PAPER • OPEN ACCESS

Accelerating gravitational wave parameter estimation with multi-band template interpolation

View the article online for updates and enhancements.

Related content

- Results from the first NINJA project
  Benjamin Aylott, John G Baker, William D Boggs et al.

- Frequency domain reduced order models for gravitational waves from aligned-spin compact binaries
  Michael Pürrer

- Black hole binaries: high-accuracy simulations and TaylorF2 template efficiency
  Michael Boyle, Duncan A Brown and Larne Pekowsky

Recent citations

- Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA
  B. P. Abbott *et al*

- An architecture for efficient gravitational wave parameter estimation with multimodal linear surrogate models
  Richard O'Shaughnessy *et al*

Accelerating gravitational wave parameter estimation with multi-band template interpolation

Serena Vinciguerra, John Veitch and Ilya Mandel

Institute of Gravitational Wave Astronomy & School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, United Kingdom

E-mail: serena@star.sr.bham.ac.uk

CrossMark

Abstract

Parameter estimation on gravitational wave signals from compact binary coalescence (CBC) requires the evaluation of computationally intensive waveform models, typically the bottleneck in the analysis. This cost will increase further as low frequency sensitivity in later second and third generation detectors motivates the use of longer waveforms.

We describe a method for accelerating parameter estimation by exploiting the chirping behaviour of the signals to sample the waveform sparsely for portions where the full frequency resolution is not required. We demonstrate that the method can reproduce the original results with a waveform mismatch of $\leqslant 5 \times 10^{-7}$, but with a waveform generation cost up to ~ 50 times lower for computationally costly frequency-domain waveforms starting from below 8 Hz.

Keywords: LIGO, parameter estimation, waveform optimisation, Einstein telescope

(Some figures may appear in colour only in the online journal)

1. Introduction

The discovery of gravitational waves from coalescing binary black hole systems made by Advanced LIGO in its first observing run opened the door to gravitational wave astronomy [1]. As the second generation of ground based detectors continues to evolve towards their design sensitivities the rate of detections is expected to increase, leading eventually to the detection of lower mass binary systems such as binary neutron star (BNS) and neutron star—black hole (NSBH) binaries [2].

The characterisation of these sources involves the use of Bayesian parameter estimation and model selection algorithms based on stochastic sampling of the posterior probability

distribution for the model parameters conditioned on the observed data. This process involves repeated comparisons of the data with template waveforms through evaluation of the likelihood function. Previous implementations (e.g. `LALInference` [3]) have required millions of likelihood evaluations, which implies that a similar number of template waveforms must be generated. In the case of sophisticated waveform models this template generation dominates the computational cost of the analysis, with the cost scaling linearly with the length of the waveform τ, which in turn scales with the low frequency starting point of the waveform as $f_{\min}^{-8/3}$. As the low-frequency sensitivity of the second-generation instruments improves, $f_{\min}$ is expected to reduce from $\sim$30 Hz to $\sim$10 Hz or lower. The issue becomes even greater in the case of subterranean third-generation instruments such as the Einstein Telescope which are expected to reduce this further to 5 Hz or lower [4, 5]. This improvement in low-frequency sensitivity should translate to much more accurate estimation of key parameters. However, taking full advantage of this improvement in a timely and computationally efficient manner is a challenge. We present a method that leverages the frequency evolution of the waveform to effectively reduce the number of waveform samples that must be computed. This has the potential to asymptotically reduce the computational cost of template generation by a factor that is proportional to $f_{\min}^{-1}$. Here we give details of a practical implementation which does not compromise the accuracy of parameter estimation and study the computational cost scaling in a realistic analysis.

Several methods have been developed previously to overcome the need to evaluate the waveform and likelihood at each point in the Fourier domain. Reduced order quadrature (ROQ) methods, first introduced for CBC waveforms in [6] and developed for the purpose of parameter estimation (PE) in [7–10], seek to represent the waveform in an alternative basis from the standard Fourier components. A waveform for a particular point in parameter space is represented as the linear combination of a number of these basis templates. By projecting the data into the same basis the likelihood function can be computed using a sum over bases rather than a sum over Fourier components, where the number of bases is far smaller than the number of Fourier components. This method significantly accelerates the likelihood computation. However, it has the drawback of requiring the basis to be constructed in advance for each waveform family, a process which is costly in terms of both computation and memory requirements to store the input waveforms, with a cost that grows rapidly as the dimensionality of the model is increased to include misaligned spins. The large intrinsic volume of the mass parameter space requires that it be subdivided into patches of manageable size, with each patch having a different set of bases. The ROQ likelihood calculation is also dependent on the particular noise curve used through the ROQ integration weights, which must be computed for the particular characteristics of the data at the time of the event of interest. Furthermore, severing the link between frequency and the representation of the waveform makes it difficult to model the effect of frequency-dependent detector calibration errors, which were included in the analysis of binary black hole systems in O1 [11, 12].

A different approach has been developed in the context of low-latency searches for gravitational waves. In this context the incoming data-stream is filtered against a pre-determined bank of templates which is chosen to cover the mass parameter space with a certain maximum guaranteed loss of signal-to-noise ratio (SNR). Although here the filtering can proceed in parallel it is still desirable to reduce the cost of the search by reducing the volume of data that has to be processed. The MBTA [13, 14] and gstlal [15] pipelines divide the templates into bands, which are chosen to exploit the chirping nature of the inspiral signal. Each band has a certain maximum signal frequency $f < f_{\max}$, so both the template and the data can be down-sampled to a lower sampling rate, reducing the cost of the filtering process for each band. The original

high-bandwidth SNR time-series can be reconstructed from the output of the banded filters by subsequent up-sampling, which can be done selectively on data stretches which have significant SNR in the banded filters. A similar approach has been advocated for LISA data analysis, employing two bands for each template [16], as is also the case for MBTA.

In this paper we pursue an approach inspired by the latter method of subdividing the waveform into band-limited pieces, with the aim of using it for PE rather than searching. This places some additional constraints on the accuracy of waveform reconstruction required to reproduce the results from a full-bandwidth analysis without adding systematic or statistical errors. Our method is currently limited to the computation of the template (likelihood evaluation is still performed in the full Fourier basis), but nevertheless can produce large reductions in computational cost for long duration signals when the more sophisticated (and costly) waveform models are employed. Unlike the ROQ, this allows us to maintain the link with frequency and easily include calibration error modelling in the analysis. Also, because the method requires no pre-computation of a new basis it can be applied without modification to any frequency-domain waveform model, including modifications to the signal such as tidal effects [17, 18] and parameterised deviations from general relativity [19, 20]. This flexibility is the main advantage of the method, which makes it especially suitable for analyses where an ROQ model is not available, or where its production would be too costly. An implementation is provided in the open source `LALInference` PE software [3]. We describe the method in detail in section 2 and demonstrate its efficacy when applied to the analysis of simulated signals in section 3. We discuss possible future developments in section 4.

