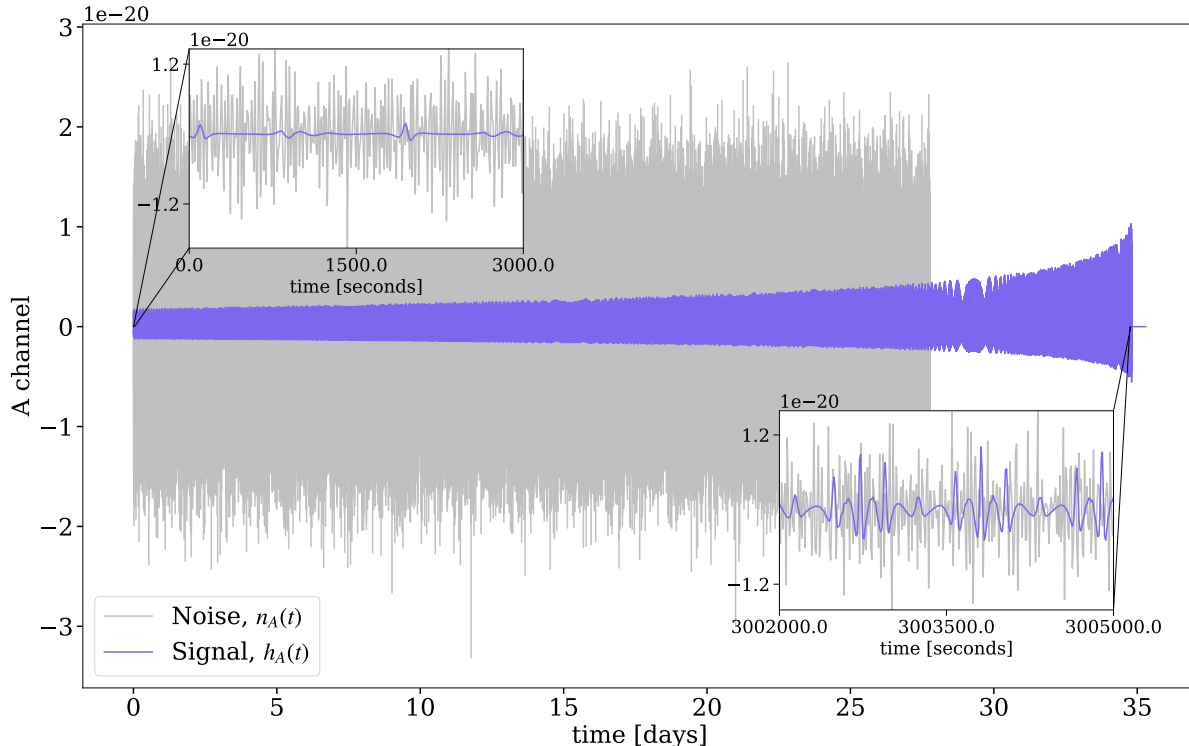


FIG. 1. Injected time-domain signal (in purple) and noise (in grey) in the TDI A-channel for the fiducial system analysed in this work which has $\text{SNR} \simeq 640$. The two insets show zoom-in regions of the signal and noise at the beginning of the segment on the left and close to plunge on the right, with durations of 3000 seconds each.

B. Comparisons

We now compare our results with those achievable with traditional likelihood-based methods in order to understand how simulation-based inference strategies could be useful in both the EMRI search and parameter-estimation space.

Firstly, we comment on the fact that the TMNRE proposal distributions are broader than some of the posteriors obtained with MCMC methods in the EMRI literature, see e.g. Refs. [64, 101, 114]. Very narrow posteriors, with relative half-widths of $10^{-3} - 10^{-2}\%$ on the intrinsic parameters $\{m_1, m_2, p_0, e_0\}$ [101], are achievable in reasonable computing time with MCMC methods if either narrow priors around the true values are used or the chains are initialised close to the true value. This allows the the region of high posterior density to be identified immediately and avoids the complicated multi-modal distribution that must otherwise be explored. In this work, since we have used wide priors and a sequential truncation method that only bounds from the upper and lower limit of the prior, the extreme multi-modality ubiquitous with EMRI systems is one aspect of the problem which could be impacting the precision we can achieve. This is because the parameter space where there is zero weight

in the proposal distribution in between two large peaks is not removed at the truncation step. We leave to future work a full investigation of a more sophisticated truncation method that would combat this multi-modality, as well as data-compression strategies that could better capture degeneracies between parameters and allow for the analysis of longer duration signals that exhibit more orbits. Improvements in these directions will help to achieve narrower proposal distributions that get closer to the true posteriors.

Note, however, that assuming wide priors as done here brings us significantly closer to a real data-analysis scenario compared to previous EMRI studies. We demonstrate the success of this pipeline for narrowing down the prior volume drastically and its possible use as part of a search strategy, as well as its performance with respect to an MCMC that is not initiated with narrow priors and/or close to the true injected values.

One strategy for narrowing down the parameter space is to identify the parameter solution vector that best maximises the log-likelihood function before running, e.g., MCMC. This can be done with an multivariate optimisation algorithm such as differential evolution [115], which is stochastic in nature and useful for searching large parameter spaces. We run differential evolution as

