Bayesian inference for gravitational waves from binary neutron star mergers in third-generation observatories

Rory Smith,[1, 2, *] Ssohrab Borhanian,[3, †] Bangalore Sathyaprakash,[3, 4, 5, ‡] Francisco Hernandez Vivanco,[1, 2] Scott E. Field,[6] Paul Lasky,[1, 2] Ilya Mandel,[1, 2] Soichiro Morisaki,[7] David Ottaway,[8] Bram Slagmolen,[9] Eric Thrane,[1, 2] Daniel Töyrä,[9] and Salvatore Vitale[10, 11]

[1]*School of Physics and Astronomy, Monash University, VIC 3800, Australia*
[2]*OzGrav: The ARC Centre of Excellence for Gravitational Wave Discovery, Clayton VIC 3800, Australia*
[3]*Institute for Gravitation and the Cosmos, Department of Physics,*
Pennsylvania State University, University Park, PA, 16802, USA
[4]*Department of Astronomy and Astrophysics, Penn State University, University Park PA 16802, USA*
[5]*School of Physics and Astronomy, Cardiff University, Cardiff CF24 3AA, UK*
[6]*Department of Mathematics and Center for Scientific Computing and*
Visualization Research, University of Massachusetts, Dartmouth, MA 02747, USA
[7]*Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA*
[8]*OzGrav, University of Adelaide, Adelaide, SA 5005, Australia*
[9]*OzGrav, ANU Centre for Gravitational Astrophysics, Research Schools of Physics, and*
Astronomy and Astrophysics, The Australian National University, ACT, 2601, Australia
[10]*LIGO Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[11]*Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology,*
77 Massachusetts Ave, Cambridge, MA 02139, USA

Third-generation (3G) gravitational-wave detectors will observe thousands of coalescing neutron star binaries with unprecedented fidelity. Extracting the highest precision science from these signals is expected to be challenging owing to both high signal-to-noise ratios and long-duration signals. We demonstrate that current Bayesian inference paradigms can be extended to the analysis of binary neutron star signals without breaking the computational bank. We construct reduced order models for ~ 90 minute long gravitational-wave signals, covering the observing band $(5 - 2048\,\mathrm{Hz})$, speeding up inference by a factor of $\sim 1.3 \times 10^4$ compared to the calculation times without reduced order models. The reduced order models incorporate key physics including the effects of tidal deformability, amplitude modulation due to the Earth's rotation, and spin-induced orbital precession. We show how reduced order modeling can accelerate inference on data containing multiple, overlapping gravitational-wave signals, and determine the speedup as a function of the number of overlapping signals. Thus, we conclude that Bayesian inference is computationally tractable for the long-lived, overlapping, high signal-to-noise-ratio events present in 3G observatories.

Introduction.— Third-generation (3G) gravitational-wave detectors such as Cosmic Explorer (CE) [1] and the Einstein Telescope (ET) [2] will observe hundreds of thousands to millions of binary neutron star (BNS) mergers a year [3, 4]. Many of the observed signals will be extremely loud, with signal-to-noise ratios (SNRs) $\sim \mathcal{O}(100 - 1000)$. These signals will provide exquisite measurements of neutron star masses, tidal deformability, and spins, facilitating breakthroughs in cosmology and fundamental physics [5, 6]. Analyzing signals in the 3G era will require scaling data analysis methods by orders-of-magnitude beyond their current capabilities: signals will be in band up to around 40 times longer than in Advanced LIGO/Virgo, the event rate will be thousands of times higher, and multiple signals will be in band at any one time [1, 2].

Bayesian inference is the gold standard for measuring the properties of gravitational-wave signals [7–10]. In Bayesian inference, the posterior probability density of source parameters Θ given experimental data d, and an hypothesis for the data $\mathcal{H}$, is:

$$p(\Theta|d, \mathcal{H}) = \frac{\pi(\Theta|\mathcal{H})\mathcal{L}(d|\Theta, \mathcal{H})}{\mathcal{Z}(d|\mathcal{H})}, \qquad (1)$$

where $\pi(\Theta|\mathcal{H})$ is the prior distribution, $\mathcal{L}(d|\Theta, \mathcal{H})$ is the likelihood function and $\mathcal{Z}(d|\mathcal{H})$ is the evidence. The posterior, $p(\Theta|d, \mathcal{H})$, is the target of parameter estimation, and the evidence is the target for hypothesis testing/model selection. As research and development of 3G instruments ramps up, there is increasing interest in the posterior density of gravitational-wave source properties because it is fundamental to answering interesting questions about the astrophysics capabilities of the detectors. However, because of the high cost of computing the posterior density for BNSs, approximate methods are often used to study the capabilities of 3G detectors; see, e.g., [11]. Fisher-matrix analyses have been used to approximate the width of $p(\Theta|d, \mathcal{H})$, assuming the likelihood is well approximated by a Gaussian distribution. While this assumption is valid for some projections of the posterior, it is not generally valid—even when the SNR is in

---
*Electronic address: rory.smith@monash.edu
†Electronic address: sub284@psu.edu
‡Electronic address: bss25@psu.edu

the thousands—and must be carefully vetted [9]. Hence, a Bayesian treatment of parameter estimates is timely in order to reliably study topics in neutron star astrophysics with networks of 3G detectors.

