Accelerating the evaluation of inspiral–merger–ringdown waveforms with adapted grids

Cecilio García-Quirós*, Sascha Husa, Maite Mateu-Lucena and Angela Borchers

Departament de Física, Universitat de les Illes Balears, IAC3 - IEEC, Crta. Valldemossa km 7.5, E-07122 Palma, Spain

E-mail: cecilio.garcia@uib.es

Received 16 April 2020, revised 16 October 2020
Accepted for publication 21 October 2020
Published 27 November 2020

Abstract
This paper presents an algorithm to accelerate the evaluation of inspiral–merger–ringdown waveform models for gravitational wave data analysis. While the idea can also be applied in the time domain, here we focus on the frequency domain, which is most typically used to reduce computational cost in gravitational wave data analysis. Our work extends the idea of multibanding Vinciguerra S et al (2017 Class. Quantum Grav. 34 115006), which has been developed to accelerate frequency domain waveforms, to include the merger and ringdown and spherical harmonics beyond the dominant quadrupole spherical harmonic. The original method of Vinciguerra S et al (2017 Class. Quantum Grav. 34 115006) is based on a heuristic algorithm based on the inspiral to de-refine the equi-spaced frequency grid used for data analysis where a coarser grid is sufficient for accurate evaluation of a waveform model. Here we use a different criterion, based on the local interpolation error, which is more flexible and can easily be adapted to general waveforms, if their phenomenology is understood. We discuss our implementation in the LIGO Algorithms Library (The LIGO Scientific Collaboration 2015 https://lsc-group.phys.uwm.edu/daswg/projects/lalsuite.html) for the IMRPhenomXHM García Quirós C et al (2020 arXiv:2001.10914) frequency domain model, and report the acceleration in different parts of the parameter space of compact binary systems.

Keywords: LIGO, parameter estimation, waveform model acceleration, Einstein Telescope, LISA

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

*Author to whom any correspondence should be addressed.
1. Introduction

The field of gravitational wave astronomy has been born through discoveries of coalescences of compact binary systems consisting of black holes and neutron stars [4–6]. For such systems, very successful programs are being carried out to model the gravitational waveforms expected according to general relativity (and possibly alternative theories) across the astrophysically plausible parameter space of observable binary systems (see e.g. [7–14]). These models are based on synthesising perturbative results, e.g. from post-Newtonian theory [15], black hole perturbation theory [16] and more recently the self-force approach [17], with numerical solutions of the Einstein equations. An important role is also played by the effective-one-body approach [18, 19], which extends post-Newtonian results to a non-perturbative description.

Gravitational wave data analysis as applied to compact binary coalescence is typically split into two steps: searches and Bayesian parameter estimation.

Searches can be performed using matched filter techniques and a fixed set of template waveforms [19, 20], but also in a model-agnostic way [21]. Bayesian parameter estimation [22, 23] is based on a likelihood function that compares the detector data with template waveforms. This is computationally expensive, since robust estimates of Bayesian posterior probability distributions may require several million template waveform evaluations or even significantly more, see also the discussion in [1]. In this paper we discuss one approach to speeding up the evaluation of waveform models, with the main aim of reducing the computational cost of Bayesian parameter estimation. A complementary approach to directly speed up the evaluation of the likelihood function is reduced order quadrature [24, 25]. Some disadvantages of reduced order quadrature are discussed in [1]. We anticipate that future gravitational wave data analysis codes will combine several techniques for speeding up calculations. Our approach is ready for the routine use of state of the art waveform models which contain subdominant harmonics [3] within the LALSuite data analysis framework [2]. One interesting application will be to reduce the time required to compute distance estimates for detected events which improves accuracy by taking into account subdominant harmonics and thus benefits the identification of multi-messenger counterparts.

While the ideas we present in this paper can also be applied in the time domain, we focus on waveform models constructed in the frequency domain. Such models are naturally adapted to matched filtering with detector noise spectral densities that are characterised in the frequency domain, and several such models have been developed in recent years with the goal of computational efficiency, including reduced order models such as [26, 27] or the family of phenomenological waveform models (see e.g. [3, 9–13, 28, 29]). The latter split the wave form for each spherical harmonic into a small number (typically 2–4) of regions based on physical intuition, where the model is then constructed as closed form expressions. We will in particular work with the latest incarnation of such models for quasi-circular non-processing binary systems of black holes, the IMRPhenomX family [3, 28].

In order to model simple non-oscillatory functions, the waveform \( h_{\ell m}(x, \Xi) \) for spherical harmonic \((\ell, m)\) is split into a real amplitude \( A_{\ell m}(x, \Xi) \) and phase \( \phi_{\ell m}(x, \Xi) \). Here \( h \) would typically be the gravitational wave strain when working in the time domain, or its Fourier transform for the purposes of the present paper, and \( x \) the time or frequency, respectively. The quantity \( \Xi \) is a shorthand for all the intrinsic parameters of the waveform, such as masses and spins. We then compute the waveform of each spherical harmonic as

\[
    h_{\ell m} = A_{\ell m} \ e^{i \phi_{\ell m}}. 
\]  

The evaluation of matched filters (e.g. due to optimization over time of arrival) typically requires the evaluation of fast Fourier transforms, which commonly utilize equispaced grids.
In order to resolve all the features of the waveform with such an equispaced grid, a small grid spacing is required. One would then expect that it is possible to construct a computationally much cheaper interpolant without significant loss of accuracy, if the coarse grid that defines the interpolant is chosen appropriately. In such a procedure it is highly advantageous to construct the interpolant in terms of the amplitudes $A_{\ell m}$ and phases $\phi_{\ell m}$, which usually have a simpler morphology than the complex and oscillatory waveforms $h_{\ell m}$. Below we will present a specific prescription to achieve this goal. Following [1] we seek to accelerate not only the evaluation of $A_{\ell m}$ and $\phi_{\ell m}$ as defined by equation (1.1), but also the calculation of the complex exponential $e^{i\phi_{\ell m}}$, through an appropriate choice of coarse grid points and interpolation algorithm. For simplicity, we will use the term ‘multibanding’ to refer to this type of algorithm, and we also use the same two core ideas:

- We split the complete frequency or time range where we want to evaluate a mode $h_{\ell m}$ (see equation (1.1)) into $n$ sub-regions, where each region has a constant grid spacing $\Delta x_n$, which will be chosen such that linear interpolation is sufficiently accurate to maintain the difference between the amplitude and phase with and without multibanding below a certain threshold. Below we will provide a prescription of how to choose $\Delta x_n$ as a function of a freely specifiable threshold parameter. The final waveform can then be evaluated by simple linear interpolation to the fine grid with constant grid-spacing $d_x$, which is determined by the requirements of gravitational wave data analysis. Since interpolation is a faster operation than evaluating a waveform model in new grid points, this step speeds up the computation of the waveform.

- The phase enters the waveform through the complex exponential in equation (1.1). Directly evaluating this exponential at each point of the fine grid would be computationally expensive, but can be avoided by taking into account that we approximate the phase at the fine grid points by linear interpolation between coarse grid points. We can then evaluate the exponential of the phase at the fine grid points by a computationally cheap iterative procedure without further approximations: the value of the phase exponential at a fine grid point is evaluated by simply multiplying the value of the previous fine grid point by a complex number defined only in terms of neighbouring coarse grid points, as discussed in detail in section 2.2.