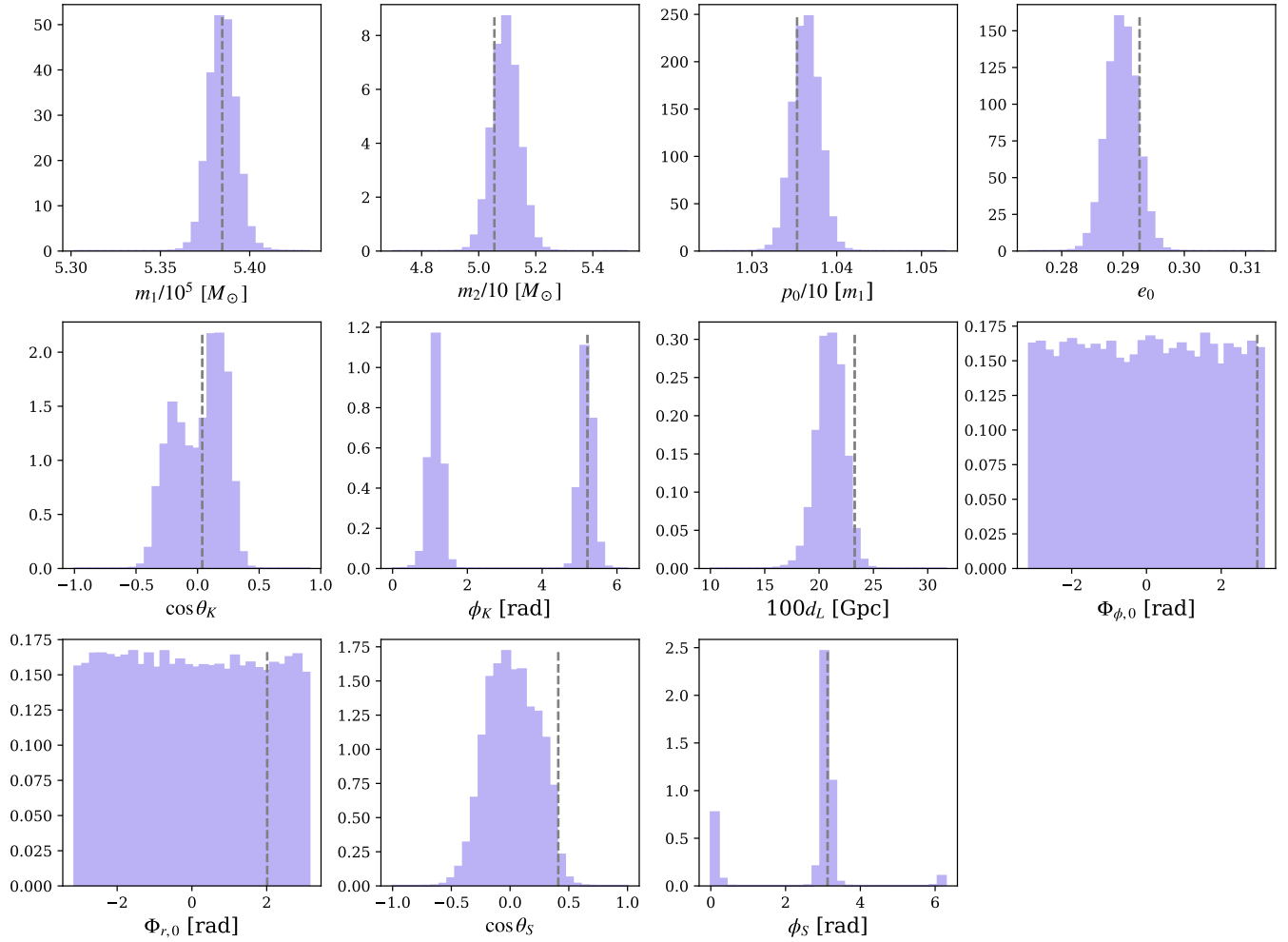


FIG. 2. Marginal proposal distributions, i.e. high probability regions identified by the neural network, for the parameters of a non-spinning EMRI in eccentric orbits. The injected signal considered is indicated by the grey dashed lines, and has $\text{SNR} \sim 640$ with masses $m_1 = 5.385 \times 10^5 M_\odot$ and $m_2 = 50.55 M_\odot$, initial semi-latus rectum $p_0 = 10.35 m_1$, initial eccentricity $e = 0.2927$, cosine of the polar angle in the detector frame $\cos \theta_K = 0.0384$, azimuthal angle in the detector frame $\phi_K = 5.212$, luminosity distance to the source $d_L = 232.8$ Mpc, initial azimuthal phase $\Phi_\phi = 2.966$, initial radial phase $\Phi_r = 2.014$, cosine of the polar sky location coordinate $\cos \theta_S = 0.4107$, and azimuthal sky location coordinate $\phi_S = 3.124$. The signal is produced and analysed in the time domain, with a duration of 3.15×10^6 s, a sampling cadence of $dt = 10$ s and is projected onto the A and E TDI channels. Noise according to the analytical LISA PSD is added. We ran TMNRE for six rounds, with 1.5×10^5 simulations in each round.

implemented in `scipy` [116] with 32 individuals drawn from the prior for 7 rounds, allowing a maximum of 150000 likelihood evaluations in each round (to match like-for-like our simulation budget). In Fig. 3, we show the bounding on each parameter after sequential rounds of simulation and training with TMNRE versus the parameter solutions found by minimising the negative log-likelihood with differential evolution. TMNRE performs very effectively, especially for m_1 , m_2 , p_0 , e and d_L . Note that its performance for the angular parameters which exhibit multi-modality in the proposal distributions is not well expressed in Fig. 3 because it does not show the low-probability regions in-between large peaks.

We can quantify the reduction in the volume of the

parameter space by comparing the initial 11-dimensional prior volume V_i with the remaining volume after 6 rounds of TMNRE V_f . The ratio V_f/V_i is 7×10^5 , meaning we have narrowed down the parameter space volume by almost a factor of a million. Differential evolution, meanwhile, is not able to identify the correct regions of the parameter space that contain the true values, likely getting stuck in local minima of the negative log-likelihood rather than finding the true global minimum. These results demonstrate a very promising use case for sequential SBI as part of a search strategy for EMRIs. We show additional examples of narrowing down the parameter space for the intrinsic values in Appendix B, including starting from an even larger prior, with a larger injected value of

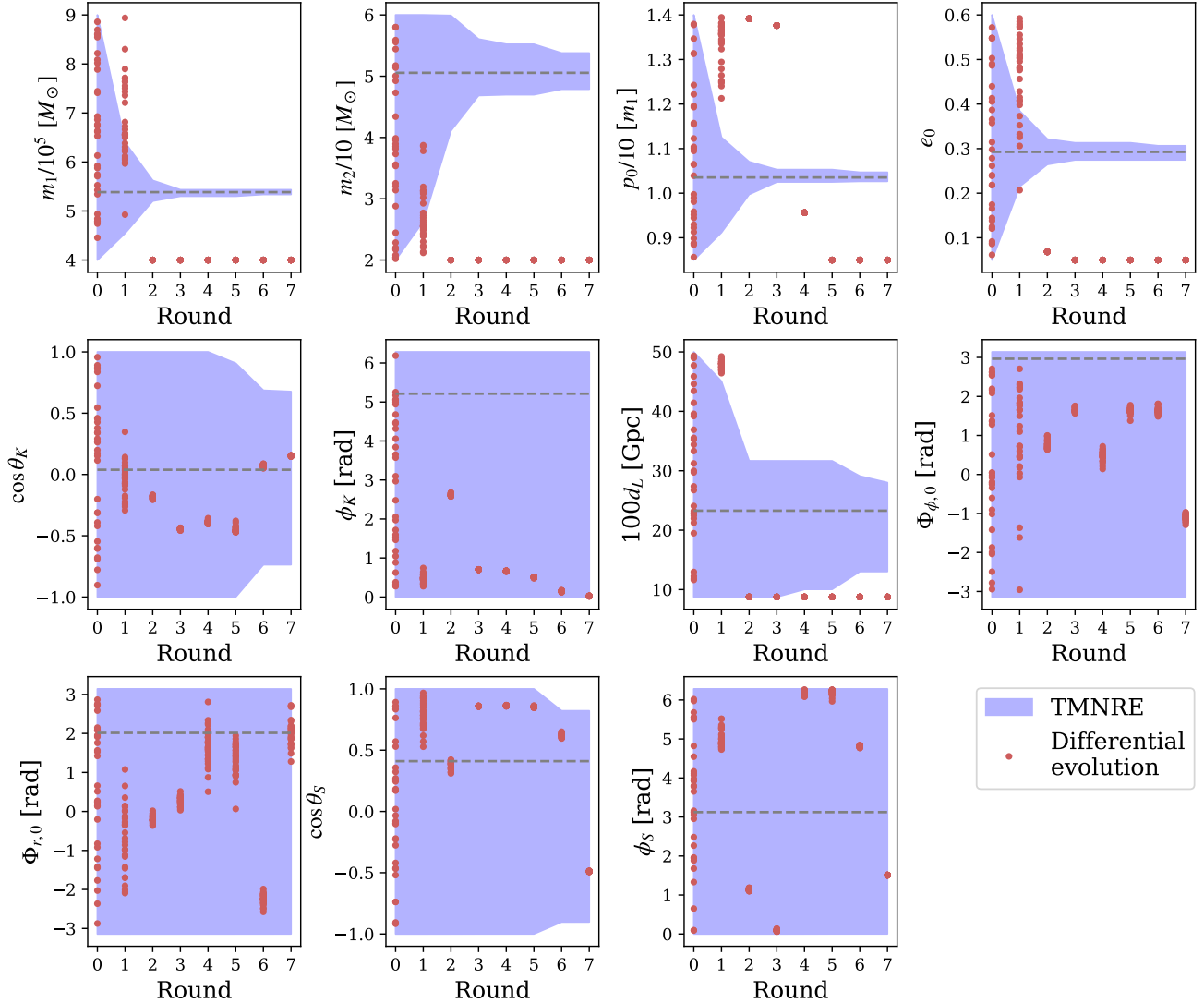


FIG. 3. Comparison of TMNRE and differential evolution for narrowing down the parameter space over 7 sequential rounds. Each round of TMNRE uses 150000 simulations and the purple shaded region is defined by the upper and lower bounds that are placed on the parameter space at the end of each round, to be used for the following round's prior. The value of the likelihood-to-evidence ratio at parameter values outside of the shaded region is less than 10^{-5} . For each round, the red points mark the solutions found by the differential-evolution algorithm for 32 population members. We limit the number of function evaluations in each round to be a maximum of 150000. The true injected values shown by the grey dashed lines are the same as for the system analysed in Fig. 2.

m_1 and lower p_0 .

We now compare our results against those from a traditional MCMC approach in a regime where we can put the two methods on an equal footing. We run an MCMC as implemented in `eryn` [97] with the affine-invariant proposal `StretchMove` [117, 118], using the truncated priors from the sixth round of our TMNRE pipeline, with 32 walkers and a step limit of 4,687, ensuring that the number of likelihood evaluations remains (conservatively) comparable to that of the final TMNRE round (150,000 simulations). We show in Fig. 4 that with these

much wider priors than are usually used for MCMC and with the same number of waveform evaluations as the TMNRE pipeline, the MCMC chains do not converge. For many of the parameters, TMNRE does significantly better at estimating the proposal distributions. This MCMC run took 49 hours on a GPU, compared to 14 hours for the final round of TMNRE.