In this *Letter* we demonstrate how Bayesian inference can be performed on BNS signals using reduced order models (ROMs) [12–14] of gravitational waveforms. Our work extends previous applications of reduced order modeling to signals that are up to 90 minutes in duration from a low frequency of 5 Hz, which is close to the expected low-frequency cutoff for 3G detectors [1, 2]. In addition, the ROMs incorporate effects of the Earth's rotation on GW signals, tidal deformability of neutron stars, and spin-induced orbital precession. We show that our ROMs are accurate representations of the original waveforms at around the $\sim 10^{-12}$ level, ensuring that the ROM approximation is valid up to SNR $\approx 10^6$ [15]— good enough for essentially all foreseeable 3G science.

The ROMs form highly efficient approximations of the likelihood function—the most expensive part of inference. This approximation is known as a *reduced order quadrature* (ROQ) [12, 13, 16]. We show that the ROQ can speed up the evaluation of the likelihood function by a factor of around 13,000 on individual BNS signals. This makes inference on these signals tractable. Additionally, we show that the ROM/ROQ framework can be applied to accelerate likelihood calculations on multiple overlapping in-band signals. At any given time, there are likely to be hundreds of BNSs in the Universe emitting gravitational waves in the 3G observing band [17]. However, signals are sufficiently separated in time that they can usually be analysed separately, though, sometimes two or more signals are sufficiently close that a simultaneous analysis is required [4]. We determine the theoretical speedup factor for multiple in-band signals and show that the ROM/ROQ framework can speed up inference by a factor of around 10,000 for several overlapping signals.

We perform Bayesian inference on a 90 min-long signal at a similar luminosity distance to GW170817, added into synthetic data of a 3G network consisting of CE, ET, and a southern-hemisphere CE-like detector which we call CE-South. The signal has SNR$\approx$2400. We are able to perform Bayesian inference in around 1600 CPU hours. Without the ROQ, the run time of the analysis would be around 10^7 CPU hours. We overcome limitations of the Fisher information matrix by accurately determining the uncertainties of source parameters whose posterior densities are highly non-Gaussian. Our results demonstrate that even loud 3G signals can be analysed with modest computational resources. However, reduced-order methods are essential for controlling the computational cost.

The likelihood function and reduced order models.—
The most expensive part of evaluating the posterior probability is the likelihood function because it involves computing gravitational waveforms. The log likelihood function is [18]

$$\ln \mathcal{L} \propto -\frac{1}{2}\Big\langle d - h,\, d - h \Big\rangle$$
$$= \langle d, h \rangle - \frac{1}{2}\langle h, h \rangle - \frac{1}{2}\mathcal{Z}_n \,, \qquad (2)$$

where the constant $\mathcal{Z}_n = \langle d, d \rangle$ is known as the "noise evidence," and quantifies the likelihood of the data under the hypothesis that it is Gaussian noise [19]. The angle brackets $\langle a, b \rangle$ denote the usual noise-weighted inner product [20].
In the frequeny domain, the gravitational waveforms h have the general form [11]

$$h(f; \Theta) = \frac{1}{D_L}\Big[F_+(f; \xi)\, h_+(f; \theta, \iota, \phi_c, t_c)$$
$$+ F_\times(f; \xi)\, h_\times(f; \theta, \iota, \phi_c, t_c) \Big] , \qquad (3)$$

where $h_{(+,\times)}(f; \theta, \iota, \phi_c, t_c)$ are the individual gravitational wave polarizations and are a function only of the *intrinsic parameters* θ, orbital inclination ι, and phase and time at coalescence ϕ_c and t_c. D_L is the luminosity distance to the source. The quantities $F_{(+,\times)}$ are the detector response functions, which depend on the binary's right ascension α, declination δ and polarization phase ψ, which define $\xi = (\alpha, \delta, \psi)$. The full set of parameters Θ which appears on the left-hand side is the combined set of intrinsic and extrinsic parameters, i.e., $\Theta = (\theta, \xi, \iota, \phi_c, t_c, D_L)$. The detector response functions are also functions of time/frequency due to the Earth's rotation, which cannot be neglected for BNS signals starting from 5 Hz [11]. Since the time evolution of the response functions is slow compared to that of the gravitational-wave signal, we can use the stationary phase approximation which allows us to separate $F_{(+,\times)}$ and $h_{(+,\times)}$—see [11] for details, and the explicit form of the frequency-dependent $F_{(+,\times)}$.