2. Multi-banding approach: the method

2.1. Motivation

In gravitational wave PE, the aim is to explore the posterior probability distribution of the source model,

$$p(\vec{\theta}|\vec{d}, H) = \frac{p(\vec{\theta}|H)\, p(\vec{d}|\vec{\theta}, H)}{p(\vec{d}|H)} \tag{1}$$

where $\vec{\theta}$ are the physical parameters of the source such as the masses, spins, position and orientation [3]. The likelihood function for a single detector under the assumption of Gaussian noise depends on the data $\vec{d}$ and the parameter $\vec{\theta}$, as well as the particular waveform model used H, as

$$p(\vec{d}|\vec{\theta}, H) \propto \exp\left[-2\sum_i^N \frac{|h_i(\vec{\theta}) - \vec{d}_i|^2}{\tau S_n(f_i)}\right] \tag{2}$$

where $S_n(f_i)$ is the power spectral density of the detector, $\tau = \delta f^{-1}$ is the duration of the data segment to be analysed, and $N = \tau/(2\delta t)$ is the number of Fourier components in the frequency-domain complex representation of the modeled signal $h_i(\vec{\theta})$ as it would be observed in the detector. Since the details of the detector responses are not important for what follows we refer the reader to [3] for a full description of how the extrinsic parameters are used to construct the observed signal in each detector. In order to accurately capture the waveform we must choose τ and δt such that the entire signal duration, from the time it enters the sensitive band of the instrument at frequency $f_{\min}$, is contained in τ, and the sampling resolution $\delta t < (2f_{\max})^{-1}$ is sufficient to capture the highest frequency components of the signal at $f_{\max}$.
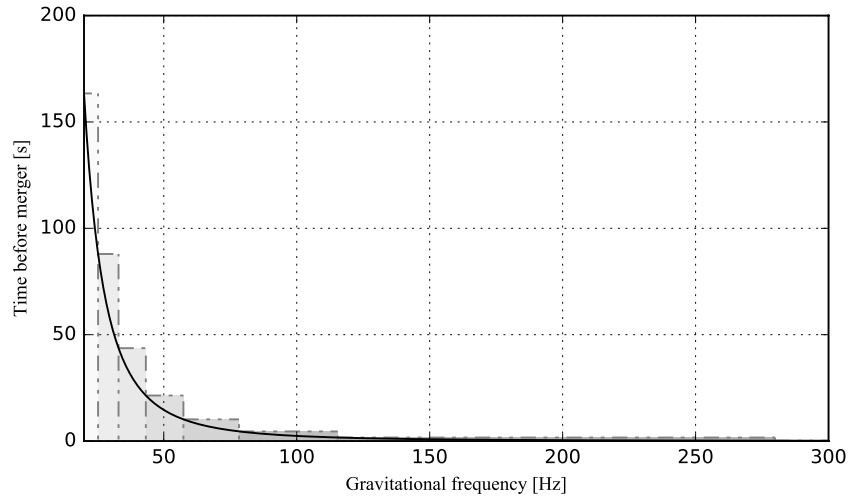
Figure 1. Time from a given frequency to coalescence for a fiducial binary neutron star signal. Coloured boxes indicate the subdivision of the waveform into bands with adaptive frequency resolution, as determined by the time before coalescence; see section 2.2.

To leading order, the duration of an inspiral signal from a certain frequency f to the formal time of coalescence is [21] (in geometrical units $G = c = 1$)

$$t(f) \approx 5 \left[8\pi f \right]^{-8/3} \mathcal{M}^{-5/3}, \tag{3}$$

where $\mathcal{M} = M_1^{3/5} M_2^{3/5} (M_1 + M_2)^{-1/5}$ is the chirp mass of a binary with component masses $M_{1,2}$ and mass ratio $q = M_2 / M_1 \leqslant 1$. During the inspiral, the gravitational-wave frequency monotonically increases until the merger and ring-down phases. An example is shown in figure 1, where we put frequency on the abscissa to emphasize that we are working in the frequency domain.

In the standard calculation, there is a fixed frequency resolution of $\delta f = \tau^{-1}$ between frequency bins, and the total number of frequency-domain samples required to describe the signal is

$$\begin{aligned} N_{\text{fix}} &= \int_{f_{\text{min}}}^{f_{\text{max}}} \delta f^{-1} \mathrm{d}f \\ &= (f_{\text{max}} - f_{\text{min}}) \tau \\ &\approx 5(8\pi)^{-8/3} \mathcal{M}^{-5/3} (f_{\text{max}} - f_{\text{min}}) f_{\text{min}}^{-8/3} . \end{aligned} \tag{4}$$

We can see from the figure that this frequency resolution is necessary to contain the full length waveform starting at time τ before merger, but as the frequency increases the time before merger $t(f)$ decreases and the waveform is over-sampled in frequency. Our aim is to take advantage of this to increase δf as a function of frequency without losing any information about the waveform phasing, thereby reducing the total number of points at which the waveform must be evaluated.

We now consider the asymptotic limit of multi-banding. In the idealized limit, the frequency step $\delta f = t(f)^{-1}$ can vary continuously throughout the signal. We then have

$$N_{\text{min}} = \int_{f_{\text{min}}}^{f_{\text{max}}} t(f)\mathrm{d}f$$
$$= -3(8\pi)^{-8/3}\mathcal{M}^{-5/3}(f_{\text{max}}^{-5/3} - f_{\text{min}}^{-5/3}). \tag{5}$$

The relative number of points required for the standard case compared to the ideal case is then

$$\frac{N_{\text{fix}}}{N_{\text{min}}} = \frac{5}{3}\frac{(f_{\text{max}} - f_{\text{min}})f_{\text{min}}^{-8/3}}{f_{\text{min}}^{-5/3} - f_{\text{max}}^{-5/3}}, \tag{6}$$

which for $f_{\text{max}} \gg f_{\text{min}}$ indicates an asymptotic reduction in number of points $5f_{\text{max}}/3f_{\text{min}}$. For a binary neutron star waveform which enters the detector at 20 Hz and terminates at 1500 Hz, the potential reduction in number of points is therefore a factor of $\sim$125.

2.2. Choice of bands

Rather than taking the continuously varying δf case, in our practical implementation we work with a pre-determined set of frequencies which divides the total frequency span into several bands with constant δf within each band. We position the bands in frequency space so that δf changes by a factor of 2 between neighboring bands, while ensuring that the Nyquist sampling criterion is always met.

