Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Received: *May 29, 2024*
Revised: *October 24, 2024*
Accepted: *December 9, 2024*
Published: *February 25, 2025*

Testing gravitational waveforms in full General Relativity

Fabio D'Ambrosio,[a] **Francesco Gozzini,**[b] **Lavinia Heisenberg**[b], **Henri Inchauspé,**[b] **David Maibach**[b] **and Jann Zosso**[a]

[a]*Institute for Theoretical Physics, ETH Zurich, Wolfgang-Pauli-Strasse 27, CH-8093 Zurich, Switzerland*
[b]*Institute for Theoretical Physics, University of Heidelberg, Philosophenweg 16, D-69120 Heidelberg, Germany*

E-mail: fabio.dambrosio@gmx.ch, gozzini@thphys.uni-heidelberg.de, l.heisenberg@thphys.uni-heidelberg.de, h.inchauspe@thphys.uni-heidelberg.de, d.maibach@thphys.uni-heidelberg.de, zosso.jann@bluewin.ch

Abstract: We perform a comprehensive analysis of state-of-the-art waveform models, focusing on their predictions concerning kick velocity and inferred gravitational wave memory. Recent advancements in gravitational wave instrumentation have established new benchmarks for the precision of future measurements. To fully exploit the potential of upcoming gravitational wave surveys, it is crucial to concurrently enhance both semi-analytical and numerical waveform models used for analyzing interferometer data. In this work, we establish a pathway towards achieving this goal by developing and implementing a pipeline to assess the accuracy of waveform models using energy-momentum balance laws derived in full, non-linear General Relativity. The numerical accuracy assessment is performed for precessing as well as non-precessing simulations for models belonging to the *EOB*, *Phenom*, and *Surrogate* families. Our analysis reveals statistically significant deviations, which we trace back to inaccuracies in modeling subdominant modes and inherent systematic errors in the chosen models. We corroborate our findings through analytical considerations regarding the mixing of harmonic modes in the computed kick velocities and inferred memories. The methodology developed and validated in this article provides a foundational approach for future waveform assessments and a selection guide for waveform models in practical applications.

Keywords: gravitational waves / theory, gravity, astrophysical black holes, GR black holes

ArXiv ePrint: 2402.19397

https://doi.org/10.1088/1475-7516/2025/02/060

JCAP02(2025)060

Contents

1 Introduction **1**

2 Overview of waveform models **4**

3 Waveform preparation **6**
- 3.1 Alignment and residual ambiguities 7
- 3.2 Comments on alignment and precessing mergers 9

4 Waveform simulations under consideration **11**
- 4.1 Non-precessing binary mergers 12
- 4.2 Precessing binary mergers 14

5 Energy-momentum balance laws **14**
- 5.1 The balance laws 15
- 5.2 Remnant mass and velocity 18
- 5.3 Gravitational memory 19
- 5.4 Connections to black hole kinematics 20

6 Waveform assessment **21**
- 6.1 Mode mismatch 21
- 6.2 Computing physical quantities 23

7 Discussion **33**

A Numerical artifacts **36**

JCAP02(2025)060

1 Introduction

The first direct detection of gravitational waves (GWs) from the merger of two black holes in 2015 [1] marked a milestone in the field of astrophysics, confirming a key prediction of Einstein's General Theory of Relativity (GR) and ushering in a new era of observational astronomy. The waveforms matched against the GW signal encode a wealth of information, including the masses and spins of the binaries, the distances to the sources, and the geometry of their motion. By analyzing these waveforms, astrophysicists can uncover the underlying physical processes, discern the properties of exotic objects, and validate theoretical models with unprecedented precision.

To date, the process of parameter estimation and the testing of GR necessitate numerical modeling of gravitational waveforms across a wide range of source parameters, including masses, spins, and other relevant factors pertaining to the merging objects. These modeled waveforms are used as fitting templates with respect to actual data. The effectiveness of this template-to-signal match hinges on the precision of the estimated waveforms. Therefore, in order to extract meaningful information from signals, it is crucial to construct comprehensive

and accurate waveform templates that faithfully capture the physics of GR. This requirement is further substantiated by the expected increase in signal-to-noise-ratio of future GW observatories such as LISA [2], the Einstein Telescope [3], and the Cosmic Explorer [4]. Their increased resolution and sensitivity opens them up to more subtle effects like the gravitational memory [5–11], and they may even reveal physics beyond GR [12].

Deviations between templates and the actual waveforms contained in the observational data introduce systematic biases, compromising the reliability of the information extracted from the signal. To counteract such biases, a diverse set of template waveforms is employed when analyzing data from GW instruments. Among these template waveforms, the ones obtained from Numerical Relativity (NR) simulations are the most reliable. The computation of NR templates, however, consumes vast amounts of computational resources per waveform generation. This poses a significant challenge, particularly as the volume of data to be processed is expected to increase drastically in the upcoming years with the advent of multiple ground- and space-based instruments like LISA [13] and LIGO/Virgo [14, 15]. Furthermore, as the measuring precision advances, deviations from GR [12] may reveal themselves in the observed data. Detecting such deviations necessitates an expanded parameter space to account for alternative descriptions of gravity, which consequently amplifies the number of waveforms against which the data must be tested.

The efficiency challenges associated with NR waveform generation are not the sole motivation for the development of alternative waveform models. Semi-analytical waveform models offer a more transparent understanding of the fundamental physics underlying gravitational wave sources. These models allow for the incorporation of essential physical effects and approximations directly into the simulations of inspiral, merger, and ringdown phases of binary black hole coalescences. Their design ensures flexibility and applicability across a broad spectrum of astrophysical scenarios, including binary black holes, neutron star mergers, and mixed systems, by adjusting specific parameters within the models. In practical applications, gravitational wave observatories such as LIGO and Virgo depend on data analysis pipelines that necessitate the rapid generation and comparison of waveform templates. Semi-analytical models are well-suited for integration into these pipelines, facilitating real-time or near-real-time analysis of incoming data. Additionally, due to the commonly employed analytical Post-Newtonian (PN) ansatz in computing the inspiral waveform, semi-analytical models can be extended to earlier stages in the evolution of a binary system, where the system exhibits nearly constant frequency oscillations long before the merger. This capability is typically limited in NR templates due to computational constraints, highlighting the complementary strengths of semi-analytical modeling approaches.

Prominent representatives of alternative waveform models are the *Surrogate* models [16–18], phenomenological models [19–22] and effective-one-body simulations [23–31], where the latter two rely on semi-analytical techniques. To obtain reliable results within reasonable timescales, the models adopt distinct strategies to compute the gravitational strain for inspiral, merger and ringdown phase. Each model focuses on different physical aspects of compact binary coalescence and is capable of producing a waveform within certain parameter ranges in an efficient manner.

JCAP02(2025)060

In the light of the increasing demands on waveform models in future data analysis runs, it is crucial to continuously test and improve the GW waveform templates. This includes adapting them corresponding to the relevant challenges in current data analysis pipelines (e.g., adding memory where missing) and establishing efficient performance assessments. All of the above are currently under active investigation by the waveform communities and constitute fundamental challenges of GW physics.

In this work, we address these issues by providing an extensive and quantitative comparison of state-of-the-art waveform models based on their prediction of physical quantities such as the remnant's kick velocity and the inferred gravitational wave memory. We compare the models to templates generated from NR simulations, as well as between themselves for simulated mergers beyond the scope of the available NR catalog where no NR-benchmarking is possible. Our analysis prioritizes non-precessing merger simulations for the evaluation of kick velocities but extends the evaluation of the memory by including precessing as well as non-precessing binary mergers. The primary tools for effecting these comparisons are the so-called *energy-momentum balance laws* [32]. These exact mathematical results were derived within full, non-linear GR and provide an infinite tower of constraints for waveform models. After performing a decomposition of the balance laws into spherical harmonics, we obtain a simple constraint equation as well as equations for directly computing kick velocity and inferred memory for any given waveform model. A proof of concept that the balance laws can be used as diagnostic tools for assessing waveform accuracy was provided in [33, 34]. Here, our objective is to fully leverage the strength of these tools for a comparative analysis between different waveform models. Our study builds upon prior research, which concentrated exclusively on analytical models [34], the kick velocity [35, 36], or gravitational memory [33, 37]. We extend these earlier investigations, including a thorough investigation of the parameter space and the applied analytical tools, and reveal systematic trends in waveform templates.

Our paper is structured as follows: in section 2, we provide an overview of the families of waveform generators that are central to our analysis, along with a detailed discussion of the specific features involved in their numerical implementation. Section 3 outlines our methodology for preparing simulated waveforms prior to comparison, covering key aspects such as alignment procedures, residual artifacts, and the equalization of time grids in the waveform time series. In section 4, we list the binary black hole (BBH) merger simulations that enter our comparative analysis, categorizing them into two groups: those generated from NR (referred to as "cataloged") and those falling outside the NR parameter space (referred to as "non-cataloged"). Section 5 revisits the framework and formulation of the energy-momentum balance laws, with a focus on a qualitative understanding and an emphasis on their power for assessing the accuracy of waveform models. The core of our investigations is found in section 6, where we analyze kick velocity, memory, and the relative differences between the selected models, based on the equations derived in section 5. Special attention is given to the impact of mode content and selection, as well as parameter space-dependent performance differences of waveform models. Finally, in section 7, we discuss the outcomes of the model comparison, considering aspects such as alignment strategies, mode selection, and overall performance.

JCAP02(2025)060

2 Overview of waveform models

We begin by introducing the alternative models under investigation in this study, which have played a significant role in GW research. We highlight the key features of these models and discuss aspects of their numerical implementation. For more details on the inner workings of each model, we have provided relevant references below. A comprehensive overview of these models can be found in [38].

To remain as closely as possible to the data analysis pipelines used in GW reserach, in this investigation, we focus on an ensemble of different waveform families available through the LIGO-issued *LALSuite* software collaboration [39]. The latter does not access the models directly but builds on a framework developed by the LIGO collaboration for optimal streamlining of analysis involving waveform models. The list of models contains four main families, which we label *NR*, *Surrogate*, *Phenom*, *EOB* according to their generation mechanisms. Further, we include an EOB-related *TEOB* model into our consideration. The motivation for this particular choice will be elaborated in detail below.

Among these five models, we focus more extensively on individual representatives. Specifically, we analyze the following implementations: *NR/SXS* and its *LALSuite* interface [40] (based on the catalog produced by the SXS collaboration [41] which relays on simulations using the Spectral Einstein Code (SpEC) [42]), the *Surrogate* model *NRSur7dq4* [18], *EOB* being represented by *SEOBNRv4PHM* [27], *Phenom* by *IMRPhenomTPHM* [22] as the phenomenological model, and *TEOB* implemented as *TEOBResumS* [29]. For simplicity, in the remainder of this work, we shall often identify the specific model (e.g. *NRSur7dq4*) with its family name (e.g. *Surrogate*).

The chosen families represent distinct approaches to waveform simulations. The *NR* family represents full numerical relativity simulations, offering the most comprehensive and reliable waveforms against which the other models are competing. The alternative "approximant" models are based on the phenomenology of binary mergers (*Phenom* [19–22, 43]), effective-one-body simulations (*EOB* [23–31, 44, 45]/*TEOB* [29–31]), and interpolation of NR simulations (*Surrogate* [16–18]). In table 1, we present an overview of the selected models, specifying for each the included (spin-weighted) harmonic modes, the applicable domains concerning mass-ratio $q = M_1/M_2 \geq 1$, the maximum time before merger, and spin magnitude $|\chi_i|$. Note that these features, except for the mode content, are inherent to the models' *LALSuite* implementation.

For *SXS*, the duration of the waveform—and consequently the maximum time before merger—varies significantly across different simulations. The median number of BBH cycles before merger in *SXS* waveforms is 39, with the shortest cataloged simulation comprising 7 cycles and the longest reaching 351.3 cycles [41].[1]

The features of the *LALSuite* implementation of *Surrogate*, as listed in table 1, are directly derived from the specific model *NRSur7dq4*, which is trained on *NR* waveforms with corresponding characteristics, namely $q \leq 4$ and $|\chi_i| \leq 0.8$. The initial time for the relevant *NR* simulations ranges from 4693 to 5234 M, resulting in an *LALSuite*-implemented time before merger that does not exceed 4500 M for this alternative waveform model [18].

[1]The number of cycles is measured with respect to the dominant $\ell = m = 2$ mode.

JCAP02(2025)060

The *EOB* representative in this study, *SEOBNRv4PHM*, inherently has no restrictions on the starting time of the waveform or, equivalently, on the minimum GW frequency (reference frequency) $f_{\rm ref}$ of a simulation. This flexibility stems from the PN ansatz employed in the waveform building during the early inspiral phase. However, despite the model's seemingly extensive parameter space coverage, *SEOBNRv4PHM* has been predominantly validated against *NR* simulations within the range of $q \leq 4$ and $|\chi_i| \leq 0.8$, with a reference frequency $f_{\rm ref} \approx 20$ Hz at a total mass of 50 $M_\odot$ (corresponding to $15 - 30$ cycles before merger) [27]. Therefore, the accuracy of this *EOB* model should be considered reliable only within this region of the waveform parameter space.[2] Despite these caveats, in table 1, we list the vastly larger parameter space values numerically accessibly in *EOB*'s *LALSuite* implementation.

Similar constraints regarding the faithfulness of simulated waveforms apply to the phenomenological model *Phenom*. The selected representative, *IMRPhenomTPHM*, is derived from its non-precessing counterpart, *IMRPhenomTHM* [46], through the so called "twisting-up" procedure. For further details, we refer to [22]. The underlying model, *IMRPhenomTHM*, is calibrated against *NR* solutions up to $q = 18$, covering nearly the entire spin spectrum ($|\chi_i| \in [0, 0.99]$). Similar to *EOB*, the gravitational wave reference frequency $f_{\rm ref}$ of *IMRPhenomTPHM* can technically be extended to arbitrarily low frequencies. Again, the values listed for *IMRPhenomTPHM* in table 1 represent the maximum accepted input parameter range for *Phenom* when loaded via the *LALSuite* environment.

For the second representative of the EOB family, *TEOBResumS*, the parameter space restrictions arise from model stability considerations and previous evaluations that revealed physically inconsistent results beyond the specified bounds. It is important to note that the calibration of *TEOBResumS* was performed using *NR* simulations with mass ratios up to $q = 20$. The *TEOB* model is limited to the non-precessing regime and exclusively outputs the $\ell = m = 2$ mode. In our analysis, *TEOBResumS* serves as a benchmark to assess how the inclusion of additional mode content impacts waveform precision, as quantified by the balance flux laws. Similar to the *EOB* and *Phenom* models, the *LALSuite* function for *TEOB* allows for arbitrarily low initial frequencies. On the other hand, it is important to note that the *LALSuite* environment generally imposes a maximal initial frequency, which depends on input parameters such as mass ratio, individual masses, and spins. This constraint ensures the minimal inclusion of the inspiral phase in the simulated waveform.

At this point, we emphasize that, for the trained eye, there may be disagreements between the intrinsic properties of waveform models, such as mode content, and their corresponding features within the *LALSuite* framework listed in table 1. For example, while it is possible to load the $\ell = 3, m = 2$ mode for both *Phenom* and *EOB* within *LALSuite*, these harmonics, when provided by *LALSuite*, do not contain any practical information and can be disregarded as artifacts. Generally, *LALSuite* does not map the simulated waveform modes directly from the original models in a one-to-one correspondence. Instead, it applies time-interpolation and frame-correction methods [39], which may slightly modify the strain harmonics compared to their "original" form. Verification procedures have been conducted for specific versions of the waveform models, and only negligible differences have been observed. For details regarding

[2] It should be noted that higher spins and mass ratios significantly increase the computational cost required to maintain a given level of simulation accuracy, due to the disparate length scales within the binary system [27].

JCAP02(2025)060

Family	**Implementation**	**Branch**	**Mode Content**	q**-range**	$\lvert\chi_i\rvert$**-range**	t_{init}
Numerical Relativity	*SXS* (NRhdf5)	precessing	$\{(\ell,m)\lvert\ell\leq 8\}$	$\lesssim 10$	$\lesssim 0.998$	-
Surrogate	NRSur7dq4	precessing	$\{(\ell,m)\lvert\ell\leq 4\}$	≤ 4	≤ 0.8	$\leq 4500M$
Effective One Body	SEOBNRv4PHM	precessing	$\{(2,\pm2),(2,\pm1),(3,\pm3),(4,\pm4),(5,\pm5)\}$	≤ 100	≤ 1	-
Phenomenological	IMRPhenomTPHM	precessing	$\{(2,\pm2),(2,\pm1),(3,\pm3),(4,\pm4),(5,\pm5)\}$	≤ 200	≤ 1	-
"Tidal" Effective One Body	TEOBResumS	non-precessing	$\{(2,2)\}$	≤ 30	≤ 0.99	-

Table 1. Summary of the model families, specific implementations, mode content and parameter space coverage of the waveform models considered in this work. Note that the maximal time before merger, t_{ref}, is given in units of total mass M, where 0 marks the merger time. In practice, t_{ref} implies a parameter-dependent lower bound for the minimum GW (reference) frequency f_{ref} of the waveform. The blank spaces for t_{ref} indicate that the corresponding model does not admit a uniform boundary regarding the minimal frequency valid for all simulations. All quantities, except for the mode contend, are displayed as listed in the *LALSuite* documentation [39].

the verified versions of the individual models, we refer to the *LALSuite* documentation [39]. Therefore, we can confidently assume that the framework, which standardizes the waveform models with respect to coordinate and time reference frames—the *LALSuite* environment—provides a highly accurate representation of the actual models listed in table 1.