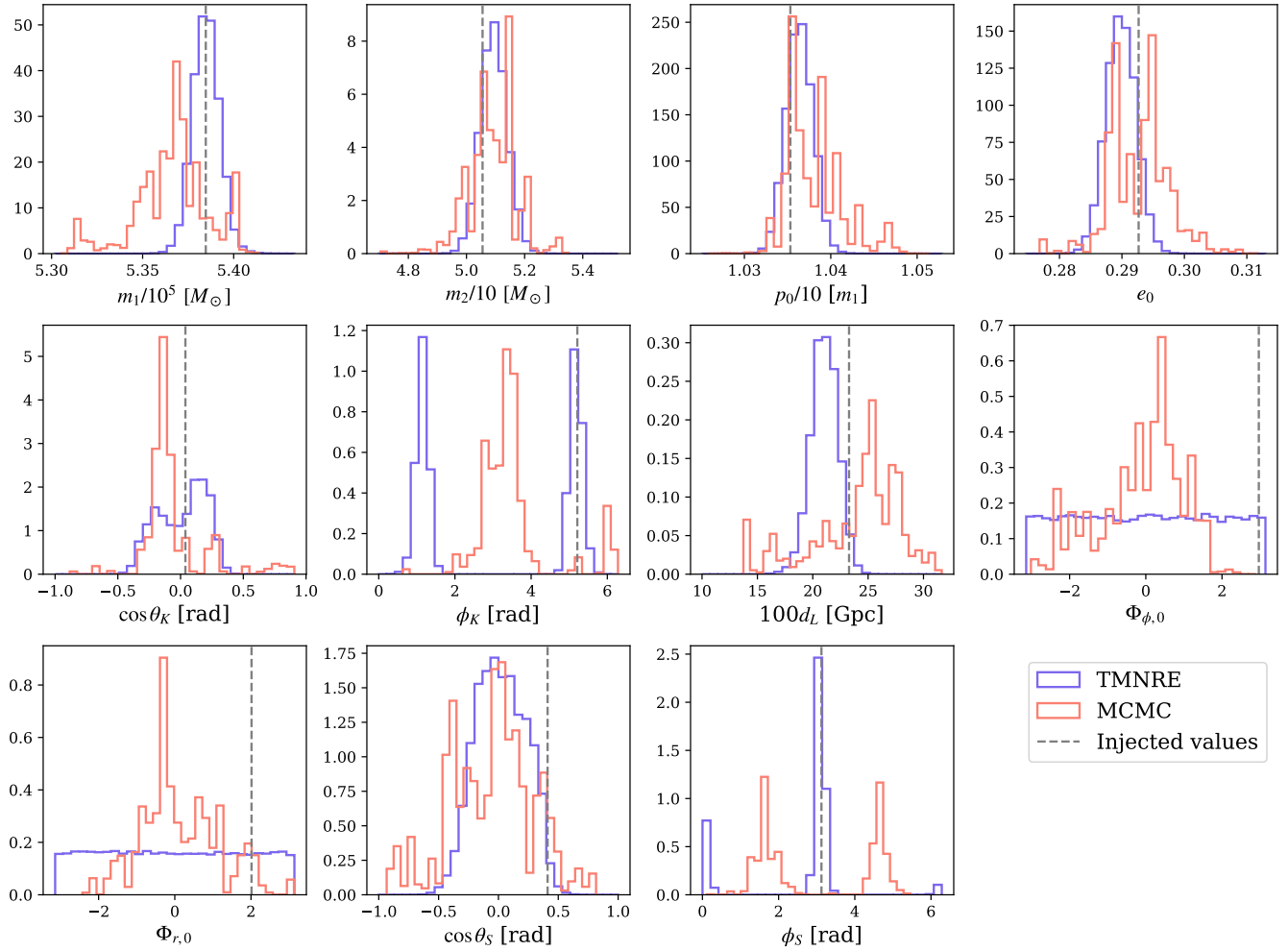


FIG. 4. Comparison of TMNRE and MCMC approaches using the same priors identified by the truncation after the 5th round of inference by TMNRE. The blue distributions show the estimated proposal distributions from the 6th round of TMNRE, trained on 150000 simulations. The red distributions are computed from an MCMC run where the starting points are initialised randomly from the prior, with 32 walkers and 4687 steps. This conservatively sets the number of waveform evaluations to be approximately 150000 such that it is a fair comparison with the 6th round of TMNRE. On inspecting the MCMC chains, they are clearly far from convergence. The true injected values are shown by the grey dashed lines.

C. Additional results

It is also possible to estimate the higher-dimensional marginal proposal distributions with this method. Since each n D marginal is trained independently, there is no guarantee that given a particular network architecture, the proposal distributions will be consistent with each other given that they are not lower-dimensional projections of a full joint posterior. However, having zoomed in on the parameter space with 6 rounds of truncation, we train the 2D marginals for the masses and eccentricities, and show in Fig. 5 that despite being trained independently of each other, they are consistent with having been projected from the full joint distribution. We can quantify this by checking the relative difference between the 2σ credible intervals for the 1d and 2d distributions, and

find that for m_1 the difference is less than $50 M_\odot$ for each distribution, for m_2 it is less than $0.1 M_\odot$ and for e_0 it is less than 0.002. The 2D marginals were trained on the same set of simulations as were used to produce the 1D marginals in Fig. 2, and the training and validation losses are summed over all marginals.

However, we caution that our set-up appears to break down for some of the other parameters, especially those exhibiting multi-modality in their 1D marginals, resulting in 2D marginals that are much broader than their 1D counterparts. This may require different summaries to be shown to the n D logratio estimator, and we leave a full diagnosis of this issue to future work. We stress that the marginal aspect of this pipeline has many benefits in terms of simulation-efficiency when narrowing down the parameter space as the number of simulations

required to reach a particular level of precision scales more favourably with the number of parameters than likelihood-based methods. This could also be extremely valuable when considering the more complicated LISA data analysis problem of dealing with many overlapping sources and non-stationary, non-Gaussian noise. Not only will the dimensionality of the problem increase dramatically, but there will also likely be many nuisance parameters that appear in the forward model. Marginal neural ratio estimation enables effortless marginalisation over nuisance parameters [61] that would otherwise be a bottleneck for many methods that estimate the full joint posterior.

VII. CONCLUSIONS AND OUTLOOK

We have explored the use of sequential simulation-based inference for search and parameter estimation of extreme mass ratio inspirals in future space-based gravitational wave detector data. We showed that TMNRE can be employed as a tool to very efficiently narrow down the search parameter space, meaning it could form a useful part of the search pipeline. In its current form, this method also estimates informative proposal distributions for all parameters, with the exception of the initial phases of the system.

We found that the proposals estimated by this set-up are much more informative than those achievable with MCMC, if compared over the same number of simulations or waveform evaluations with the MCMC chains initialised from the same final round priors as the TMNRE approach. However, we note that the extremely narrow posteriors that are in principle realisable for EMRIs, seen, for example, when MCMC chains are initialised very close to the true value, were not achieved in this TMNRE set-up. We believe this to be in part due to the truncation scheme saturating conservatively early in the presence of multi-modal proposal distributions, as well as the need for a more sophisticated data-compression scheme in order for the network to identify differences between waveforms that are dominated by noise. To expand on this point, we believe that the saturation in the sensitivity could be attributed to the dimensionality of the data input. In particular, in the current form, the network must learn subtle features of the signal vs. noise distribution in each data bin. For a large number of bins, this requires a very large number of data examples and potentially significantly larger network capacity, even if the information could be obtained from a significantly lower dimensional summary. This motivates exploring pre-conditioned data summaries for these long duration signals. This will also be vital for handling longer and more distant signals without compromising on sampling rate, which has been shown to be important for EMRI data analysis [64]. Furthermore, a different parametrisation of the search space might reduce the multimodality of the problem and hence improve the accuracy of this

setup. We also comment here on the other SBI options that are available beyond the TMNRE setup. For example, after using the current algorithm to efficiently narrow down the parameter space, the learned proposal distribution could be used directly in a density estimation (such as neural posterior estimation [119]) or a higher-dimensional autoregressive estimation setup. This would allow us to maintain the benefits of simulation efficiency in the search phase, but obtain full information about the joint posterior in later stages of the pipeline. Additionally, the TMNRE proposal could be used to inform some form of data preconditioning (see e.g. Ref. [59]) to optimise the compression step that we argue is required to obtain full sensitivity.