Figure 1 shows a schematic of the basic idea. We must choose our bands such that they are able to accurately represent the longest waveform in our allowed mass prior. This can be determined automatically at run-time of the PE code; e.g. a $1 + 1M_\odot$ binary neutron star signals lasts 281 s from 20 Hz to coalescence. Starting at the lowest frequency f_{min}, the frequency resolution necessary to contain the waveform is $\delta f_0 \leqslant t(f_{\text{min}})^{-1}$. Each subsequent band has a sampling rate $\delta f_b = 2\delta f_{b-1}$ and so the time at the start of the new band is a factor of two closer to coalescence, $t(f_b) = t(f_{b-1})/2$. The frequencies at which to place the band boundaries are then determined by inverting equation (3) and solving for the series of δf_b. To summarise, we can specify the frequencies at which the waveform is evaluated via the following algorithm

$b = 0, i = 0$
$\delta f_b = t(f_{\text{min}})^{-1}, f_i = f_{\text{min}}$
while $f_i < f_{\text{max}}$ **do**
  while $t(f_i) > (2\delta f_b)^{-1}$ **do**
    $f_{i+1} = f_i + \delta f_b$
    $i = i + 1$
  end
  $\delta f_{b+1} = 2\delta f_b$
  $b = b + 1$
end

2.3. Up-sampled waveform

Having defined the reduced set of frequencies at which the waveform is to be calculated, we now outline the procedure for reconstructing the full waveform. Note that unlike in reduced order quadrature methods, we still compute the likelihood using the fully sampled dataset. A naïve decimation or averaging of the frequency-domain detector data leads to a loss of information relative to the fully-sampled results. We therefore use an interpolation scheme to reconstruct the waveform at the full sampling rate in order to match filter the original data.

Direct linear interpolation of the reduced waveform $\tilde{h}(f_j)$ does not accurately reproduce the original waveform as the oscillatory behaviour is not captured by the interpolating straight line segments. We therefore work with the waveform represented in amplitude and phase as $\tilde{h}(f_j) = A_j \exp(\mathrm{i}\phi_j)$, where j labels the reduced set of frequencies. Within each coarse bin, we linearly interpolate the amplitude A and phase ϕ to obtain estimates of the amplitude $\hat{A}_k = \hat{A}(\hat{f}_k)$ and phase $\hat{\phi}_k = \hat{\phi}(\hat{f}_k)$ at the dense set of frequencies $\hat{f}$ labeled with k:

$$\hat{A}_k = A_j + \frac{\hat{f}_k - f_j}{f_{j+1} - f_j}(A_{j+1} - A_j),\tag{7}$$

$$\hat{\phi}_k = \phi_j + \frac{\hat{f}_k - f_j}{f_{j+1} - f_j}(\phi_{j+1} - \phi_j),\tag{8}$$

where f_j is the nearest coarse frequency point below $\hat{f}_k$ and $\hat{f}_{k+1} - \hat{f}_k = \delta f_0$. The up-sampled waveform after multi-banding and interpolation (hereafter MB-Interpolation) is then $\tilde{h}(\hat{f}_k) = \hat{A}_k \exp(\mathrm{i}\hat{\phi}_k)$.

One practical problem with applying this formula is that the exact estimation of $\exp \mathrm{i}\hat{\phi}_k$ is computationally expensive. To avoid this we use the recursive property $\mathrm{e}^{\mathrm{i}\hat{\phi}_{k+1}} = \mathrm{e}^{\mathrm{i}\hat{\phi}_k}\mathrm{e}^{\mathrm{i}\delta f_0(\phi_{j+1} - \phi_j)/(f_{j+1} - f_j)}$. The last term needs to be computed only once for each coarse bin [22]. The recursion relation can be expressed in terms of the real and imaginary parts of the complex frequency-domain signal as

$$\Re(\hat{h}_{k+1}) = \left[1 + \frac{(A_{j+1} - A_j)\delta f_0}{\hat{A}_k(f_{j+1} - f_j)}\right]\left[\Re(\hat{h}_k)\left(1 - 2\sin^2\frac{\delta\phi_j}{2}\right) - \Im(\hat{h}_k)\sin\delta\phi_j\right],$$

$$\Im(\hat{h}_{k+1}) = \left[1 + \frac{(A_{j+1} - A_j)\delta f_0}{\hat{A}_k(f_{j+1} - f_j)}\right]\left[\Im(\hat{h}_k)\left(1 - 2\sin^2\frac{\delta\phi_j}{2}\right) + \Re(\hat{h}_k)\sin\delta\phi_j\right],$$

where $\delta\phi_j \equiv \delta f_0(\phi_{j+1} - \phi_j)/(f_{j+1} - f_j)$; therefore we only need to compute $\sin(\delta\phi_j)$ and $\sin^2(\delta\phi_j/2)$.

2.4. Accuracy

The waveform accuracy required for parameter estimation is determined by the condition that systematic bias in parameter estimates from imperfect waveforms should be much smaller than the statistical measurement uncertainty of inference on data with finite signal-to-noise ratios (e.g. [23]). Therefore, the shift in the log likelihood due to the use of MB-Interpolation waveforms in lieu of the original waveforms, $\delta\log L_{\mathrm{MB-Interpolation}}$, should be smaller than the spread in the log likelihood over the posterior $\sigma_{\log L}$:

$$\delta\log L_{\mathrm{MB-Interpolation}} \ll \sigma_{\log L} \sim \sqrt{\frac{N_{\mathrm{param}}}{2}},\tag{9}$$

where N_{param} is the number of parameters in the model. This condition on the log likelihood can be expressed in terms of the match between the original waveform h_0 and the MB-Interpolation waveform h [24–26]:

$$\frac{\langle h_0 - h | h_0 - h\rangle}{\langle h_0 | h_0 \rangle} \ll \frac{\sqrt{2N_{\mathrm{param}}}}{\rho^2},\tag{10}$$
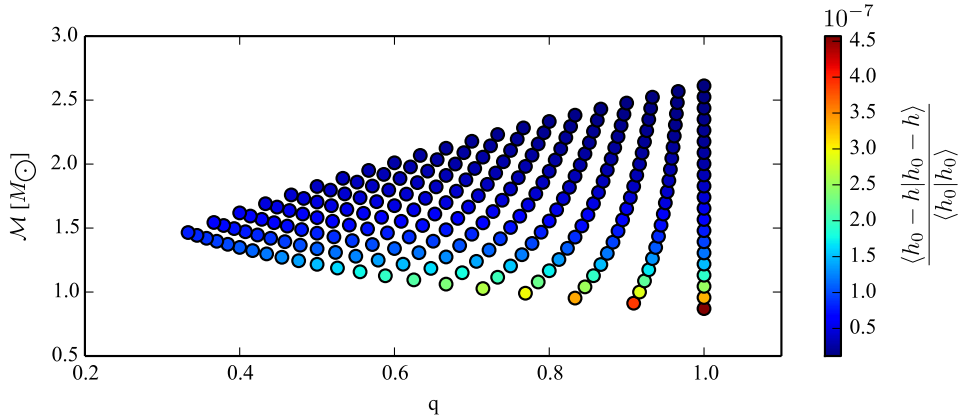
Figure 2. Mismatch of MB-Interpolation waveforms (h) against waveforms computed with the standard procedure (h_0) as a function of chirp mass and mass ratio. The mismatch is calculated up to 1024 Hz.