Finally, it is important to note that none of the waveform models tested here includes the GW memory effect in any form. However, tools to incorporate memory, either manually or as an integrated feature of the waveform model, are already available. For *SXS* waveforms, memory can be extracted (and added) using Cauchy-characteristic extraction (CCE) [47, 48] which, so far, has only been implemented in a few publicly available simulations but can in principle be applied to any cataloged merger using the software package *Scri* [49]. For other models, similar techniques, involving the flux balance laws, can be applied [37]. Moreover, recent updates to some alternative waveform models have begun to include memory in their simulations (see references for the corresponding model families). In this study, we use memory-free waveforms, as the computation of memory is employed as a quality metric for evaluating the individual strain modes. This aspect will be thoroughly demonstrated in section 6.

3 Waveform preparation

For each point in the waveform parameter space considered in this investigation, the corresponding waveforms from each model are loaded via *LALSuite*. Our primary motivation for using this framework lies in the waveform model-dependent choice of reference frames in each GW simulation. Since different waveform models employ various numerical techniques to generate the strain and its harmonics, the resulting outputs may be scaled to different reference frames. To avoid frame-related mismatches, we utilize the *LALSuite* environment, which applies appropriate frame transformations to ensure that the output waveforms (and their harmonics) are aligned within the same reference frame, except for a possible global phase shift. Addressing, among others, these reference frame alignment residuals, in what

JCAP02(2025)060

follows, we present the preparation procedure for GW waveforms employed in this work, which constitutes the first step towards a comprehensive comparison of different waveform models.

3.1 Alignment and residual ambiguities

As mentioned above, the waveforms generated through *LALSuite* are already aligned in a common frame, up to a rotation in the orbital plane, or equivalently, up to a phase factor denoted hereafter as $\phi_{\rm ref}$. Provided the waveforms of the approximants are of equal length (i.e., start with a common reference frequency $f_{\rm ref}$[3]), the orbital phase remains the only obstacle in the alignment of non-precessing waveforms.

Complications occur when waveform models are compared to NR data from the *SXS* waveform catalog: for a simulation associated with a specific set of spins and initial masses, it is possible that the maximal waveform domain of each approximant varies due to different restrictions on the reference frequency $f_{\rm ref}$ (in particular, $f_{\rm ref}$ varies between *SXS* simulations, see the discussion around table 1 in the previous section). In this case, we re-generate the corresponding waveform for each approximant using a new minimal common $f_{\rm ref}$, dictated by the most stringent constraint throughout the considered models for the given BBH system. While the *LALSuite* approximants can produce waveforms with different initial or reference frequencies $f_{\rm ref}$ for a given simulation (within intrinsic constraints, see table 1), the *SXS* waveforms come with a fixed reference frequency that cannot be altered. As a result, the necessity of adjusting NR waveforms "by hand" (shortening the waveform's inspiral part) may arise, which in turn requires an additional alignment step with respect to the phase factor $\phi_{\rm ref}$. Both the *LALSuite* residual phase shift and the latter correction to $\phi_{\rm ref}$ due to waveform cutting can be handled through a single optimization procedure.

For precessing mergers, analogous arguments apply. Here, the ambiguity to be resolved by the alignment extends to the initial directions of the individual spin vectors of the coalescing black holes, introducing four more independent parameters (the angles subtended by the spin vectors) to the alignment process, in addition to $\phi_{\rm ref}$. Potential challenges associated with the additional parameters are discussed in greater detail below.

Regarding the optimization procedure, both the adjustments of $\phi_{\rm ref}$ and the initial spin vectors are achieved through minimizing a mismatch function $\mathcal{M}(\phi_{\rm ref}, \Omega_1, \Omega_2)$, where $\Omega_i = (\phi_{\chi_i}, \theta_{\chi_i})$ are the pairs of angles for each spin vector. Once minimization is accomplished, we posit that any discrepancies identified when comparing physical quantities on the aligned waveforms can be attributed to inherently distinct evaluations of the binary configuration for each individual model.[4]

Before addressing the removal of phase ambiguities in detail, it is important to note that the time axis of the strain and its corresponding harmonics must also be synchronized, meaning the waveform time series need to be placed on a common time grid. This synchronization is largely managed by the *LALSuite* environment, which allows for loading waveforms from

[3]Here, we refer to the reference frequency as the frequency of the waveform at the initial time step from which on the generated waveform is non-trivial. This frequency can be translated to the number of orbits included before the merger, which generally differs between waveform models.

[4]Naturally, this is a crude assumption. However, discarding it would require an in-depth analysis of potential shortcomings in the alignment procedure beyond what is discussed in this section. The latter lays out of the scope of this work.

JCAP02(2025)060

different models with consistent time- or frequency-binning.[5] The fixed point (origin) of the time grid is automatically set to the merger time, defined as the moment when

$$\max \sqrt{\sum_{\ell,m} |h_{\ell,m}(t)|^2}\,, \tag{3.1}$$

reaches its maximum. With equal binning and a uniform origin for all waveforms, the remaining task in terms of time optimization is to trim the time grids for each waveform so that they align on the initial time t_i and final time t_f. This step is carried out prior to any alignment procedure and after loading the *LALSuite* waveforms for a given merger simulation.

Now, turning to the residual phase shift determined through the mismatch, the numerically implemented mismatch function $\mathcal{M}$ requires an input that is related to the gravitational strain or its harmonics. Our primary strategy is to align with respect to the dominant $h_{2,\pm2}$ mode such that the mismatch function is computed as

$$\mathcal{M}(\phi_\mathrm{ref}, \Omega_1, \Omega_2) \coloneqq 1 - \frac{\langle \tilde{h}^\mathrm{ref}_{2,2}, \tilde{h}^\mathrm{align}_{2,2} \rangle}{\|\tilde{h}^\mathrm{ref}_{2,2}\|\,\|\tilde{h}^\mathrm{align}_{2,2}\|}\,. \tag{3.2}$$

Here, $\tilde{h}_{2,2}$ corresponds to the Fourier transformed ($\ell = m = 2$)-mode of the strain in the orthogonal decomposition

$$h = \sum_{\ell,m} h_{\ell,m}(u)\, {}_{-2}Y_{\ell,m}(\theta,\phi)\,, \tag{3.3}$$

where ${}_{-2}Y_{\ell,m}$ are the spin-weighted spherical harmonics of weight -2. The inner product $\langle\cdot,\cdot\rangle$ is defined in the frequency domain as

$$\langle \tilde{h}_1, \tilde{h}_2 \rangle \coloneqq 4\,\mathrm{Re} \int_{f_\mathrm{min}}^{f_\mathrm{max}} \mathrm{d}f\, \tilde{h}_1(f)\, \tilde{h}_2^*(f)\,, \tag{3.4}$$

where $\tilde{h}^*$ is the complex conjugate of $\tilde{h}$. Naturally, this inner product induces a norm, which we denote as $\|\cdot\|$. The integration domain results from the Fourier transform of the common time grid waveforms. Specifically, we integrate over the entire frequency spectrum computed by fast Fourier transforming the $h_{2,2}$ time series. The modes $\tilde{h}^\mathrm{ref}_{2,2}$ and $\tilde{h}^\mathrm{align}_{2,2}$ are extracted from the reference model and the model to be aligned, respectively. Throughout the analysis, we adopt *SXS* waveforms as reference. For regions in parameter space where no *SXS* waveforms are available, we take *Surrogate* as reference instead. We emphasize that, for calculating the mismatch (3.2) and its norm as defined by (3.4), it is not strictly necessary to transform the modes into the frequency domain. A time domain mismatch computation is equally feasible. However, to facilitate better comparability with the mismatch functions used by the models themselves (see model reference but, for instance, [18]) and with other studies on waveform comparison (e.g., [35, 36]), we compute the mismatch in Fourier space.

An inherent limitation of the alignment strategy outlined by equation (3.2) is that phase corrections are confined to the $h_{2,2}$, meaning that the procedure adjusts the phase of only the

[5] We have evaluated the effects of varying time (frequency) sampling on a small test ensemble of BBH simulations. Our assessment revealed no significant impact on the calculations presented in section 6, provided that all waveform features and oscillations are adequately resolved within the specified time (frequency) interval.

JCAP02(2025)060

dominant mode. It is important to note that subdominant harmonic modes with uneven-m values can still experience a π-phase shift after $h_{2,2}$ has been aligned according to (3.2). To address this well-known effect, we incorporate an additional step in the waveform preparation process following the alignment of the dominant $h_{2,2}$ mode on a common time grid: for each uneven-m mode, we calculate the mismatch with respect to its corresponding reference, as described in (3.2), twice: once for the original mode and once for the mode shifted by $\exp(i\pi)$. If the π-shift results in a lower mismatch, it is adopted for the corresponding mode throughout the rest of the analysis. This ensures that all alternative waveform models are accurately aligned across all relevant harmonic modes with respect to the reference waveforms provided by *NR* or *Surrogate*.

3.2 Comments on alignment and precessing mergers

Due to its dominance in the decomposition (3.3), it is established common practice to align waveforms with respect to the harmonic mode $h_{2,2}$ in frequency domain. This way, one automatically ensures a sufficient alignment of the full strain that can be computed in similar fashion provided an orientation/line-of-sight for the simulation is chosen. A potential caveat to this approach arises when considering the integration boundaries of equation (3.4): when minimizing $\mathcal{M}$ for the $h_{2,2}$ mode across the entire Fourier-space waveform, i.e., over the frequency range $f \in [f_{\min}, f_{\max}]$ (including the merger and ringdown phase), the resulting alignment can potentially obscure systematic differences between the models. While most of the alternative waveform models use PN techniques for the inspiral, the merger and ringdown phases are modeled quite differently, and systematic differences between the models become more pronounced during these phases. These differences naturally contribute to the mismatch calculated by equation (3.4). To avoid obscuring these fundamentally distinct features by matching waveforms over the full frequency domain, a more appropriate mismatch calculation could involve integrating only up to $f_{\text{end of PN-phase}}$ instead of $f_{\max}$. The latter would ensure that potential systematic differences fully remain in the aligned waveforms and only the PN inspiral phase is precisely phase-synchronized. For the analysis of this work, the only alignment-dependent quantity tested (for non-precessing binary mergers) is the direction of the remnant velocity. Since this is only one of many assessment metrics, the alignment issues discussed here are only of secondary interest in this work. Therefore, we adopt the commonly used integration bounds as in equation (3.2) and leave the thorough testing of alternative alignment strategy to future work.

Finally, a few comments on precessing systems are in order: aligning waveforms for precessing systems remains challenging, and there is currently no universally optimal strategy. In fact, optimizing alignment procedures for these systems is an active area of research. The primary difficulty for precessing systems lies in the fact that the spin vectors of both merging black holes have time-dependent orientations. A proper alignment would require finding the right instant of time where these spins χ_1, χ_2 each align across waveform models, which can be extremely resource demanding, especially since, depending on the model, this instance of time may lay beyond the minimal initial time step of a given BBH simulation. Even if all waveform models provided a time evolution of the individual black holes' spins, finding a match is not guaranteed due to differing internal evolution procedures. In literature, it has

JCAP02(2025)060

thus established common practice to minimize the mismatch (as above) with respect to $\phi_{\rm ref}$ and the four spin angles of the binary system prior to the merger simultaneously. The latter is implemented as follows: across the considered approximants, for a given BBH merger, the waveforms are loaded with the corresponding simulation parameters, such as initial spins (magnitude and direction), reference frequency, and initial masses. Once the waveforms are set on a common time grid, the spin amplitudes of the targeted simulation are used as input parameters for the mismatch function. The numerical implementation of the alignment then employs an optimization scheme designed to minimize the mismatch function (3.2). In each iteration of the optimization scheme, the alternative models' waveforms are re-simulated for a different set of spin angles at the common initial time step yielding a slightly smaller mismatch. This process is repeated until the mismatch is sufficiently low, i.e., comparable in order of magnitude to mismatch values of non-precessing simulations, for more details see section 6. As a result, this scheme yields multiple waveforms on a common time grid with equal spin magnitudes but potentially different initial spin orientations. These variations in spin orientation reflect the differences in how each waveform model handles the precession dynamics, even though the spin magnitudes are consistent across the models.

Since the spin orientation significantly affects the direction and magnitude of the remnant velocity, calculating the kick for aligned precessing simulations becomes meaningless. Thus, in section 6, we focus exclusively on the (non-linear) memory contribution from precessing mergers. Unlike the kick, the memory effect is unaffected by phase shifts but can still respond to changes in the initial spin configuration, albeit in a much less sensitive manner. In fact, for the merger simulations analyzed in this work (detailed further in section 4), we found that the alignment procedure for precessing systems, as described above, significantly enhances the accuracy of the memory component in the corresponding waveforms.

The kick's and memory's dependence on the initial spin configurations stems from the asymmetry between ℓ, m and $\ell, -m$ strain modes for precessing mergers. The fact that the two quantities demonstrate differing sensitivities with respect to these mode asymmetry can be analytically elucidated and will be addressed thoroughly in section 5. At this point, note that the ℓ, m and $\ell, -m$ discrepancy is explicitly modeled only in *Surrogate* and *SXS*. The versions of *EOB* and *Phenom* considered in this study do not account for this effect. For recent phenomenological models including the mentioned asymmetry, see for instance [50, 51].

The effectiveness of adapted alignment strategy is illustrated in figures 1–4. For better readability, we only display the *Surrogate* and *NR*, where *NR* serves as a reference model. Figures 1, 2 show the alignment of the dominant harmonic strain mode for precessing and non-precessing mergers, respectively, generally indicating slightly worse alignment for precessing binaries. Moreover, figures 3, 4 demonstrate the necessity of the residual π-shifts for subdominant uneven-m modes. Note again that subdominant modes do not experience any alignment-related optimization scheme but solely a π-shift if necessary. The waveforms depicted in these figures are selected to highlight this effect in specifically selected subdominant modes.

JCAP02(2025)060

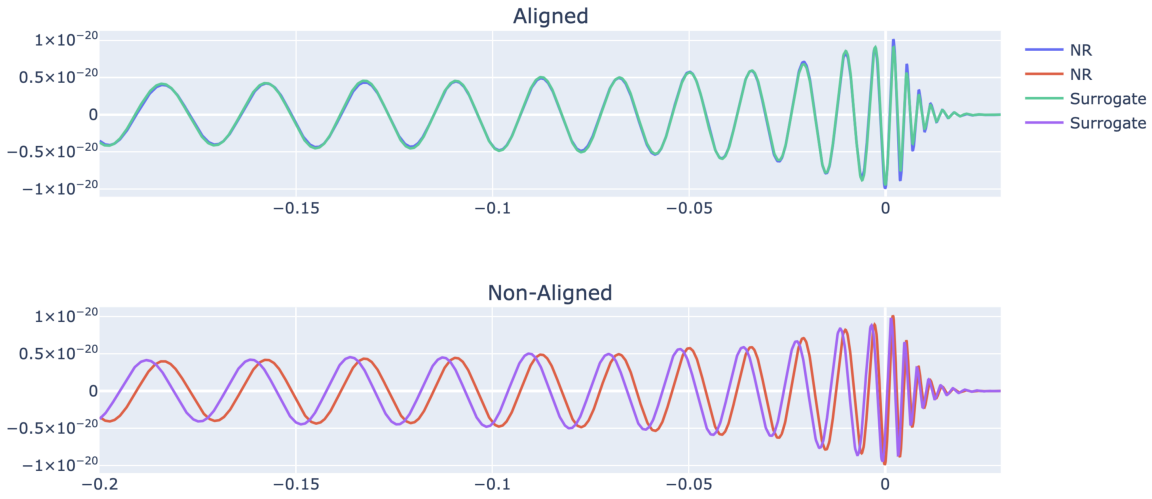

Figure 1. Aligned (top) and non-aligned (bottom) $h_{2,2}$ waveform mode for a precessing binary merger (SXS:BBH:1011).