In the frequency domain, the ROMs of h (Eq. 3) have the following general form [13]:

$$h^{\mathrm{ROM}}(f; \Theta) = \sum_{J=1}^{N} h(\mathcal{F}_J; \Theta)\, B_J(f) . \qquad (4)$$

The quantities $B_J(f)$ are a basis set which span the space of the signal. The $h(\mathcal{F}_J; \Theta)$ which appear on the right-hand side of Eq. 4 are the unapproximated waveform (Eq. 3) evaluated at a frequency $\mathcal{F}_J$ from a reduced set of N frequencies $\{\mathcal{F}_J\}_{J=1}^N$. Previous work on reduced order modeling for gravitational-wave parameter estimation also constructed ROMs for the waveform amplitude h^*h which is used to approximate the $\langle h, h \rangle$ term in the likelihood (Eq. 2) [13]. We find this unnecessary, and this term can be computed at negligible cost which we discuss below.

The ROM requires M/N fewer waveform evaluations than the unapproximated expression for the waveform,

where $M = T \times [f_{\max} - f_{\min}]$ and T is the signal duration. This is given by the Nyquist theorem, assuming a fixed sampling rate. For 3G detectors, we assume $f_{\max} = 2048\,\mathrm{Hz}$, $f_{\min} = 5\,\mathrm{Hz}$, $T = 90\,\mathrm{min}\,(5400\,\mathrm{s})$. When ROMs are substituted into the likelihood function, they form a compressed inner product known as a reduced order quadrature (ROQ). The ROQ speeds up the likelihood function by a factor [12, 13]

$$S \approx M/N\,, \qquad (5)$$

and in general $S \gg 1$. This speedup assumes that the waveforms have a closed-form expression, which frequency-domain waveforms typically do.

ROM construction.—The ROMs are constructed in three steps: (i) make a representative "training-space" of gravitational waveforms which span the parameter range of interest; (ii) select basis elements from the training set; and (iii) determine the reduced set of frequency nodes. All steps are achieved using a greedy algorithm [21–23]. We construct ROMs of $90\,\mathrm{min}$-long gravitational-wave signals including spin precession and tidal deformability. That we can build ROMs for $90\,\mathrm{min}$-long BNS signals should not be taken for granted. Previous studies [12, 13, 24] have not established whether ROMs for such signals can be made in practice or if they would be practical for data analysis. In [12, 13, 24] various scalings for ROM basis sizes are given as a function of low-frequency of the signals, or parameter space ranges. However, there has been no systematic study on the size of ROM bases on both low-frequency *and* parameter space size. The fundamental issue is whether the parameter space can be made small enough to be both effective and efficient for long-duration signals. Below, we show that it is indeed the case.

We focus on individual signals, and target a small region of intrinsic parameter space on which we build a training set. The parameters are the chirp mass $\mathcal{M}_c$, symmetric mass ratio η, tidal deformabilities (λ_1, λ_2), spin components projected along the orbital angular momentum axis (χ_1^z, χ_2^z), effective-precession spin and the initial value of the azimuthal precession angle (χ_p, α_0), and orbital inclination ι. We use the waveform model IMRPHENOMPV2_NRTIDALV2 [25–27], which is parameterized by the vector $\theta = \{\mathcal{M}, \eta, \chi_1^z, \chi_2^z, \chi_p, \alpha_0, \iota, \lambda_1, \lambda_2\}$. Inclination appears here because it evolves during the inspiral due to spin-induced orbital precession.

The size of the ROM basis is sensitive to the range in chirp mass. We pick a fiducial chirp mass value of $\mathcal{M}_* = 1.385\,M_\odot$ and restrict the width of the chirp mass of the training set to be $\pm 5 \times 10^{-4} M_\odot$. This mass range is approximately $1 \times 10^3 \Delta\mathcal{M}$, where $\Delta\mathcal{M}$ is the Fisher-matrix error, estimated using `gwbench` [28]. Following the Fisher-matrix error treatment in [29], we find that signals with SNRs of around 10 have Fisher errors $\Delta\mathcal{M} \sim 10^{-4} M_\odot$. Hence, our chirp mass range ensures that we can analyze signals with SNRs around 10 without artificially railing against prior bounds in mass. In

practice, we may want to use broader priors in mass than are possible with a single parameter-space patch. Broader prior ranges can be employed simply by utilizing multiple ROM bases that individually span small parameter-space ranges. Around 1000 such patches in $\mathcal{M} - \eta$ would be needed to cover the full BNS mass space, assuming BNSs have chirp masses approximately in the range $1 M_\odot$–$2 M_\odot$.