On the signal modeling side, our current analysis focused on Schwarzschild EMRIs in vacuum, but the next step is to include higher-order effects, such as spin, which introduce additional complexities to the waveform. Given the flexibility of SBI methods, which only require simulation data, incorporating these effects should be straightforward in terms of implementation. We plan to test our framework on the upcoming version 2.0.0 of FEW that models Kerr EMRI systems [7, 63].

We have assumed Gaussian noise for simplicity in this work. However, real LISA data will exhibit non-stationary and non-Gaussian noise features due to instrument drifts [40], gaps in data acquisition [46], and overlapping signals. This aspect of the LISA data-analysis problem is where SBI could prove to be extremely powerful, since computing a likelihood for an EMRI signal that incorporates these noise complications is challenging. On the other hand, SBI solely requires a computational recipe to simulate the noise, which can then be fed directly into the neural network in the way we have described for the Gaussian noise in this work. We would like to explore the potential that this framework has for combatting this issue, as it could be uniquely useful in this EMRI regime.

Ultimately, the goal is to integrate EMRI data analysis into a broader LISA global fit [39, 120–122]. This involves searching for and measuring the parameters of multiple populations of sources simultaneously. The SBI approach explored here has the potential to play a role in such a pipeline, either by providing refined proposal distributions for EMRIs after an initial search stage or by contributing directly to a hierarchical global fit.

More broadly, our pipeline can serve as a building block for further investigation into the use of sequential simulation-based inference for space-based gravitational wave data. Future improvements will incorporate increasing complexity on both the signal and noise modeling sides, as well as dealing with the fact that EMRIs are just one species of the broader population of systems that will be present in the data.

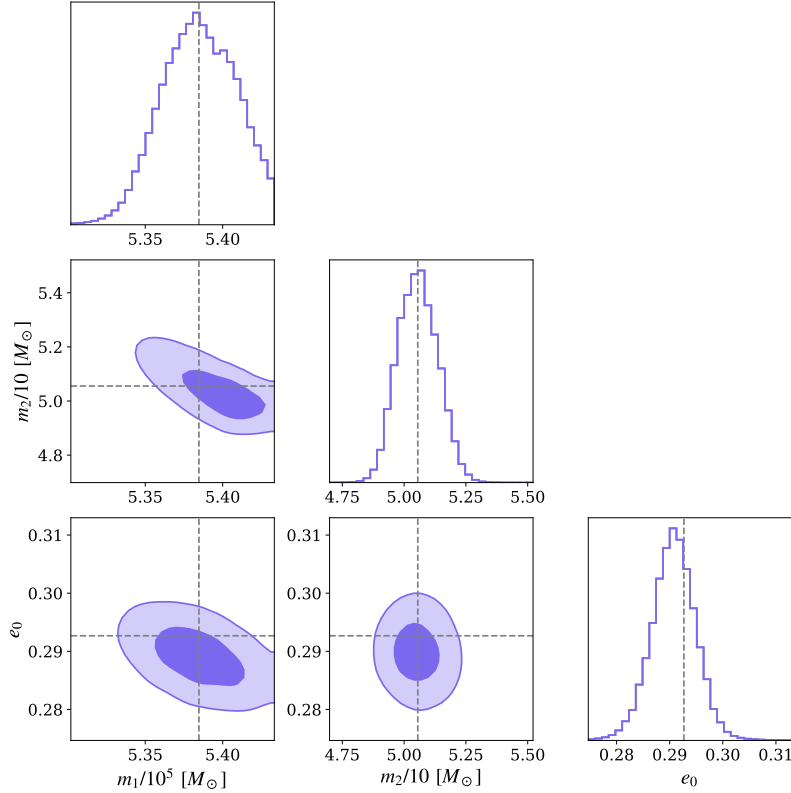


FIG. 5. 2d marginal proposal distributions for the black-hole masses and eccentricity, trained in parallel with the 1d marginals. The light shaded regions are the 2σ contours while the dark shaded regions are the 1σ contours. We used the same 1.5×10^5 simulations from the 6th round of TMNRE that produced the 1d marginals in Fig. 2. The true injected values are shown by the grey dashed lines.

ACKNOWLEDGMENTS

P.C. and D.G. are supported by ERC Starting Grant No. 945155–GWmining, Cariplo Foundation Grant No. 2021-0555, MUR PRIN Grant No. 2022-Z9X4XS, MUR Grant “Progetto Dipartimenti di Eccellenza 2023-2027” (BiCoQ), and the ICSC National Research Centre funded by NextGenerationEU. D.G. is supported by MSCA Fellowship No. 101064542–StochRewind, MSCA

Fellowship No. 101149270–ProtoBH, and MUR Young Researchers Grant No. SOE2024-0000125. Computational work was performed at CINECA with allocations through INFN, Bicocca, and an ISCRA Class C project grant. J.A. is supported by a fellowship from the Kavli Foundation. This work is part of a project that has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 864035).

-
- [1] M. Colpi *et al.*, LISA Definition Study Report (2024), [arXiv:2402.07571 \[astro-ph.CO\]](#).
 - [2] J. Luo *et al.* (TianQin), *Class. Quantum Grav.* **33**, 035010 (2016), [arXiv:1512.02076 \[astro-ph.IM\]](#).
 - [3] W.-R. Hu and Y.-L. Wu, *Natl. Sci. Rev.* **4**, 685 (2017).
 - [4] S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. L. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein, *Phys. Rev. D* **95**, 103012 (2017), [arXiv:1703.09722 \[gr-qc\]](#).
 - [5] W.-B. Han and X. Chen, *Mon. Not. R. Astron. Soc.* **485**, L29 (2019), [arXiv:1801.07060 \[gr-qc\]](#).
 - [6] P. Gupta, L. Speri, B. Bonga, A. J. K. Chua, and T. Tanaka, *Phys. Rev. D* **106**, 104001 (2022), [arXiv:2205.04808 \[gr-qc\]](#).
 - [7] L. Speri, S. Barsanti, A. Maselli, T. P. Sotiriou, N. Warburton, M. van de Meent, A. J. K. Chua, O. Burke, and J. Gair, (2024), [arXiv:2406.07607 \[gr-qc\]](#).
 - [8] S. Kejriwal, L. Speri, and A. J. K. Chua, *Phys. Rev. D* **110**, 084060 (2024), [arXiv:2312.13028 \[gr-qc\]](#).
 - [9] A. Cárdenas-Avendaño and C. F. Sopuerta, in *Recent Progress on Gravity Tests* (Springer, 2024) [arXiv:2401.08085 \[gr-qc\]](#).