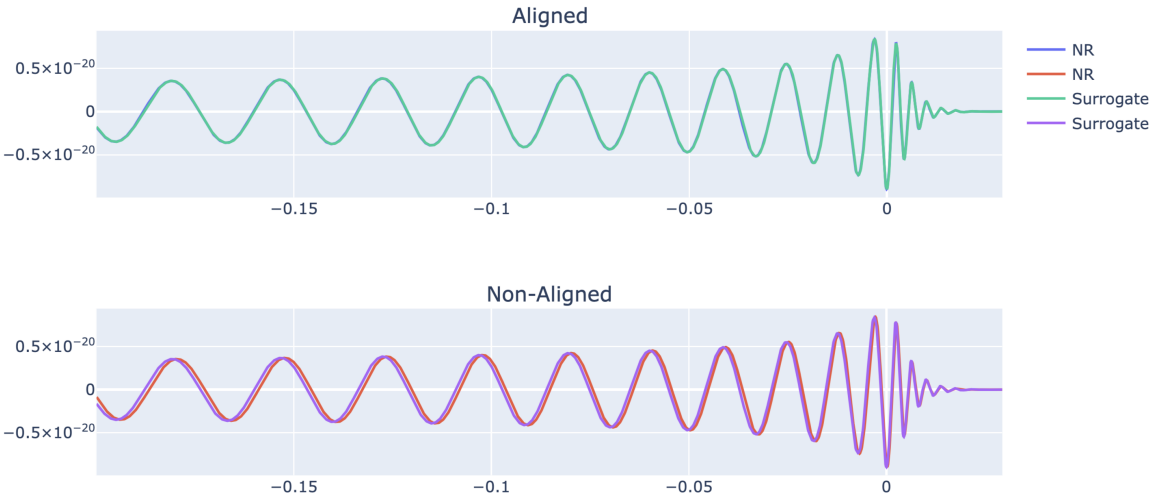

Figure 2. Aligned (top) and non-aligned (bottom) $h_{2,2}$ waveform mode for a non-precessing binary merger (SXS:BBH:0191).

4 Waveform simulations under consideration

Given the substantial number of binary merger simulations considered in this study, it is instructive to discuss the covered regions in parameter space. This allows us to establish meaningful connections between our findings and phenomenological insights, especially in cases where one or more approximants exhibit large deviations with respect to the reference model. Additionally, by carefully spreading the considered mergers over a large parameter space, we avoid introducing selection biases. A connection between errors in physical quantities for a given merger and its position in parameter space is drawn in section 6.

For each case, that is precessing and non-precessing BBH mergers, we discuss the considered waveforms for which NR counterparts are available and simulate additional ones to diversify the parameter space under investigation.

JCAP02(2025)060

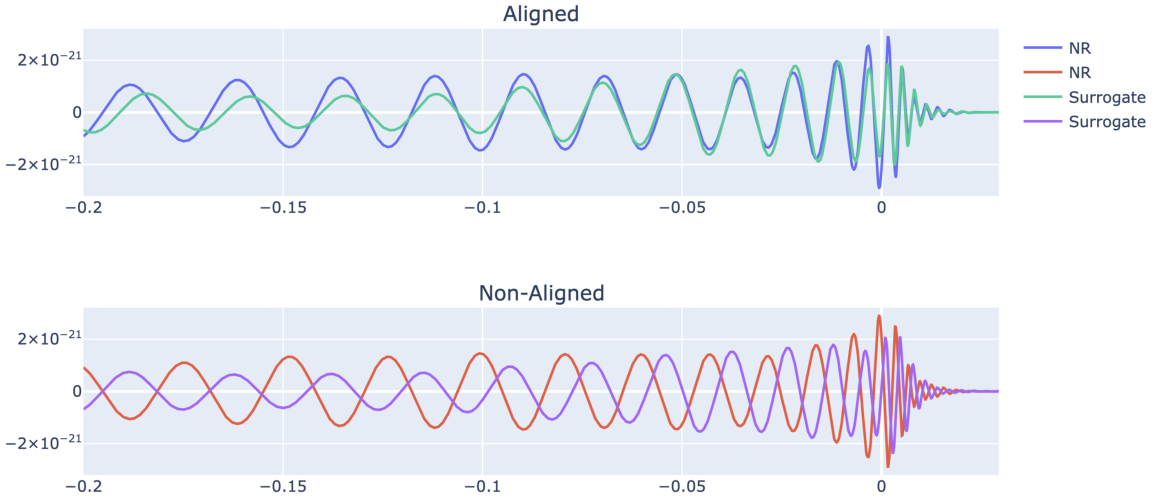

Figure 3. Aligned (top) and non-aligned (bottom) $h_{2,1}$ waveform mode for a precessing binary merger (SXS:BBH:1011). The aligned waveform is displayed.

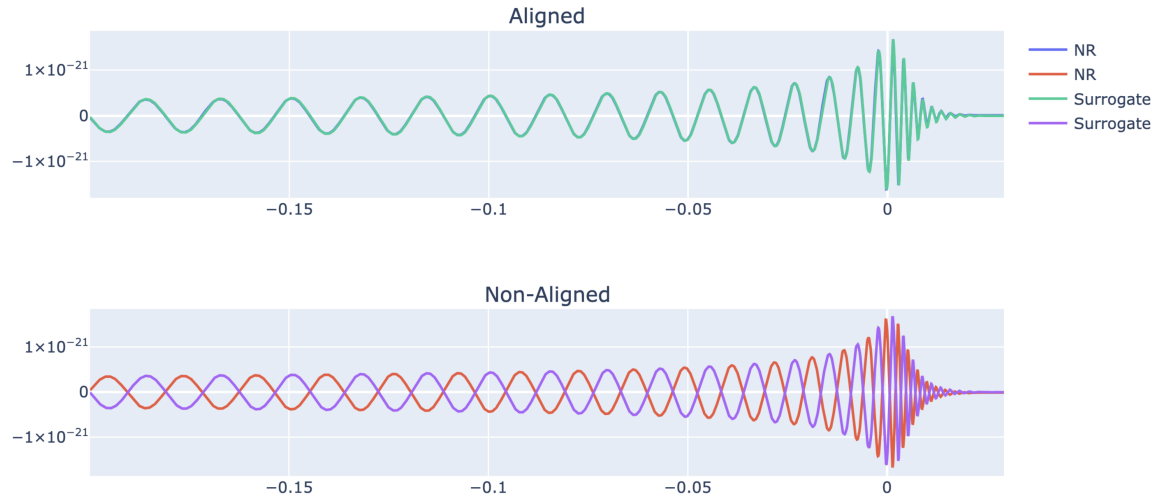

Figure 4. Aligned (top) and non-aligned (bottom) $h_{3,3}$ waveform mode for a non-precessing binary merger (SXS:BBH:0191).

4.1 Non-precessing binary mergers

The majority of our analysis resides in examining non-precessing BBH simulations within the *SXS* catalog. In figure 5 we illustrate the distribution across parameter space, distinguishing between aligned and anti-aligned spin configurations and employing color-coding based on the mass ratio $\eta := \frac{q}{(1+q)^2}$ with $q := M_1/M_2 \geq 1$, corresponding to the ratio of final and initial mass. While a total of 175 mergers with non-negligible kick, i.e., $v > 20$km/s, is selected for the non-precessing case, the parameter space displayed in figure 5 seems to be only sparsely covered. This is due to many simulations depicted in figure 5 overlapping in the sense that either only their mass ratio or the alignment of spins changes.

From figure 5 it is evident that regions characterized by lower spins χ_1 and χ_2 are insufficiently represented within this excerpt of cataloged data. To address sparsely populated

JCAP02(2025)060

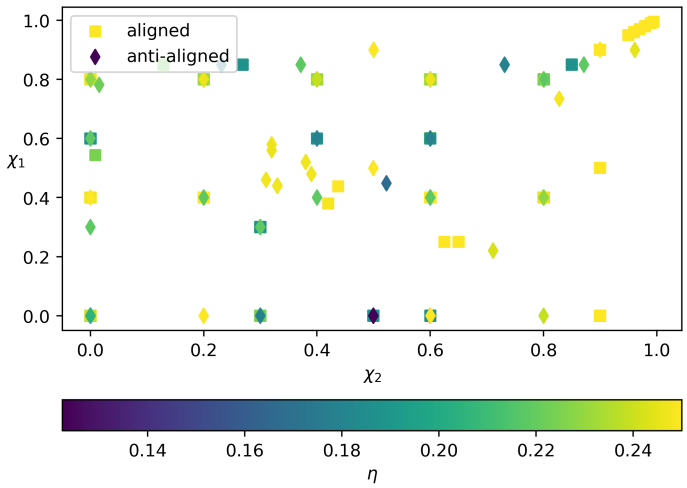

Figure 5. Parameter space for the non-precessing *SXS* data.

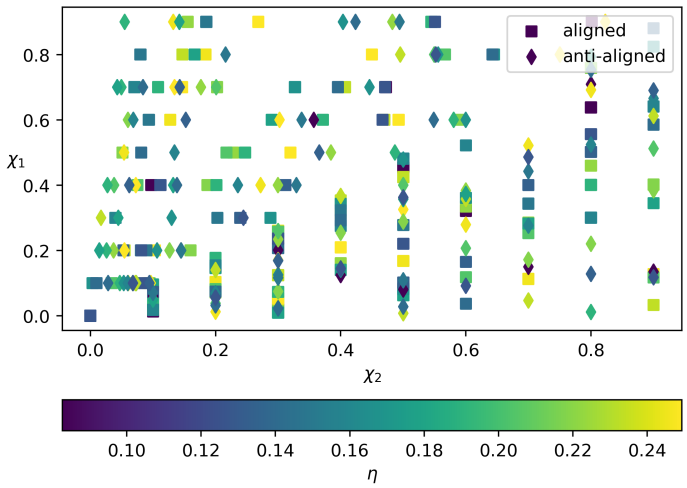

Figure 6. Parameter space for the non-precessing merger simulations without NR counterpart.

areas within the parameter space, we introduce additional simulated BBH mergers without a *SXS* counterpart. These mergers feature aligned as well as anti-aligned spins and span across an extensive range of mass ratios η, as depicted in figure 6. This augmented set adds an additional 220 instances to our set of non-precessing waveforms.

Together, cataloged and non-cataloged non-precessing simulations form a well-distributed set, which is suitable for an unbiased systematic investigation. Despite all efforts of accurately modeling GW waveforms, for some instances of non-precessing mergers, certain models may deviate strongly from the reference model and the other approximants in terms of remnant velocity or gravitational memory (again anticipating results of section 6). A closer investigation of such instances reveals that in these cases one of the modes of the deviating model (e.g. the $h_{2,1}$ mode) significantly differs from the corresponding mode of the remaining approximants and the reference model. This behavior predominantly occurs when one of the normalized spin components $\chi_{1,2}$ is extremal, i.e., either close to 0 or 1. Examples of

JCAP02(2025)060

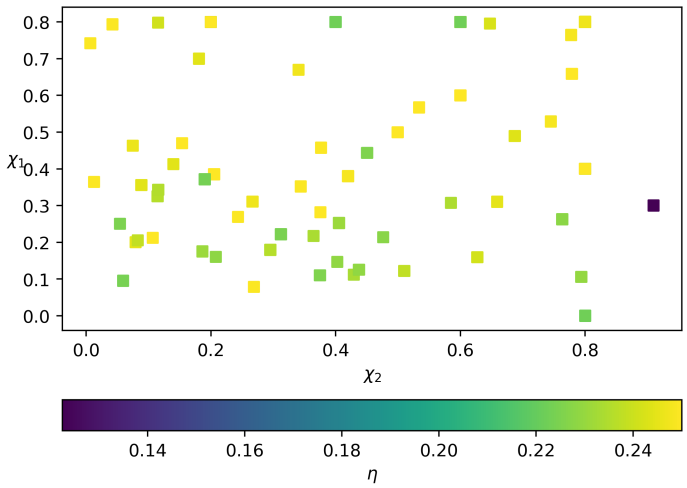

Figure 7. Parameter space for the precessing *SXS* data.

these types include for instance *SXS:BBH:0222*, *SXS:BBH:0223*, and *SXS:BBH:0251*, for which *EOB* results in a kick velocity much larger or smaller than for the remaining waveform models. In order to not affect our statistical considerations, such simulated waveforms are excluded from our investigation.

4.2 Precessing binary mergers

For the analysis of precessing BBH, we make use of 130 simulations with *SXS* counterparts, as displayed in figure 7. In terms of the two spin vectors, the binary merges are homogeneously distributed over the available parameter space. Notably, the cataloged simulations do not extend to lower mass ratios except for the *SXS:BBH:0165*, which involves two black holes of masses $51.4\,M_\odot$ and $8.6\,M_\odot$. As this lack of low mass ratio binaries ($\eta < 0.2$) potentially demonstrates a parameter bias, we supplement these precessing *SXS* simulations with a set of 75 additional instances. The added BBH merger simulatons are displayed in figure 8. The difference in the number and randomness of the added instances in the precessing and non-precessing case is a direct result of the computational complexity of aligning precessing waveforms. For these, we optimize the mismatch with respect to 5 variables instead of only ϕ_{ref}, resulting in a significantly larger data generation time. Due to time constraints, the added simulations in the precessing case are less numerous and more grid-like distributed.

5 Energy-momentum balance laws

To assess a waveform model's performance against a chosen reference model, an adequate measure of comparison has to be selected. We choose the remnant's kick velocity and the gravitational wave memory, which are both physical observables that can be calculated solely from the strain of a GW. In what follows we derive explicit expressions for the kick velocity and the memory from constraint equations which are known as *energy-momentum balance laws* [32]. These laws are exact mathematical results derived from full, non-linear GR when applied to the coalescence of compact binaries. For an exhaustive review and

JCAP02(2025)060

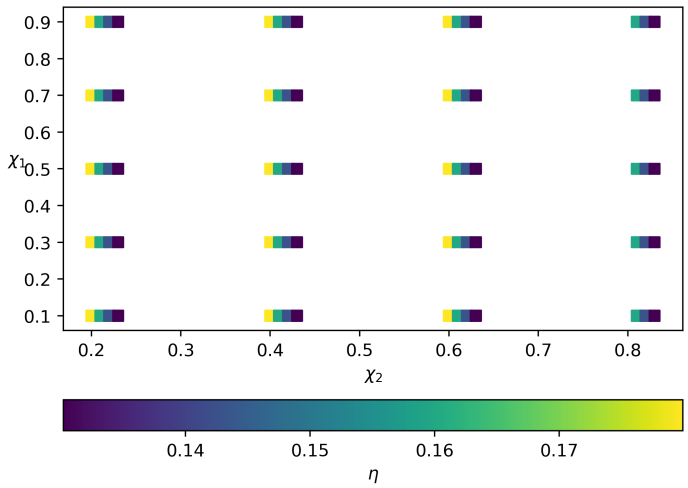

Figure 8. Parameter space for the precessing simulations without NR counterpart. For every bundle of data points, the spin configuration is identical. For illustrative reasons, the plot shows a slight misplacement.

derivation of the balance laws, we refer to [32, 52]. For a proof of concept that the balance laws can be used as diagnostic tools for assessing waveforms, we refer to [34], which uses a simple analytical waveform model, and to [33]. Here, we will simply state the main result, explain its physical content, and perform a decomposition into spherical harmonics in order to derive explicit expressions for kick velocity and memory. These expressions then serve as the basis of our comparative analysis.

5.1 The balance laws

Waveform models describe GWs emitted from isolated systems composed of two compact objects that orbit each other and coalesce. The coalescence is caused by a loss of orbital energy, due to the emission of GWs. When the two compact objects finally merge, the remnant can be subjected to a "kick". This is a consequence of the fact that GWs are emitted anisotropically and because they do not only carry energy but also momentum. Based on this simple qualitative picture of the physics of compact binary coalescence, one would expect that there is a mathematical law that describes the energy loss of the system and balances it against the energy carried away by GWs. One may furthermore expect that this law establishes a relation between *(i)* the initial and final mass of the system, *(ii)* the kick velocity, and *(iii)* the strains h_+, $h_\times$ of the GW. Such a mathematical law does indeed exist, and it can be derived from full, non-linear GR when applied to compact binary coalescence [32]. As a matter of fact, there is not only one law but infinitely many. Namely, one per point on the 2-sphere, i.e., one so-called energy-momentum balance law per choice of (θ, ϕ). These laws can be brought into the form

$$c^2 \left(\frac{M_{\text{remnant}}}{\gamma^3 \left(1 - \frac{\vec{v}}{c} \cdot \hat{x}\right)^3} - M_{\text{binary}} \right) = -\frac{1}{4} \frac{D_L^2 \, c^3}{G} \int_{-\infty}^{\infty} \left(\dot{h}_+^2 + \dot{h}_\times^2\right) \mathrm{d}t + \frac{1}{2} \frac{D_L \, c^4}{G} \mathrm{Re} \left[\eth^2 \left(h_+ - i \, h_\times\right) \right] \Big|_{t=-\infty}^{t=+\infty}, \tag{5.1}$$

JCAP02(2025)060

where $\vec{v}$ stands for the kick velocity, $\gamma \equiv \gamma(v)$ is the usual Lorentz factor from Special Relativity, $\hat{x} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ is the unit radial vector in spherical coordinates, D_L is the luminosity distance between source and observer, and $\eth$ is a differential operator called "eth" whose precise definition can be found in [32, 52].