The key technical difference between our work and [1] is that we change the criterion to compute the grid spacings $\Delta x_n$ in the $n$ coarse grids to use the standard estimate of the local interpolation error derived according to Taylor’s theorem of basic calculus instead of a heuristic algorithm based on the relation between the duration of a data segment and the frequency spacing in the Fourier domain. Our new criterion has several advantages: the use of Taylor’s theorem provides a solid mathematical basis for our algorithm, which allows us to introduce the local numerical error as a free parameter. This parameter can then be used to adjust the accuracy and speed of the algorithm, and will be referred to as the multibanding threshold. Our new criterion can easily be extended to spherical harmonic modes beyond the dominant quadrupolar one, which were not considered in [1]. This allows us to apply our algorithm to the frequency domain IMRPhenomXHM model [3], which describes the gravitational wave signal of non-precessing binary black holes including modes beyond the quadrupolar one: ((2,2), (2,1), (3,3), (3,2), (4,4)). Furthermore, our criterion allows us to extend the analysis from the inspiral to the merger and ringdown, and thus consider waveforms that describe the complete coalescence. We note that it would be challenging to compare the speedup achieved in our work and in [1]: our free multibanding threshold parameter allows us to adjust the speed and accuracy of the algorithm, while for the algorithm in [1] a comparable error estimate would first have to be obtained. Mapping the algorithm in [1] to an appropriate value of the...
parameter in our code would likely be ambiguous and depend on the parameter space of signals. It would also be necessary to investigate the robustness of [1] in the presence of subdominant harmonics and when the merger contributes significantly to the signal-to-noise-ratio. A statement on the relative speedup would thus not be straightforward to interpret. Finally, an adequate comparison would also require a more homogeneous implementation of both algorithms within the LALSuite [2] framework, which goes beyond the scope of this paper.

We will first carry out the analysis separately for the amplitude and phase of the different modes incorporated in IMRPhenomXHM, and then define coarse grids that are appropriate for both phase and amplitude for each mode. For an overview of how interpolation is incorporated in the context of reduced order models see [30].

This paper is organized as follows: in section 2 we discuss the details of this algorithm, and how it is applied to quasi-circular non-precessing frequency domain waveforms for the inspiral, merger and ringdown. In section 3 we present our results for computational efficiency and accuracy, and we conclude with a summary and comments on possible future work in section 4.

The masses of the two objects are defined by \( m_1, m_2 \) so the total mass of the system is \( M = m_1 + m_2 \). We define the mass ratio \( q = m_1/m_2 \geq 1 \) as the ratio of the masses of the bigger black hole over the smaller one. The symmetric mass ratio is computed as \( \eta = m_1 m_2 / M^2 \). In the non-precessing case under consideration, the dimensionless spins of the two black holes only have \( z \) components which are parallel to the orbital angular momenta and we denote by \( \chi_{1,2} \). The set \( (q, M, \chi_1, \chi_2) \) or \( (\eta, M, \chi_1, \chi_2) \) then defines the intrinsic parameters of a non-precessing binary, where for the purposes of waveform modelling the total mass \( M \) does not have to be taken into account, as it takes the role of a scale factor.

### 2. Algorithms

#### 2.1. Interpolation error

In the following, we will restrict ourselves to linear interpolation for simplicity. In section 3 we find that this leads to a significant speedup of waveform evaluation. More sophisticated higher order schemes may offer further improvements, but may also require further study, e.g. in order to avoid computational bottlenecks due to the increased complexity of the code.

A real-valued differentiable function \( g(x) \) can be approximated at a point \( x_0 \) by a linear approximation in the following sense: there exists a function \( h(x) \) such that

\[
g(x) = g(x_0) + g'(x_0)(x - x_0) + h(x)(x - x_0), \quad \lim_{x \to x_0} h(x) = 0. \tag{2.1}
\]

The error \( R(x) \) of the approximation is

\[
R(x) = h(x)(x - x_0). \tag{2.2}
\]

According to standard refinements of Taylor’s theorem of basic calculus, the error term \( R(x) \) can be estimated using the second derivative \( g''(x) \) of the function \( g \) we want to approximate by the statement that there exists a \( \xi \), \( x_0 \leq \xi \leq x \), such that

\[
R(x) = \frac{g''(\xi)}{2}(x - x_0)^2. \tag{2.3}
\]
If we apply this result to our problem of interpolating to a fine grid from a coarse grid with grid spacing $\Delta x_n$, then

$$ R(x) \leq \max_{s \leq \xi \leq r} \frac{g''(\xi)}{2} \Delta x_n^2. $$

Consequently we can choose our coarse grid spacing $\Delta x_n$ to satisfy a given error threshold $R$ as

$$ \Delta x_n \leq \sqrt{\frac{2R}{\max_{s \leq \xi \leq r} g''(\xi)}}. $$

Our application of interpolation will initially be guided by the requirements of phase accuracy, and we will then discuss in which sense these criteria also lead to a sufficiently small amplitude error. Below we will develop the details of constructing a hierarchy of grids as appropriate for linear interpolation of both the frequency domain phase and amplitude for different spherical \[31\] or spheroidal \[32\] harmonic modes, and describe how to efficiently evaluate complex exponentials of the phase on such a grid hierarchy.

The hierarchy of grids is determined by the behaviour of the second derivative of the phase as a function of the frequency according to equation (2.5). We distinguish between three main regions: inspiral, merger and ringdown. As shown in figure 1, the behaviour of the phase derivative ($\phi'_{lm} = d\phi_{lm}/df$) is sharper and changes very drastically, then more points will be needed in this region. However, the merger and ringdown parts are ‘flatter’ and less points will be necessary to describe these parts.

2.1.1. **Inspiral in the frequency domain.** In order to derive an appropriate frequency grid spacing $\Delta f$ ($f$ is the dimensionless frequency in geometric units $G = c = 1$) for the Fourier domain phase during inspiral, we will approximate the phase by the standard leading TaylorF2 phase
expression [33],
\[ \Phi_{\ell m} = c_0 + c_1 f + \frac{m}{2} \frac{3}{128 \eta} \left( \frac{2\pi f}{m} \right)^{-5/3}, \]
(2.6)

where \( c_0, c_1 \) are constants of integration which can be identified with the freedom to rotate the source and with the time of coalescence respectively (see e.g. equation (3.18) of [33]). The constants do not affect the second derivative, which reads
\[ \Phi''_{\ell m} = \frac{m}{2} \frac{5}{48 \eta} \left( \frac{2\pi}{m} \right)^{-5/3} f^{-11/3}. \]
(2.7)

The phase and phase derivatives become singular as the frequency \( f \) approaches zero, and the magnitude of the second derivative increases toward decreasing frequency. Note that the singularity in the phase at vanishing frequency is due to the fact that an infinite number of cycles is described by a finite frequency interval. We can therefore estimate the maximal second phase derivative as the second derivative of the phase at the start of each frequency interval for which we want to use interpolation, assumed at frequency \( f \), and obtain
\[ \Delta f_{\text{phase}}(f) = \sqrt{\frac{2R}{\Phi''_{\ell m}(f)}} = \sqrt{\frac{2R}{c_{f,\text{insp}}}} f^{11/6}, \]
(2.8)
where
\[ c_{f,\text{insp}} = \frac{m}{2} \frac{5}{48 \eta} \left( \frac{2\pi}{m} \right)^{-5/3}. \]
(2.9)

In section 2.3 we will use equation (2.8) to split the calculation of the phase into frequency bins, where in each bin the grid spacing \( \Delta f \) is kept constant, but it increases from bin to bin with increasing start frequency of the bin.

We now turn to the inspiral amplitude. For the modes considered in this work, the known leading order terms of the Fourier domain post-Newtonian description are as follows, see e.g. [3], where we use the definitions \( v = (2\pi f / m)^{1/3} \) and \( \delta = \sqrt{1 - 4\eta} \):
\[ A_{\ell m} = \pi \sqrt{2 \eta/3} v^{-7/2} a_{\ell m}, \]
(2.10)
\[ a_{22} = 1 + O(v^2), \]
(2.11)
\[ a_{21} = v \delta \sqrt{2/3} + O(v^2), \]
(2.12)
\[ a_{33} = v \delta \frac{3}{4} \sqrt{\frac{5}{7}} + O(v^3), \]
(2.13)
\[ a_{32} = v^2 \frac{3}{5} \sqrt{\frac{5}{7}} (1 - 3\eta) + O(v^4), \]
(2.14)
\[ a_{44} = v^2 \frac{4}{5} \sqrt{\frac{10}{7}} (1 - 3\eta) + O(v^5). \]
(2.15)

The leading order terms already show that the amplitude of all the modes decreases for higher mass ratios, approaching zero for the extreme-mass-ratio limit as seen from equation (2.10). From equations (2.11)–(2.15) we see however that for higher mass ratios the
amplitudes of the higher modes relative to the (2, 2) mode are more important. As a consequence, the higher modes are more significant for highly asymmetric systems, and they have indeed recently been observed for the GW190412 and GW190814 events [34, 35].