All other intrinsic parameters are chosen to have physically-motivated ranges. The symmetric mass ratio is restricted to $0.2 \leq \eta \leq 0.25$. Assuming a minimum neutron star mass of $1 M_\odot$, this range ensures we describe neutron stars with masses up to $2.6 M_\odot$ (around the maximum plausible non-rotating neutron star mass [30–32]). For all other parameters, we consider the following ranges: $0 < \chi_p \leq 0.1$, $-0.1 \leq \chi_{1,2}^z \leq 0.1$, $0 \leq \lambda_{1,2} \leq 5000$. Additionally, the ROM is constructed to be valid for all values of sky location parameters $\xi = (\alpha, \delta, \psi)$, luminosity distance D_L, phase at coalescence ϕ_c, ι, and α. We consider three starting frequencies $f_{\min} = 5, 10, 20\,\mathrm{Hz}$, maximum frequency $f_{\max} = 2048\,\mathrm{Hz}$, and signal duration of $T = 90\,\mathrm{min}$. These values of $f_{\min}$ test how the size of the ROM bases scales with the low-frequency cut off.

We construct a training set of waveforms for the parameter space defined above. The basis and reduced frequency nodes are selected using a greedy algorithm. Details about the training set and greedy algorithm are described in the supplementary material. For signals starting in band from $5, 10, 20\,\mathrm{Hz}$, the ROMs have $N = 522, 291, 179$ basis elements. The basis size only increases by a factor of three when going from $20\,\mathrm{Hz}$ to $5\,\mathrm{Hz}$, despite the signals being over 40 times longer in duration. Bases of around ~ 500 elements should be typical for ROMs of BNS signals starting from $5\,\mathrm{Hz}$ with parameter ranges similar to those used here. Reducing the chirp mass to that of a $1 M_\odot + 1 M_\odot$ binary will change the signal duration only by a factor of 2, much less than the difference in the duration of signals starting from $20\,\mathrm{Hz}$ vs $5\,\mathrm{Hz}$. Hence, the basis size should be roughly constant for lower-mass systems.

The computational cost of building the ROM is relatively small. We require 160 16-core 2.20GHz Intel Xeon E5-2660 CPUs running for around 7 minutes, and then a single CPU running for around 2 hours to complete the basis construction – see *Step-(ii)* of the ROM building strategy in the Supplementary material. The memory footprint of the basis is around 90GB. Thus, it would be feasible to build reduced order models covering the full chirp-mass range of BNSs.

Likelihood speedup.—The most efficient use of ROMs in Bayesian inference is to compress the large inner products in the likelihood function. The compressed inner products are known as a ROQ integration rule. We obtain the ROQ likelihood by substituting the ROM (Eq. 4) into the likelihood (Eq. 2). The ROQ likelihood is

$$\ln \mathcal{L}_{\mathrm{ROQ}} \propto L(\Theta) - \frac{1}{2} Q(\Theta) - \frac{1}{2} \mathcal{Z}_n\,, \qquad (6)$$

where the functions $L(\Theta)$ and $Q(\Theta)$ are given by

$$L(\Theta) = \Re \sum_{J=1}^{N} h(F_J; \Theta)\, \omega_J(t_c)\,, \qquad (7)$$

$$Q(\Theta) = \sum_{I=1}^{N} \sum_{J=1}^{N} h^*(F_I; \Theta) h(F_J; \Theta) \psi_{IJ}\,. \qquad (8)$$

The quantities $\omega_J(t_c)$ and ψ_{IJ} are integration weights that depend only on the basis functions, data, and noise power spectral density, and are defined in the supplementary material.

The computational cost of the ROQ likelihood scales as

$$\ln \mathcal{L} \sim \mathcal{O}(N \times W)\,, \qquad (9)$$

where W is the number of operations required to evaluate the waveform at a given frequency. Unlike previous work [13], we have chosen to write the Q term without the use of an explicit basis for the waveform amplitude h^*h. The scaling of Eq. 8 is independent of W because waveforms at the reduced frequencies have already been computed as part of L (Eq. 7). Thus, Eq. 8 scales like $\sim O(N^2)$ and we find that N is small enough such that $N^2 \ll N \times W$. For our basis starting from 5 Hz (which contains $N = 522$ basis elements), the theoretical speedup (Eq. 5) is $S = 5400\text{s} \times (2048 - 5)\text{Hz}/522 \approx 21{,}000$. Empirically, we find a speed up of around 13,000. The degradation in performance is due to fixed overheads, such as allocating data structures for the waveforms. The integration weights $\omega_{IJ}(t_c)$ and ψ_{IJ} are dependent on the data and noise power spectral density and have to be computed before data analysis can take place. The cost of computing both is negligible in practice.

