Note

Length dependence of waveform mismatch: a caveat on waveform accuracy

Keefe Mitman[1,*] ⓘ **, Leo C Stein**[2] ⓘ **, Michael Boyle**[1] ⓘ **,**
Nils Deppe[1,3,4] ⓘ **, Lawrence E Kidder**[1] ⓘ **, Harald P Pfeiffer**[5] ⓘ
and Mark A Scheel[6] ⓘ

[1] Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, NY 14853, United States of America
[2] Department of Physics and Astronomy, University of Mississippi, University, Oxford, MS 38677, United States of America
[3] Department of Physics, Cornell University, Ithaca, NY 14853, United States of America
[4] Laboratory for Elementary Particle Physics, Cornell University, Ithaca, NY 14853, United States of America
[5] Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-14467 Potsdam, Germany
[6] Theoretical Astrophysics 350-17, California Institute of Technology, Pasadena, CA 91125, United States of America

E-mail: kem343@cornell.edu

CrossMark

Abstract
The Simulating eXtreme Spacetimes Collaboration's code SpEC can now routinely simulate binary black hole mergers undergoing ~ 25 orbits, with the longest simulations undergoing nearly ~ 180 orbits. While this sounds impressive, the mismatch between the highest resolutions for this long simulation is $\mathcal{O}(10^{-1})$. Meanwhile, the mismatch between resolutions for the more typical simulations tends to be $\mathcal{O}(10^{-4})$, despite the resolutions being similar to the long simulations'. In this note, we explain why mismatch alone gives an incomplete picture of code—and waveform—quality, especially in the context of providing waveform templates for LISA and 3G detectors, which require

---

* Author to whom any correspondence should be addressed.

templates with $\mathcal{O}(10^3) - \mathcal{O}(10^5)$ orbits. We argue that to ready the GW community for the sensitivity of future detectors, numerical relativity groups must be aware of this caveat, and also run future simulations with at least three resolutions to properly assess waveform accuracy.

Keywords: gravitational wave, black hole, binary black hole, numerical relativity, mismatch

1. Introduction

Gravitational wave (GW) science, whether it be analysis of observed data, or quantifying waveform models' predictions, requires an inner product on the space of waveforms. The standard choice is a weighted $L^2(\mathbb{R})$ inner product[7] called the *overlap*, where for two waveforms h_1 and h_2, the overlap is

$$\mathcal{O}(h_1, h_2) = \langle h_1, h_2 \rangle \equiv 4 \text{Re} \int_0^\infty \frac{\tilde{h}_1(f)^* \, \tilde{h}_2(f)}{S_h(f)} \mathrm{d}f, \tag{1}$$

where tildes represent the Fourier transform, * represents complex conjugation, and $S_h(f)$ is the power spectral density of the strain noise for some GW detector [2–5]. This is a natural choice for an inner product, appearing in the Gaussian log-likelihood, and giving the optimal[8] signal-to-noise ratio (SNR) for a waveform $h(t)$,

$$\rho^2 = ||h||^2 = \langle h, h \rangle = 4 \int_0^\infty \frac{|\tilde{h}(f)|^2}{S_h(f)} \mathrm{d}f. \tag{2}$$

When comparing waveforms, however, the overlap is often replaced by the *mismatch*

$$\mathcal{M}(h_1, h_2) = 1 - \frac{\langle h_1, h_2 \rangle}{\sqrt{\langle h_1, h_1 \rangle \langle h_2, h_2 \rangle}}. \tag{3}$$

Note that this is often defined with an implicit optimization over time and phase, which we do not explicitly include (see appendix for more discussion). Whereas overlap measures the *similarity* between two waveforms, mismatch measures the *dissimilarity* of two waveforms in a scale-invariant way [7]. Mismatch is a useful measure because by inserting equation (2) into equation (3) and expanding to leading order in $\delta h \equiv h_1 - h_2$ with $||\delta h|| \ll ||h_{1,2}||$, one finds

$$\mathcal{M}(h_1, h_2) \simeq \frac{\langle \delta h_\perp, \delta h_\perp \rangle}{2\rho^2} \leqslant \frac{\langle \delta h, \delta h \rangle}{2\rho^2} \tag{4}$$

where

$$\delta h_\perp \equiv \delta h - \frac{h_1}{||h_1||} \left\langle \frac{h_1}{||h_1||}, \delta h \right\rangle. \tag{5}$$

---

[7] Strain waveforms with memory effects are infrared (IR) divergent and hence not square-integrable with the standard L^2 norm, i.e. with $S_h = 1$. With detectors, this is not a problem since $1/S_h \to 0$ as $f \to 0$. Moreover, since the overlap is computed with some cutoff $f_{\min} > 0$, this is not an issue. Alternatively, one can instead use the news (the first time derivative of the strain) which is IR-finite. See [1] for improved methods of Fourier transforming waveforms with memory.