For the amplitude, it is natural to define a threshold for the relative error of the interpolation, which we denote by $r$. The frequency dependent coarse grid resolution $\Delta f_{\text{amp}}^\ell m(f)$ which results from specifying a relative error threshold $r$ is then independent of $\eta$ and depends linearly on the frequency,

$$\Delta f_{\text{amp}}^\ell m(f) = \sqrt{\frac{2rA_{\text{lin}}}{A_{\text{lin}}^n}} = c_{\ell m} f \sqrt{r}, \quad (2.16)$$

where

$$c_{22} = 6 \sqrt{\frac{2}{91}}, \quad (2.17)$$

$$c_{21} = c_{33} = 6 \sqrt{\frac{2}{55}}, \quad (2.18)$$

$$c_{44} = c_{32} = 2 \sqrt{\frac{2}{3}}. \quad (2.19)$$

Combining expressions (2.8) and (2.16) we can then write the ratio of coarse grid spacing required for the phase to stay below a phase error of $R$ radians to the coarse grid spacing required for the amplitude to guarantee a relative amplitude error below $r$ as

$$\frac{\Delta f_{\text{phase}}^\ell m}{\Delta f_{\text{amp}}^\ell m} = \frac{\sqrt{\eta}}{c_{\ell m}} \sqrt{\frac{R}{r}} (\pi f)^{5/6}, \quad (2.20)$$

where $\beta_{\ell m}$ are constant factors straightforwardly obtained by computing the quotient between expressions (2.8) and (2.16), and are given by

$$\beta_{22} = 2 \sqrt{\frac{91}{55}}, \quad \beta_{21} = 2^{1/3} \sqrt{\frac{11}{3}}, \quad (2.21)$$

$$\beta_{33} = 2^{7/3} \sqrt{\frac{11}{3}}, \quad \beta_{32} = 6 \sqrt{\frac{21}{5}}, \quad \beta_{44} = 3 \sqrt{\frac{21}{5}}. \quad (2.22)$$

Choosing e.g. $r = R$, the expression (2.20) is always smaller than unity up to the minimum energy circular orbit (MECO) frequency [36]. The step size restriction for the phase is thus more restrictive than the one for the amplitude. For simplicity we will use the phase criteria to build just one coarse frequency array and use this for both the phase and amplitude. We discuss the merger and ringdown in the next section.

The values of $R$ and $r$ can in principle be specified freely, e.g. to suite the accuracy requirements for specific events, depending on factors such as their signal-to-noise ratio, we set some default values with the help of the tests carried out in section 3, but the user has the ability to change this behaviour. The default value was chosen to be $10^{-3}$ for both $R$ and $r$, meaning that the phase will have a maximum absolute error of $10^{-3}$ radians with respect to the phase without multiband and the amplitude will have a maximum relative error of 0.1% with respect to the amplitude without multiband.
2.1.2. Merger and ringdown in the frequency domain. The merger–ringdown phase exhibits a morphology that is rather different from the inspiral. A detailed phenomenological description for the $\ell = |m| = 2$ is provided in the contexts of the construction of the IMRPhenomD [12] and IMRPhenomXAS [28] waveform models, and for subdominant modes the IMRPhenomXHM model [3]. This allows us to identify the crucial features of the merger–ringdown regime, and to adapt the estimate (2.5) for the step size as we have done for the inspiral.

The first ingredient will be to identify the end of the inspiral. In [3, 28] we confirm that the MECO [36] provides a good approximation for the transition between inspiral and merger for comparable masses. In the merger–ringdown part, the Fourier domain phase derivative is given by a superposition of a Lorentzian function and a background term [3, 28]. The Lorentzian dominates the phase derivative and is given by (see equation (6.3) in [3])

$$\Phi'(f) = \frac{a}{(f - f_0)^2 + b^2}, \quad (2.23)$$

and the second derivative by

$$\Phi''(f) = \frac{-2a(f - f_0)}{(f - f_0)^2 + b^2}, \quad (2.24)$$

where we have introduced the shorthands $a = \alpha_\lambda f_{\text{damp}}^\text{inst} \lambda f_{\text{damp}}$ (note this coefficient is not related to those in equations (2.11)–(2.15), $b = f_{\text{damp}}^\text{inst}$, and $f_0 = f_{\text{damp}}^\text{inst}$. Thus, $\alpha_\lambda$ is a term that determines the overall amplitude of the Lorentzian, $f_0$ is the frequency at which the ‘dip’ of the Lorentzian happens, and $b$ is a measure of the width of the dip. Inserting the expression for the Lorentzian into equation (2.5) we obtain

$$\Delta f(f) = \sqrt{\frac{2R}{\Phi'(f)}} = \sqrt{\frac{R}{a|f - f_0|}} \left((f - f_0)^2 + b^2\right), \quad (2.25)$$

which replaces equation (2.8) for computing the spacing of the coarse frequency grid in the merger and ringdown.

The spacing computed according to (2.5) depends on the absolute value of the second derivative, and we note that the second derivative of the Lorentzian phase function, $\Phi''$, has two local maxima for $f_0 \pm b/\sqrt{3}$, with identical absolute value

$$\left|\Phi'' \left(f_0 \pm \frac{b}{\sqrt{3}}\right)\right| = 3\sqrt{3}a \frac{2}{8b^3}. \quad (2.26)$$

While for the inspiral the number of frequency bins depends on the start frequency, as we will discuss in more detail below in section 2.3, for the merger and ringdown we choose two bins which we call the merger and ringdown bins. The merger bin is defined by the frequency interval $(f_{\text{Lorentzian}}, f_{\text{Lorentzian}}^\text{inst})$, and captures the frequency regime where high resolution is required to capture the shape of the Lorentzian. The frequency $f_{\text{Lorentzian}}$ marks the end of the inspiral region of the IMRPhenomXAS model for the $\ell = |m| = 2$ mode and the IMRPhenomXHM model for the other harmonics, and is chosen approximately at the MECO frequency (see [3, 28] for details and a further discussion of waveform morphology). We set the frequency $f_{\text{Lorentzian}}^\text{inst}$ to $f_{\text{Lorentzian}}^\text{inst} + 2f_{\text{damp}}^\text{inst}$, which is chosen to approximate the lowest frequency where the second phase derivative of the Lorentzian can be neglected, and the first phase derivative is approximately constant. We find that this pragmatic choice works well, however a more sophisticated choice might be found advantageous in the future. The ringdown bin starts at this frequency, and is the highest frequency bin in our procedure. It is characterized by low
resolution requirements for the phase due to neglecting \( \Phi'' \) and ends at the highest frequency of the waveform, the frequency range of this last bin is thus \( (f_{\text{Lorentzian}}, f_{\text{max}}) \).

We compute the grid spacing of both bins by evaluating the maximum value of \( |\Phi''| \) in these two intervals according to equations (2.8) and (2.26) and inserting it into equation (2.5). For the merger bin this is the maximum between the inspiral value and the value for the Lorentzian, and we thus obtain

\[
\Delta f_{\text{phase}}^{\text{merger}} = \min \left( \sqrt{\frac{2R}{c_{f,\text{insp}}}} \frac{h_{\text{hcp}}}{h_{\text{insp}}}, \frac{4f_{\text{damp}}^{\text{lm}}}{3^{3/4}} \sqrt{\frac{R}{|\alpha\lambda|}} \right),
\]

(2.27)

For the ringdown bin, the second phase derivative \( |\Phi''| \) decreases monotonically to zero, we thus take the value at the start of the region \( f_{\text{Lorentzian}} \), which yields

\[
\Delta f_{\text{phase}}^{\text{RD}} = 5f_{\text{damp}}^{\text{lm}} \sqrt{\frac{R}{2|\alpha\lambda|}}.
\]

(2.28)

Again we turn to the amplitude now. We approximate the amplitude falloff in the ringdown bin as

\[
h \approx e^{-\Lambda f},
\]