where ρ is the signal-to-noise ratio. Considering $\rho \sim 20$, typical for a moderately loud signal [27], the threshold on the mismatch is $\sim 10^{-3}$.

Figure 2 show that the mismatch of MB-Interpolation waveforms against the original waveforms is a factor of a thousand smaller than this requirement over the binary neutron star region, decreasing for more massive systems. This is expected, as in the frequency domain the density of cycles at low frequency increases with the time duration of the waveform, so the most demanding case is that of the lowest mass considered in a particular analysis (in our case a 1–1 $M_\odot$ binary). Therefore, we conclude that this procedure provides sufficient accuracy for unbiased inference at all masses above 1–1 $M_\odot$.

3. Results

We implemented the MB-Interpolation approach (section 2), including the waveform interpolation procedure (section 2.3) within `LALInference` [3]. We performed several tests in order to validate MB-Interpolation. We first checked the effectiveness of the MB-Interpolation by verifying the reduction of the number of frequencies at which the template is evaluated when multibanding. We then measured the speedup in the waveform computation following multibanding and interpolation. Finally, we tested the overall acceleration of the complete PE analysis with MB-Interpolation and confirmed its accuracy.

3.1. Reduction of template evaluations

To measure the speedup in waveform generation we first defined the frequency set at which the multibanded template is evaluated according to the algorithm in section 2.2. The in-band signal duration is set by a BNS with both component masses equal to $1M_\odot$ as a reference system, corresponding to the lowest limit of the component mass prior adopted in the analysis. The number of frequencies at which the waveform is evaluated is shown in figure 3 as a function of the starting frequency f_{min} for both MB-Interpolation and the standard algorithm. This figure clearly demonstrates the effectiveness of the approach in reducing template evaluations: the number of frequencies defining the two sets, N_{fix} and N_{MB} respectively for the standard
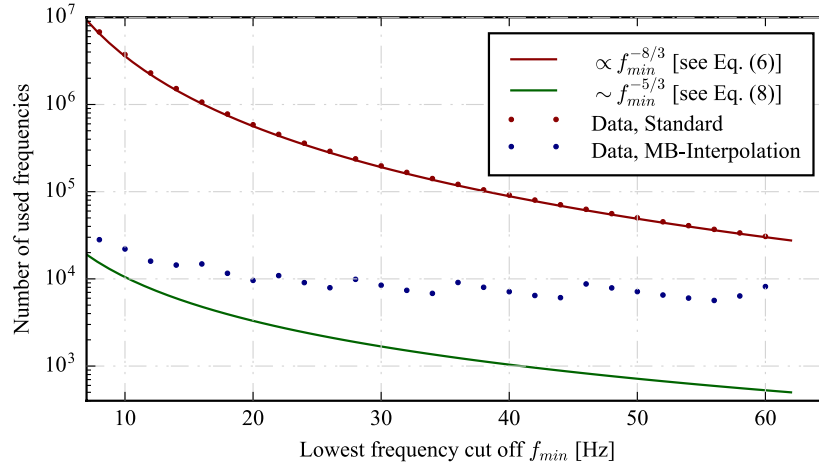
Figure 3. Number of frequencies at which the waveform is evaluated when using the standard (N_{fix}, red dots) and MB-Interpolation (N_{MB}, blue dots) algorithms as a function of the lower frequency limit. The red curve corresponds to equation (4) while the green curve shows the number of frequency samples in the theoretical limit of continuously adapted sampling steps, equation (5). MB-Interpolation is sub-optimal but approaches the asymptotic case in the limit $f_{\text{min}} \to 0$, as the templates become very long and δf_0 approaches 0; the number of frequency bands increases from 3 at $f_{\text{min}} = 60$ Hz to 8 at $f_{\text{min}} = 20$ Hz and 11 at $f_{\text{min}} = 8$ Hz.

and the MB-Interpolation algorithm, differs by an order of magnitude or more for starting frequencies below 40 Hz.

The evident segmented structure of N_{MB} reflects the varying number of frequency bands used in MB-Interpolation. Within each band, δf is constant and the number of frequencies follows the same $\sim f_{\text{min}}^{-8/3}$ scaling as for the standard algorithm. As expected, this yields sub-optimal behavior relative to the theoretical limit of a continuously varied sampling frequency, as clearly demonstrated by the ideal case (green line) falling well below the actual N_{MB} points in the same figure.

3.2. Speedup of waveform generation

We measured the reduction in the total waveform generation time, including both multibanding and interpolation, for compact binary systems with chirp-mass of $\sim 1.48\, M_\odot$. The waveforms were generated up to a frequency f_{max} of 2048 Hz with a time domain sampling rate of 4096 Hz. We used two different waveform models for both generating and analysing injections to test the efficacy of our approach: *TaylorF2* (see for example [28]) and *IMRPhenomPv2* [29]. The former is one of the simplest and most common waveform models available for the coalescence of compact binaries. It analytically describes the inspiral stage of the coalescence using the *stationary phase approximation*. Meanwhile, the analytical *IMRPhenomPv2* model includes the inspiral, merger and ringdown phases, calibrated to numerical relativity simulations. The *IMRPhenomPv2* waveform family has been used to characterize the BBH systems discovered during O1, the first science run of Advanced LIGO [30]. *IMRPhenomPv2* waveforms are more sophisticated and more computationally expensive than *TaylorF2* ones. Since the main effect of the proposed method is reducing the number of template evaluations, it is for computationally expensive cases that we expect to benefit the most from its application.