Validation and accuracy.—The accuracy of the ROQ likelihood (Eq. 6) is limited by the accuracy of the ROM. We validate the accuracy by computing the mismatch M between the ROM representation of h (Eq. 4) and its unapproximated form (Eq. 3):

$$M(h) = 1 - \frac{\langle h^{\mathrm{ROM}}, h \rangle}{\sqrt{\langle h^{\mathrm{ROM}}, h^{\mathrm{ROM}} \rangle \langle h, h \rangle}}\,. \qquad (10)$$

In the noise-weighted inner products, we assume a flat power spectrum, meaning our mismatches are more conservative than if one used a gravitational-wave detector noise power spectral density. In Fig. 1 we show the mismatch $M(h)$ for 2×10^6 random parameter values Θ that were not included in the training space. We include random sky locations, inclinations, luminosity distances, and phases at coalescence. We also include the frequency-dependent response functions in the mismatch calculations, demonstrating that the ROM is accurate describing signals with amplitudes modulated by the Earth's rotation. The mismatch is strongly peaked around 10^{-12}, ensuring that parameter estimation will be unbiased up to SNR $\approx 10^6$ (so that twice the mismatch multiplied by the SNR squared is less than unity [15, 33]).
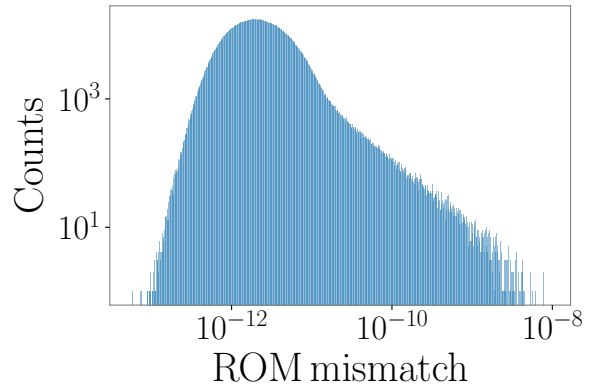

FIG. 1: Accuracy of the ROM approximation for 90-minute long BNS signals starting from 5 Hz. The signals include amplitude modulation due to the Earth's rotation.

Inference with a high-SNR signal.—As an illustrative example, we consider inference on a signal, which we nickname GW370817. The parameters are: $\Theta_{370817} = \{\mathcal{M} = 1.3854\, M_\odot, \eta = 0.24925, \chi_1^z = -0.0113, \chi_2^z = 0.01070, \chi_p = 0.03, \alpha_0 = 1.1, \iota = 0.785, \lambda_1 = 422.5, \lambda_2 = 839.4, D_L = 38.77\,\text{Mpc}, \alpha = 1\,\text{hours}\,57\,\text{min}\,20.5\,\text{s}, \delta = -14.9\,\text{deg}, \psi = 2.012, \phi_c = 0\}$. This signal has a luminosity distance consistent with GW170817. We add the signal into synthetic data of a three-detector network consisting of CE, ET, and CE-South. We use a "zero-noise" realization of Gaussian noise [34], which (statistically) is the most likely realization. This noise realization has the added convenience that if we use flat priors, the posterior peaks at exactly the true parameter values which serves as a useful diagnostic check. The signal has SNR = 2400. We use flat priors on all parameters, and the ranges are given by the range of validity of the ROM. In addition, we impose a physically motivated prior constraint on the component tidal deformability: $\lambda_2 > \lambda_1$. Lastly, we use a uniform prior over a 0.2s interval centered on the true trigger time. In general it is not necessary to restrict the chirp mass prior to such a narrow range. Provided that ROMs exist in local patches covering an extended chirp-mass region, a wide prior can be utilized by building ROQ weights from multiple ROM bases.