(2.29)

with \( \Lambda = \lambda/(f_{\text{damp}}^{\text{lm}}/\sigma) \), where these coefficients correspond to those used in the ringdown ansatz for the IMRPHENOMXHM model (see [3] for more details):

\[
A_{\text{RD}}^{\text{lm}} \propto \frac{1}{(f - f_{\text{ring}}^{\text{lm}})^2 + \left(f_{\text{damp}}^{\text{lm}}/\sigma \right)^2} e^{-\left(f_{\text{damp}}^{\text{lm}}/\sigma \right)^\lambda}.
\]

(2.30)

The grid spacing required to guarantee a relative error smaller than \( r \) is then given by

\[
\Delta f_{\text{amp}}^{\text{RD}}(r) = \frac{\sqrt{2 rf}}{\Lambda},
\]

(2.31)

which is independent of the frequency \( f \). For \( r = R \) this condition is typically more restrictive than the condition (2.28) derived from the phase, the dependence across parameter space is however complicated. We therefore always compute the two frequency spacings, and then use the more restrictive one. We believe that this choice is quite conservative and that the choice could be relaxed in the future, since our ringdown bin only starts at frequencies where the amplitude is already quite small. Note that the start frequency of our ringdown region is either significantly higher than the ringdown frequency, or, for very high spins, the exponential falloff is significantly steeper than for moderate spins. In consequence we could always use the phase criterion (2.27) to set the grid spacing in the ringdown region without worrying too much about loss of accuracy. If greater amplitude accuracy for the ringdown would be required, it would be also possible to switch from linear interpolation to the fine grid to third order spline interpolation for the amplitude.

In the merger bin, the functional dependence of the mode amplitudes is more complex (see [3]). In this case we compute numerically the grid spacing \( \Delta f \) for the amplitude as

\[
\Delta f_{\text{amp}}^{\text{merger}} = \sqrt{2 \frac{r |h_{\text{insp}}(f)|}{|h_{\text{insp}}(f)|''}}.
\]

(2.32)
We evaluate this quantity for the merger bin across our parameter space with the choice \( r = R \) and compare with the grid spacing derived for the phase given by equation (2.27). We find that the ratio \( \Delta f_{\text{merger}}^{\text{phase}} / \Delta f_{\text{merger}}^{\text{amp}} \) is typically lower than one so the criteria for the phase is more restrictive than the one for the amplitude. We find that for some cases with comparable masses and high positive spins the ratio is between a value of one and two, but for simplicity we will always choose the criterion for the phase. This means that in some cases the grid for the amplitude will be up to twice coarser than it should, and given the quadratic relation between the grid spacing and the threshold given by equation (2.5) we will interpret this choice such that the actual relative amplitude and phase errors will be bounded by the thresholds within a factor of four. We leave refinements of the simple strategy to set \( r = R \) to future work.

Below we will study the mismatch between the original model and different levels of error threshold to arrive at a more practical evaluation of error than to check for local deviations between model and approximation, and in section 3.3 we will perform a parameter estimation exercise and find that all choices of the value of \( R = (0.1, 0.01, 0.001) \) lead to indistinguishable results for the case considered. Note that for higher signal-to-noise ratios more stringent criteria may be desired.

2.2. Efficient evaluation of complex exponentials

The evaluation of the complex exponential function when constructing the strain from amplitude and phase as in equation (1.1) is one of the most time consuming operations in the C-code of the LALSuite [2] implementation of our model. The number of required evaluations of the complex exponential (or, equivalently, of trigonometric functions), can however be reduced drastically by implementing the method described in [1] (adapted from [37]). Instead of interpolating the phase on the uniform fine grid and computing the complex exponential, we compute the complex exponential in the non-uniform coarse grid and then rewrite the interpolation of this quantity in terms of an iterative algorithm. In this way we reduce the number of points where the complex exponential has to be evaluated. The gain in speed using this method can be approximated by the ratio of the number of points of the coarse grid over the fine grid times the evaluation time of the complex exponential.

Let \( \Phi_j \) be the phase at one coarse frequency point \( f_j \) and let \( \hat{\Phi}_k, \hat{f}_k \) be the estimated phase and the frequency at one point of the final uniform frequency grid, the spacing of the uniform grid is therefore \( df = \hat{f}_{k+1} - \hat{f}_k \). Then we use the simple identity

\[
e^{i\hat{\Phi}_{k+1}} = e^{i\hat{\Phi}_k} e^{idf \frac{\Phi_{j+1} - \Phi_j}{f_{j+1} - f_j}}
\]

(2.33)

to compute the complex exponential recursively on the fine frequency grid points that lay between two coarse frequency points \( j \) and \( j + 1 \). The first of the fine points is given by \( e^{i\hat{\Phi}_0} = e^{i\Phi_j} \).

2.3. Complete multibanding algorithm in the frequency domain

We will now describe our final algorithm for accelerated waveform evaluation, which is based on our previous results. Our final results will be the strain, evaluated on a uniform frequency grid, with a resolution \( df \) that is adapted to the requirements of some given data analysis application. The motivation for uniform grid spacing stems for the typical context of matched filtering, where an inverse Fourier transform is used to optimize a match over the time shift between a signal and a template. We will refer to this uniform frequency grid as the fine grid.
In order to accelerate the waveform evaluation we will however only evaluate our model waveform on a coarser non-uniform grid, and then use the iterative evaluation described above in section 2.2 to evaluate the complex exponential of the phase.

By default we will use linear interpolation for the amplitude, with optional cubic spline interpolation. Both interpolation algorithms are currently using the open source GSL library [38], however, based in some profiling studies, we expect a further speedup of up to 10%–20% by replacing the GSL implementation by adding a standalone implementation of the required interpolations to our code.

We will now first discuss how to construct the non-uniform coarse frequency grid, and then the details of how to evaluate the waveform on the fine grid, first for spherical harmonics without mode mixing, and then for modes with mode mixing [39–41], which for the current IMRPhenomXHM model concerns only the $\ell = 3, |m| = 2$ modes.

### 2.3.1. Building the coarse frequency grid.

We assume that we are given an input frequency range $(f_{\text{min}}, f_{\text{max}})$ where we need to evaluate the spherical harmonic modes of the waveform. We wish to construct a non-uniform frequency grid, such that for every two successive frequency points the grid spacing $\Delta f(f)$ between them is sufficiently small to guarantee that the local phase error resulting from using linear interpolation between the coarse frequency points is smaller than a given threshold value $R$. We can then use equations (2.8) and (2.25) to compute $\Delta f$ as a function of the threshold $R$, the frequency $f$, the intrinsic parameters $(q, \chi_1, \chi_2)$, and the spherical harmonic mode $(\ell, m)$ under consideration. The coarse grid will also depend on the desired grid spacing $df$ for the final uniform grid, since we build the coarse frequency grid such that the coarse points also belong to the fine grid. This simplifies the interpolation procedure for the complex exponential.

Lower values for the threshold $R$ result in smaller errors, but higher computational cost. In section 3 we will compare different threshold settings and evaluate the computational cost and compare the actual errors with the chosen threshold $R$.

As mentioned above in section 2.1 we split the frequency range into three regions corresponding to the inspiral, merger and ringdown. For the practical implementation, instead of using the continuously varying $\Delta f(f)$ of expressions (2.8) and (2.25), we work with a series of frequency bins where $\Delta f$ is fixed in each bin. The merger and ringdown parts have a much smaller dynamic range for $\phi'(f)$ than the inspiral part (the phase ‘flattens out’ from inspiral toward merger), and we just use one frequency bin for each region. Their spacings $\Delta f_{\text{insp}}$ and $\Delta f_{\text{RD}}$ are given by equations (2.27) and (2.28)–(2.31) respectively.