[8] For the matched filtering SNR, see, e.g. equations (109) and (110) in [6].

Consequently, one can use the mismatch as a kind of distinguishability criterion when comparing two waveforms or two models' predictions. That is, two waveforms cannot be distinguished if (see appendix G of [8] or [9])

$$\mathcal{M} \lesssim \frac{D}{2\rho^2}, \tag{6}$$

where D is the number of dimensions that your two waveform models depend on, e.g. 9 for binary black hole (BBH) mergers[9]. This criterion can also be used for testing Einstein's equations, in which case one should study the mismatch between a beyond-GR model and a set of approximate GR solutions. It is crucial that the approximate solutions' errors (e.g. analytical or numerical truncation errors) are also smaller than this criterion, in the sense explained in section 2.

The mismatch is clearly foundational in GW science. However, it comes with many caveats—see appendix for a brief discussion. In this paper, we highlight one issue: how mismatch depends on the length of a numerical relativity (NR) waveform.

2. Results

One area of GW science where mismatch is used extensively is NR, where the mismatch is used to determine if an NR simulation is accurate enough to be used to analyze GW observations. To quantify our truncation error, we run multiple NR simulations of the same system at two or more resolutions, resulting in two waveforms, $h_{\text{res. }1,2}$, and then compute $\mathcal{M}(h_{\text{res. }1}, h_{\text{res. }2})$. How small must this be? As an example, suppose a detector is expected to observe BBH mergers with SNRs of $\rho = 10$; then equation (6), with $D = 9$, says the truncation error is sufficiently small when

$$\mathcal{M}\left(h_{\text{res.1}}, h_{\text{res.2}}\right) \lesssim 4.5 \times 10^{-2}. \tag{7}$$

Consequently, it is often unfortunately assumed in the NR community that if a simulation can be run that achieves this mismatch, then the NR code is ready to produce waveforms for the detector in question. In reality, an NR code must achieve this mismatch *over the detector's entire frequency range*[10]. Furthermore, as we will show now, even if one can run an NR simulation that achieves a target mismatch, this *does not* imply that the same mismatch can be achieved for a *longer* simulation.

We consider the regime where our simulations are converging well enough that we can treat the amplitude and phase errors as small. We write the waveform of one resolution in terms of another,

$$h_{\text{res.2}}\left(t\right) = \left(1 + \frac{\delta A}{A}\left(t\right)\right) e^{i\delta\phi(t)} h_{\text{res.1}}\left(t\right). \tag{8}$$

This parameterization is completely general, but particularly useful when the waveforms are similar. Here the phase error $\delta\phi(t)$ as well as the fractional amplitude error $(\delta A/A)(t)$ are both real functions. Amplitude errors tend to be bounded, while phase errors are secular and

---

[9] These 9 parameters are the mass ratio, two three-dimensional spin vectors, and the eccentricity and mean anomaly. The extrinsic parameters are not included in D, as they are model-independent.

[10] There is a slight caveat to this that we explain at the end of this section.

steadily accumulate as simulations run for a long time[11]. Thus, phase errors dominate for sufficiently long binary simulations. Now we estimate the time-domain overlap between these two resolutions, using a flat noise curve $S_h = 1$, over the time interval $[t_1, t_2]$. While the overlap in equation (1) is defined in the frequency domain, we stress that because of Parseval's theorem, performing this integral in the time domain is equally valid and we do so to avoid tapering complexities that are necessitated by Fourier transforms. This overlap, distinguished from its frequency-domain counterpart by a subscript t, is

$$\mathcal{O}_t\left(h_{\text{res.1}}, h_{\text{res.2}}\right) = 4\text{Re} \int_{t_1}^{t_2} \left(1 + \frac{\delta A}{A}\right) e^{i\delta\phi} |h_{\text{res.1}}|^2 \, dt. \tag{9}$$

As the phase error is supposed to be notably small, we Taylor expand the exponential. Keeping in mind that we take the real part of the integral, and everything else in the integrand is real, we must go to second order,

$$\mathcal{O}_t\left(h_{\text{res.1}}, h_{\text{res.2}}\right) \approx 4\text{Re} \int_{t_1}^{t_2} \left[1 + \frac{\delta A}{A}(t) + i\delta\phi(t) - \frac{1}{2}\left(\delta\phi(t)\right)^2\right] |h_{\text{res.1}}|^2 \, dt. \tag{10}$$