We use the `dynesty` nested sampling package to infer the posterior density. In order to obtain well-converged posteriors, we set the number of live points to 5000, and use a random-walk proposal from the `bilby` [10, 35] inference library, which takes a number of steps equal to 70 times the running estimate of the autocorrelation length. The analysis is parallelized over 160 cores. The analysis takes 10 hours (1600 CPU hours) on a cluster of 16-core 2.20GHz Intel Xeon E5-2660 CPUs. The large CPU time occurs because the implementation of the nested sampling algorithm in [10, 35] is extremely slow to converge when the SNR is in the thousands. However, only a handful of events are likely to be detected at these SNRs, with the vast majority of signals having "moderate" SNRs
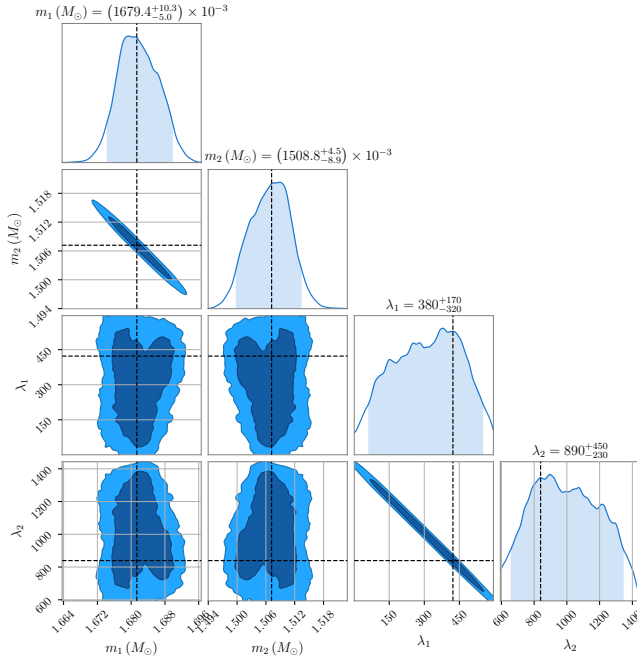
FIG. 2: One- and two-dimensional posterior densities for component masses and component tidal deformabilities. Dark/light shading indicates the one-sigma/two-sigma credible interval respectively. True parameter values are indicated by dashed lines.

less than 100. Analysis on signals with moderate SNRs takes on the order of a day using a single CPU using ROM/ROQ techniques [13]. In contrast, the CPU time without ROM/ROQ methods would be on the order of 20×10^6 hours, i.e., prohibitively expensive. This analysis highlights the need for improvements to the convergence of stochastic-sampling based approaches to inference.

In Fig. 2, we show the one- and two-dimensional posterior densities for the component masses and tidal deformabilities. The component masses can be constrained to the $\sim 5 \times 10^{-3} M_\odot$ level at the 90% credible intervals, which is consistent with Fisher information estimates. The tidal deformabilities have broad uncertainty and are highly non-Gaussian, demonstrating the importance of full Bayesian inference for understanding how well tidal effects—and hence nuclear physics—can be constrained by 3G observatories.

Overlapping signals.—When multiple signals are present simultaneously, the log likelihood function is

$$\ln \mathcal{L} \propto -\frac{1}{2} \Big\langle \big(d - \sum_{i=1}^{n} h_i\big), \, \big(d - \sum_{j=1}^{n} h_j\big) \Big\rangle$$
$$= \Big\langle d, \sum_{i=1}^{n} h_i \Big\rangle - \frac{1}{2} \sum_{ij} \langle h_i, h_j \rangle - \frac{1}{2} \mathcal{Z}_n, \qquad (11)$$

where the sums run over the independent gravitational-wave signals. The double sum $\sum_{ij}$ runs over all pairs (i, j). In the multiple-signal case, the ROQ likelihood

and its scaling are

$$\ln \mathcal{L}_{\mathrm{ROQ}} = \sum_{i=1}^{n} L(\Theta_i) - \sum_{i=1}^{n} \frac{1}{2} Q(\theta_i)$$
$$+ \sum_{\{j,k\}}^{n^2/2} R_{jk}(\Theta_j, \Theta_k, \Delta t^{jk}) - \frac{1}{2} \mathcal{Z}_n \qquad (12)$$
$$\sim \mathcal{O}\Big(N \times n \times W + N^2 \times (n + n^2/2) \Big),$$

where L and Q are given by Eq. 7 and Eq. 8. The R_{jk} term sums over products of all pairs of waveforms $\{j, k\}$, with $j > k$, and is also a function of the relative time offset between two signals, Δt^{jk}. The function $R_{jk}(\Theta_j, \Theta_k, \Delta t^{jk})$ is

$$\Big\langle h(\Theta_j), h(\Theta_k) \Big\rangle \approx R_{jk}(\Theta_j, \Theta_k, \Delta t^{jk}) =$$
$$\Re \sum_{K=1}^{N} \sum_{L=1}^{N} h^*(F_K; \Theta_j) h(F_L; \Theta_k) \Gamma_{KL}(\Delta t^{jk}), \qquad (13)$$

where the matrix $\Gamma_{KL}(\Delta t^{jk})$ is a set of integration weights, given in the supplementary material. The overall speedup of the multiple-signal ROQ likelihood Eq. 12 with respect to non-ROQ likelihood (Eq. 11) is

$$S \approx \frac{MnW + Mn^2/2}{NnW + Nn^2/2}, \text{ with } n > 1, \qquad (14)$$

where we have kept terms at $\mathcal{O}(W)$ and $\mathcal{O}(n^2)$. The $N^2 n^2$ scaling is potentially problematic if the number of in-band signals is large. However, most overlapping signals—roughly between 96% to 99.5%—are well separated in time so that they can be analysed separately (tables I and IV of [4]). Thus, we only consider the speedup for the simultaneous analysis of a few in-band signals. We empirically determine a speed-up of around $\sim \mathcal{O}(10^4)$ compared to the calculation time without ROM for up to 10 in-band signals. We show the speedup as a function of the number of in-band signals in the supplementary material.