However, the inspiral part has a large dynamic range, and $\Delta f$ given by (2.8) changes with a power law of $f^{11/6}$ so it also changes fast. The spacing that would accurately describe the whole inspiral part would be $\Delta f(f_{\text{min}})$. However, if we used this spacing for the whole region, we would be using many more points than what are really needed since $\Delta f$ increases so much for frequencies above $f_{\text{min}}$. Therefore we use a varying number of frequency bins $N$, and we build each of them with a spacing $\Delta f_i$ twice larger than the previous bin. For the first bin we set $\Delta f_0 = \Delta f(f_{\text{min}})$, thus

$$\Delta f_0 = \sqrt{\frac{2R}{c_{\text{insp}}} f_{\text{min}}^{11/6}}, \quad (2.34)$$

$$\Delta f_i = 2^i \Delta f_0, \quad i = 1, 2, \ldots, N. \quad (2.35)$$

In practice we require that between two coarse points there is an integer number of fine frequency points, in consequence we modify $\Delta f_0$ such that
\[ \Delta f'_i = \int \frac{\Delta f_i}{df} df. \]  
\hspace{1cm} (2.36)

Now that we have computed the spacing of each frequency bin, we need to compute the final frequency of each bin \( f_{i,\text{end}} \), which is the frequency that doubles the spacing \( \Delta f_i \) of the current bin, i.e. we have to solve the equation \( \Delta f(f_{i,\text{end}}) = 2\Delta f_i \). Inserting this into equation (2.8) we obtain

\[ \frac{f_{i,\text{end}}}{f_{i,\text{start}}} = 2\frac{6}{11}. \]  
\hspace{1cm} (2.37)

\[ f_{i,\text{end}} = 2 f_{i,\text{start}}, \]  
\hspace{1cm} (2.38)

We require that in a frequency bin there must be an integer number of coarse frequency points, so we modify the end frequencies of each bin to

\[ f'_{i,\text{end}} = \int \left[ f_{i,\text{end}} - f_{i,\text{start}} \right] \Delta f_i. \]  
\hspace{1cm} (2.39)

With the above frequency factor we can estimate the number \( N \) of bins that will be used in the inspiral. Since the inspiral regions ends at \( f_{\text{insp}} \), \( N \) has to satisfy the relation

\[ \frac{f_{\text{insp}}}{f_{\min}} = \left( \frac{2}{\alpha} \right)^N, \]  
\hspace{1cm} (2.40)

and therefore we obtain

\[ N = \log_2 \left( \frac{f_{\text{insp}}}{f_{\min}} \right) = \frac{11}{6} \log_2 \left( \frac{f_{\text{insp}}}{f_{\min}} \right). \]  
\hspace{1cm} (2.41)

Since \( N \) is however the number of constant frequency bins for the inspiral, it has to be an integer number, and we replace \( f_{\text{insp}} \) by \( f'_{\text{insp}} \) such that

\[ N = \frac{11}{6} \log_2 \left( \frac{f'_{\text{insp}}}{f_{\min}} \right) = \int \left[ \frac{11}{6} \log_2 \left( \frac{f_{\text{insp}}}{f_{\min}} \right) \right]. \]  
\hspace{1cm} (2.42)

For the merger and ringdown regions, we proceed analogously to the inspiral region, and ensure that an integer number of fine grid points aligns with the coarse grid. Since this algorithm depends on the input values for \( f_{\min}, f_{\max} \) and \( df \), we perform several sanity checks to ensure that there is not any overlapping between regions. For example, if \( f_{\text{insp}} > f_{\text{Lorentzian}} \) we skip the merger bin or if \( f_{\text{Lorentzian}} > f_{\text{max}} \) we skip the ringdown bin.

In figure 2 we compare the final non-uniform coarse grid with the uniform grid. In the top panel we can see how the frequency spacing \( \Delta f \) increases for subsequent bins that constitute the inspiral part. In the case shown, the merger bin has a slightly lower \( \Delta f \) than the last inspiral bin in order to resolve the Lorentzian feature of the phase derivative. For other cases where the Lorentzian is less pronounced the limiting factor will be the derivative at the end of the inspiral and then the merger will have exactly twice the spacing of the last inspiral bin. The ringdown bin is the one with a coarser \( \Delta f \) since there the phase derivative is practically flat. In the bottom panel we show the number of frequency points for the inspiral, merger and ringdown parts. The uniform grid has most of its frequency points in the merger–ringdown part, which leads to an excessive computational cost in these regions, where far fewer points are required to capture the flatter behaviour of the phase derivative (see figure 1). For the non-uniform grid, most of
the points are located in the inspiral part, where high resolution is needed to describe the phase, which becomes singular as the frequency tends to zero.

Now that we have described the non-uniform coarse frequency grid, the next step is to evaluate the model in this grid and carry out the interpolation to the fine grid. In this next step however, different procedures need to set up for the modes with and without mixing, as we discuss below.

2.3.2. Evaluate the modes on the fine grid, with and without mode-mixing. For the modes without mixing (at the moment all modes except $\ell = 3$, $|m| = 2$), the waveform modes are evaluated on the fine grid as follows:

(a) First the amplitude and phase are evaluated separately for the coarse frequency grid, which yields two 1D arrays, one for the amplitude and one for the phase.

(b) The complex exponential $e^{i\phi_{lm}}$ is computed on the coarse grid.
(c) The fine uniform frequency grid is constructed with spacing \( df \).
(d) The complex exponential is interpolated to the final uniform frequency grid following the procedure described in section 2.2.
(e) The amplitude is interpolated to the fine grid by using linear (optionally third) order interpolation (using the GSL library).
(f) The complex waveform \( \tilde{h}_{lm} \) is constructed by multiplying the arrays for the amplitude and the complex exponential on the fine grid.

For the modes with mixing [39–41] (in our present implementation of IMRPhenomXHM these are only the \( \ell = 3, |m| = 2 \) modes) our procedure is slightly different from the modes without mixing. To handle mode mixing, in the ringdown region the model is built in terms of spheroidal harmonics [32] instead of spherical harmonics [31], to simplify the waveform and avoid sharp features in the phase derivative and in the amplitude, as discussed in detail in [3]. After building the model waveform in terms of spheroidal harmonics, it is then rotated back to spherical harmonics and connected with the inspiral part, which is directly modelled in terms of spherical harmonics. Performing our interpolation of the ringdown part in terms of the spherical harmonics as for the other modes would require significantly higher resolution and increase computational cost. We thus use the same strategy as we have employed to construct the modes without mixing, and apply our multibanding algorithm separately to the inspiral region expressed in spherical harmonics, and to the ringdown part expressed in spheroidal harmonics, and then transform the latter to spherical harmonics once the fine grid values have been computed. Our detailed procedure is as follows:

(a) We split the coarse frequency array into the spherical part, where we will perform the model evaluation and multibanding in terms of the spherical harmonics, and the spheroidal part, where we transform from the spheroidal to the spherical representation in the ringdown region.

The start frequencies of the ringdown region for the phase and amplitude, \( f_{\text{RD}}^{\text{phase}}, f_{\text{RD}}^{\text{amp}} \), are given in equation (5.2) in [3]. Note that \( f_{\text{RD}}^{\text{phase}} < f_{\text{RD}}^{\text{amp}} \). For our multibanding algorithm we split between the ‘spherical’ and ‘spheroidal’ coarse grids, where the spherical and spheroidal amplitude and phase are computed. There is some overlap between the frequency ranges of both in the interval \((f_{\text{RD}}^{\text{phase}}, f_{\text{RD}}^{\text{amp}})\), since the spherical array goes up to \( f_{\text{RD}}^{\text{amp}} \), but the spheroidal one starts at \( f_{\text{RD}}^{\text{phase}} \), see step (h) below.

(b) Evaluate the spherical amplitude and phase in the spherical coarse array and evaluate the spheroidal amplitude and phase in the spheroidal coarse array, we get therefore four one-dimensional arrays.

(c) Compute the complex exponential for the two coarse arrays of phases.

(d) Build the uniform frequency grid with spacing \( df \) and split it into spherical and spheroidal parts as above.

(e) Interpolate the two arrays of complex exponential in their respective regions using the iterative procedure described in section 2.2.

(f) Interpolate the two arrays of amplitude in their respective regions using linear (optionally third) order spin interpolation using the GSL library [38].

(g) We have thus obtained four arrays: spherical amplitude and complex exponential evaluated in the spherical fine grid, and spheroidal amplitude and complex exponential in the spheroidal fine grid.

(h) Finally we combine amplitude and phase with different procedures in three frequency ranges:
1. \( f_{\text{min}} \leq f < f_{\text{phase}}^{\text{RD}} \): we directly multiply spherical harmonic amplitude and complex exponential.