- [10] K. Eda, Y. Itoh, S. Kuroyanagi, and J. Silk, *Phys. Rev. Lett.* **110**, 221101 (2013), [arXiv:1301.5971 \[gr-qc\]](#).
- [11] C. F. B. Macedo, P. Pani, V. Cardoso, and L. C. B. Crispino, *Astrophys. J.* **774**, 48 (2013), [arXiv:1302.2646 \[gr-qc\]](#).
- [12] B. J. Kavanagh, D. A. Nichols, G. Bertone, and D. Gaggero, *Phys. Rev. D* **102**, 083006 (2020), [arXiv:2002.12811 \[gr-qc\]](#).
- [13] H. Khalvati, A. Santini, F. Duque, L. Speri, J. Gair, H. Yang, and R. Brito, *Phys. Rev. D* **111**, 082010 (2025), [arXiv:2410.17310 \[gr-qc\]](#).
- [14] R. Vicente, T. K. Karydas, and G. Bertone, (2025), [arXiv:2505.09715 \[gr-qc\]](#).
- [15] D. Baumann, G. Bertone, J. Stout, and G. M. Tomaselli, *Phys. Rev. D* **105**, 115036 (2022), [arXiv:2112.14777 \[gr-qc\]](#).
- [16] S. Barsanti, A. Maselli, T. P. Sotiriou, and L. Gualtieri, *Phys. Rev. Lett.* **131**, 051401 (2023), [arXiv:2212.03888 \[gr-qc\]](#).
- [17] C. Zhang, Y. Gong, D. Liang, and B. Wang, *J. Cosmology Astropart. Phys.* **06**, 054 (2023), [arXiv:2210.11121 \[gr-qc\]](#).
- [18] G. M. Tomaselli, T. F. M. Spieksma, and G. Bertone, *Phys. Rev. Lett.* **133**, 121402 (2024), [arXiv:2407.12908 \[gr-qc\]](#).
- [19] P. S. Cole, G. Bertone, A. Coogan, D. Gaggero, T. Karydas, B. J. Kavanagh, T. F. M. Spieksma, and G. M. Tomaselli, *Nat. Astron.* **7**, 943 (2023), [arXiv:2211.01362 \[gr-qc\]](#).
- [20] L. Speri, A. Antonelli, L. Sberna, S. Babak, E. Barausse, J. R. Gair, and M. L. Katz, *Phys. Rev. X* **13**, 021035 (2023), [arXiv:2207.10086 \[gr-qc\]](#).
- [21] F. Duque, S. Kejriwal, L. Sberna, L. Speri, and J. Gair, *Phys. Rev. D* **111**, 084006 (2025), [arXiv:2411.03436 \[gr-qc\]](#).
- [22] L. Copparoni, L. Speri, L. Sberna, A. Derdzinski, and E. Barausse, (2025), [arXiv:2502.10087 \[gr-qc\]](#).
- [23] M. Volonteri, *Astron. Astrophys. Rev.* **18**, 279 (2010), [arXiv:1003.4404 \[astro-ph.CO\]](#).
- [24] N. A. Wittek, A. Pound, H. P. Pfeiffer, and L. Barack, *Phys. Rev. D* **110**, 084023 (2024), [arXiv:2403.08864 \[gr-qc\]](#).
- [25] S. A. Hughes, N. Warburton, G. Khanna, A. J. K. Chua, and M. L. Katz, *Phys. Rev. D* **103**, 104014 (2021), [Erratum: *Phys. Rev. D* 107, 089901 (2023)], [arXiv:2102.02713 \[gr-qc\]](#).
- [26] A. Buonanno and T. Damour, *Phys. Rev. D* **59**, 084006 (1999), [arXiv:gr-qc/9811091](#).
- [27] N. Yunes, A. Buonanno, S. A. Hughes, Y. Pan, E. Barausse, M. C. Miller, and W. Thorne, *Phys. Rev. D* **83**, 044044 (2011), [Erratum: *Phys. Rev. D* 88, 109904 (2013)], [arXiv:1009.6013 \[gr-qc\]](#).
- [28] A. Albertini, R. Gamba, A. Nagar, and S. Bernuzzi, *Phys. Rev. D* **109**, 044022 (2024), [arXiv:2310.13578 \[gr-qc\]](#).
- [29] E. Poisson, A. Pound, and I. Vega, *Living Rev. Relativ.* **14**, 7 (2011), [arXiv:1102.0529 \[gr-qc\]](#).
- [30] A. Pound, *Fund. Theor. Phys.* **179**, 399 (2015), [arXiv:1506.06245 \[gr-qc\]](#).
- [31] B. Wardell, *Fund. Theor. Phys.* **179**, 487 (2015), [arXiv:1501.07322 \[gr-qc\]](#).
- [32] L. Barack and A. Pound, *Rep. Prog. Phys.* **82**, 016904 (2019), [arXiv:1805.10385 \[gr-qc\]](#).
- [33] A. Pound, B. Wardell, N. Warburton, and J. Miller, *Phys. Rev. Lett.* **124**, 021101 (2020), [arXiv:1908.07419 \[gr-qc\]](#).
- [34] A. Pound and B. Wardell, in *Handbook of Gravitational Wave Astronomy* (Springer, 2021) [arXiv:2101.04592 \[gr-qc\]](#).
- [35] O. Burke, G. A. Piovano, N. Warburton, P. Lynch, L. Speri, C. Kavanagh, B. Wardell, A. Pound, L. Durkan, and J. Miller, *Phys. Rev. D* **109**, 124048 (2024), [arXiv:2310.08927 \[gr-qc\]](#).
- [36] L. Speri, N. Karnesis, A. I. Renzini, and J. R. Gair, *Nat. Astron.* **6**, 1356 (2022).
- [37] K. Dey, N. Karnesis, A. Toubiana, E. Barausse, N. Korsakova, Q. Baghi, and S. Basak, *Phys. Rev. D* **104**, 044035 (2021), [arXiv:2104.12646 \[gr-qc\]](#).
- [38] Q. Baghi, N. Korsakova, J. Slutsky, E. Castelli, N. Karnesis, and J.-B. Bayle, *Phys. Rev. D* **105**, 042002 (2022), [arXiv:2112.07490 \[gr-qc\]](#).
- [39] T. B. Littenberg and N. J. Cornish, *Phys. Rev. D* **107**, 063004 (2023), [arXiv:2301.03673 \[gr-qc\]](#).
- [40] A. Spadaro, R. Busicchio, D. Vetrugno, A. Klein, S. Vitale, R. Dolesi, W. J. Weber, and M. Colpi, *Phys. Rev. D* **108**, 123029 (2023), [arXiv:2306.03923 \[gr-qc\]](#).
- [41] N. Houbba, L. Ferraioli, and D. Giardini, *Phys. Rev. D* **109**, 083027 (2024), [arXiv:2401.00846 \[astro-ph.IM\]](#).
- [42] N. Houbba, J.-B. Bayle, and M. Vallisneri, (2024), [arXiv:2412.20793 \[astro-ph.IM\]](#).
- [43] O. Sauter, P. Wass, W. Sanchez, and H. Inchauspé, (2025), [arXiv:2501.06315 \[gr-qc\]](#).
- [44] R. Mao, J. E. Lee, and M. C. Edwards, *Phys. Rev. D* **111**, 024067 (2025), [arXiv:2410.05571 \[gr-qc\]](#).
- [45] E. Castelli, Q. Baghi, J. G. Baker, J. Slutsky, J. Bobin, N. Karnesis, A. Petiteau, O. Sauter, P. Wass, and W. J. Weber, *Class. Quantum Grav.* **42**, 065018 (2025), [arXiv:2411.13402 \[gr-qc\]](#).
- [46] O. Burke, S. Marsat, J. R. Gair, and M. L. Katz, (2025), [arXiv:2502.17426 \[gr-qc\]](#).
- [47] J. Crowder and N. J. Cornish, *Phys. Rev. D* **70**, 082004 (2004), [arXiv:gr-qc/0404129](#).
- [48] A. Antonelli, O. Burke, and J. R. Gair, *Mon. Not. R. Astron. Soc.* **507**, 5069 (2021), [arXiv:2104.01897 \[gr-qc\]](#).
- [49] S. H. Strub, L. Ferraioli, C. Schmelsbach, S. C. Stähler, and D. Giardini, *Phys. Rev. D* **106**, 062003 (2022), [arXiv:2204.04467 \[astro-ph.IM\]](#).
- [50] R. Niu and W. Zhao, *Fundamental Research* **11**, 007 (2024), [arXiv:2406.12789 \[gr-qc\]](#).
- [51] K. Chatziioannou, T. Dent, M. Fishbach, F. Ohme, M. Pürrer, V. Raymond, and J. Veitch, (2024), [arXiv:2409.02037 \[gr-qc\]](#).
- [52] J. R. Gair, L. Barack, T. Creighton, C. Cutler, S. L. Larson, E. S. Phinney, and M. Vallisneri, *Class. Quantum Grav.* **21**, S1595 (2004), [arXiv:gr-qc/0405137](#).
- [53] C. J. Moore, D. Gerosa, and A. Klein, *Mon. Not. R. Astron. Soc.* **488**, L94 (2019), [arXiv:1905.11998 \[astro-ph.HE\]](#).
- [54] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, *Phys. Rev. Lett.* **127**, 241103 (2021), [arXiv:2106.12594 \[gr-qc\]](#).
- [55] M. Crisostomi, K. Dey, E. Barausse, and R. Trotta, *Phys. Rev. D* **108**, 044029 (2023), [arXiv:2305.18528 \[gr-qc\]](#).
- [56] U. Bhardwaj, J. Alvey, B. K. Miller, S. Nissanke, and C. Weniger, *Phys. Rev. D* **108**, 042004 (2023), [arXiv:2304.02035 \[gr-qc\]](#).