By recalling our qualitative description of the physics of compact binary coalescence, we can give a tentative interpretation of the balance laws. Since $E = M\,c^2$, according to Einstein, we can interpret the left-hand side of (5.1) as the difference of two energies measured at two different instants of time. Namely, the energy of the remnant, which corresponds to the energy of the system after the merger, minus the energy of the binary system, which is the energy measured well before the merger. The denominator containing the Lorentz factor and the kick velocity can be understood as a correction due to the fact that the energy of the binary system is measured in its instantaneous rest frame, while the remnant is, in general, moving with respect to that frame with a velocity $\vec{v}$. As we will see below, when we decompose (5.1), for the $\ell = 0$ mode this factor reduces precisely to the special relativistic formula for the kinetic energy of a body moving relative to an inertial frame, $E = \gamma(v)Mc^2$. For a derivation of this factor and technical details, we refer the reader to [32]. Since the binary system is losing energy due to the emission of GWs, it is natural to expect that the difference in energies is negative. Indeed, on the right-hand side of (5.1) we find an integral over a manifestly positive quantity, $\left(\dot{h}_+^2 + \dot{h}_\times^2\right)$, with a minus sign in front. This integral computes the energy carried by GWs, and it is a well-known result of linearized GR. However, we stress again that these balance laws hold well beyond the linear regime, and, in fact, no linearization is ever used. Finally, the second term on the right-hand side represents the GW memory. Because the left side of (5.1) depends on $\hat{x}$, which is a function of (θ, ϕ), and because the strains h_+, $h_\times$ are not only functions of time but also functions of the angular variables (θ, ϕ), we find that there is indeed one balance law per choice of (θ, ϕ).

The strength and diagnostic capabilities of the balance laws derive from the fact that they are a precise mathematical result that is valid in full GR. The quantities which appear in their formulation, i.e., the initial and final mass of the binary system, kick velocity, and the GW strains, are all quantities that either enter as parameters or are provided in an approximate fashion by waveform models. Thus, the exact balance laws can be used to test the accuracy of waveform models.

The balance laws are based on the pioneering work of Bondi, Metzner, Sachs, and van der Burg, who laid the foundations for describing GWs beyond the linear approximation [53, 54]. Their work was subsequently extended and formalized by Newman, Penrose, Geroch, Ashtekar, and others, who introduced precise mathematical definitions for "asymptotically flat" and "asymptotically Minkowski" spacetimes [55–57], which, qualitatively speaking, capture the idea of having an isolated system contained within a finite spacetime region. Another key element in the derivation of the balance laws is the symmetry of asymptotic Minkowski spacetimes. These are described by the so-called Bondi-Metzner-Sachs (BMS) group [53]. By definition, this group leaves the asymptotic structure on null infinity $\mathscr{I}^+$ (read: scri plus) invariant. A variant of Noether's theorem can then be applied, which leads to a description of the (non-)conservation of Noether charges: the energy-momentum balance laws [58, 59]. The balance laws (5.1) shown here are an adaptation to the special case of compact binary

JCAP02(2025)060

coalescence. We refer to [52] for a didactic introduction and recent review of the fascinating mathematical subjects (asymptotic Minkowski spacetimes, the BMS group, balance laws), which here we only brushed over.

In what follows, we will need the mode-decomposition of equation (5.1) with respect to spherical harmonics $Y_{\ell m}(\theta, \phi)$. In order to compactify the notation, we set[6]

$$M_{\text{binary}} \coloneqq M_{i^\circ} \qquad \text{and} \qquad M_{\text{remnant}} \coloneqq M_{i^+}\,, \tag{5.2}$$

and we also introduce the asymptotic shear

$$h(t, \theta, \phi) \coloneqq \frac{1}{2}\left(h_+ + i\, h_\times\right)(t, \theta, \phi)\,, \tag{5.3}$$

where i is the imaginary unit. This last definition in particular allows us to re-write the memory term compactly as $\eth^2 \Delta \bar{h}$, with $\bar{h}$ denoting the complex conjugate of h and where

$$\Delta \bar{h} \coloneqq \int_{-\infty}^{+\infty} \dot{\bar{h}}\, \mathrm{d}t\,. \tag{5.4}$$

Re-writing the memory term in this fashion is possible thanks to the fact that $\mathrm{Im}(\eth^2 \bar{h})$ vanishes at $t = \pm\infty$ (cf. [32]).

Using these definitions and expanding both sides of equation (5.1) into spherical harmonics, we obtain a tower of constraints, namely one for each pair of (ℓ, m) with $|m| \leq \ell$. Concretely, this decomposition reads

$$\left(M_{i^\circ} - \frac{M_{i^+}}{\gamma^3\left(1 - \frac{\vec{v}}{c}\cdot\hat{x}\right)^3}\right)_{\ell,m} = \frac{D_L^2 c}{G}\int_{-\infty}^{+\infty}\left(|\dot{h}|^2\right)_{\ell,m}\mathrm{d}t - \frac{D_L c^2}{G} C_\ell \Delta \bar{h}_{\ell,m}\,, \tag{5.5}$$

where the C_ℓ coefficients are defined as

$$C_\ell \coloneqq \sqrt{(\ell-1)\ell(\ell+1)(\ell+2)}\,. \tag{5.6}$$

To arrive at this expression, we used well-known properties of spherical harmonics and the explicit expression of $\eth^2 \bar{h}$, which can for instance be found in [52]. Note that C_ℓ tells us that not each term in the balance laws (5.1) contributes to each mode. In fact, C_ℓ vanishes for $\ell = 0$ and $\ell = 1$. Thus, the memory term only contributes to $\ell \geq 2$. Treating the $\ell < 2$ and $\ell \geq 2$ modes of the balance laws separately allows us to devise a method for computing the remnant mass, in case it is not provided by the waveform model, the kick velocity, and the memory solely based on the GW strain.

For carrying out the mode-decomposition of the left-hand side of equation (5.5), it is convenient to first align the z-axis of the rest frame in which M_{i° is measured with $\vec{v}$ using a

[6] This is standard notation in the balance law literature and it hints at its origin: the balance laws formally hold on null infinity, $\mathscr{I}^+$. If one takes the limit into the distant past, i.e., well before the binaries merge, one reaches spacelike infinity, i°. On the other hand, if one moves along $\mathscr{I}^+$ into the distant future, well after the binaries have merged, one reaches timelike infinity, denoted by i^+.

JCAP02(2025)060

rotation in the plane spanned by $\vec{v}$ and $\hat{z}$. The first few (ℓ, m) coefficients then read

$$\begin{aligned}
(\ell = 0, m = 0): &\qquad 2\sqrt{\pi}(\gamma M_{i^+} - M_{i^\circ}), \\
(\ell = 1, m = 0): &\qquad 2\sqrt{3\pi}\gamma M_{i^+} v, \\
(\ell = 2, m = 0): &\qquad \frac{\sqrt{5\pi}\gamma M_{i^+}\left(5v^3 + 3\gamma^{-4}\tanh^{-1}(v) - 3v\right)}{v^3}.
\end{aligned}$$

Observe that the first term is, up to a factor of c^2, simply the difference between the rest energy M_{i° of the binary and the kinetic energy $\gamma\, M_{i^+}$ of the remnant, while the second term is the (z-component) of the momentum of the remnant. The third term does not admit a simple physical interpretation, but we emphasize that the inverse of the hyperbolic tangent of v caused numerical instabilities in the code we used for our analysis. We therefore Taylor expanded $\tanh^{-1}(v)$ up to sixth order in v/c. To obtain the corresponding modes in the original frame, we transform the above modes using Wigner D-matrices $D^{\ell}_{m,m'}$ for the inverse rotation used to align $\vec{v}$ and $\hat{z}$.

The decomposition of the GW energy integral in (5.5) can easily be carried out. In fact, the integrand can be written as

$$|\dot{h}|^2 = \sum_{\ell,m} \alpha_{\ell m} Y_{\ell m}(\theta, \phi)\,. \tag{5.7}$$

A straightforward computation reveals that the $\alpha_{\ell m}$ coefficients are given by (see also [33])

$$\begin{aligned}
\alpha_{\ell m} = \sum_{\ell_1=2}^{\infty}\sum_{\ell_2=2}^{\infty}\sum_{|m_1|\le\ell_1}\sum_{|m_2|\le\ell_2} (-1)^{m_2+m}\dot{h}_{\ell_1 m_1}\bar{\dot{h}}_{\ell_2 m_2}\sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell+1)}{4\pi}} \\
\times \begin{pmatrix} \ell_1 & \ell_2 & \ell \\ m_1 & -m_2 & -m \end{pmatrix}\begin{pmatrix} \ell_1 & \ell_2 & \ell \\ 2 & -2 & 0 \end{pmatrix}.
\end{aligned} \tag{5.8}$$

The $\left(\begin{smallmatrix} \ell_1 & \ell_2 & \ell \\ m_1 & m_2 & m \end{smallmatrix}\right)$ denotes the Wigner-3j symbol, which effectively determines which strain modes $\dot{h}_{\ell_1,m_1}\bar{\dot{h}}_{\ell_2,m_2}$ couple to each other in (5.8).

5.2 Remnant mass and velocity

For astrophysically realistic kick velocities one can assume $\gamma \approx 1$. This approximation reduces the $\ell = 0$ mode of the balance laws to an energy conservation equation of the form

$$c^2(M_{i^\circ} - M_{i^+}) = \frac{D_L^2 c^3}{16\pi G}\int_{-\infty}^{\infty} \mathrm{d}t \oint \mathrm{d}\Omega\, |\dot{h}|^2\,, \tag{5.9}$$

where $\oint \mathrm{d}\Omega$ is an integral over the unit 2-sphere. In agreement with our intuition, the mass loss of the system is accounted for by the energy radiated away by GWs. For waveform models that provide the mass of the remnant, this formula can be used as a consistency check. On the other hand, when a waveform model does not provide M_{i^+}, the above formula can be used to determine the remnant's mass using only the total mass of the binary system, the

JCAP02(2025)060

luminosity distance, and the GW strain as input.[7] Upon using the expansion coefficient (5.8) for $\ell = m = 0$, the energy conservation equation can also be expressed as

$$c^2(M_{i^\circ} - M_{i^+}) = \frac{D_L^2 c^3}{8\sqrt{\pi}G} \int_{-\infty}^{\infty} \mathrm{d}t\, \alpha_{0,0}. \tag{5.10}$$

The $\ell = 1$ mode of the balance laws encodes the conservation of linear momentum, as one might have expected. This allows us to solve the momentum equations for the components of the kick velocity, which can be expressed as

$$\vec{v}_{\rm kick} = \frac{D_L^2 c^2}{16\pi G M_{i^+}} \int_{-\infty}^{\infty} \mathrm{d}t \oint \mathrm{d}\Omega\, \hat{x}_i\, |\dot{h}|^2, \tag{5.11}$$

in the original reference frame (i.e., the frame where $\vec{v}$ and $\hat{z}$ are not aligned). Here, $\hat{x}_i$ stand for the i-th component of the radial unit vector $\hat{x} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. The kick velocity components can of course also be expressed in terms of the $\alpha_{\ell m}$ coefficients for $\ell = 1$ and $m = -1, 0, +1$. One obtains

$$v_1 = \frac{D_L^2 c^2}{16\pi G M_{i^+}} \sqrt{\frac{2\pi}{3}} \int_{-\infty}^{\infty} \mathrm{d}t\, (\alpha_{1,-1} - \alpha_{1,1})\,, \tag{5.12}$$

$$v_2 = \frac{-iD_L^2 c^2}{16\pi G M_{i^+}} \sqrt{\frac{2\pi}{3}} \int_{-\infty}^{\infty} \mathrm{d}t\, (\alpha_{1,-1} + \alpha_{1,1})\,, \tag{5.13}$$

$$v_3 = \frac{D_L^2 c^2}{8\pi G M_{i^+}} \sqrt{\frac{\pi}{3}} \int_{-\infty}^{\infty} \mathrm{d}t\, \alpha_{1,0}\,. \tag{5.14}$$

Equations (5.12)–(5.14) determine the kick velocity using luminosity distance, remnant mass, and strain as input. If M_{i^+} is not provided by a waveform model, it can be determined via the energy conservation equation (5.9).

5.3 Gravitational memory

As we have seen above, modes with $\ell \geq 2$ contain contributions from the GW memory. The memory term can be decomposed into two contributions,

$$\Delta\bar{h}_{\ell,m} = \Delta\bar{h}^{\rm lin}_{\ell,m} + \Delta\bar{h}^{\text{non-lin}}_{\ell,m}\,. \tag{5.15}$$

Following standard conventions in the literature, we define the first contribution as

$$\Delta\bar{h}^{\rm lin}_{\ell,m} \coloneqq \frac{G}{C_\ell D_L c^2} \left(\frac{M_{i^+}}{\gamma^3 (1 - \frac{\vec{v}}{c}\cdot\hat{x})^3} - M_{i^\circ} \right)_{\ell,m}, \tag{5.16}$$

and call it the *linear memory* [5, 60]. The second term, defined as

$$\Delta\bar{h}^{\text{non-lin}}_{\ell,m} \coloneqq \frac{D_L}{4C_\ell c} \int_{-\infty}^{\infty} \mathrm{d}t\, \alpha_{\ell m}\,, \tag{5.17}$$

[7] It is important to note that the accuracy of analytical predictions is inherently linked to the accuracy of the strain, which includes the number of cycles prior to the merger captured in the waveform time series. In numerical simulations, these waveforms are typically truncated at a certain point in time, introducing a natural margin of error for all kinematic predictions. However, since this truncation affects all waveforms uniformly, it does not impact the comparative analysis conducted in this work.

JCAP02(2025)060

is known as the *non-linear memory* [6, 7]. Other names for these two terms are *ordinary memory* for the former and *null memory* for the latter [37]. More importantly, we interpret the $\ell \geq 2$ modes of the balance laws (5.5) as constraints on the memory terms. The same constraints have been used in previous studies to either add memory to waveform models or correct waveform models which did not accurately incorporate the memory effect [33, 37]. In this work, we instead use the linear and non-linear memory inferred from (5.16) and (5.17), respectively, as a means of comparison between different waveform models.

5.4 Connections to black hole kinematics

The significant utility of the balance flux laws lies not only in their mathematical simplicity but also in their predictive power concerning the kinematic properties of the BBH merger's remnant compact object. As demonstrated in this section, the final mass and velocity of the remnant can be accurately computed based on the GW strain alone. However, especially for precessing systems, decomposing physical quantities in terms of the components specified by equation (5.8) can further elucidate specific subtleties of the merger process.

To provide a brief insight, we consider the selection rules (i.e, the Wigner-3j symbols) that govern the harmonic strain components entering the relevant $\alpha_{\ell m}$'s. Generally, the α's are dominated by terms including at least one dominant strain mode $h_{2,\pm 2}$. For illustration, consider $\alpha_{1,0}$, which determines the out-of-plane kick of the remnant body after the merger, i.e., the velocity component v_3. Evaluating the sums in (5.8) for $\alpha_{1,0}$, one finds that two contributions proportional to $h_{2,-2}\bar{h}_{2,-2}$ and $h_{2,2}\bar{h}_{2,2}$[8] cancel each other exactly,effectively nullifying any significant contribution to the out-of-plane kick for non-precessing BBHs. However, in the presence of an asymmetry between $h_{2,2}$ and $h_{2,-2}$, as it is the case for precessing binaries, these two contributions to $\alpha_{1,0}$ no longer cancel out, leading to a non-trivial contribution for the remnant velocity v_3. Although the asymmetry is generally expected to be small, and thus the difference between the $h_{2,-2}\bar{h}_{2,-2}$ and $h_{2,2}\bar{h}_{2,2}$ terms is small as well, the out-of-plane contribution to the kick can become comparable in magnitude to the in-plane kick.

The relatively small asymmetry between the ℓ, m and $\ell, -m$ modes is indeed crucial when assessing the impact on various physical quantities, particularly the non-linear memory effect. This memory effect is especially sensitive to the $\alpha_{2,0}$ component which plays a dominant role because it permits the coupling of two dominant modes, such as $h_{2,-2}\bar{h}_{2,-2}$ and $h_{2,2}\bar{h}_{2,2}$. In fact, in contrast to $\alpha_{1,0}$, these terms appear with equal sign. Thus, independent of any precession in the merging system, the terms proportional to $h_{2,-2}\bar{h}_{2,-2}$ and $h_{2,2}\bar{h}_{2,2}$ reinforce each other, leaving only minor traces of any potential asymmetry between $h_{2,2}$ and $h_{2,-2}$ in their sum. This constructive interference of these two contributions to $\alpha_{2,0}$ is a key factor in the build-up of the non-linear memory. While the contribution of the asymmetry accounts for the entire out-of-plane kick, the (non-linear) memory is modified only by a fraction of the asymmetry's magnitude relative to the integral of the dominant mode terms $h_{2,\pm 2}\bar{h}_{2,\pm 2}$. Therefore, unlike the kick, the non-linear memory is only mildly affected by the asymmetries between the ℓ, m and $\ell, -m$ modes. This mild sensitivity implies that the memory effect can still be meaningfully estimated even for waveform models that do not explicitly account for the differences between these modes in precessing merger waveforms.