2. \( f_{\text{phase}}^{\text{RD}} \leq f < f_{\text{amp}}^{\text{RD}} \): we rotate to spherical the spheroidal complex exponential term (which requires the spheroidal amplitude), and then multiply the resulting spherical complex exponential with the spherical amplitude.

3. \( f_{\text{amp}}^{\text{RD}} < f \leq f_{\text{max}} \): we first multiply the spheroidal amplitude and complex exponential and then transform to the spherical basis.

3. Results

3.1. Computational performance

We first test the gain in speed due to multibanding and compare the results for different threshold values and for different spacings of the fine frequency grid. Note that the frequency spacing \( df \) of the grid in the Fourier domain is related to the duration \( T \) of the time segment that is analyzed by

\[
df = \frac{1}{T},
\]

and thus longer signals require a smaller grid spacing. To illustrate this dependency, in figure 3 we show the approximate duration of a binary black hole coalescence signal as a function of mass and mass ratio. To leading post-Newtonian order the duration in dimensionless units is given by

\[
T/M = \frac{5}{256\eta (\pi M f_0)^{8/3}},
\]

where \( f_0 \) is the frequency where the dominant spherical harmonic mode, \( |\ell| = |m| = 2 \) enters the frequency band of the detector. Lower start frequencies thus imply much longer signals. In figure 3 we show results for two values of the lower frequency cutoff of the detector, \( f_0 = 10 \) Hz, 20 Hz, the latter is what is typical for current compact binary parameter estimation, see e.g. [42, 43]. The coalescence time is approximated with the non-precessing TaylorT2 approximant at second post-Newtonian order, and extreme Kerr values, adding a very rough time of 500 \( M \) in geometric units to roughly account for typical lengths of merger and ringdown [12], in order to obtain an approximate upper limit on the duration. Figure 3 focuses on short signals, where a time duration of 4 s is appropriate, and show the range of signals and templates in mass and mass ratio that fit into this time window.

We will now discuss an example case of a non-spinning system of black holes with total mass of 50 solar masses and mass ratio \( q = 1.5 \), and evaluate the computational cost as a function of frequency spacing \( df \). In figure 4 we show the evaluation time of one waveform versus the spacing of the final uniform frequency grid. The frequency range spans from 10 to 4096 Hz and we fix the mass of the system to 50 \( M_\odot \). The dashed lines represent the waveforms generated without multibanding while the solid lines correspond to the multibanding version with different values of the threshold: \( 10^{-4}, 10^{-2}, 10^{-3}, 10^{-4} \). First, we focus on the no-multibanding results, in principle we would expect that the higher modes model is 5 times slower than the (2, 2)-mode-only model because IMRPhenomXHM has 5 modes instead of just one. However it is a bit more expensive due to some particularities that are only present in the higher modes code, like the checks for the amplitude vetoes and mainly the extra steps needed to describe the mode-mixing of the \((3, \pm 2)\) modes.
Figure 3. The approximate merger time observed by a detector with lower frequency cutoff at 20 Hz (upper panel) and 10 Hz (lower panel) is shown as a function of the total mass and the mass ratio of the system. The blue horizontal surface marks a duration of 3 s, which would allow for 1 s of buffer time between the signal duration and the length of the data segment. For a start frequency of 10 Hz only very high mass signals fit the time window.

Focusing now on the results with multiband, we notice that when $d f$ is coarser the multibanding tends to equalize the no-multibanding. This is expected since for coarser $d f$ we have less frequency points and then the coarse and fine grid tends to be similar and there is no gain by using the interpolation. Also, it can happen that the input $d f$ may be larger than the $\Delta f$ of the coarse grid given by the multiband criteria, in these cases we just evaluate in the frequency points of the fine grid and there is no much gain in speed. In current LIGO–Virgo parameter estimation the highest $d f$ that is used is 0.25 Hz, since $d f$ is the inverse of the time duration of the signal as in equation (3.1), and in practice the smallest duration considered,
Figure 4. Evaluation time of the LAL code for waveforms with and without multiband- ing (solid and dashed lines respectively) for a total mass of $50M_\odot$. PhXAS denotes the quadrupole model, and the different ‘PhXHM_MB’ items in the legend correspond to different values of the threshold $R$: $10^{-1}$, $10^{-2}$, $10^{-3}$ and $10^{-4}$. The evaluation time is averaged over 100 repetitions. The results are shown as a function of the spacing of the fine uniform frequency grid $d_f$. They were obtained with the LIGO cluster CIT.

Figure 5. Evaluation time versus total mass of the system, choosing a fine grid spacing $d_f$ that corresponds approximately to the inverse merger time at this mass ratio, as is common in data analysis applications (we denote this behaviour by writing $d_f = 0$). The evaluation time is averaged over 100 repetitions. Results were obtained with the LIGO cluster CIT.

e.g. for high mass events with very short duration, is 4 s. Note however that as low frequency noise is reduced in detectors, and the lower cutoff frequency for data analysis can be lowered, waveforms get longer and frequency spacings are reduced. On the contrary, when $d_f$ is very small we have a lot of points in the fine grid, then the interpolation is much more efficient and the multibanding has the highest gain in speed.
Figure 6. Maximum absolute error for the phase of the \((l, m)\) mode between the multibanding and no-multibanding waveforms for four values of the threshold \(R\). The threshold \(R\) can be interpreted as an approximate upper limit for the maximum absolute error introduced in the phase.

The different values of the threshold behave as expected: larger values of the threshold are less accurate, but allow faster evaluation. For small \(d_f\) we observe however that the evaluation speed is almost independent of the threshold value. This is due to the fact that for small \(d_f\) the evaluation of the model at the coarse grid points is computationally much cheaper than the subsequent interpolation to the fine grid points. Future optimization of our code will be required to address this issue and intend to reduce the computational cost of the interpolation to the fine grid.

We now show the dependence of the evaluation time on the total mass of the system. In this case the spacing of the fine grid \(d_f\) is computed by the LAL function \(\text{XLALSimInspiralFD}\) which adapts the \(d_f\) accordingly to an internal estimation of the time duration of the signal which depend on the lower cutoff \(f_{\text{min}}\), chosen here as \(f_{\text{min}} = 10\) Hz, and the mass of the system. This is similar in spirit to the estimate of the merger time that we have used in figure 3. In figure 5 we see qualitatively the same results than when simply scaling \(d_f\) as in figure 4. The multibanding is more efficient for lower masses where the duration of the signal is longer and therefore a smaller \(d_f\) is used.

3.2. Accuracy

In this section we discuss the accuracy of the multibanding algorithm as well as compare different choices of the threshold and motivate the choice of the default value for the model. In section 2.1 we explained that the non-uniform coarse grid is built such that the error in the phase (of a single mode) is below a threshold \(R\). To check this we compute the waveform with and without multibanding for 270 000 random configurations in the parameter range...
Figure 7. Absolute error for the phase averaged over the frequency array for the \((l, m)\) modes between the multiband and no-multiband waveforms for four values of the threshold \(R\). The threshold \(R\) can be interpreted as an approximate upper limit for the absolute error introduced in the phase.

For the multiband we compute again four different threshold values. In figure 6, we first show the maximum absolute error for the whole uniform frequency array for all the modes except \(\ell = 3, |m| = 2\), where mode mixing needs to be taken into account for interpreting results as discussed below. We see that for most cases the maximum error is indeed below the threshold. However, there can be special configurations where a few frequency points may give an error above the threshold. These few cases correspond typically to configurations where the approximations employed by the algorithm are less accurate, e.g. using the \textsc{TaylorF2} phase to approximate the phase in the inspiral to compute \(\Delta f\) for extreme spins or for cases with higher masses, where much of the signal-to-noise ratio is accumulated in the merger and ringdown, which are not described by the post-Newtonian \textsc{TaylorF2}. To better see the general behaviour of the absolute error over the frequency array, in figure 7 we show the mean error, averaged over the frequency array, and we see how it remains below the thresholds.

Now we consider the \((3, 2)\) mode, where mode mixing with the \((2, 2)\) mode is present. In figure 8 we show the results for the same test as shown in figure 6 for the other modes. The interpretation of the results is however different now, since the ringdown of the \((3, 2)\) mode, where mode mixing is present, is modelled and interpolated in terms of the spheroidal harmonics, see [3], where the waveform phenomenology is much simpler than in terms of spherical harmonics.