Figure 4 shows the speedup in the template generation as a function of the starting frequency, for the *TaylorF2* waveform model in the left panel, and for *IMRPhenomPv2* in the right one. The length of the data segments was set by calculating the duration of a BNS signal with $1 M_\odot$ components starting from the chosen f_{min}. The template generation speed was calculated by averaging the time necessary to construct one waveform over 3000 (300) trials for the *TaylorF2* (*IMRPhenomPv2*) model. We define the gain in speed (blue points in figure 4) as the ratio between the average time required by the standard and the MB-Interpolation algorithms to compute one template.

For comparison, the right panel of figure 4 includes the reduction in the number of frequencies at which the waveform is evaluated when using MB-Interpolation, N_{fix}/N_{MB} (red points). The gains for MB-Interpolation are smaller than the ratio N_{fix}/N_{MB} because of the additional cost of interpolating between the N_{MB} frequency samples.

We find that the MB-Interpolation scheme yields a dramatic gain in computational speed for smaller values of f_{min}. At $f_{min} = 20$ Hz *TaylorF2* templates were accelerated by a factor of 10. The slower *IMRPhenomPv2* family shows significantly greater gains than the faster *TaylorF2* family, as illustrated by the difference in ordinate scales between the two panels of figure 4. Thus, *IMRPhenomPv2* template generation was around 25 times faster with MB-Interpolation at $f_{min} = 20$ Hz. While both waveform families show a greater gain for smaller values of f_{min}, the same segmentation is present as in figure 3, although this is less obvious for the *TaylorF2* family. This reflects the dependence of the time required to compute one waveform, T, on the number of frequency bands.

Within the standard approach, this time $T_{Standard}$ can be approximated as the product of the time necessary to calculate the waveform at a given frequency t^w with the number frequencies N_{fix}:

$$T_{Standard} \sim N_{fix} \cdot t^w. \tag{11}$$

To estimate the same time in the MB-Interpolation algorithm we need to take into account two different contributions: the template generation applied to a reduced set of frequencies, and the calculations necessary for the waveform interpolation t^i. This leads to the following approximation:

$$T_{MB-Int.} \sim N_{fix} \cdot t^i + N_{MB} \cdot \left(t^w + \delta t^i\right) \tag{12}$$

Here δt^i represents the time required to compute the quantities necessary for the interpolation (such as phase, derivatives, etc); typically $\delta t^i \ll t^w$.

According to equation (12), the time required to compute a complete waveform via MB-Interpolation depends on f_{min} only through N_{fix} and N_{MB}. However, the first term in equation (12) becomes increasingly dominant as f_{min} decreases, since $N_{fix} \propto f_{min}^{-8/3}$. For sufficiently small f_{min}, $N_{fix} \cdot t^i \gg N_{MB} \cdot \left(t^w + \delta t^i\right)$ and the speedup asymptotes to a fixed factor $T_{Standard}/T_{MB-Int.} \to t^w/t^i$, independent of f_{min}. The frequency at which this happens depends in general on the computational cost of the particular waveform model. The results reported in figure 4 suggest gains exceeding ~ 16 (~ 50) for starting frequencies below 8 Hz for the *TaylorF2* (*IMRPhenomPv2*) waveform models.

3.3. Inference

To verify that the results obtained with the MB-Interpolation algorithm remain accurate, and to measure the speedup of an end-to-end inference run, we also performed several complete PE analyses. We injected a gravitational wave signal emitted by a neutron star binary with
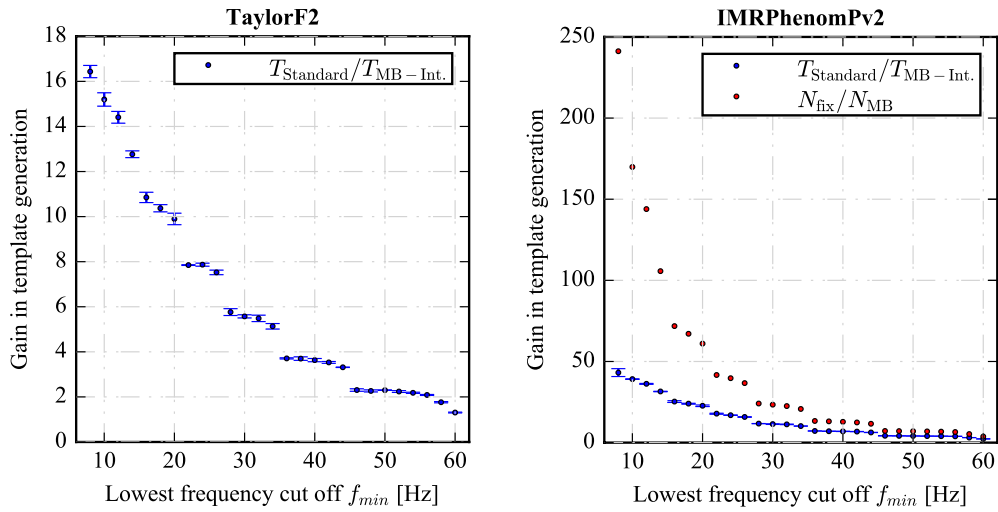
Figure 4. Gain factor in computational speed of template generation as a function of f_{min}. Blue points: ratio of the average waveform computation cost for the standard procedure versus MB-Interpolation as a function of the starting frequency for *TaylorF2* (left panel) and *IMRPhenomPv2* (right panel) waveform families. Red points in right panel: ratio between the number of frequencies at which the waveform is evaluated when using the standard procedure versus MB-Interpolation.

component masses $M_1 = M_2 = 1.4 \, M_\odot$ into stationary Gaussian noise, coloured according to the design sensitivity curves of advanced LIGO and Virgo [31]. The signal was always injected at a distance of $D_L \approx 200$ Mpc so that the SNR at $f_{min} = 40$ Hz source was 15; signals with lower f_{min} have correspondingly higher SNR.

The PE analyses were performed with `LALInferenceNest`, using the same maximum frequency $f_{max} = 2048$ Hz and time-domain sampling rate (4096 Hz) adopted in section 3.2. Priors on companion masses were uniform in the range 1–3 $M_\odot$ and the prior on distance was uniform in volume with a maximum distance of 500 Mpc. We chose this region of mass space as it is the most challenging in terms of computational cost and has the strictest accuracy requirements for waveform interpolation.

3.3.1. PE consistency The analysis of the mock data with MB-Interpolation templates produced posterior distributions statistically identical to the ones obtained with a standard analysis. As a representative case, in figure 5 we show the marginalized posterior probability density functions for chirp mass, mass ratio and luminosity distance, the quantities most sensitive to phase and amplitude errors. We confirmed the visual agreement between the marginal probability distributions obtained with the standard and the MB-Interpolation algorithms by performing a Kolmogorov–Smirnov test, which showed that the two sets of samples are consistent with random draws from the same distribution.