For sufficiently long waveforms, phase errors dominate over amplitude errors, and keeping only the real part yields,

$$\mathcal{O}_t\left(h_{\text{res.1}}, h_{\text{res.2}}\right) \approx 4 \int_{t_1}^{t_2} \left[1 - \frac{1}{2}\left(\delta\phi(t)\right)^2\right] |h_{\text{res.1}}|^2 \, dt, \tag{11}$$

$$= \mathcal{O}_t\left(h_{\text{res.1}}, h_{\text{res.1}}\right) - 2 \int_{t_1}^{t_2} \left(\delta\phi(t)\right)^2 |h_{\text{res.1}}|^2 \, dt. \tag{12}$$

Converting this to a mismatch, we find

$$\mathcal{M}_t\left(h_{\text{res.1}}, h_{\text{res.2}}\right) \approx \frac{2}{\rho^2} \int_{t_1}^{t_2} \left(\delta\phi(t)\right)^2 |h_{\text{res.1}}(t)|^2 \, dt, \tag{13}$$

where $\rho^2 = \mathcal{O}_t\left(h_{\text{res.1}}, h_{\text{res.1}}\right) = 4\int_{t_1}^{t_2} |h_{\text{res.1}}(t)|^2 \, dt$. The integrals in the numerator and denominator both grow with the length of the interval $|t_2 - t_1|$. However, the numerator has the extra factor of $(\delta\phi(t))^2$, making it grow faster. As a simple estimate, suppose we take the phase error to be a linear function of time, $\delta\phi(t) = (t - t_a)(d\,\delta\phi/dt)|_{t_a}$, where $t_a \in [t_1, t_2]$ is some alignment time. Now our mismatch is approximated as

$$\mathcal{M}_t\left(h_{\text{res.1}}, h_{\text{res.2}}\right) \approx \frac{2}{\rho^2} \left(\frac{d\,\delta\phi}{dt}\bigg|_{t_a}\right)^2 \int_{t_1}^{t_2} (t - t_a)^2 |h_{\text{res.1}}(t)|^2 \, dt$$

$$\propto \left(\frac{d\,\delta\phi}{dt}\bigg|_{t_a}\right)^2 |t_2 - t_1|^2, \tag{14}$$

where we kept the asymptotic scaling as $|t_2 - t_1|$ grows very large. The integral can be performed asymptotically with, e.g. a post-Newtonian (PN) waveform. This shows that, assuming the main error in NR simulations is a small phase error, the *mismatch will tend to increase with the square of the length of the simulation.*

---

[11] In the study of orbital dynamics it is typical to use symplectic integration methods that are able to drastically reduce such errors (often $\mathcal{O}(\Delta t^2)$ smaller than the integration method) because they preserve the Hamiltonian dynamics exactly [10–15]. It is not currently known if or how to do this for NR simulations of BBH mergers.
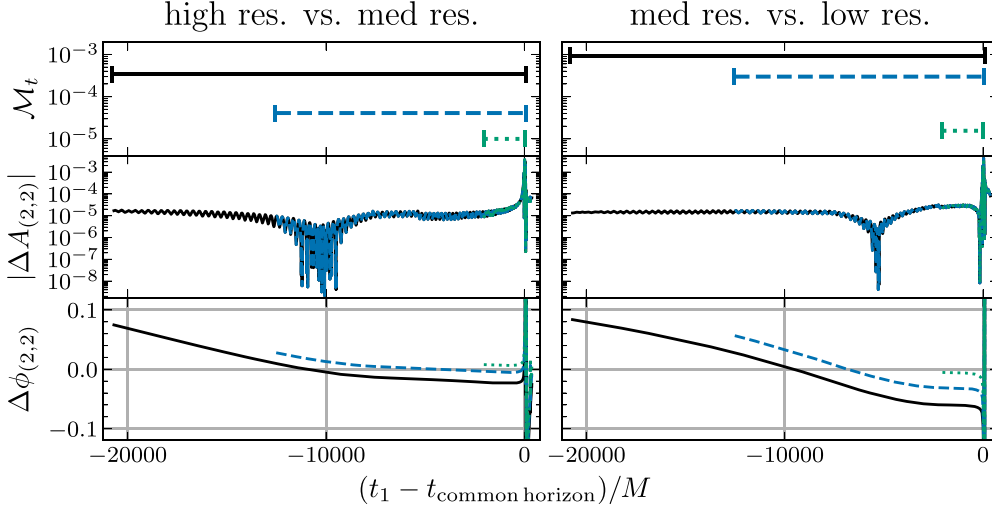
Figure 1. Comparison of waveforms from different resolutions for SXS:BBH:1132: a quasi-circular, equal-mass, and non-spinning simulation from the SXS catalog. The left panels are for the two highest resolutions, while the right panels are for the next two highest resolutions. The top panels show the mismatch between the waveforms when truncated at time t_1, while the middle and bottom panels show the $(2, 2)$ mode's amplitude and phase error. The full simulation (black) has length $\sim 25\,000\,M$, and we truncate to lengths of $\sim 15\,000\,M$ (blue, dashed) and $\sim 2500\,M$ (green, dotted) to demonstrate the effects of simulation length. Waveforms between different resolutions are aligned to each other by optimizing over a time translation and an $SO(3)$ rotation.