In figure 9 we show some typical behaviour for mode mixing in the ringdown: here the complex waveform comes very close to or crosses zero, visible as a sharp feature in the logarithm of the amplitude. Near the zero-crossing splitting the waveform into a spherical harmonic amplitude and phase creates artefacts when computing phase differences or relative amplitude.
errors between two waveforms, even if they are very close. Comparing our theoretical thresholds with the phase only makes sense in the spheroidal basis, but not in the spherical one. We omit a comparison of the phase errors in the ringdown as computed in the spheroidal picture in order to avoid excess baggage in our LALSuite implementation. We thus arrive at the following interpretation of figure 8: while phase errors are typically small and below the threshold, a significant number of outliers arise due to the phenomenon shown in figure 9. They are however not due to problems of the multibanding algorithm, but due to keeping our test simple and uniformly comparing in the spherical harmonic picture for all modes and across the whole frequency range.

To truly understand the accuracy of the algorithm we must compute the mismatch between the two waveforms. In the following, we evaluate the multimode waveform and compute the mismatch for the $h_+$ polarization. We carry out an extensive study across the whole parameter space also to test the robustness of the algorithm and evaluate one million of random configurations in the parameter space. The results are shown in figure 10. As expected, the threshold $10^{-4}$ has the lowest mismatches since it is the most accurate and the threshold 0.1 has the worst mismatches since it is the less accurate. The reader may be wondering why there are a significant number of cases with mismatch $10^{-16}$ when we would expect that the number of
Figure 9. Example case that produces a maximum phase error above the threshold for the (3, 2) mode. The parameters for this case are $m_1 = 35.7 M_\odot$, $m_2 = 16.6 M_\odot$, $\chi_1 = 0.33$, $\chi_2 = -0.54$. Top panel: absolute phase error between the no-multibanding and multibanding waveform with $R = 0.1$ versus the frequency. Bottom panel: amplitude of the (3, 2) mode for the no-multibanding and multibanding. Notice the correspondence of the maximum phase error with the deep in the amplitude of the (3, 2) mode.

cases decreases with the higher accuracy. The explanation is that for such cases the randomly chosen fine grid spacing $d_f$ is in fact larger than the coarse grid spacing $\Delta f$ that the multibanding criterion provides. In such cases no interpolation to accelerate the waveform evaluation is necessary. Our criterion to choose $\Delta f$ appropriately for interpolation therefore does not apply, and we simply replace $\Delta f$ with $d_f$ to avoid to evaluate the model in more frequency points than needed and perform an interpolation to a coarser grid. There is then no difference between the multibanding and no-multibanding and we reach machine precision.

Figure 10 is also very useful to decide which threshold we want to use as default value in the LALSuite code. We consider that $R = 10^{-3}$ performs well in accuracy and given that is faster than the $10^{-4}$ we set this as the default value.
Figure 10. Mismatches between the no-multibanding and multibanding waveforms for one million random configurations across parameter space. We show the results for four different values of threshold. The configurations are choosen randomly with \( q \in [1, 1000] \), \( \chi_{1,2} \in [−1, 1] \), \( M_t \in [1,500]M_\odot \), \( df \in [0.01, 0.3] \) Hz, \( \iota \in [0, \pi] \), \( f_{\text{min}} = 10 \) Hz, \( f_{\text{max}} = 1024 \) Hz. The mismatch is computed for the \( h_+ \) polarization.

3.3. Parameter estimation

In order to illustrate our algorithm in a parameter estimation application, we compare the performance of the original model and different choices for the threshold parameter \( R \).

We select a publicly available numerical relativity data set from the SXS waveform catalogue [44], SXS:BBH:0264, which corresponds to a binary black hole merger at mass ratio 3, with individual spins of \( −0.6 \) anti-aligned with the orbital momentum. Then we inject this numerical relativity simulation into zero noise as a way to get a non-precessing and non-eccentric strain of 4 s of duration, with 100\( M_\odot \) total mass, near edge-on with \( \pi/3 \) rad of inclination. We use a relatively close source at 430 Mpc, which implies a signal-to-noise ratio of 28. Recovery of the signal uses the advanced LIGO zero detuning high power noise curve [45]. We choose the parameters in order to challenge our approximations in the regime where higher modes are particularly relevant, not in order to demonstrate significant computational gains. We have chosen here a lower cutoff frequency of 20 Hz to compute the likelihood function in our Bayesian inference algorithm (see e.g. [22, 23] for details of Bayesian inference for compact binary coalescence signals). Note that the start frequency of the numerical relativity waveform we choose here is approximately 9 Hz at \( M = 100M_\odot \).

For our analysis we use a sampling method called ‘nested sampling’ [46], in particular the CPNest sampler [47] as implemented in the Python-based Bayesian inference framework Bilby [23]. For each waveform model used, we carry out runs with five different seeds and 2048 ‘live points’ in the language of nested sampling, and we merge the results from the five seeds to a single posterior result.

We define prior distributions as follows: the mass ratio is assumed to be uniform between 0.125 and 1, and the chirp mass prior is assumed uniform between 15 and 60\( M_\odot \). The luminosity distance is uniform in volume with a maximal allowed distance at 1500 Mpc. Finally, the magnitudes of the dimensionless black hole spins are uniform with an upper limit at 0.99.

Our main results concern the comparison of the IMRPhenomXHM model, evaluated with different values of the threshold parameter, \( R = (10^{-1}, 10^{-2}, 10^{-3}) \) as well as without multibanding, which corresponds to \( R = 0 \). The IMRPhenomXHM model includes the spherical harmonic modes \((l, |m|) = (2, 2), (2, 1), (3, 2), (3, 3), (4, 4)\). Differences between the recovered
value of parameters and the injected parameters may arise due to the approximations in our multibanding algorithm, errors in the IMRPhenomXHM model, errors in the numerical relativity waveform, and the absence of modes in the model, which are present in the numerical relativity data set (which contains all modes up to $l = 8$). We also compare with the IMRPhenomXAS model, which corresponds to IMRPhenomXHM with only the $(l, |m|) = (2, 2)$ modes and no multibanding. The latter serves as a comparison in terms of the errors in recovering the injection parameters.

Our results are presented in figure 11. In the case of the higher modes model, the injected values are recovered by the most probable regions of the posterior distributions. However, for the dominant mode model, a significant bias in the recovered parameters can be observed. This confirms the importance of the higher mode contributions for the case we have chosen. All the results for IMRPhenomXHM are consistent within the statistical errors implied by our finite sampling. As expected, in the case presented here the sampling time only decreases weakly when increasing the threshold value. We attribute the observed parameter bias for IMRPhenomXHM to the incomplete set of modes described by the model, as well as modelling errors. Future work will investigate the effect of dropping modes in the model in more detail.

4. Conclusions

We have presented a simple way to accelerate the evaluation of frequency domain waveforms by first evaluating on a coarse grid, and then interpolating to a fine grid with an iterative...
scheme to evaluate complex exponential functions (or equivalently trigonometric functions). This work builds upon the method presented in [1], but replaces the heuristic criterion used there to determine the spacing of the coarse grid by the standard estimate for first order interpolation error, and then extends the criterion for the coarse frequency spacing to the merger and ringdown. Several extensions of our algorithm are possible: first, similar techniques can also be developed for the time domain. The simple estimates to determine the appropriate coarse grid spacing given a threshold parameter could be improved, e.g. by adding low order spin terms. The amplitude could be treated in a similarly careful way as the phase. Second, we have already applied a simple version of multibanding acceleration to the Euler angle descriptions used in the modelling of precessing waveforms [29]. Further work is required to refine this technique to properly account for the precessing phenomenology at merger and ringdown, and other details due to precession dynamics, such as oscillations in different quantities. A better understanding of how to best apply multibanding acceleration to precession is also expected to inform how to treat eccentric waveforms.