[8] Time derivatives are omitted for simplicity.

JCAP02(2025)060

The latter include the *EOB* and *Phenom* model considered in this work. Naturally, other mode combinations within the α's are also affected by the potential asymmetry between ℓ, m and $\ell, -m$ modes. However, these contributions are generally subdominant and thus their effects are only of secondary concern.

The kinematic mode analysis of the α components encompasses a broad range of intricate aspects that go beyond the brief demonstration provided above. Although it is valuable to delve deeper into these details, offering a comprehensive discussion without empirical data is not productive at this point. Instead, we will revisit and expand upon this analysis in section 6. There, we will illustrate how state-of-the-art waveform models can fully exploit the analytical bounds provided by the balance flux laws. This will be supported by numerical data, which will offer concrete evidence to reinforce the theoretical insights presented here.

To summarize this section, we decomposed the balance laws (5.1) into spherical harmonic modes to obtain analytical equations for the remnant mass, the kick velocity, and the GW memory. The remnant mass M_{i^+} is inferred from the $\ell = 0$ mode of the balance laws, while the kick velocity is deduced from the conservation of linear momentum implied by the $\ell = 1$ mode. The $\ell \geq 2$ modes are used to infer linear and non-linear GW memory. Connections to the kinematics of the remnant BH (or remnant compact object) have been pointed out. Equations (5.9), (5.11), (5.16), and (5.17) are implemented numerically and applied to all simulations across all waveform models considered in this work. The resulting waveform assessment is the subject of the following section.

6 Waveform assessment

Computing the remnant velocity, following equations (5.12)–(5.14), and evaluating the gravitational memory, encoded in equations (5.16) and (5.17), we assess the performance of the chosen waveform approximants across the parameter space specified in 3. We also examine the impact of the subdominant strain modes on the kick and inferred GW memory. The analysis of alignment-sensitive quantities is restricted to non-precessing siumlations only. Before examining large sets of data, we calibrate our numerical pipeline by testing the computations of selected physical quantities against the metadata contained in the *SXS* database. In particular, we measure deviations of the kick's magnitude computed across approximants using (5.11) with respect to the *SXS* data. The substantial agreement found in these comparisons validates the applicability of the balance laws as numerically implemented in the context of this work.

6.1 Mode mismatch

We start by considering the non-precessing mergers detailed in section 4, for which NR waveforms are accessible. A natural measure of comparability between waveforms (or individual modes) is the mismatch function (3.2). In the alignment procedure, the mismatch function $\mathcal{M}$ is minimized for the dominant $h_{2,2}$ mode. The remaining modes are not exposed to any direct mismatch minimization. Thus, a first intuition about potential shortcomings of waveform models can be gained based on the residual mismatch of the subdominant modes, and how the latter compares to the mismatch in the dominant $h_{2,2}$. In figure 9, we present

JCAP02(2025)060

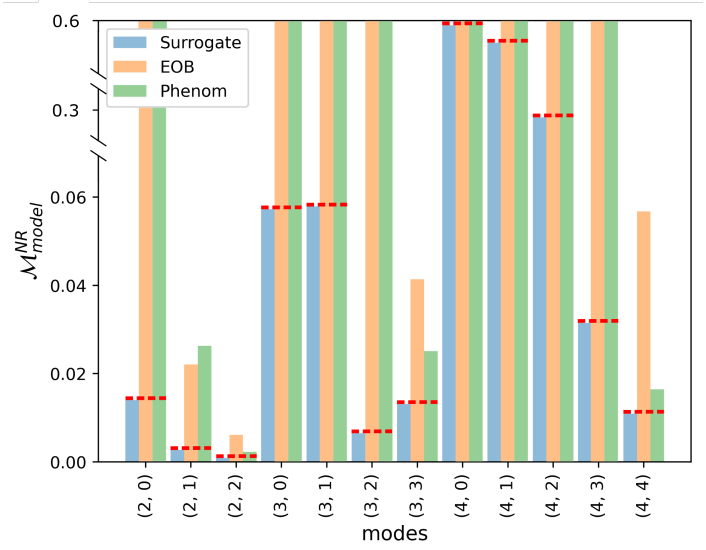

Figure 9. Mode mismatch with respect to NR, averaged over all non-precessing BBH simulations. We limit the mismatch axis at the highest value for *Surrogate* within the selected modes. The modes exceeding the mismatch cut-off are not simulated by *EOB* and *Phenom* but accessible in their *LALSuite* implementation. Thus, their precise mismatch values are insignificant.

the mismatch calculated via (3.2) for the mode content common to all waveform models,[9] averaged over all simulations under consideration. As expected, we observe the lowest mismatch for the dominant $h_{2,2}$ mode throughout approximants, with *Surrogate* performing best in this metric. For subdominant ℓ modes, the mismatch becomes larger. Particularly for $\ell = 4$ modes but also for $h_{2,0}$, $h_{3,2}$, $h_{3,1}$, $h_{3,0}$ and their complex conjugates, the mismatch is multiple orders of magnitude larger than for $h_{2,2}$. The higher mismatch between *Surrogate* and NR in subdominant modes is attributed to *Surrogate* prioritizing the modeling accuracy of the dominant $h_{2,2}$ in its interpolation procedure [16, 17, 61, 62].

As outlined in table 1, multiple subdominant modes are not modeled by *EOB* and *Phenom* (despite being accessible via the *LALSuite*) and thus their mismatch is close to one. For better readability we choose not to extend the y-axis beyond the largest *Surrogate* mismatch in figure 9. The absence of non-trivial contributions for said modes roughly translates to precessing simulations as well, despite internal model procedures shifting some mode power across subdominant strain modes via, for instance, the twisting-up procedure of *Phenom* models [22]. Thus, for precessing mergers, a similar pattern regarding the mode-by-mode mismatches as in figure 9 is found. Note, however, that for precessing simulations the mismatch computed for an ℓ, m mode and the $\ell, -m$ mode can differ due to the asymmetry for such counterparts present in *NR* and *Surrogate* waveforms.

It should be noted that the *TEOB* model is omitted from figure 9 due to only providing the $h_{2,2}$ mode. On average, the mismatch for this mode is only marginally higher for *TEOB* compared to *EOB*, featuring *TEOB*, as well, an excellent approximant for the dominant GW strain mode. The mismatch of *TEOB* with respect to NR improves further if only the

[9] Note that we display only up to $\ell = 4$ due to the restricted mode content of *Surrogate*.

JCAP02(2025)060

inspiral is considered. This directly follows from the intrinsic modeling procedure of the NR informed *TEOB* waveform perfecting precision in the inspiral at the cost of marginal deviations around the merger phase [29, 63].

We conducted a similar analysis for non-cataloged merger simulations, i.e., waveforms for which no NR waveform is available. In this case, we chose *Surrogate* as the reference model and observe an analogous relationship between the mismatch of *EOB* and *Phenom* versus *Surrogate*. Specifically, for $\ell = 2$, the mismatch with respect to *Surrogate* for both *EOB* and *Phenom* is low. However, for $h_{3,3}$ and $h_{4,4}$, *EOB* waveforms exhibit visibly stronger dissimilarities with respect to *Surrogate* than *Phenom* does.

In addition to intrinsic differences in the waveform modes, the mismatch might be influenced by a broad class of nonphysical signatures that we term "numerical artifacts", an instance of which is displayed in appendix A. These artifacts, particularly prominent in higher modes, require separate consideration. In literature, some artifacts similar in nature to what is displayed in figure 16 have been addressed and mitigated, for instance, through the alignment of BMS charges [64, 65]. Here, we will not delve into a detailed discussion about their origin or impact but simply point out that even with a well-defined alignment procedure, the numerical nature of the analysis can potentially affect the mismatch results.

Despite the higher mismatch of certain modes, the full waveforms provide an adequate approximation to NR due to the dominance of the well-modeled $h_{2,2}$ mode for the majority of BBH simulations. At the current level of precision, addressing the individual subdominant mode mismatch is thus only of secondary interest. However, in the perspective of future GW probes, the precision benchmark will increase significantly, potentially rendering higher modes more relevant [66]. Hence, it is imperative for future GW waveform models to confront the evident challenges depicted in figure 9, such as incomplete mode contents and decreasing precision of higher-order (subdominant) modes. In the subsequent section, we establish the groundwork for addressing these issues by examining (alignment-independent) physical properties of the BBH computed solely based on the asymptotic strain and its harmonic modes.

6.2 Computing physical quantities

In order to investigate the impact of an approximant's mode content with regard to its overall waveform quality, we compute the relevant physical (strain-based) quantities of BBH mergers using two sets of strain modes. The first set includes all available modes for each individual model, the second set only includes low-mismatch subdominant modes, e.g. we exclude modes belonging to the set[10]

$$H_{\rm sub} = \{h_{2,0}, h_{3,\pm2}, h_{3,\pm1}, h_{3,0}, h_{4,\pm3}, h_{4,\pm2}, h_{4,\pm1}, h_{4,0}\}.$$

Excluding $H_{\rm sub}$ sets all waveform approximants on roughly equal footings and allows for a quantitative comparison by including only the modes that are meaningfully simulated throughout the waveform models.

[10]Note that we do not exclude modes that exceed the mode content of *Surrogate* model, i.e., $\ell > 4$ are not excluded for the remaining models (including NR).

JCAP02(2025)060

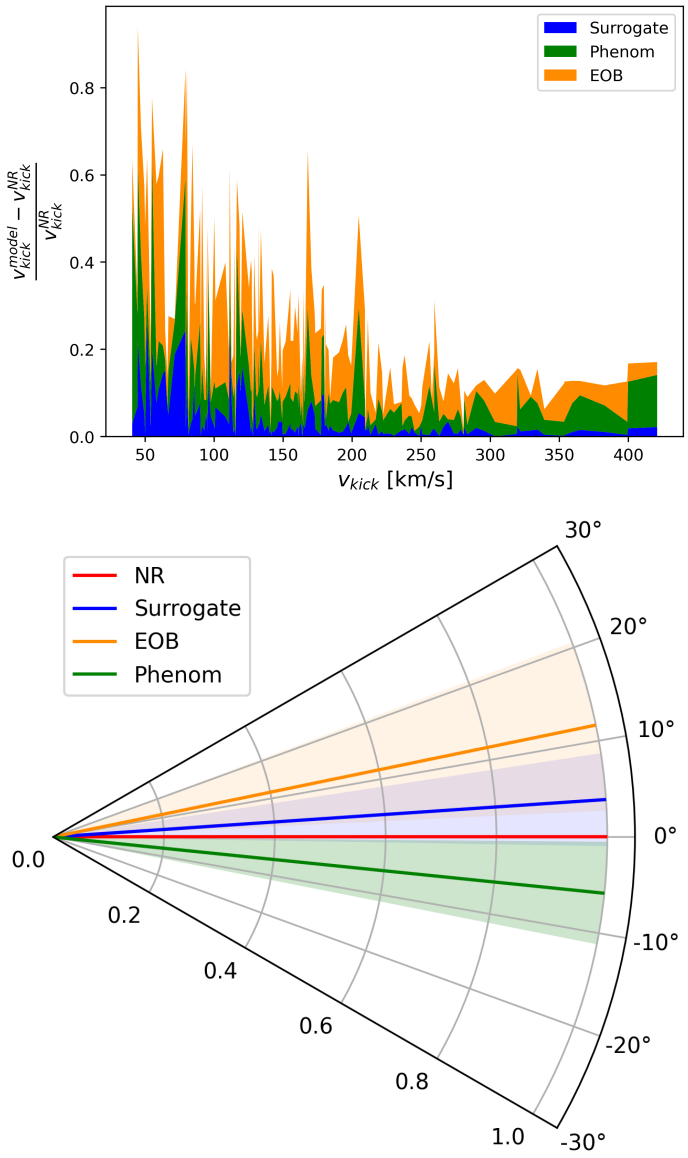

Figure 10. Relative error for the kick magnitude (top) and absolute deviation angle (bottom) of the selected approximants with respect to NR for non-precessing merger simulations. The top stack plot shows the relative errors for each individual waveform labeled by its resulting kick magnitude. The bottom plot shows the deviation in kick direction with respect to orientation predicted by NR, averaged over all simulated waveforms. A 1σ-interval above and below the mean is colorized correspondingly. Despite displaying the mean absolute deviation, for better readability, the presented value for *Phenom* is mirrored by the 0° line.

JCAP02(2025)060

6.2.1 Remnant velocity

For the mentioned set of modes, we first consider the recoil velocity of the remnant BH of non-precessing simulations. Our attention thereby is limited to mergers with kick velocities varying between 20 and 400 km/s. Merger simulations with a negligible kick (i.e., $v < 20$ km/s) are excluded from the analysis. We compare the magnitude and direction of the remnant velocities (5.11) by computing each model's relative error with respect to NR.

Computed using the full mode content, the relative error for the kick magnitude and the directional deviation of the remnant velocity vectors are displayed in figure 10. In the top plot, we stack the relative errors and plot them against the corresponding kick magnitude for all simulations. The bottom plot displays the average across simultations of the absolute deviation in terms of kick direction with respect to NR. Negative angles for *Phenom* are assigned for better visibility in the plot. Around the average deviation, we shade the 1σ-neighborhood in the corresponding color, where σ represents the standard deviation of the deviation from the *NR* kick direction.

From a quantitative point of view, figure 10 shows a good agreement between *Surrogate* and NR, as expected. In particular, for larger velocities, the relative errors are small and the directional deviation is less than 5 degrees on average. *EOB* and *Phenom* waveforms cannot reproduce the reference velocity vector with comparable accuracy. They also display a similar trend of larger relative errors for lower remnant velocities. The strong fluctuations indicate a higher dependency on the individual simulation and its corresponding intrinsic parameters. For both the kick magnitude and direction, *EOB* exhibits on average the largest errors averaged over the tested parameter space. In total, these results agree with the findings of previous assessments based on kick velocities (cf. to figure 4, 5 in [36], but note the different waveform model instances and paramter space). For *TEOB*, the computed kicks are trivial throughout the parameter space tested in this analysis. This is a direct result of providing only the dominant strain mode and will be clarified below.

Computing similar statistics without the subdominant high-mismatch modes H_{sub}, we find that neither the kick magnitude nor its direction exhibits significant differences compared to the values computed using the full mode content. For NR and *Surrogate*, the change in directional deviation and the relative magnitude error, on average, corresponds to only a few percent of the corresponding full-mode content values. For *EOB* and *Phenom* we do not find statistically significant differences between excluding or including H_{sub}. The same applies to non-cataloged non-precessing merger simulations.

The results indicate that quantities related to the remnant velocity of the final BH are not sensitive to the subdominant mode content of the strain generated by a GW waveform model. This observation extends as well to the computation of the final mass of the remnant black hole (cf. equation (5.9)). The insensitivity of the remnant's mass and kick velocity with respect to the set of modes H_{sub} can be readily explained by the analytical expressions (5.8) for the $\alpha_{\ell,m}$-coefficients and equations (5.12)–(5.14): The Wigner-3j symbols appearing in equation (5.8) single out individual terms $h_{\ell_1,m_1}\bar{h}_{\ell_2,m_2}$[11] which contribute to each $\alpha_{\ell m}$. These coefficients, in turn, determine physical quantities such as the kick vector. According to equations (5.12)–(5.14), the coefficients relevant to the kick are $\alpha_{1,\pm1}$ and $\alpha_{1,0}$, where the

[11] We suppress time derivative for ease of notation.

JCAP02(2025)060

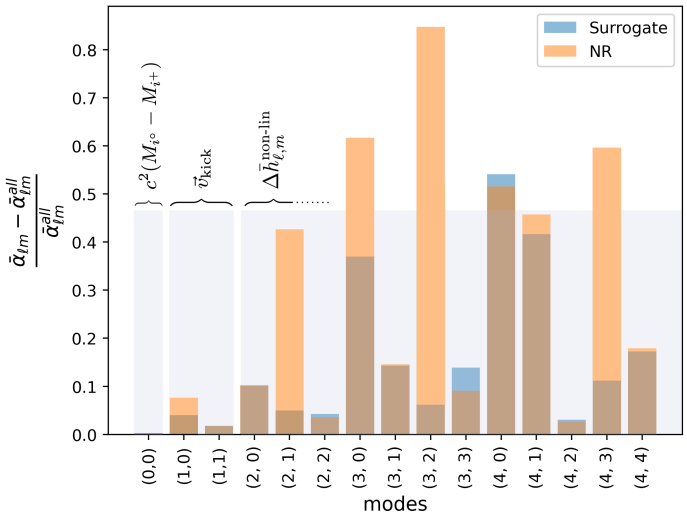

Figure 11. Time-integrated, relative error for $\alpha_{\ell m}$ (5.8). The error compares the α computed including all strain modes against the same coefficient calculated excluding $H_{\rm sub}$ for NR and *Surrogate*. We mark the modes (ℓ, m) relevant for energy, kick, and non-linear memory.

former two $\alpha_{\ell,m}$'s determine the in-plane kick and the latter dictates the out-of-plane kick. As mentioned in the section 5.4, this component becomes non-trivial only for precessing mergers.