Discussion.—Reduced order models of gravitational-wave signals from BNS mergers can be used to accelerate parameter estimation in 3G observatories, thereby removing a computational hurdle. This work lays the groundwork for detailed studies of BNS systems in the 3G era. Further, the ROM/ROQ framework can be used to efficiently carry out inference on data containing multiple overlapping signals. Further avenues to pursue include ROMs of more sophisticated waveforms, e.g., with higher-order gravitational-wave modes, which can place tighter constraints on parameter estimates [36]. Bayesian inference on very loud signals—SNR $\sim \mathcal{O}(1000)$—requires significantly more likelihood evaluations than in analyses on LIGO/Virgo signals. Sampling-based methods for Bayesian inference will have to be significantly adapted and scaled up in order to efficiently

analyze data in the 3G era. For instance, Hamiltonian Monte Carlo methods [37] (which exploit gradients of posterior densities) and machine learning techniques, e.g., [38–41] (which provide rapid approximations to reduced order models and posterior densities) may be promising avenues to explore.

[1] D. Reitze, R. X. Adhikari, S. Ballmer, B. Barish, L. Barsotti, G. Billingsley, D. A. Brown, Y. Chen, D. Coyne, R. Eisenstein, et al., *Cosmic explorer: The u.s. contribution to gravitational-wave astronomy beyond ligo* (2019), 1907.04833.

[2] M. Maggiore, C. V. D. Broeck, N. Bartolo, E. Belgacem, D. Bertacca, M. A. Bizouard, M. Branchesi, S. Clesse, S. Foffa, J. García-Bellido, et al., Journal of Cosmology and Astroparticle Physics **2020**, 050–050 (2020), ISSN 1475-7516, URL http://dx.doi.org/10.1088/1475-7516/2020/03/050.

[3] V. Baibhav, E. Berti, D. Gerosa, M. Mapelli, N. Giacobbo, Y. Bouffanais, and U. N. Di Carlo, Physical Review D **100** (2019), ISSN 2470-0029, URL http://dx.doi.org/10.1103/PhysRevD.100.064060.

[4] A. Samajdar, J. Janquart, C. van den Broeck, and T. Dietrich, In Prep. (2021), dCC/P2100030.

[5] S. Vitale and M. Evans, Physical Review D **95** (2017), ISSN 2470-0029, URL http://dx.doi.org/10.1103/PhysRevD.95.064052.

[6] R. X. Adhikari, P. Ajith, Y. Chen, J. A. Clark, V. Dergachev, N. V. Fotopoulos, S. E. Gossan, I. Mandel, M. Okounkova, V. Raymond, et al., Classical and Quantum Gravity **36**, 245010 (2019), ISSN 1361-6382, URL http://dx.doi.org/10.1088/1361-6382/ab3cff.

[7] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, et al., Physical Review D **91** (2015), ISSN 1550-2368, URL http://dx.doi.org/10.1103/PhysRevD.91.042003.

[8] G. Ashton et al., Astrophys. J. Suppl. **241**, 27 (2019), 1811.02042.

[9] M. Vallisneri, Physical Review D **77** (2008), ISSN 1550-2368, URL http://dx.doi.org/10.1103/PhysRevD.77.042001.

[10] R. J. E. Smith, G. Ashton, A. Vajpeyi, and C. Talbot, Monthly Notices of the Royal Astronomical Society **498**, 4492–4502 (2020), ISSN 1365-2966, URL http://dx.doi.org/10.1093/mnras/staa2483.

[11] W. Zhao and L. Wen, Physical Review D **97** (2018), ISSN 2470-0029, URL http://dx.doi.org/10.1103/PhysRevD.97.064031.

[12] P. Canizares, S. E. Field, J. Gair, V. Raymond, R. Smith, and M. Tiglio, Phys. Rev. Lett. **114**, 071104 (2015), 1404.6284.

[13] R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, M. Pürrer, V. Raymond, and P. Schmidt, Phys. Rev. D **94**, 44031 (2016).

[14] M. Pürrer, Class. Quant. Grav. **31**, 195010 (2014), 1402.4146.

[15] L. Lindblom, B. J. Owen, and D. A. Brown, Phys. Rev. D **78**, 124020 (2008), URL https://link.aps.org/doi/10.1103/PhysRevD.78.124020.