3.3.2. PE speedup To test the effect of the speedup in the waveform generation on the overall PE analysis, we measured the computational time required to perform end-to-end PE runs. We performed PE analyses from different values of the starting frequency f_{min}, consequently changing the lengths of the data segments. The results are reported in table 1 and figure 6.

For each starting frequency and for both standard and MB-Interpolation algorithms, the times of 4 runs have been averaged. The ratio between the average time required to complete
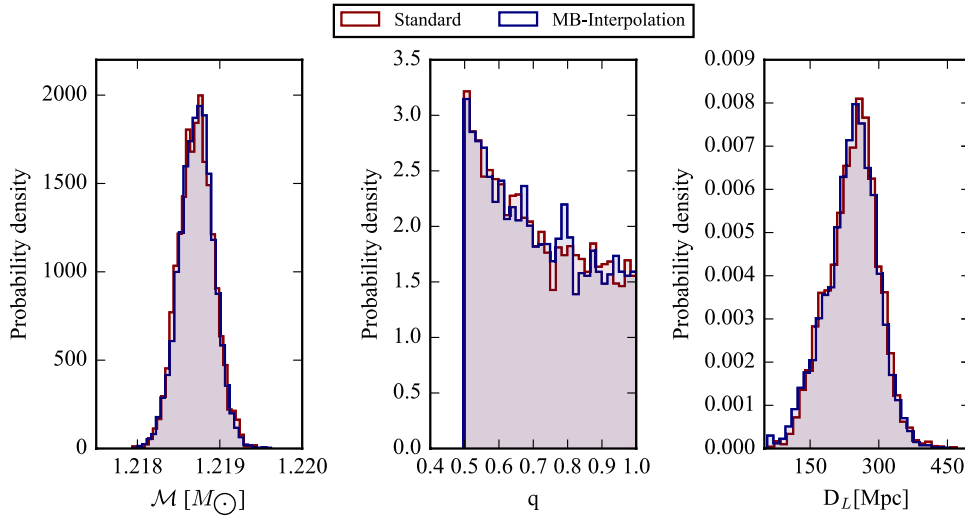
Figure 5. Posterior distributions of chirp mass $\mathcal{M}$, mass ratio q and luminosity distance D_L with the *TaylorF2* waveform model.

a PE analysis adopting the standard and the MB-Interpolation algorithms has been used to define the overall speedup gain G_{PE}. Each group of 4 identical analyses has been run at the same time on a Dual-Core AMD Opteron 2218 Processor with a clock speed of 2.6 GHz.

Table 1 reports the measured speed gain for the whole PE analysis with the *TaylorF2* (abbreviated as TF2) waveform model in the third column. The fourth column contains the speedup in the template calculation (see figure 4). In the last two columns we also report the ratio between the number of frequencies at which the waveform is evaluated in the standard and the MB-Interpolation algorithms N_{fix}/N_{MB}, as well as the idealized improvement in the limit of continuously adapted sampling rates N_{fix}/N_{min} (equation (6)).

As can be seen from table 1, we do not obtain the full idealized gain that may be expected from multi-banding for three reasons. Firstly, the actual number of frequencies at which the waveform is computed in our multi-banding algorithm is larger than the theoretical limit, so $N_{fix}/N_{MB} < N_{fix}/N_{min}$. Secondly, the template computation speedup is less than the reduction in the set of frequencies due to multi-banding, $G_{template} < N_{fix}/N_{MB}$, because of the additional cost of interpolation. Thirdly, the PE speedup is smaller than the speedup in template generation, $G_{PE} < G_{template}$, because template generation is only one component of the PE algorithm. Although the waveform computation is the dominant computational cost for computationally expensive templates, the cost of evaluating the likelihood still grows with the number of frequency bins even when using MB-Interpolation, and along with interpolation this can become the most expensive step when using MB-Interpolation with very long waveforms.

We did not repeat the end-to-end parameter estimation calculations across the full range of starting frequencies with *IMRPhenomPv2* waveforms because the computational cost was unacceptably high when using the standard procedure. Nonetheless, it is possible to estimate the computational cost gain one would achieve with *IMRPhenomPv2* waveforms when starting with low values of f_{min}. For computationally expensive waveforms, the parameter estimation cost is dominated by the waveform computation cost; this is a factor of $\sim$3 higher for *IMRPhenomPv2* waveforms than for *TaylorF2* waveforms. (This factor is independent of the waveform duration or starting frequency, and reflects the difference in the cost of computing the two waveforms at a given frequency point.) Therefore, we expect that the total PE

Table 1. The table reports the results obtained for different values of starting frequency $f_{\min}$ (first column) and corresponding sampling steps δf_0 for standard template generation (second column). The values $G_{\text{PE}}^{\text{TF2}}$ are the actually measured speed gains in the complete PE analyses with *TaylorF2* due to using MB-Interpolation. $G_{\text{template}}^{\text{TF2}}$ is the gain in the waveform generation speed ($T_{\text{Standard}}/T_{\text{MB-Int.}}$). $N_{\text{fix}}/N_{\text{MB}}$ is the ratio between the number of frequencies at which the standard and multibanded waveforms are evaluated, while $N_{\text{fix}}/N_{\min}$ is the limiting case for the reduction in waveform evaluations when continuously adapting frequency steps.

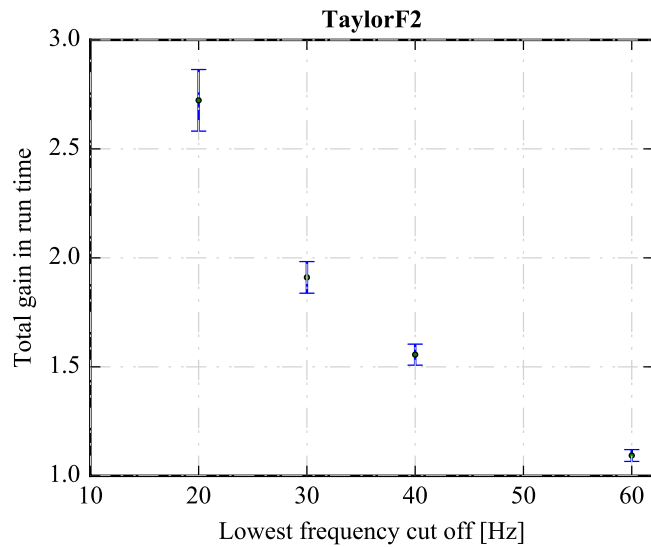| $f_{\min}$ (Hz)	δf_0 (Hz)	$G_{\text{PE}}^{\text{TF2}}$	$G_{\text{template}}^{\text{TF2}}$	$N_{\text{fix}}/N_{\text{MB}}$	$N_{\text{fix}}/N_{\min}$
60	1/16	1.09 ± 0.03	1.31 ± 0.01	3.76	55.4
40	1/64	1.56 ± 0.05	3.8 ± 0.1	12.82	83.8
30	1/128	1.91 ± 0.07	5.5 ± 0.1	23.40	112.2
20	1/300	2.72 ± 0.14	8.8 ± 0.2	61.01	169.1