- [57] V. Raymond, S. Al-Shammari, and A. Göttel, (2024), [arXiv:2406.03935 \[gr-qc\]](#).
- [58] C. Pacilio, S. Bhagwat, and R. Cotesta, *Phys. Rev. D* **110**, 083010 (2024), [arXiv:2404.11373 \[gr-qc\]](#).
- [59] M. Dax, S. R. Green, J. Gair, N. Gupte, M. Pürner, V. Raymond, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, *Nature* **639**, 49 (2025), [arXiv:2407.09602 \[gr-qc\]](#).
- [60] J. Alvey, U. Bhardwaj, V. Domcke, M. Pieroni, and C. Weniger, *Phys. Rev. D* **109**, 083008 (2024), [arXiv:2309.07954 \[gr-qc\]](#).
- [61] B. K. Miller, A. Cole, P. Forré, G. Louppe, and C. Weniger, in *35th Conference on Neural Information Processing Systems* (2021) [arXiv:2107.01214 \[stat.ML\]](#).
- [62] B. K. Miller, A. Cole, C. Weniger, F. Nattino, O. Ku, and M. W. Grootes, *J. Open Source Softw.* **7**, 4205 (2022).
- [63] M. L. Katz, A. J. K. Chua, L. Speri, N. Warburton, and S. A. Hughes, *Phys. Rev. D* **104**, 064047 (2021), [arXiv:2104.04582 \[gr-qc\]](#).
- [64] L. Speri, M. L. Katz, A. J. K. Chua, S. A. Hughes, N. Warburton, J. E. Thompson, C. E. A. Chapman-Bird, and J. R. Gair, (2023), [10.3389/fams.2023.1266739](#), [arXiv:2307.12585 \[gr-qc\]](#).
- [65] A. J. K. Chua, M. L. Katz, N. Warburton, and S. A. Hughes, *Phys. Rev. Lett.* **126**, 051102 (2021), [arXiv:2008.06071 \[gr-qc\]](#).
- [66] M. L. Katz, L. Speri, A. J. K. Chua, C. E. A. Chapman-Bird, N. Warburton, and S. A. Hughes, [10.5281/zenodo.8190418](#) (2023).
- [67] L. Barack and C. Cutler, *Phys. Rev. D* **69**, 082005 (2004), [arXiv:gr-qc/0310125](#).
- [68] A. J. Chua, C. J. Moore, and J. R. Gair, *Phys. Rev. D* **96**, 044005 (2017), [arXiv:1705.04259 \[gr-qc\]](#).
- [69] A. J. Chua and J. R. Gair, *Class. Quant. Grav.* **32**, 232002 (2015), [arXiv:1510.06245 \[gr-qc\]](#).
- [70] L. C. Stein and N. Warburton, *Phys. Rev. D* **101**, 064007 (2020), [arXiv:1912.07609 \[gr-qc\]](#).
- [71] R. Fujita and M. Shibata, *Phys. Rev. D* **102**, 064005 (2020), [arXiv:2008.13554 \[gr-qc\]](#).
- [72] A. J. Chua, C. R. Galley, and M. Vallisneri, *Phys. Rev. Lett.* **122**, 211101 (2019), [arXiv:1811.05491 \[astro-ph.IM\]](#).
- [73] S. E. Field, C. R. Galley, F. Herrmann, J. S. Hesthaven, E. Ochsner, and M. Tiglio, *Phys. Rev. Lett.* **106**, 221102 (2011), [arXiv:1101.3765 \[gr-qc\]](#).
- [74] M. Tinto and S. V. Dhurandhar, *Living Rev. Relativ.* **8**, 4 (2005), [arXiv:gr-qc/0409034](#).
- [75] J. W. Armstrong, F. B. Estabrook, and M. Tinto, *Astrophys. J.* **527**, 814 (1999).
- [76] M. Tinto and O. Hartwig, *Phys. Rev. D* **98**, 042003 (2018), [arXiv:1807.02594 \[gr-qc\]](#).
- [77] M. Tinto, F. B. Estabrook, and J. W. Armstrong, *Phys. Rev. D* **69**, 082001 (2004), [arXiv:gr-qc/0310017](#).
- [78] M. Muratore, O. Hartwig, D. Vetrugno, S. Vitale, and W. J. Weber, *Phys. Rev. D* **107**, 082004 (2023), [arXiv:2207.02138 \[gr-qc\]](#).
- [79] O. Hartwig, M. Lilley, M. Muratore, and M. Pieroni, *Phys. Rev. D* **107**, 123531 (2023), [arXiv:2303.15929 \[gr-qc\]](#).
- [80] M. L. Katz, J.-B. Bayle, A. J. K. Chua, and M. Vallisneri, *Phys. Rev. D* **106**, 103001 (2022), [arXiv:2204.06633 \[gr-qc\]](#).
- [81] W. Martens and E. Joffre, *J. Astronaut. Sci.* **68**, 402 (2021), [arXiv:2101.03040 \[gr-qc\]](#).
- [82] S. Babak and A. Petiteau, (2021), <https://lisa-ldc.lal.in2p3.fr/static/data/pdf/LDC-manual-002.pdf>.
- [83] A. Nitz, I. Harry, D. Brown, C. M. Biwer, and J. Willis, [10.5281/zenodo.10013996](#) (2023).
- [84] A. J. K. Chua and C. J. Cutler, *Phys. Rev. D* **106**, 124046 (2022), [arXiv:2109.14254 \[gr-qc\]](#).
- [85] N. J. Cornish, *Class. Quant. Grav.* **28**, 094016 (2011), [arXiv:0804.3323 \[gr-qc\]](#).
- [86] J. R. Gair, E. K. Porter, S. Babak, and L. Barack, *Class. Quantum Grav.* **25**, 184030 (2008), [arXiv:0804.3322 \[gr-qc\]](#).
- [87] S. Babak, J. R. Gair, and E. K. Porter, *Class. Quantum Grav.* **26**, 135004 (2009), [arXiv:0902.4133 \[gr-qc\]](#).
- [88] A. J. K. Chua, *Phys. Rev. D* **106**, 104051 (2022), [arXiv:2205.08702 \[gr-qc\]](#).
- [89] Y. Wang, Y. Shang, S. Babak, Y. Shang, and S. Babak, *Phys. Rev. D* **86**, 104050 (2012), [arXiv:1207.4956 \[gr-qc\]](#).
- [90] C.-Q. Ye, H.-M. Fan, A. Torres-Orjuela, J.-d. Zhang, and Y.-M. Hu, *Phys. Rev. D* **109**, 124034 (2024), [arXiv:2310.03520 \[gr-qc\]](#).
- [91] Q. Yun, W.-B. Han, Y.-Y. Guo, H. Wang, and M. Du, *Sci. China Phys. Mech. Astron.* **68**, 210413 (2025), [arXiv:2311.18640 \[gr-qc\]](#).
- [92] X.-T. Zhang, C. Messenger, N. Korsakova, M. L. Chan, Y.-M. Hu, and J.-d. Zhang, *Phys. Rev. D* **105**, 123027 (2022), [arXiv:2202.07158 \[astro-ph.HE\]](#).
- [93] C. Badger, J. A. Font, M. Sakellariadou, and A. Torres-Forné, *Phys. Rev. D* **110**, 064074 (2024), [arXiv:2407.02908 \[gr-qc\]](#).
- [94] A. Ali, N. Christensen, R. Meyer, and C. Rover, *Class. Quantum Grav.* **29**, 145014 (2012), [arXiv:1301.0455 \[gr-qc\]](#).
- [95] S. Babak *et al.*, *Class. Quantum Grav.* **25**, 184026 (2008), [arXiv:0806.2110 \[gr-qc\]](#).
- [96] I. D. Saltas and R. Oliveri, *SciPost Phys. Codebases*, **44** (2023), [arXiv:2311.17174 \[gr-qc\]](#).
- [97] N. Karnesis, M. L. Katz, N. Korsakova, J. R. Gair, and N. Stergioulas, *Mon. Not. R. Astron. Soc.* **526**, 4814 (2023), [arXiv:2303.02164 \[astro-ph.IM\]](#).
- [98] S. Kejriwal, F. Duque, A. J. K. Chua, and J. Gair, (2025), [arXiv:2503.01120 \[gr-qc\]](#).
- [99] M. Toscani, O. Burke, C. Liu, N. B. Zamel, N. Tamanini, and F. Pozzoli, *Phys. Rev. D* **109**, 063505 (2024), [arXiv:2307.06722 \[astro-ph.CO\]](#).
- [100] B. Liang, H. Guo, T. Zhao, H. wang, H. Evangelinelis, Y. Xu, C. liu, M. Liang, X. Wei, Y. Yuan, P. Xu, M. Du, W.-L. Qian, and Z. Luo, (2024), [arXiv:2409.07957 \[physics.comp-ph\]](#).
- [101] Y. Liu, T. Zhao, and T. Xiang, *Phys. Rev. B* **110**, 195137 (2024), [arXiv:2406.11195 \[cond-mat.str-el\]](#).
- [102] J. Alvey, U. Bhardwaj, S. Nissanke, and C. Weniger, (2023), [arXiv:2308.06318 \[gr-qc\]](#).
- [103] J. Alvey, U. Bhardwaj, V. Domcke, M. Pieroni, and C. Weniger, (2024), [arXiv:2408.00832 \[gr-qc\]](#).
- [104] G. Franco Abellán, G. C. n. Herrera, M. Martinelli, O. Savchenko, D. Sciotti, and C. Weniger, *J. Cosmology Astropart. Phys.* **11**, 057 (2024), [arXiv:2403.14750 \[astro-ph.CO\]](#).
- [105] J. Alvey, M. Gerdes, and C. Weniger, *Mon. Not. R. Astron. Soc.* **525**, 3662 (2023), [arXiv:2304.02032 \[astro-ph.GA\]](#).