Acceleration is more significant for smaller spacings of the fine grid, as is appropriate for smaller masses, and for detectors with broader sensitivity in frequency, e.g. future detectors such as the upgrades of the current generation of the advanced detector network, the Einstein Telescope [48] or LISA [49]. For total masses around three solar masses, as is appropriate for binary neutron star masses, the current speed of the multi-mode IMRPhenomXHM roughly equals the speed of IMRPhenomXAS for the $\ell = |m|$ modes. Detailed profiling of the code reveals this rough equality as a coincidence, and performance is limited by a small number of bottlenecks, e.g. evaluating the spline interpolation for the amplitude, for which we use the GSL library [38]. Future optimization work will focus on these bottlenecks. Another possible avenue for further speedup would be an implementation on GPUs or similar highly parallel hardware.

The availability of a threshold parameter that regulates accuracy and speed also allows future applications to tune codes for parameter estimation, where the threshold parameter could be set depending on the information associated with the detection in a search (such as the signals rough parameter estimate from the search and its signal-to-noise ratio), or the threshold could be changed dynamically, and could be relaxed in the burn-in-phase of a parameter estimation simulation, or in the early stages of a nested sampling run. The coupling of strategies to accelerate the evaluation of individual waveforms and Bayesian parameter estimation simulations as a whole may also have implications on the development of future waveform models, which could introduce further parameters to tune accuracy and evaluation speed.

Acknowledgments

We thank G Pratten, M Colleoni for discussions on the development of the IMRPhenomXAS and IMRPhenomXHM [3, 28] models, and the internal reviewers of the LIGO–Virgo collaboration for their careful checking of our LALSuite code implementation and their valuable feedback. This work was supported by European Union FEDER funds, the Ministry of Science and Innovation and the Spanish Agencia Estatal de Investigación grants PID2019-106416GB-I00/AEI/10.13039/501100011033, FPA2016-76821-P, RED2018-102661-T, RED2018-102573-E, FPA2017-90687-REDC, Vicepresid'encia i Conselleria d’Innovació, Recerca i Turisme, Conselleria d’Educació, i Universitats del Govern de les Illes Balears i Fons Social Europeu, Generalitat Valenciana (PROMETEO/2019/071), EU COST Actions CA18108, CA17137, CA16214, and CA16104, and the Spanish Ministry of Education, Culture and Sport Grants FPU15/01319. The authors thankfully acknowledge the computer resources at MareNostrum and the technical support provided by Barcelona
Supercomputing Center (BSC) through Grants No. AECT-2019-2-0010, AECT-2019-1-0022, AECT-2018-3-0017, AECT-2018-2-0022, AECT-2018-1-0009, AECT-2017-3-0013, AECT-2017-2-0017, AECT-2017-1-0017, AECT-2016-3-0014, AECT-2016-2-0009, from the Red Española de Supercomputación (RES) and PRACE (Grant No. 2015133131). 

Bilby and LALInference simulations were carried out on the BSC MareNostrum computer under RES (Red Española de Supercomputación) allocations and on the FONER computer at the University of the Balearic Islands. The authors are also grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459. Comunitat Autònoma de les Illes Balears through the Direcció General de Política Universitaria i Recerca with funds from the Tourist Stay Tax Law ITS 2017-006 (PRD2018/24).

ORCID iDs

Cecilio García-Quirós https://orcid.org/0000-0002-8059-2477

References

[1] Vinciguerra S, Veitch J and Mandel I 2017 Class. Quantum Grav. 34 115006
[2] The LIGO Scientific Collaboration 2015 LALSuite: LSC algorithm library suite https://lsc-group.phys.uwm.edu/daswg/projects/lalsuite.html
[3] García Quirós C, Colleoni M, Estellís H, Pratten G, Ramos-Buades A, Mateu-Lucena M and Jaume R 2020 arXiv:2001.10914
[4] Abbott B P et al (LIGO Scientific Collaboration and Virgo Collaboration) 2016 Phys. Rev. Lett. 116 061102
[5] Abbott B P et al (LIGO Scientific Collaboration and Virgo Collaboration) 2017 Phys. Rev. Lett. 119 161101
[6] Abbott B P et al (LIGO Scientific, Virgo) 2018 arXiv:1811.12907
[7] Taracchini A, Buonanno A, Khanna G and Hughes S A 2014 Phys. Rev. D 90 084025
[8] Bohí A et al 2017 Phys. Rev. D 95 044028
[9] Ajith P et al 2007 Class. Quantum Grav. 24 S689
[10] Ajith P et al 2011 Phys. Rev. Lett. 106 241101
[11] Hannam M, Schmidt P, Bohí A, Haegel L, Husa S et al 2014 Phys. Rev. Lett. 113 151101
[12] Husa S, Khan S, Hannam M, Pürrer M, Ohme F, Jiménez Forteza X and Bohí A 2016 Phys. Rev. D 93 044006
[13] Khan S, Husa S, Hannam M, Ohme F, Pürrer M, Jiménez Forteza X and Bohí A 2016 Phys. Rev. D 93 044007
[14] Blackman J, Field S E, Galley C R, Szilagyi B, Scheel M A et al 2015 arXiv:1502.07758
[15] Blanchet L 2014 Living Rev. Relativ. 17 78–160
[16] Kokkotas K and Schmidt B 1999 Living Rev. Relativ. 2 7–22
[17] Barrac L and Pound A 2019 Rep. Prog. Phys. 82 016904
[18] Buonanno A and Damour T 1999 Phys. Rev. D 59 084006
[19] Usman S A et al 2016 Class. Quantum Grav. 33 215004
[20] Sachdev S et al 2019 arXiv:1901.08580
[21] Klimenko S et al 2016 Phys. Rev. D 93 042004
[22] Veitch J et al 2015 Phys. Rev. D 91 042003
[23] Ashton G et al 2019 Astrophys. J. Suppl. 241 27
[24] Canizares P, Field S E, Gair J R and Tiglio M 2013 Phys. Rev. D 87 124005
[25] Smith R, Field S E, Blackburn K, Haster C-J, Pürrer M, Raymond V and Schmidt P 2016 Phys. Rev. D 94 044031
[26] Pürrer M 2016 Phys. Rev. D 93 064041
[27] Cotesta R, Marsat S and Pürrer M 2020 Phys. Rev. D 101 124040
[28] Pratten G, Husa S, García-Quirós C, Colleoni M, Ramos-Buades A, Estellés H and Jaume R 2020 arXiv:2001.11412
[29] Pratten G et al 2020 arXiv:2004.06503
[30] Pürrer M 2014 Class. Quantum Grav. 31 195010
[31] Thorne K S 1980 Rev. Mod. Phys. 52 299
[32] Teukolsky S A 1972 Phys. Rev. Lett. 29 1114
[33] Buonanno A, Iyer B, Ochsner E, Pan Y and Sathyaprakash B S 2009 Phys. Rev. D 80 084043
[34] LIGO Scientific Collaboration and Virgo Collaboration 2020 Phys. Rev. D 102 043015
[35] T L S Collaboration and the Virgo Collaboration 2020 Astrophys. J. 896 L44
[36] Cabero M, Nielsen A B, Lundgren A P and Capano C D 2017 Phys. Rev. D 95 064016
[37] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 2007 Numerical Recipes: The Art of Scientific Computing 3rd edn (Cambridge: Cambridge university press)
[38] Galassi M, Davies J, Theiler J, Gough B, Jungman G, Alken P, Booth M, Rossi F and Ulmerich R 2019 The gnu scientific library version 2.6 https://gnu.org/software/gsl
[39] Berti E and Klein A 2014 Phys. Rev. D 90 064012
[40] Kelly B J and Baker J G 2013 Phys. Rev. D 87 084004
[41] London L 2020 arXiv:2006.11449
[42] Abbott B P et al (Virgo, LIGO Scientific) 2016 Phys. Rev. Lett. 116 241102
[43] Chatziioannou K et al 2019 Phys. Rev. D 100 104015
[44] http://black-holes.org/waveforms
[45] https://dcc.ligo.org/LIGO-T0900288/public
[46] Skilling J 2006 Bayesian Anal. 1 833
[47] Veitch J and Pozzo W D 2017 CPNEST https://github.com/johnveitch/cpnest/
[48] Punturo M et al 2010 Class. Quantum Grav. 27 084007
[49] Amaro-Seoane P et al 2017 arXiv:1702.00786