Figure 6. Observed gains in the end-to-end PE run-time, $G_{\text{PE}}^{\text{TF2}}$, as a function of the starting frequency for analyses carried out with *TaylorF2* waveforms.

computational cost with *IMRPhenomPv2* waveforms to be about the same factor of 3 larger than for *TaylorF2* waveforms when starting at low frequencies and using the standard procedure. Meanwhile, for sufficiently low starting frequencies, the MB-Interpolation waveform computation cost is dominated by interpolation, so that the template computation cost with *IMRPhenomPv2* and *TaylorF2* waveforms when using MB-Interpolation asymptotes to the same value—and so does the full PE computational cost. Thus, we expect that end-to-end PE gains from using MB-Interpolation with *IMRPhenomPv2* will be a factor of $\sim$3 greater than with *TaylorF2* waveforms.

MB-Interpolation is most effective for smaller values of the starting frequency. Figure 4 shows that with $f_{\min} = 8$ Hz the template speed-up factor is $\sim$16 for *TaylorF2* waveforms and $\sim$50 for *IMRPhenomPv2* waveforms, where we conservatively assume that there are no further significant gains in waveform computation at lower starting frequencies because of fixed interpolation costs. As discussed above, the total PE speed-up will not be as large as the speed

gain in template generation because of other fixed costs. Nevertheless, as next generation interferometers (such as the Einstein Telescope [4], KAGRA [32] and the Cosmic Explorer [33]) take advantage of low-frequency data, MB-Interpolation should improve parameter estimation costs by factors of tens, or more for more expensive waveform models.

4. Conclusion

Parameter estimation has played an important role in the opening of the field of gravitational wave astronomy, as demonstrated in the analysis of Advanced LIGO's first observations [11, 12]. The stochastic sampling algorithms used in these analyses require the generation of millions of template waveforms which are compared to the data, a computational task that becomes more expensive as the in-band signal duration increases: for signals from lower mass binaries and for detectors with improved sensitivities at lower frequencies. The generation of computationally expensive template waveforms is the bottleneck in the PE analysis, limiting our ability to obtain results quickly. In this paper we proposed an alternative approach to reduce this cost and consequently the overall time required to produce a result. The procedure is inspired by the same *multi-banding* approach already adopted for low-latency algorithms dedicated to gravitational wave searches [14, 34]. It consists in reducing the set of frequencies at which to evaluate waveforms by dividing the spectral range into different bands and optimising the sampling procedure. However, the greater accuracy required in the context of PE demanded an additional up-sampling of the waveform when computing the likelihood function, which led us to apply a linear interpolation in phase and amplitude.

We have demonstrated the effectiveness of the method by implementing it in the `LALInference` PE code and comparing the results to inference with the full waveform. We found negligible differences between the results at a greatly reduced computational cost. We showed that the MB-Interpolation algorithm reduces the number of frequencies at which the waveform is evaluated by more than an order of magnitude for $f_{min} < 40$ Hz. This leads to an acceleration of the waveform generation and, consequently, the whole analysis. For a fixed chirp mass of the binary, the speed-up factor depends on the complexity of the model and on the starting frequency f_{min}. We studied the most challenging case of binary neutron stars, adopting the *TaylorF2* and *IMRPhenomPv2* waveform families. In section 3 we reported speedup factors in the template generation which reached $\sim$50 for the most sophisticated waveform model (*IMRPhenomPv2*) at $f_{min} \sim 10$ Hz. Although the overall decrease in the computational cost of end-to-end PE is more modest than the improvement in template generation because of fixed costs, we expect factors of tens in speed gain when using *IMRPhenomPv2* templates with starting frequencies of a few Hz. The considerable speedup gains reached by the implementation of the MB-Interpolation method demonstrates the effectiveness of the approach.

Our method is related to the reduced order quadrature models of gravitational waveforms introduced for the simple TaylorF2 model in [6] and later created for more sophisticated SEOBNR and IMRPhenomP models [10]. These methods also result in a large acceleration of PE, by factors of 70 for a TaylorF2 waveform [9] or 300 for IMRPhenomPv2 [10] from 20 Hz. The two methods are conceptually similar in that the number of points in frequency at which the waveform is evaluated is reduced. However, for ROQ the interpolation makes use of a different set of bases, which means that interpolation must be performed across the parameter space of the signals in addition to interpolation in frequency. Unlike our MB-Interpolation method, this requires significant setup costs to create the parameter space interpolants, which are difficult to produce for the higher dimensional precessing spin parameter space. In this regard our method is more flexible, since it can be used for any signal without pre-computing

a reduced order model. This is advantageous in the case of models which include additional physical parameters, such as neutron star tidal deformability, which further increase the dimensionality of the parameter space. The MB-Interpolation method can also be used in combination with ROQ, where the MB-Interpolation is used to accelerate the initial creation of the ROQ model by reducing the number of calculations and also the memory overhead. Indeed, the reduced frequency basis idea was used in [10], but without the interpolation up-sampling step.

Accelerated waveform generation techniques such as the one we have developed here are likely to be essential in future, as the detectors evolve toward greater sensitivity at low frequencies. Third-generation detectors such as the Einstein Telescope [4] or LIGO Voyager [35] will be sensitive down to a few Hz, meaning signals may be in band for hours or longer. The same MB-Interpolation procedure can also be applied to GW studies in the context of space missions [16], and in particular for phase-coherent modeling of the signal in both space-based and ground-based detectors as would be useful for joint science exploitation [36, 37].

In principle, a similar multi-banding approach could also be applied to time-domain waveforms, which could be sampled at a lower rate earlier in the waveform. However, the time domain waveforms of greatest interest use numerical integration of the waveform with an adaptive step size in time. This prevents a great speedup from being obtained as one cannot reduce the step size arbitrarily in the simple way we could for the frequency domain waveforms. This factor, in addition to the technical difficulty of efficiently reconstructing the FFT of a non-uniformly sampled time series prevented us from exploring this option in the current work.

Finally we note that despite the specialisation to gravitational wave analysis, the same technique of adapting the sampling interval could be applied to any area where the signal frequency changes monotonically with time.