Among all the strain mode pairs within these kick coefficients, the most relevant terms involve at least one dominant strain mode, $h_{2,\pm2}$. Terms proportional to $h_{2,2}\bar{h}_{2,2}$ appear exclusively in $\alpha_{1,0}$, where it is observed that the term proportional to $h_{2,2}\bar{h}_{2,2}$ cancels the term containing $h_{2,-2}\bar{h}_{2,-2}$. Therefore, the most significant contributions to the kick components arise from the mixing of $h_{2,\pm2}$ modes with subdominant ones, such as $h_{2,\pm2}\bar{h}_{2,\pm1}$in $\alpha_{1,\pm1}$. Contrarily, for $\alpha_{1,0}$ there is no non-vanishing contribution combining two $\ell = 2$ strain modes, which demonstrates analytically the comparatively small out-of-plane-component of the kick for non-precessing BBHs. Naturally, the latter statement is not true in the precessing case as already discussed in section 5.4.

In $\alpha_{1,\pm1}$, the strain mode $h_{2,2}$ also appears in combinations such as $h_{2,\pm2}h_{3,\pm1}$ and $h_{2,\pm2}h_{3,\pm3}$, which cancel in the sum (5.8). Furthermore, when combining $\alpha_{1,1}$ and $\alpha_{1,-1}$ in the v_x and v_y velocity components, the entries in x- and y-direction are determined by terms proportional $h_{2,\pm2}h_{2,\pm1}$ due cancellations in the sum $\alpha_{1,1} \pm \alpha_{1,-1}$ in equation (5.11). Computing the magnitude of the in-plane kick, we thus observe that an alignment induced phase correction of the $h_{2,\pm2}$ harmonic strain mode enters both components $\alpha_{1,\pm1}$ equally and, hence, the overall phase drops out and the in-plane kick magnitude remains invariant under phase shifts in $h_{2,\pm2}$.

Using the above arguments, it follows that, throughout the simulated BBHs, the dominant contribution in the α's relevant to the in-plane kick are proportional to $h_{2,\pm1} \notin H_{\rm sub}$ and thus only marginally affected by the exclusion of subdominant modes $H_{\rm sub}$; therefore, neither magnitude nor direction (within the plane) changes significantly when disabling $H_{\rm sub}$ in the computation. A similar conclusion applies also to the final mass, which is determined by $\alpha_{0,0}$.

JCAP02(2025)060

The qualitative analysis of the mode contributions to the kick is quantified in figure 11. It displays the relative error of the α-coefficients for NR and *Surrogate* for $H_{\rm sub}$ included versus excluded.[12] The plot clearly illustrates that the alteration in the α's relevant to the kick velocity is minimal. The more pronounced error value for NR is explained by its expanded mode content ($\ell \leq 8$) compared to *Surrogate* ($\ell \leq 4$), leading to a significantly larger number of mode mixing terms in the summation of $\alpha_{1\pm1}$. Consequently, if $H_{\rm sub}$ is disregarded, $\alpha^{\rm NR}_{\ell m}$ is more severely affected than $\alpha^{\rm Surrogate}_{\ell m}$.

In conclusion, the fundamental limit of the kick velocity in assessing the harmonics modes results from the restricted mixing appearing in α_{1m}. Based on the Wigner-3j symbols in (5.8), the allowed coupling of modes is restricted to $|\ell_1 - \ell_2| \leq \ell$ and, correspondingly, the dominant $h_{2,2}$ mode does not combine with any higher modes. This, however, changes when computing the memory.

6.2.2 Memory components

The linear and non-linear GW memory induced by the waveforms is computed and decomposed in equation (5.15). Figure 12 illustrates the relative errors for both linear and non-linear contributions to the memory for non-precessing mergers, considering the full mode content. In the upper plot, the total linear memory is normalized to the NR contribution for each waveform model. The shaded regions mark the 1σ interval above and below the memory error of the ensemble of simulations (resulting in a shaded region of total with 2σ). The solid lines represent the mean values of the memory ratio $\Delta h_{\rm model}/\Delta h_{\rm NR}$. Similar quantities are presented in the bottom plot for the non-linear memory, depicted as a time series to observe the error accumulation as the non-linear memory builds up. The time $t = 0$ corresponds to the merger time, and the normalization is performed with respect to the time-integrated non-linear NR memory for each simulation.

In terms of both linear and non-linear memory, *EOB* and *Phenom* exhibit notably poorer performance compared to the *Surrogate* approximant. These two models consistently produce smaller memory values. In contrast, *Surrogate* performs well, particularly in the case of the dominant non-linear GW memory. For *EOB* and *Phenom*, the NR values fall outside the 1σ range. *EOB* also shows a similar trend for the linear part, suggesting systematic deficiencies regarding the memory in these models. In the bottom plot, we replace *Phenom* with *TEOB*, demonstrating its capabilities of accurately modeling the memory despite containing only the dominant mode. As for the kick velocity, *TEOB*'s linear memory contribution is trivial due to the absence of the $h_{2,\pm1}$ and other subdominant modes.

In general, the linear memory is highly non-trivial to model, i.e., it requires the knowledge of additional Newman-Penrose scalars which are not provided by the majority of approximants.[13] This, however, can be avoided when extending time integrals from negative to positive time-like infinity, such as in equation (5.1). Given the finiteness of the simulated waveforms, this approximation naturally introduces a small systematic error which is irrelevant to our comparison as it is introduced for all models equally and only relative numerical

[12] *EOB* and *Phenom* are excluded from the presentation as they do not model modes contained in $H_{\rm sub}$ in the first place.

[13] An exception is given by the SXS Collaboration Caushy-Characteristic-Extraction (CCE) catalog, see [67]. Similar techniques are also adapted by other models, e.g., [68].

JCAP02(2025)060

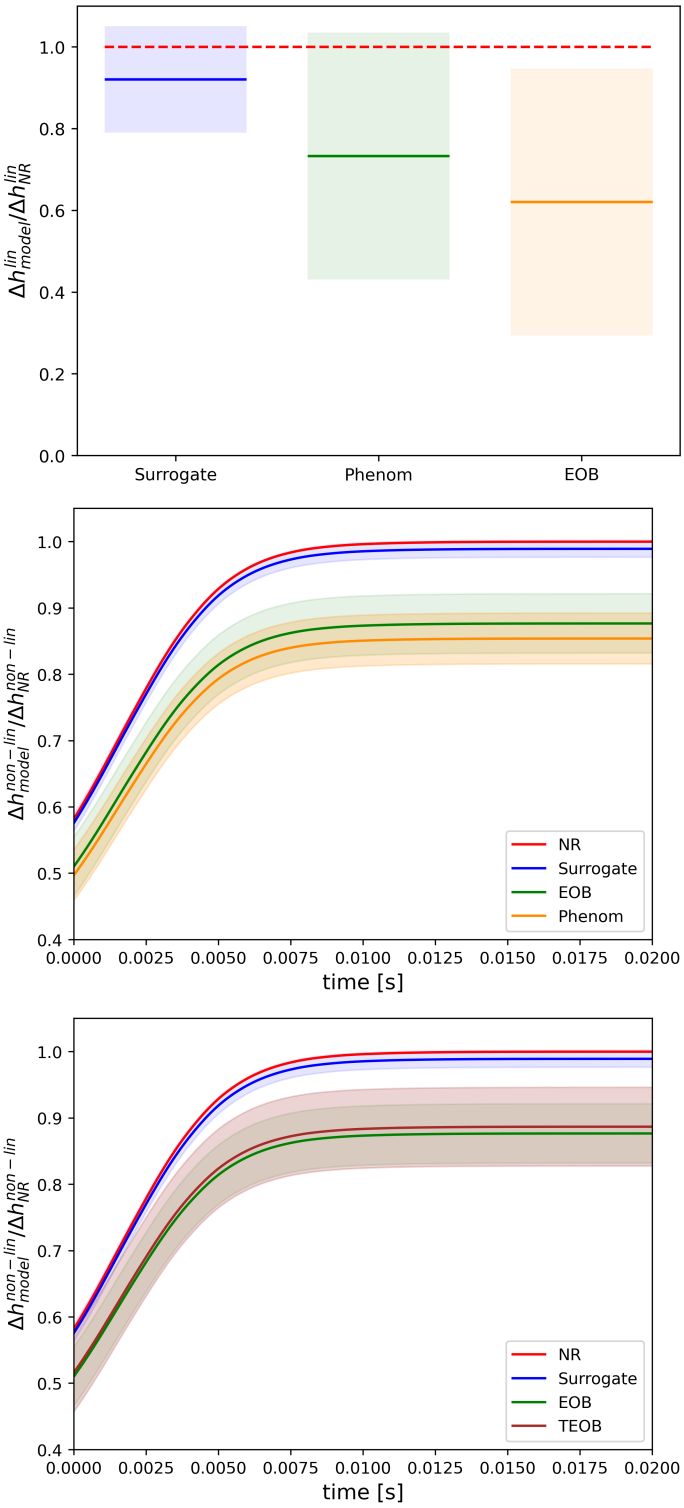

Figure 12. Deviation of waveform models' linear (top) and non-linear (middle) memory with respect to NR. Corresponding 1σ-intervals above and below the mean are colorized. The non-linear memory is plotted as a time series. Both memory contributions are normalized to the NR value and averaged over all simulated mergers. The bottom plot shows again the non-linear memory, but we replace *Phenom* with *TEOB*. For better readability, we provide two plots for the non-linear memory.

JCAP02(2025)060

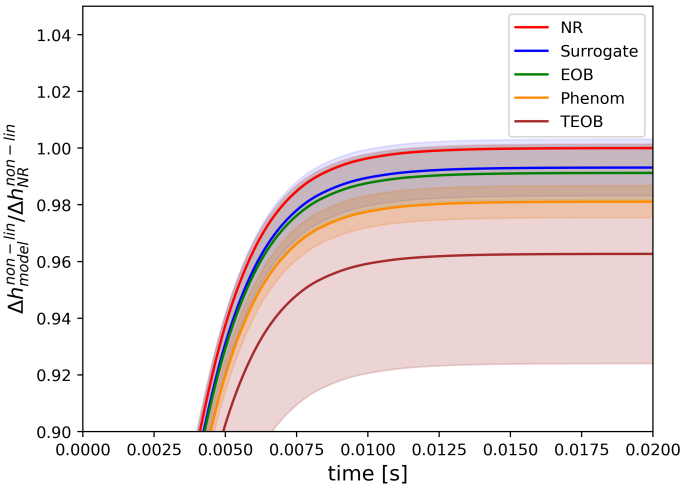

Figure 13. Non-linear memory as depicted in figure 12, here computed without high-mismatch modes, i.e., excluding H_{sub}.

quantities are considered. A more exact formulation of the linear memory is given in [37]. In addition to the considered approximation, the linear part is many orders of magnitude smaller than the non-linear part, making it more prone to numerical errors in the models. Consequently, the linear memory exhibits large fluctuations.

As before, the memory can be computed excluding the set of harmonic stain modes H_{sub}. We find that, for the linear memory component, the change of mode content, similar to the kick velocity, results only in a marginal change of the models' individual contributions. This is unsurprising, as by equation (5.16) the linear memory is solely determined by the remnant velocity. This dependency on the kick results from the approximation used in the derivation of the balance laws as we state them, i.e., (5.1), where the Newman-Penrose scalars originally appearing in the balance laws can be replaced by functions of the remnant velocity in certain limits.[14] Consequently, as the linear memory is determined by the kick, the discussion regarding the strain modes entering the computation of the linear memory follows similar arguments as for the kick in the previous subsection. Notably, the kick velocity, and thus the linear memory, is an alignment-dependent quantity which suggest that the displayed dissimilarities visible figure (5.16) throughout approximants may occur (partially) due to suboptimal alignment.

A completely different picture is drawn by the non-linear memory when disabling the set of modes H_{sub}. As depicted in figure 13, the relative errors throughout models are converging towards *NR* when the mode content is reduced to a common set. While the absolute values for *NR* and *Surrogate* decrease by $\mathcal{O}(10)$ percent, *EOB*, *TEOB*, and *Phenom*, as expected, do not change. Consequently, after excluding H_{sub}, the non-linear memory contribution of all considered approximants is roughly equal, and, by the definition of the balance laws, so is their energy flux. This statement is independent of the alignment procedure as the non-linear memory is not sensitive to the chosen reference phase. The variances of the relative errors

[14] For a detailed discussion on the relevant transformations and the validity of physical assumptions we refer to [32].

JCAP02(2025)060

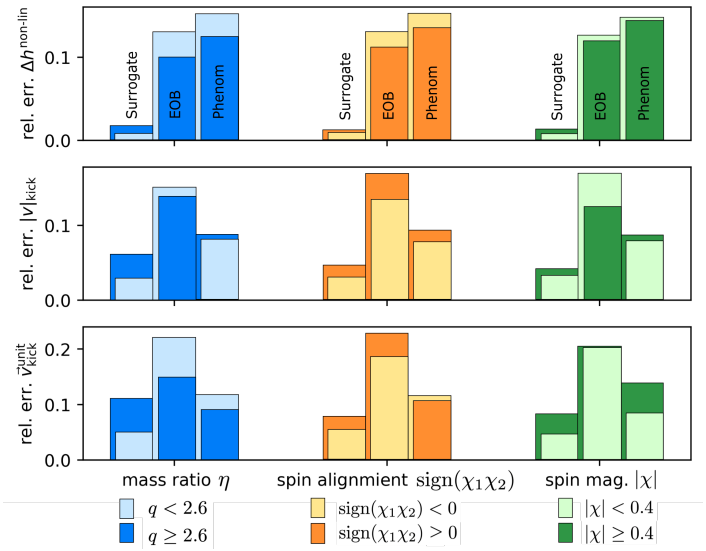

Figure 14. Relative errors for kick direction, magnitude, and non-linear memory with respect to NR waveforms averaged over all simulations selected by the specific marker (x-axis). For each marker and model, we display two error bars corresponding to the categorizations below the x-axis. The smaller errors are plotted in front of larger ones. To convert to the mass parameter η used in 4, we use $\eta = q/(1+q)^2$, where $q = M_1/M_2$ (M_1 being the heavier of the BBH system).

for *EOB* and *Phenom* visibly reduce due to fewer modes being included in the memory computation of NR, introducing fewer fluctuations with respect to these models. For *TEOB* however, this trend is not displayed since considerable variations are still present within the set of subdominant strain modes included in NR but not in *TEOB* or H_{sub}.

As anticipated by previous literature (e.g. [36], and cf. with findings of [33]), *EOB* and *Phenom* models do not reproduce the full information available within *NR*-simulations with their generated strain modes. For the harmonic strain modes generated by these model, however, the proximity of the non-linear memory estimates suggests a very good agreement with respect to the reference model(s). Comparing figures 12 and 13, it becomes apparent that the modes contained in H_{sub} roughly constitute the discrepancy in (memory) information content transmitted through the waveforms of *EOB* and *Phenom* with respect to NR (and *Surrogate*). To explicitly demonstrate this, we once again turn to the analysis of strain mode mixing in the memory components (5.8): according to equation (5.17), the total non-linear memory contains all α-coefficients with $\ell > 1$, where $\alpha_{2,0}$ dominates the sum of coefficients. This, again, is a consequence of the mode mixing present in the $\alpha_{\ell m}$'s. As we already elaborated, two dominant strain modes $h_{2,2}$, $h_{2,-2}$ can only mix for α-coefficients such as $\alpha_{\ell,0}$. However, the $h_{2,2}h_{2,2}$ and $h_{2,-2}h_{2,-2}$ terms cancel for all ℓ, except for $\alpha_{2,0}$ and $\alpha_{4,0}$ (for non-precessing BBH merger). For $\alpha_{\ell,m}$ with $\ell > 4$ such couplings are forbidden by virtue of the Wigner-3j symbols in (5.8). For $\alpha_{4,0}$ the prefactor in (5.8) is of order $\mathcal{O}(10^{-2})$, suppressing its contribution to the full non-linear part. Consequently, on average, $\alpha_{2,0}$ accounts for roughly 96% of the non-linear memory.[15]

[15]Physically, this results from the fact that the majority of energy flux away from the binary is transported by isotropic gravitational radiation. The azimuthally symmetric flux is sourced by $h_{2,\pm 2}$. Its manifestation in

JCAP02(2025)060

Naturally, the major contribution to $\alpha_{2,0}$ is due to the term including two dominant modes. However, there is a significant part of the $\alpha_{2,0}$ component as well as the contributions from all other α-coefficients that are due to subdominant modes coupling to $h_{2,\pm2}$. As many more α's enter the non-linear memory and for $\ell > 1$ more mixing of the strain modes is permitted by the Wigner-3j symbols, the impact of subdominant modes is non-negligible (in contrast to what was observed for the kick velocity) and grows with the number of strain modes included in the computation of the non-linear memory promoting it to a potent waveform quality metric sensitive to all harmonic modes provided by a model. The difference in the non-linear memory between figures 12 and 13 is, hence, largely due to the coupling of $h_{2,2}$ with subdominant modes in α's with $\ell > 1$, demonstrating their non-negligible effect. Switching off the (for *EOB* and *Phenom*) missing subdominant modes in NR and *Surrogate*, reduces the memory mainly to the $h_{2,\pm2}h_{2,\pm2}$ couplings. Due to the well-modeled dominant strain mode throughout approximants, the resulting non-linear memory contributions are roughly equal to *EOB* and *Phenom* and the mismatch is, thus, minimal, as indicated by figure 13. The dominance of the $h_{2,\pm2}h_{2,\pm2}$ in the memory computation is again undermined by the performance of *TEOB* displayed in figures 13 and 12. Despite *TEOB* containing only $h_{2,\pm2}$ its non-linear memory contributions reside in close proximity to the other approximants' values.