- [106] N. Anau Montel and C. Weniger, in *Conference on Neural Information Processing Systems* (2022) [arXiv:2211.04291 \[astro-ph.IM\]](#).
- [107] A. Cole, B. K. Miller, S. J. Witte, M. X. Cai, M. W. Grootes, F. Nattino, and C. Weniger, *J. Cosmology Astropart. Phys.* **09**, 004 (2022), [arXiv:2111.08030 \[astro-ph.CO\]](#).
- [108] I. Loshchilov and F. Hutter, in *International Conference on Learning Representations* (2019) [arXiv:1711.05101 \[cs.LG\]](#).
- [109] N. Anau Montel, J. Alvey, and C. Weniger, *Mon. Not. R. Astron. Soc.* **530**, 4107 (2024), [arXiv:2308.08597 \[astro-ph.IM\]](#).
- [110] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, *Mon. Not. R. Astron. Soc.* **488**, 4440 (2019), [arXiv:1903.00007 \[astro-ph.CO\]](#).
- [111] M. Deistler, P. J. Goncalves, and J. H. Macke, in *Conference on Neural Information Processing Systems* (2022) [arXiv:2210.04815 \[stat.ML\]](#).
- [112] K. He, X. Zhang, S. Ren, and J. Sun, (2015), [10.1109/CVPR.2016.90](#), [arXiv:1512.03385 \[cs.CV\]](#).
- [113] J. R. Gair, *Class. Quant. Grav.* **26**, 094034 (2009), [arXiv:0811.0188 \[gr-qc\]](#).
- [114] I. Sáez-Casares, Y. Rasera, T. R. G. Richardson, and P. S. Corasaniti, *Astron. Astrophys.* **691**, A323 (2024), [arXiv:2410.05226 \[astro-ph.CO\]](#).
- [115] R. Storn and K. Price, *J. Glob. Optim.* **11**, 341 (1997).
- [116] P. Virtanen *et al.*, *Nat. Methods* **17**, 261 (2020), [arXiv:1907.10121 \[cs.MS\]](#).
- [117] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *Publ. Astron. Soc. Pac.* **125**, 306 (2013), [arXiv:1202.3665 \[astro-ph.IM\]](#).
- [118] J. Goodman and J. Weare, *Commun. Appl. Math. Comput. Sc.* **5**, 65 (2010).
- [119] G. Papamakarios and I. Murray, in *Conference on Neural Information Processing Systems* (2016) [arXiv:1605.06376 \[stat.ML\]](#).
- [120] M. L. Katz, N. Karnesis, N. Korsakova, J. R. Gair, and N. Stergioulas, *Phys. Rev. D* **111**, 024060 (2025), [arXiv:2405.04690 \[gr-qc\]](#).
- [121] S. H. Strub, L. Ferraioli, C. Schmelzbach, S. C. Stähler, and D. Giardini, *Phys. Rev. D* **110**, 024005 (2024), [arXiv:2403.15318 \[gr-qc\]](#).
- [122] S. Deng, S. Babak, M. Le Jeune, S. Marsat, E. Plagnol, and A. Sartirana, *Phys. Rev. D* **111**, 103014 (2025), [arXiv:2501.10277 \[gr-qc\]](#).

Appendix A: Convergence tests

$p-p$ plots for each parameter are shown in Fig. 6 and are calculated with 500 samples from the 6th-round priors, again with the marginals trained in parallel. The empirical coverage is an estimate of the probability mass of the learned proposal distributions that lies within the credible region around the true value. If the empirical coverage as a function of the nominal credibility lies above (below) the diagonal, then the estimated proposals are conservatively wide (optimistically narrow). Our $p-p$ plots are (close to) diagonal and demonstrate reasonable coverage across all the parameters. There is some evidence of overconfidence for m_1 , p_0 , d_L and θ_K , and some

evidence of underconfidence for ϕ_S . We expect that this is partly due to the fact that the marginals are trained in parallel, with the training and validation losses summed across the parameters. This means that individual parameters for which the network has converged could be overfit if another parameter is dominating the continued decrease in the losses.

We also show in Fig. 7 the values of the training and validation losses at the end of each epoch for the network which produced the results in Figs. 2 and 3. In this sixth round, the network trained for 28 epochs, with no decrease in the validation loss for the final 7 epochs (equal to the early stopping criterion), and hence the weights of the network were returned to the state after 21 epochs. There is some evidence of overfitting in the latter stages of training, which could also explain the overconfidence observed in Fig. 6.

Appendix B: Truncation for intrinsic parameters

Here we discuss some additional examples of narrowing down the parameter space for intrinsic parameters. When we draw new values of m_1 , m_2 and p_0 individually, keeping all other parameters fixed to the fiducial values in the main body of the paper, we find that the ability to narrow down the parameter space degrades considerably with $V_i/V_f = \{5.0, 55, 2.9\}$ for $m_1 = 8.3 \times 10^5 M_\odot$, $m_2 = 50.6 M_\odot$ and $p_0/m_1 = 13.0$ respectively. This is expected, since for these injected values, the signals are from much earlier phases of the inspiral instead of the final part before plunge where most of the information about the parameters is contained.

However for a new injected value of the orbital eccentricity, it is again the final part of the signal which is shown to the network and for $e_0 = 0.53$ we find $V_i/V_f = 1.1 \times 10^6$ with truncation bounds shown in Fig. 8. Finally, we show the truncation starting from even larger priors and for a larger value of $m_1 = 8.3 \times 10^5 M_\odot$ and lower value of $p_0/m_1 = 9.25$, such that the injection again captures the full final tenth of a year. The total prior volume for all 11 parameters is narrowed down by a factor of $V_i/V_f = 6 \times 10^7$ after 6 sequential rounds of TMNRE. The SNR of this system is 530.