Acknowledgments

References

[1] Abbott B P *et al* 2016 Observation of gravitational waves from a binary black hole merger *Phys. Rev. Lett.* **116** 061102
[2] Abadie J *et al* 2010 Predictions for the rates of compact binary coalescences observable by ground-based gravitational-wave detectors *Class. Quantum Grav.* **27** 173001
[3] Veitch J *et al* 2015 Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library *Phys. Rev.* D **91** 042003
[4] Punturo M *et al* 2011 Einstein gravitational wave telescope conceptual design study www.et-gw.eu/ etdsdocument
[5] Punturo M *et al* 2010 The Einstein Telescope: a third-generation gravitational wave observatory *Class. Quantum Grav.* **27** 194002

[6] Field S E, Galley C R, Herrmann F, Hesthaven J S, Ochsner E and Tiglio M 2011 Reduced basis catalogs for gravitational wave templates *Phys. Rev. Lett.* **106** 221102

[7] Canizares P, Field S E, Gair J R and Tiglio M 2013 Gravitational wave parameter estimation with compressed likelihood evaluations *Phys. Rev.* D **87** 124005

[8] Smith R J E, Cannon K, Hanna C, Keppel D and Mandel I 2013 Towards rapid parameter estimation on gravitational waves from compact binaries using interpolated waveforms *Phys. Rev.* D **87** 122002

[9] Canizares P, Field S E, Gair J, Raymond V, Smith R and Tiglio M 2015 Accelerated gravitational wave parameter estimation with reduced order modeling *Phys. Rev. Lett.* **114** 071104

[10] Smith R, Field S E, Blackburn K, Haster C J, Pürrer M, Raymond V and Schmidt P 2016 Fast and accurate inference on gravitational waves from precessing compact binaries *Phys. Rev.* D **94** 044031

[11] Abbott B P *et al* 2016 Properties of the binary black hole merger GW150914 *Phys. Rev. Lett.* **116** 241102

[12] Abbott B P *et al* 2016 Binary black hole mergers in the first advanced LIGO observing run *Phys. Rev.* X **6** 041015

[13] Abadie J *et al* 2012 First low-latency LIGO + Virgo search for binary inspirals and their electromagnetic counterparts *Astron. Astrophys.* **541** A155

[14] Adams T, Buskulic D, Germain V, Guidi G M, Marion F, Montani M, Mours B, Piergiovanni F and Wang G 2016 Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era *Class. Quantum Grav.* **33** 175012

[15] Cannon K *et al* 2012 Toward early-warning detection of gravitational waves from compact binary coalescence *Astrophys. J.* **748** 136

[16] Porter E K 2014 A new method of accelerated bayesian inference for comparable mass binaries in both ground and space-based gravitational wave astronomy (arXiv:1411.0598)

[17] Vines J, Flanagan É É and Hinderer T 2011 Post-1-Newtonian tidal effects in the gravitational waveform from binary inspirals *Phys. Rev.* D **83** 084051

[18] Hinderer T *et al* 2016 Effects of neutron-star dynamic tides on gravitational waveforms within the effective-one-body approach *Phys. Rev. Lett.* **116** 181101

[19] Abbott B P *et al* 2016 Tests of general relativity with GW150914 *Phys. Rev. Lett.* **116** 221101

[20] Agathos M, Del Pozzo W, Li T G F, Van Den Broeck C, Veitch J and Vitale S 2014 Tiger: a data analysis pipeline for testing the strong-field dynamics of general relativity with gravitational wave signals from coalescing compact binaries *Phys. Rev.* D **89** 082001

[21] Cutler C and Flanagan E E 1994 Gravitational waves from merging compact binaries: how accurately can one extract the binary's parameters from the inspiral waveform? *Phys. Rev.* D **49** 2658

[22] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 2007 *Numerical Recipes 3rd edn: the Art of Scientific Computing* (Cambridge: Cambridge University Press)

[23] Ohme F 2012 Analytical meets numerical relativity: status of complete gravitational waveform models for binary black holes *Class. Quantum Grav.* **29** 124002

[24] Baird E, Fairhurst S, Hannam M and Murphy P 2013 Degeneracy between mass and spin in black-hole-binary waveforms *Phys. Rev.* D **87** 024035

[25] Hannam M, Husa S, Ohme F and Ajith P 2010 Length requirements for numerical-relativity waveforms *Phys. Rev.* D **82** 124052

[26] Haster C J, Mandel I and Farr W M 2015 Efficient method for measuring the parameters encoded in a gravitational-wave signal *Class. Quantum Grav.* **32** 235017

[27] Chen H Y and Holz D E 2014 The loudest gravitational wave events (arXiv:1409.0522)

[28] Buonanno A, Iyer B R, Ochsner E, Pan Y and Sathyaprakash B S 2009 Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors *Phys. Rev.* D **80** 084043

[29] Hannam M, Schmidt P, Bohé A, Haegel L, Husa S, Ohme F, Pratten G and Pürrer M 2014 Simple model of complete precessing black-hole-binary gravitational waveforms *Phys. Rev. Lett.* **113** 151101

[30] LIGO Scientific Collaboration, Virgo Collaboration *et al* 2016 Binary black hole mergers in the first advanced LIGO observing run *Phys. Rev.* X **6** 041015

[31] Aasi J *et al* 2016 Prospects for observing and localizing gravitational-wave transients with Advanced LIGO and Advanced Virgo *Living Rev. Relat.* **19** 1

[32] Aso Y, Michimura Y, Somiya K, Ando M, Miyakawa O, Sekiguchi T, Tatsumi D and Yamamoto H 2013 Interferometer design of the KAGRA gravitational wave detector *Phys. Rev.* D **88** 043007

[33] LIGO 2015 Instrument science white paper https://dcc.ligo.org/public/0113/T1400316/004/T1400316-v5.pdf

[34] Cannon K, Hanna C and Keppel D 2012 Interpolating compact binary waveforms using the singular value decomposition *Phys. Rev.* D **85** 081504

[35] McLelland D, Cavaglia M, Evans M, Schnabel R, Lantz B, Quetschke V and Martin I 2016 The LSC-Virgo white paper on instrument science (2016–2017 edn) https://dcc.ligo.org/LIGO-T1600119/public

[36] Sesana A 2016 Prospects for multiband gravitational-wave astronomy after GW150914 *Phys. Rev. Lett.* **116** 231102

[37] Vitale S 2016 Multiband gravitational-wave astronomy: parameter estimation and tests of general relativity with space- and ground-based detectors *Phys. Rev. Lett.* **117** 051102