Prior to switching to the analysis of precessing simulations, we investigate the sensitivity of relative errors of physical properties of the merger to the parameters of non-precessing simulations, specifically the mass ratio η, the spin magnitudes $|\chi|$, and the spin alignments. The results are summarized in figure 14. Here, we display error bars for each considered physical quantity, separating simulations by mass ratio, spin-alignment, and spin magnitude. For each quantity and separation variable, the models display two error bars corresponding to the sections in which we split the parameter space. The separating value is selected to split the ensemble of simulations in roughly equal parts. For instance, for the mass ratio, we display error bars computed using BBH mergers only included in the high or low ratio regime, split at $q = 2.6$ (corresponding to $\eta \approx 0.2$). The *Surrogate* approximant demonstrates the most distinct trend across all markers, exhibiting significantly higher relative errors for mass ratios $q = M_1/M_2 \geq 2.6$ (see section 4; the conversion to η is given by $\eta = q/(1+q)^2$). A similar but less pronounced trend is observed for spin alignment and magnitude, where the trend suggests that *Surrogate* generally performs better for anti-aligned spins with lower magnitudes, $|\chi| < 0.4$, with regard to the kick velocity and non-linear memory.[16] In contrast, the *EOB* approximant shows a consistent trend only with the mass ratio, performing statistically better for $q \geq 2.6$. For other parameters, the trends are less consistent. A similar absence of tendencies is observed for the *Phenom* approximant.

6.2.3 Precessing merger simulations

Regarding the simulation of precessing mergers, similar results for the non-linear memory are obtained, as demonstrated in figure 15. In the top plot, the average relative errors for the full mode content are displayed as time series. The normalization and shading according to

terms of memory is carried by $h_{2,0}$.

[16] Linear memory is excluded from this analysis due to its (redundant) dependence on the remnant velocity, as discussed in section 5.

JCAP02(2025)060

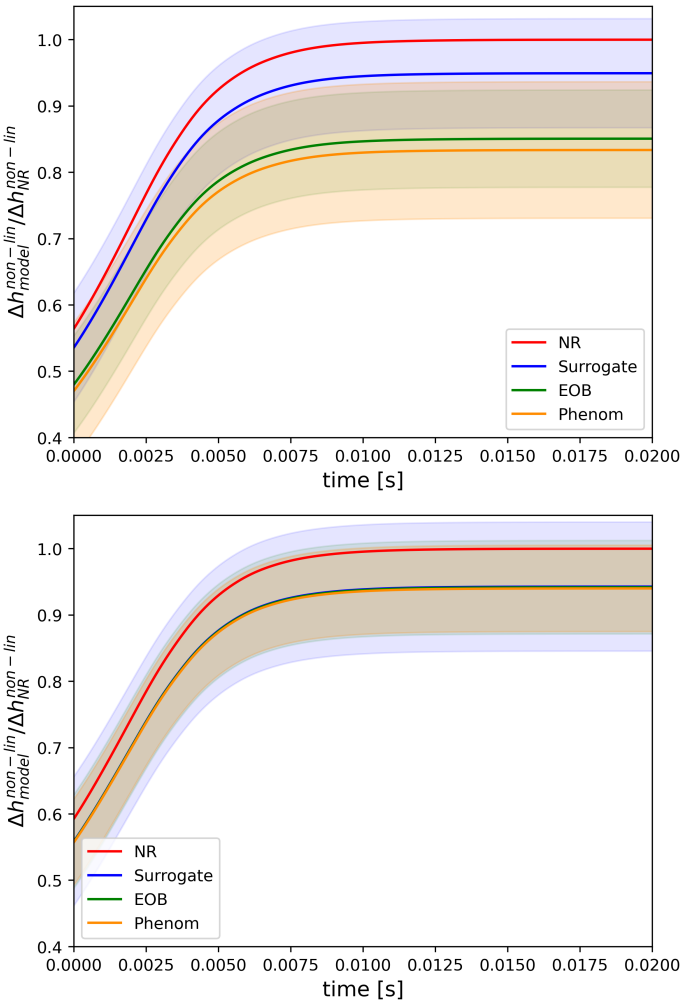

Figure 15. Non-linear memory as depicted in figure 12, here computed with (top) and without (bottom) high-mismatch modes, i.e., excluding $H_{\rm sub}$, for precessing mergers.

the standard deviation are computed as in figure 12. It is observed that *EOB* and *Phenom* exhibit similar deviations from NR as in the non-precessing case. Additionally, *Surrogate* demonstrates considerably larger fluctuations compared to non-precessing mergers manifesting in a large standard deviation. The latter persists when $H_{\rm sub}$ is excluded, as shown in the bottom plot of figure 15. Nonetheless, for the non-linear memory, it still performs better than *EOB* and *Phenom*. When the modes in $H_{\rm sub}$ are disabled, *Surrogate*, *EOB* and *Phenom* exhibit identical memory contributions. Notably here, *Surrogate* has the largest σ interval.

Although the non-linear memory contributions of the approximants closely resemble each other, we do not observe a similar convergence to the reference model as in figure 13. Instead, all models settle on average on 90% of NR's total non-linear memory. This indicates that for the precessing case, *Surrogate*, *EOB* and *Phenom* differ from each other mainly by modes included in $H_{\rm sub}$, even though *Surrogate*, as opposed to the other models, adequately simulates asymmetry between ℓ, m and $\ell, -m$ modes. However, to reproduce NR all approximants are potentially missing information beyond what is contained in $H_{\rm sub}$ as indicated by the

JCAP02(2025)060

gap between approximants and NR in the bottom plot of figure 15. In concordance with our previous discussions on alignment of precessing waveforms, it is crucial to point out that we cannot quantify the amount of deviation in the non-linear memory in figure 15 that is due to the chosen alignment strategy. An argument against alignment issues causing the larger deviation throughout approximants is given by the fact that all models seem to be equally "off". As each model is aligned individually, one would naively expect the alignment to impact the accuracy non-uniformly. Another explanation for the collective discrepancy with respect to *SXS*'s non-linear memory is given by the models vastly larger mode content (see table 1). However, we find that cutting *SXS*'s harmonic strain series at $\ell \leq 4$ in the statistics producing figure 15 does not significantly reduce the gap with respect to the alternative waveform models.

We further test if the results in figure 14 equally hold for precessing mergers. As before, we restrict this analysis to the non-linear memory and discard the spin alignment marker as it is meaningless for precessing waveforms. The boundaries of the remaining parameters, i.e., $q = 2.6$ (corresponding to $\eta = 0.2$) and $|\chi| = 0.4$, are adjusted so that the ensemble of simulations falling into each category are roughly equal, i.e., $q = 1.8$ (corresponding to $\eta = 0.23$) and $|\chi| = 0.45$. We find that for all approximants besides *Surrogate*, larger η (smaller q) yield higher relative errors in the approximants' memory contribution with respect to NR. For *Surrogate*, the opposite is the case. A similar trend for magnitude spins is found to be not significant. Comparing these findings with figure 14, it is evident that the parameter dependence of the models' performance is consistent across precessing and non-precessing mergers. Finally, note that the results regarding the parameter-dependent distribution of relative errors are insensitive to the mode content included in their computation: removing $H_{\rm sub}$ reduces the overall relative error amplitudes but does not obscure the trends exhibited in figure 14.

Leveraging the complete parameter space detailed in 3, we examine whether the conclusions concerning the memory are influenced by the choice of NR binary mergers. By considering the supplementary parameter space beyond the *SXS* data, we verify that the results outlined earlier remain consistent for both precessing and non-precessing waveforms. The relations between the memory contributions of *Surrogate*, *EOB*, and *Phenom* are validated for these instances, affirming that our observations are not affected by potential biases in the selection of source parameters.

7 Discussion

We conducted an extensive comparison of gravitational waveform models based on the calculation of the remnant's kick velocity and the gravitational memory for BBH mergers. Four state-of-the-art waveform models were considered (cf. table 1). We reviewed their numerical implementations' main features and discussed issues in the commonly used waveform alignment procedures, pointing out limitations of this analysis in the form of potential difficulties in the meaningful distinction between alignment residuals and intrinsic waveform dissimilarities. After applying a well-tested alignment strategy, we utilized the so-called balance laws to compute the kick velocity and the memory. In order to minimize a potential selection bias, we extended our analysis beyond the cataloged *SXS* BBH simulations and additionally considered

JCAP02(2025)060

binary mergers randomly picked from parameter space. We compared the computed kick velocities and memory terms with respect to reference models. In the case of cataloged simulations, we chose NR waveforms as a reference, while *Surrogate* was used for BBH mergers not contained in the *SXS* database.

Our results provide a novel perspective on the established inconsistencies between the selected approximants and the reference model regarding both the computed kick velocity and inferred memory, corroborating earlier findings in the literature [36]. Elucidating the origin of these discrepancies, it is observed that physical quantities computed via the balance flux laws are excellent probes of the dominant $h_{2,\pm2}$ mode and the distribution of information across the approximants' mode content, leading to the source of potential dissimilarities of the full waveforms. The former is effectively testable via the strain-induced remnant velocity (or linear memory as computed in this work), whereas the non-linear gravitational wave memory is sensitive to the latter.

Using the kick velocity as a metric, we showed that throughout approximants there is a substantial agreement with respect to the reference model, where *Surrogate* performs best, followed closely by *Phenom* and then *EOB*. This holds true for both the magnitude and the direction of the kick velocity. The ranking of approximants aligns with the mismatch values of the dominant strain modes computed post-alignment, stressing an $h_{2,2}$-dependence of the kick. Indeed, analytical considerations identified $h_{2,\pm1}$ and $h_{2,\pm2}$ as the primary kick-determining strain modes. However, most discrepancies in the kick are attributed to subdominant modes (i.e. $h_{2,\pm1}$ and so on), as evidenced by the remnant velocity error being approximately an order of magnitude larger than the $h_{2,2}$ mismatch across all approximants. It was demonstrated that the kick is highly alignment-dependent and, particularly for non-precessing binary mergers, sensitive to subtle features of the harmonic strain modes. For instance, a meaningful out-of-plane kick computation for precessing mergers demands an accurate simulation the asymmetry between ℓ, m and $\ell, -m$ modes inherent to precessing BBH waveforms. Dropping our working assumption of an optimal alignment, the approximants' relative errors regarding the remnant velocity may be partially due to alignment artifacts. Although we expect the residual miss-alignment to be subdominant, improvement in the alignment of GW waveforms, in particular for precessing merger simulations, should be addressed by future works, especially in the context of high-precision GW measurements.

The well-known deficiencies in the mode contents of *Phenom* and *EOB* were highlighted through the computation of the non-linear GW memory: as the non-linear memory accumulates many individual strain mode combinations (in form of the sum over $\alpha_{\ell,m}$'s, equation (5.8)), it reacts sensitively to any alteration of mode contend, demonstrated by NR and *Surrogate* approaching *Phenom* and *EOB*'s estimations after the removal of strain mode set $H_{\rm sub}$. Interestingly, for the non-linear memory, *EOB* generally yields a smaller error than *Phenom* but both are, on average, outperformed by *TEOB*. The latter demonstrates nicely that for non-linear memory estimations, an accurate modeling of solely the dominant harmonic strain mode is sufficient for practical computations. We further illustrate that the asymmetry in harmonic modes for precessing system plays only a minor role in the memory computation, via computing the non-linear memory for precessing waveforms under the exclusion of $H_{\rm sub}$.

JCAP02(2025)060

Overall, the combination of kick and memory synergize well and provide a wholesome waveform assessment that can be directly applied to any numerically accessible waveform model. For the tested models specifically, the analysis of the latter physical properties indicates that *Phenom* waveforms on average deliver a more reliable estimation of $h_{2,2}$ while the subdominant modes of *EOB* produce a more accurate non-linear memory estimate. Nonetheless, both approximants are surpassed by *Surrogate* waveform models which is, however, partially due to *Surrogate*'s larger set of non-trivial strain modes up to $\ell < 5$. Curiously, we observe that only *Surrogate* waveforms exhibit statistically significant trends regarding their performance for different regions in parameter space. Most notably, mass ratios $q < 2.6$ ($\eta > 0.2$) lead to significantly better agreement with NR data, suggesting a potential parameter bias of the interpolation procedure adapted by *Surrogate* models.

The findings of this work are supported by analytical considerations based on strain mode mixing appearing in the spherical harmonic decomposition of the memory and the kick, as described by the α-coefficients of equation (5.8). The selection rules that govern the appearance of specific mode mixing in the sum of (5.8) is determined by the fact that the strain is a spin-weight -2 function. The latter mathematical property leads to the emergence of the Wigner-3j symbols within the α-coefficients. The physical interpretation of each α term is implied by equation (5.5), expressing the relation between the physical quantities of the BBH merger, such as radiated energy or remnant mass, kick velocity, and gravitational memory. This framework explains how the azimuthal symmetry of the energy flux away from the binary correlates with the dominant memory contribution residing in the azimuthally independent $h_{2,0}$ mode.

Subtracting the dominant memory component (proportional to $\alpha_{2,0}$), our analytical considerations predict a strong subdominant mode dependence of the remaining non-linear memory. This prediction can potentially be used to evaluate the quality of individual subdominant harmonic modes for future GW waveform models. An instance of such practice was presented in section 6 with the comparision of non-linear memory contributions for different sets of harmonic modes. Physically, testing individual α-coefficients with $\ell \geq 2$ and $m \neq 0$ corresponds to analyzing the (subdominant) anisotropic non-linear memory, or equivalently, the anisotropic energy flux radiated to infinity from the binary system. Naturally, the anisotropic energy flux is highly BBH configuration-dependent and subdominant. Nevertheless, it potentially constitutes a powerful tool in high-precision tests of subdominant waveform contributions, especially for precessing BBH mergers. We leave a detailed investigation of this claim to future works.

Generally, the advantage of testing a waveform's quality by systematically accessing the physical properties rather than relying on a mode-by-mode mismatch analysis as depicted in figure 9, lies in the physically grounded nature of the former approach as well as the potential independence from alignment (for the non-linear memory, see the discussion in 3.2 and 5.4). While the mismatch function entails numerous uncertainties regarding the obscurity of intrinsic waveform differences, thoroughly discussed in 3.1, physical properties such as kick and memory possess fundamental physical significance and are computed without any approximations, i.e., in full, non-linear GR. Thus, in particular for precessing simulations for which a sufficiently accurate alignment strategy is yet to be defined, comparing the latter provides an appealing phenomenological waveform assessment.

JCAP02(2025)060

In conclusion, this analysis provides a guiding framework for future waveform model assessments, facilitating the development of more precise GW waveforms for systematic application to measurement data from future instruments. Additionally, by employing the balance flux laws, we identify distinct strengths and weaknesses of individual models concerning the computed physical quantities and their regions of applicability in parameter space. This information can serve as a selection guide for waveform applications, such as GW memory or metadata estimations. Applying the developed methodology to updated versions of the models remains a task for future work.

Acknowledgments

LH would like to acknowledge financial support from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme grant agreement No 801781. LH further acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2181/1 — 390900948 (the Heidelberg STRUCTURES Excellence Cluster). The authors thank the Heidelberg STRUCTURES Excellence Cluster for financial support. The authors acknowledge support by the state of Baden-Württemberg, Germany, through bwHPC. The authors thank Cecilio Garcia-Quiros, Keefe Mitman and Nils Deppe for fruitful discussions.