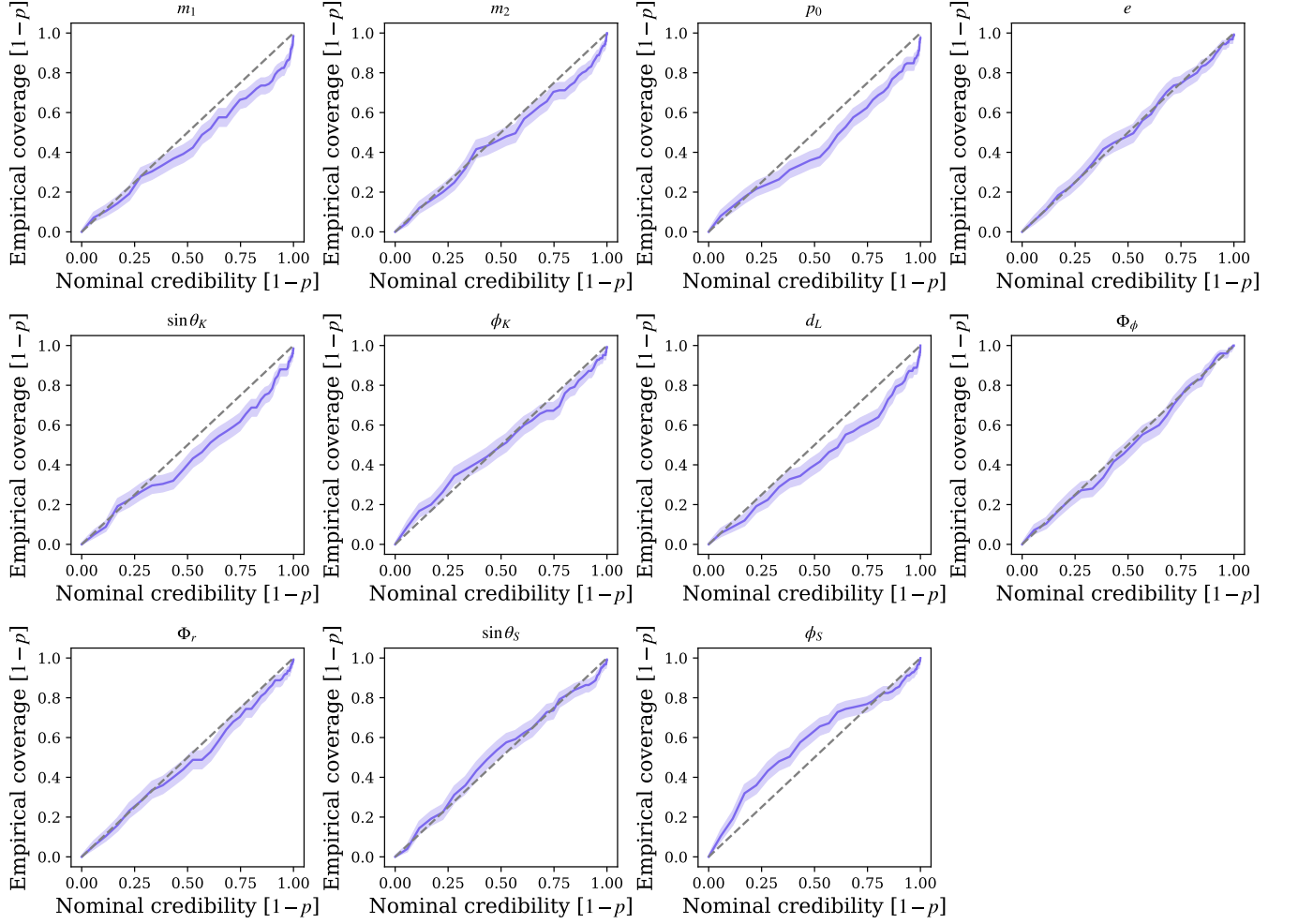


FIG. 6. $p-p$ plots for each of the 11 Schwarzschild EMRI parameters with 500 samples drawn from the 6th round priors of the TMNRE pipeline. The dashed grey diagonal lines represent perfect agreement.

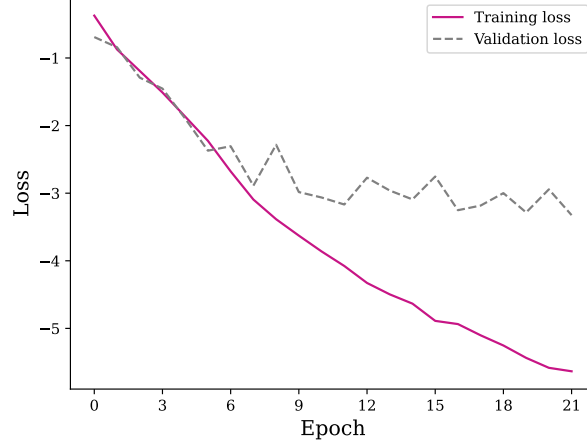


FIG. 7. The training and validation losses of the network from the sixth round of TMNRE which produced the results for the fiducial system in this paper, namely the proposal distributions and truncation bounds in Figs. 2 and 3. The training loss (pink solid) is computed on-the-fly and here we show the value at the end of the epoch, while the validation loss (grey dashed) is computed just once at the end of each epoch.

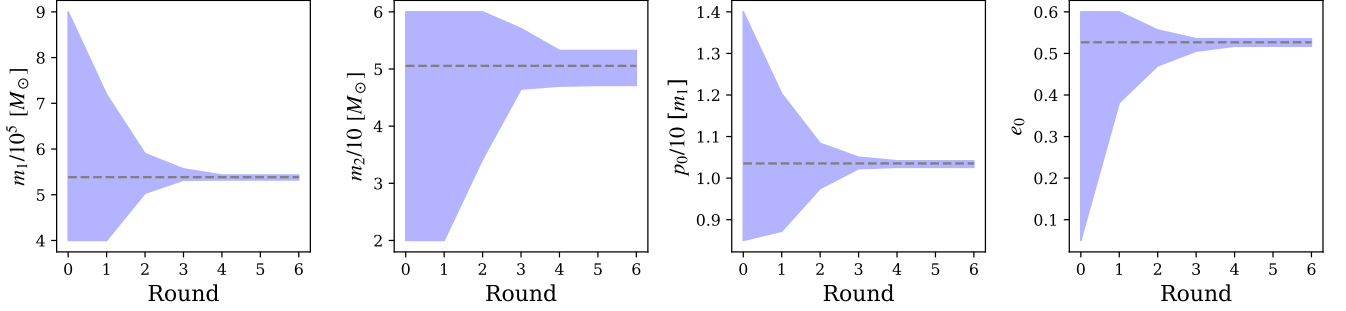


FIG. 8. Shaded regions show the parameter space that is identified by TMNRE as still containing significant likelihood-to-evidence ratio weight for the four intrinsic parameters where $e_0 = 0.53$ is different to that in the main body of the paper, in 6 sequential rounds, trained on 150000 simulations in each round. The initial priors are shown by the shaded region at round 0, and the grey lines show the injected parameters. The total prior volume for all 11 parameters is narrowed down by a factor of $V_i/V_f = 1.1 \times 10^6$ after 6 sequential rounds of TMNRE.

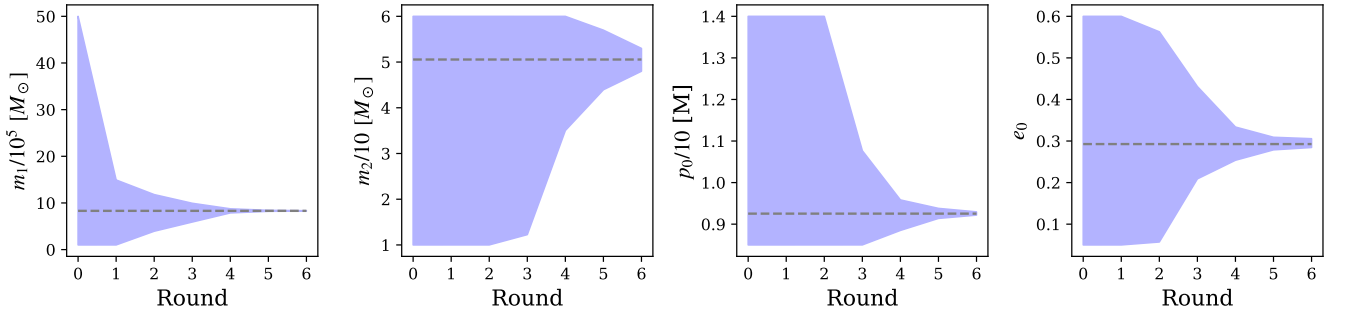


FIG. 9. Shaded regions show the parameter space that is identified by TMNRE as still containing significant likelihood-to-evidence ratio weight for the four intrinsic parameters, with $m_1 = 8.3 \times 10^5 M_\odot$ and $p_0/m_1 = 9.25$ shifted with respect to the system in the main body of the paper, in 6 sequential rounds, trained on 150000 simulations in each round. The initial priors are shown by the shaded region at round 0 and are also wider than those used in the main body of the paper, and the grey lines show the injected parameters. The total prior volume for all 11 parameters is narrowed down by a factor of $V_i/V_f = 6 \times 10^7$ after 6 sequential rounds of TMNRE.