[16] H. Antil, S. E. Field, F. Herrmann, R. H. Nochetto, and M. Tiglio, Journal of Scientific Computing **57**, 604 (2013).

[17] B. P. Abbott et al., ArXiv e-prints (2017), 1710.05837.

[18] B. J. Owen and B. S. Sathyaprakash, Phys. Rev. D **60**, 022002 (1999).

[19] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, and et al., Living Reviews in Relativity **23** (2020), ISSN 1433-8351, URL http://dx.doi.org/10.1007/s41114-020-00026-9.

[20] J. Veitch and A. Vecchio, Phys. Rev. D **81**, 062003 (2010), URL https://link.aps.org/doi/10.1103/PhysRevD.81.062003.

[21] H. Antil, D. Chen, and S. Field, Computing in Science & Engineering **20**, 10–25 (2018), ISSN 1558-366X, URL http://dx.doi.org/10.1109/MCSE.2018.042781323.

[22] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Phys. Rev. X **4**, 031006 (2014), URL https://link.aps.org/doi/10.1103/PhysRevX.4.031006.

[23] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, Comptes Rendus Mathematique **339**, 667 (2004), ISSN 1631-073X, URL http://www.sciencedirect.com/science/article/pii/S1631073X04004248.

[24] S. Morisaki and V. Raymond, Physical Review D **102** (2020), ISSN 2470-0029, URL http://dx.doi.org/10.1103/PhysRevD.102.104020.

[25] T. Dietrich, S. Khan, R. Dudi, S. J. Kapadia, P. Kumar, A. Nagar, F. Ohme, F. Pannarale, A. Samajdar, S. Bernuzzi, et al., Phys. Rev. D **99**, 024029 (2019), URL https://link.aps.org/doi/10.1103/PhysRevD.99.024029.

[26] T. Dietrich, S. Bernuzzi, and W. Tichy, Phys. Rev. D **96**, 121501 (2017), URL https://link.aps.org/doi/10.1103/PhysRevD.96.121501.

[27] T. Dietrich, A. Samajdar, S. Khan, N. K. Johnson-McDaniel, R. Dudi, and W. Tichy, Phys. Rev. D **100**, 044003 (2019), URL https://link.aps.org/doi/10.1103/PhysRevD.100.044003.

[28] S. Borhanian (2020), 2010.15202.

[29] E. Poisson and C. M. Will, Physical Review D **52**,

848–855 (1995), ISSN 0556-2821, URL http://dx.doi.org/10.1103/PhysRevD.52.848.

[30] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, et al., The Astrophysical Journal **896**, L44 (2020), ISSN 2041-8213, URL http://dx.doi.org/10.3847/2041-8213/ab960f.

[31] R. Essick, P. Landry, and D. E. Holz, Physical Review D **101** (2020), ISSN 2470-0029, URL http://dx.doi.org/10.1103/PhysRevD.101.063007.

[32] Y. Lim and J. W. Holt, European Physical Journal A **55**, 209 (2019), 1902.05502.

[33] F. Ohme, Classical and Quantum Gravity **29**, 124002 (2012), ISSN 1361-6382, URL http://dx.doi.org/10.1088/0264-9381/29/12/124002.

[34] C. Pankow, K. Chatziioannou, E. A. Chase, T. B. Littenberg, M. Evans, J. McIver, N. J. Cornish, C.-J. Haster, J. Kanner, V. Raymond, et al., Physical Review D **98** (2018), ISSN 2470-0029, URL http://dx.doi.org/10.1103/PhysRevD.98.084016.

[35] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov, et al., The Astrophysical Journal Supplement Series **241**, 27 (2019), ISSN 1538-4365, URL http://dx.doi.org/10.3847/1538-4365/ab06fc.

[36] J. C. Bustillo, T. Dietrich, and P. D. Lasky, *Higher-order gravitational-wave modes will allow for percent-level measurements of hubble's constant with single binary neutron star merger observations* (2020), 2006.11525.

[37] M. Betancourt, *A conceptual introduction to hamiltonian monte carlo* (2018), 1701.02434.

[38] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, *Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy* (2020), 1909.06296.

[39] H. Shen, E. A. Huerta, Z. Zhao, E. Jennings, and H. Sharma, *Deterministic and bayesian neural networks for low-latency gravitational wave parameter estimation of binary black hole mergers* (2019), 1903.01998.

[40] A. J. K. Chua and M. Vallisneri, Phys. Rev. Lett. **124**, 041102 (2020), URL https://link.aps.org/doi/10.1103/PhysRevLett.124.041102.

[41] A. J. K. Chua, C. R. Galley, and M. Vallisneri, Phys. Rev. Lett. **122**, 211101 (2019), URL https://link.aps.org/doi/10.1103/PhysRevLett.122.211101.