A Numerical artifacts

Numerical artifacts can significantly influence the outcome of the mismatch function discussed in section 3. The morphology with which artefacts can appear varies. One of the most prominent examples is displayed in figure 16. Here, the waveform mode carries a tail, which shall not be confused with the gravitational memory. The latter is not contained in the waveforms displayed. Instead, the non-trivial extent of the waveform after the ring down results from numerical artifacts and is to be disregarded as non-physical. Since the effect is rather small and effects only subdominant modes, we ignored it in the main analysis. In future works tackling the precision of higher strain modes, this issue has to be addressed.

JCAP02(2025)060

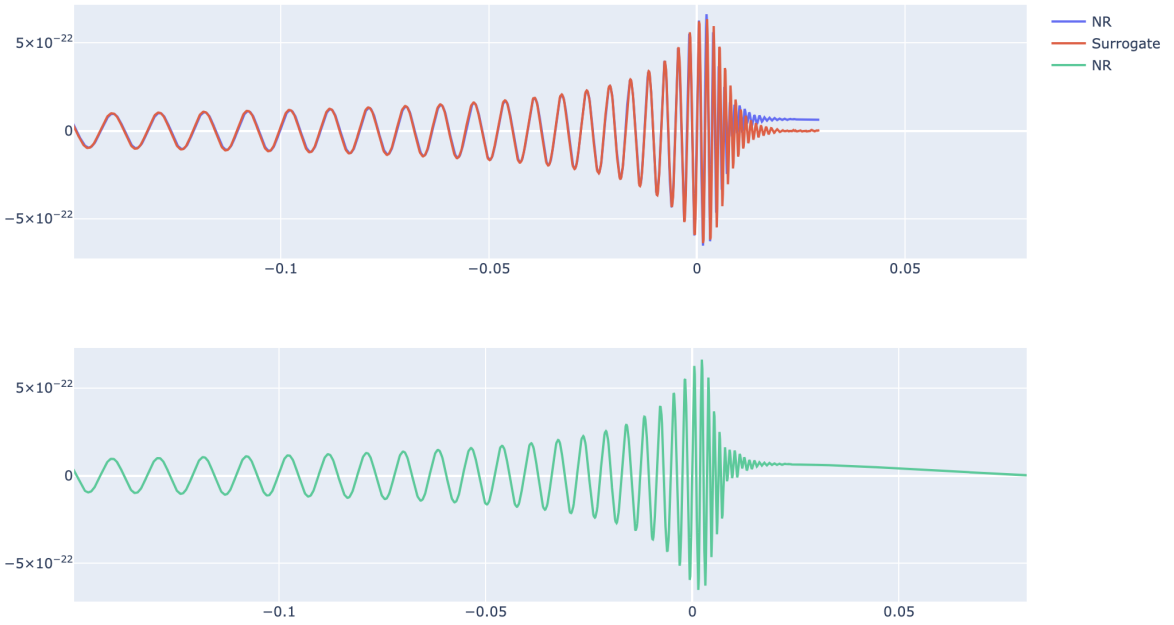

Figure 16. $(4, 4)$ waveform mode for a the non-precessing merger simulation SXS:BBH:0155. Displayed are the truncated and aligned waveforms (above) for NR and *Surrogate* as well as the "raw" full waveform for NR.

References

[1] LIGO Scientific and Virgo collaborations, *Observation of gravitational waves from a binary black hole merger*, *Phys. Rev. Lett.* **116** (2016) 061102 [arXiv:1602.03837] [INSPIRE].

[2] M. Colpi et al., *LISA definition study report*, arXiv:2402.07571 [INSPIRE].

[3] ET collaboration, *Science case for the Einstein Telescope*, *JCAP* **03** (2020) 050 [arXiv:1912.02622] [INSPIRE].

[4] M. Evans et al., *Cosmic explorer: a submission to the NSF MPSAC ngGW subcommittee*, arXiv:2306.13745 [INSPIRE].

[5] Y.B. Zel'dovich and A.G. Polnarev, *Radiation of gravitational waves by a cluster of superdense stars*, *Sov. Astron.* **18** (1974) 17 [INSPIRE].

[6] D. Christodoulou, *Nonlinear nature of gravitation and gravitational wave experiments*, *Phys. Rev. Lett.* **67** (1991) 1486 [INSPIRE].

[7] K.S. Thorne, *Gravitational-wave bursts with memory: the Christodoulou effect*, *Phys. Rev. D* **45** (1992) 520 [INSPIRE].

[8] A.G. Wiseman and C.M. Will, *Christodoulou's nonlinear gravitational wave memory: evaluation in the quadrupole approximation*, *Phys. Rev. D* **44** (1991) R2945 [INSPIRE].

[9] M. Favata, *Post-Newtonian corrections to the gravitational-wave memory for quasi-circular, inspiralling compact binaries*, *Phys. Rev. D* **80** (2009) 024002 [arXiv:0812.0069] [INSPIRE].

[10] M. Favata, *Nonlinear gravitational-wave memory from binary black hole mergers*, *Astrophys. J. Lett.* **696** (2009) L159 [arXiv:0902.3660] [INSPIRE].

JCAP02(2025)060

[11] H. Inchauspé et al., *Measuring gravitational wave memory with LISA*, arXiv:2406.09228 [INSPIRE].

[12] L. Heisenberg, *A systematic approach to generalisations of general relativity and their cosmological implications*, *Phys. Rept.* **796** (2019) 1 [arXiv:1807.01725] [INSPIRE].

[13] LISA collaboration, *Astrophysics with the Laser Interferometer Space Antenna*, *Living Rev. Rel.* **26** (2023) 2 [arXiv:2203.06016] [INSPIRE].

[14] LIGO Scientific collaboration, *Advanced LIGO*, *Class. Quant. Grav.* **32** (2015) 074001 [arXiv:1411.4547] [INSPIRE].

[15] VIRGO collaboration, *Advanced Virgo: a second-generation interferometric gravitational wave detector*, *Class. Quant. Grav.* **32** (2015) 024001 [arXiv:1408.3978] [INSPIRE].

[16] J. Blackman et al., *Numerical relativity waveform surrogate model for generically precessing binary black hole mergers*, *Phys. Rev. D* **96** (2017) 024058 [arXiv:1705.07089] [INSPIRE].

[17] J. Blackman et al., *A surrogate model of gravitational waveforms from numerical relativity simulations of precessing binary black hole mergers*, *Phys. Rev. D* **95** (2017) 104023 [arXiv:1701.00550] [INSPIRE].

[18] V. Varma et al., *Surrogate models for precessing binary black hole simulations with unequal masses*, *Phys. Rev. Research.* **1** (2019) 033015 [arXiv:1905.09300] [INSPIRE].

[19] S. Husa et al., *Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal*, *Phys. Rev. D* **93** (2016) 044006 [arXiv:1508.07250] [INSPIRE].

[20] S. Khan et al., *Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era*, *Phys. Rev. D* **93** (2016) 044007 [arXiv:1508.07253] [INSPIRE].

[21] G. Pratten et al., *Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes*, *Phys. Rev. D* **103** (2021) 104056 [arXiv:2004.06503] [INSPIRE].

[22] H. Estellés et al., *New twists in compact binary waveform modeling: a fast time-domain model for precession*, *Phys. Rev. D* **105** (2022) 084040 [arXiv:2105.05872] [INSPIRE].

[23] A. Taracchini et al., *Prototype effective-one-body model for nonprecessing spinning inspiral-merger-ringdown waveforms*, *Phys. Rev. D* **86** (2012) 024011 [arXiv:1202.0790] [INSPIRE].

[24] B.D. Lackey et al., *Extracting equation of state parameters from black hole-neutron star mergers: aligned-spin black holes and a preliminary waveform model*, *Phys. Rev. D* **89** (2014) 043009 [arXiv:1303.6298] [INSPIRE].

[25] Y. Pan et al., *Inspiral-merger-ringdown waveforms of spinning, precessing black-hole binaries in the effective-one-body formalism*, *Phys. Rev. D* **89** (2014) 084006 [arXiv:1307.6232] [INSPIRE].

[26] A. Bohé et al., *Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors*, *Phys. Rev. D* **95** (2017) 044028 [arXiv:1611.03703] [INSPIRE].

[27] S. Ossokine et al., *Multipolar effective-one-body waveforms for precessing binary black holes: construction and validation*, *Phys. Rev. D* **102** (2020) 044055 [arXiv:2004.09442] [INSPIRE].

JCAP02(2025)060

[28] A. Ramos-Buades et al., *Next generation of accurate and efficient multipolar precessing-spin effective-one-body waveforms for binary black holes*, *Phys. Rev. D* **108** (2023) 124037 [arXiv:2303.18046] [INSPIRE].

[29] A. Nagar et al., *Time-domain effective-one-body gravitational waveforms for coalescing compact binaries with nonprecessing spins, tides and self-spin effects*, *Phys. Rev. D* **98** (2018) 104052 [arXiv:1806.01772] [INSPIRE].

[30] R. Gamba, S. Akçay, S. Bernuzzi and J. Williams, *Effective-one-body waveforms for precessing coalescing compact binaries with post-Newtonian twist*, *Phys. Rev. D* **106** (2022) 024020 [arXiv:2111.03675] [INSPIRE].

[31] A. Nagar et al., *Analytic systematics in next generation of effective-one-body gravitational waveform models for future observations*, *Phys. Rev. D* **108** (2023) 124018 [arXiv:2304.09662] [INSPIRE].

[32] A. Ashtekar, T. De Lorenzo and N. Khera, *Compact binary coalescences: constraints on waveforms*, *Gen. Rel. Grav.* **52** (2020) 107 [arXiv:1906.00913] [INSPIRE].

[33] N. Khera, B. Krishnan, A. Ashtekar and T. De Lorenzo, *Inferring the gravitational wave memory for binary coalescence events*, *Phys. Rev. D* **103** (2021) 044012 [arXiv:2009.06351] [INSPIRE].

[34] L. Heisenberg, *Balance laws as test of gravitational waveforms*, *Phil. Trans. Roy. Soc. Lond. A* **382** (2023) 20230086 [arXiv:2309.12505] [INSPIRE].

[35] A. Borchers and F. Ohme, *Black-hole kicks: a tool to measure the accuracy of gravitational-wave models*, in the proceedings of the *55th Rencontres de Moriond on Gravitation*, (2021) [arXiv:2106.02414] [INSPIRE].

[36] A. Borchers and F. Ohme, *Inconsistent black hole kick estimates from gravitational-wave models*, *Class. Quant. Grav.* **40** (2023) 095008 [arXiv:2207.13531] [INSPIRE].

[37] K. Mitman et al., *Adding gravitational memory to waveform catalogs using BMS balance laws*, *Phys. Rev. D* **103** (2021) 024031 [arXiv:2011.01309] [INSPIRE].

[38] LISA Consortium Waveform Working Group collaboration, *Waveform modelling for the Laser Interferometer Space Antenna*, arXiv:2311.01300 [INSPIRE].

[39] LIGO Scientific collaboration, *LALSuite: LIGO scientific collaboration algorithm library suite*, Astrophysics Source Code Library ascl:2012.021 (2020).

[40] P. Schmidt, I.W. Harry and H.P. Pfeiffer, *Numerical relativity injection infrastructure*, arXiv:1703.01076 [INSPIRE].

[41] M. Boyle et al., *The SXS collaboration catalog of binary black hole simulations*, *Class. Quant. Grav.* **36** (2019) 195006 [arXiv:1904.04831] [INSPIRE].

[42] L.E. Kidder et al., *Black hole evolution by spectral methods*, *Phys. Rev. D* **62** (2000) 084032 [gr-qc/0005056] [INSPIRE].

[43] M. Hannam et al., *Simple model of complete precessing black-hole-binary gravitational waveforms*, *Phys. Rev. Lett.* **113** (2014) 151101 [arXiv:1308.3271] [INSPIRE].

[44] T. Damour and A. Nagar, *New effective-one-body description of coalescing nonprecessing spinning black-hole binaries*, *Phys. Rev. D* **90** (2014) 044018 [arXiv:1406.6913] [INSPIRE].

[45] A. Buonanno, Y. Chen and T. Damour, *Transition from inspiral to plunge in precessing binaries of spinning black holes*, *Phys. Rev. D* **74** (2006) 104005 [gr-qc/0508067] [INSPIRE].

JCAP02(2025)060

[46] H. Estellés et al., *Time-domain phenomenological model of gravitational-wave subdominant harmonics for quasicircular nonprecessing binary black hole coalescences*, *Phys. Rev. D* **105** (2022) 084039 [arXiv:2012.11923] [INSPIRE].

[47] N.W. Taylor et al., *Comparing gravitational waveform extrapolation to Cauchy-characteristic extraction in binary black hole simulations*, *Phys. Rev. D* **88** (2013) 124010 [arXiv:1309.3605] [INSPIRE].

[48] K. Mitman et al., *Computation of displacement and spin gravitational memory in numerical relativity*, *Phys. Rev. D* **102** (2020) 104007 [arXiv:2007.11562] [INSPIRE].

[49] M. Boyle, D. Iozzo and L.C. Stein, *moble/scri: v1.2*, Zenodo, September 2020.

[50] S. Ghosh, P. Kolitsidou and M. Hannam, *First frequency-domain phenomenological model of the multipole asymmetry in gravitational-wave signals from binary-black-hole coalescence*, *Phys. Rev. D* **109** (2024) 024061 [arXiv:2310.16980] [INSPIRE].

[51] J.E. Thompson et al., *PhenomXO4a: a phenomenological gravitational-wave model for precessing black-hole binaries with higher multipoles and asymmetries*, *Phys. Rev. D* **109** (2024) 063012 [arXiv:2312.10025] [INSPIRE].

[52] F. D'Ambrosio et al., *Gravitational waves in full, non-linear general relativity*, arXiv:2201.11634 [INSPIRE].

[53] H. Bondi, M.G.J. van der Burg and A.W.K. Metzner, *Gravitational waves in general relativity. 7. Waves from axisymmetric isolated systems*, *Proc. Roy. Soc. Lond. A* **269** (1962) 21 [INSPIRE].

[54] R.K. Sachs, *Gravitational waves in general relativity. 8. Waves in asymptotically flat space-times*, *Proc. Roy. Soc. Lond. A* **270** (1962) 103 [INSPIRE].

[55] R. Penrose, *Asymptotic properties of fields and space-times*, *Phys. Rev. Lett.* **10** (1963) 66 [INSPIRE].

[56] R. Geroch, *Asymptotic structure of space-time*, in the proceedings of the *Symposium on asymptotic structure of space-time*, Springer, U.S.A. (1977), p. 1 [DOI:10.1007/978-1-4684-2343-3_1] [INSPIRE].

[57] A. Ashtekar, *Geometry and physics of null infinity*, *Surveys Diff. Geom.* **20** (2015) 99 [arXiv:1409.1800] [INSPIRE].

[58] A. Ashtekar and M. Streubel, *Symplectic geometry of radiative modes and conserved quantities at null infinity*, *Proc. Roy. Soc. Lond. A* **376** (1981) 585 [INSPIRE].

[59] R.M. Wald and A. Zoupas, *A general definition of 'conserved quantities' in general relativity and other theories of gravity*, *Phys. Rev. D* **61** (2000) 084027 [gr-qc/9911095] [INSPIRE].

[60] V.B. Braginsky and K.S. Thorne, *Gravitational-wave bursts with memory and experimental prospects*, *Nature* **327** (1987) 123 [INSPIRE].

[61] V. Varma et al., *Surrogate model of hybridized numerical relativity binary black hole waveforms*, *Phys. Rev. D* **99** (2019) 064045 [arXiv:1812.07865] [INSPIRE].

[62] V. Varma et al., *Surrogate models for precessing binary black hole simulations with unequal masses*, *Phys. Rev. Research.* **1** (2019) 033015 [arXiv:1905.09300] [INSPIRE].

[63] T. Damour, *The general relativistic two body problem and the effective one body formalism*, *Fundam. Theor. Phys.* **177** (2014) 111 [arXiv:1212.3169] [INSPIRE].

[64] K. Mitman et al., *Fixing the BMS frame of numerical relativity waveforms with BMS charges*, *Phys. Rev. D* **106** (2022) 084029 [arXiv:2208.04356] [INSPIRE].

JCAP02(2025)060

[65] K. Mitman et al., *A review of gravitational memory and BMS frame fixing in numerical relativity*, *Class. Quant. Grav.* **41** (2024) 223001 [arXiv:2405.08868] [INSPIRE].

[66] C. Pitte et al., *Detectability of higher harmonics with LISA*, *Phys. Rev. D* **108** (2023) 044053 [arXiv:2304.03142] [INSPIRE].

[67] J. Moxon, M.A. Scheel and S.A. Teukolsky, *Improved Cauchy-characteristic evolution system for high-precision numerical relativity waveforms*, *Phys. Rev. D* **102** (2020) 044052 [arXiv:2007.01339] [INSPIRE].

[68] J. Yoo et al., *Numerical relativity surrogate model with memory effects and post-Newtonian hybridization*, *Phys. Rev. D* **108** (2023) 064027 [arXiv:2306.03148] [INSPIRE].

JCAP02(2025)060