In figure 1 we first show an example illustrating that the dominant resolution error in NR simulations tends to be the binary phase error. In figure 1 we show the $(2, 2)$ mode amplitude and phase errors between the two highest resolutions for the Simulating eXtreme Spacetimes (SXS) Collaboration's simulation SXS:BBH:1132 [16]. We also show how these errors change as one changes the effective length of the simulation, by removing inspiral data from each resolution's waveform. While the amplitude error for each of the two resolution pairs is roughly consistent and below the square root of the mismatch[12], the phase error can be notably large. Meanwhile, in the top panel of figure 1, we show the mismatch between the two resolutions' waveforms. Here, and in subsequent plots, the mismatch is computed by integrating over the two-sphere rather than by evaluating the strain waveforms at some arbitrary point. As one decreases the effective length of the simulation, the mismatch correspondingly decreases, even though the amplitude and phase errors are similar to the comparison in which the numerical waveforms are longer. Note that here, and in subsequent figures, mismatch is computed by integrating the strain over the two-sphere.

In figure 2 we further illustrate this point regarding the length dependence of the mismatch. Each point in this figure represents taking waveforms from two resolutions, truncating them so that their length prior to the simulation's common horizon time is some t_1, minimizing the L^2 norm of their residual over the two-sphere by optimizing over a time translation and an $SO(3)$ rotation[13], and then computing a mismatch between t_1 and some fixed time in the

---

[12] Note that mismatch goes as the residual squared; see equation (4).

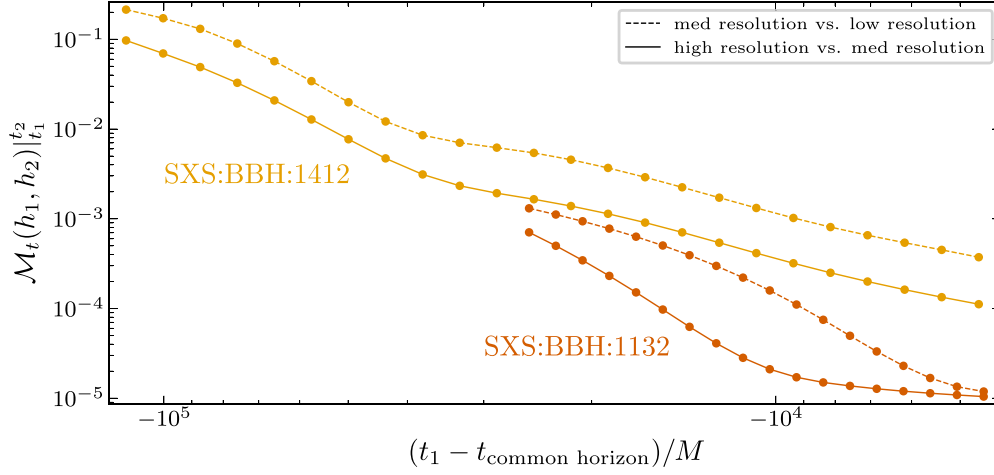[13] For an explanation of why this is done, see appendix.

Figure 2. Mismatch between waveforms from different resolutions for two simulations from the SXS catalog as a function of the length of the waveform. Solid lines correspond to the mismatch between the two highest resolutions, while dashed lines correspond to the mismatch between the next two highest resolutions. Waveforms between different resolutions and for different trunctations are aligned to each other by optimizing over a time translation and an $SO(3)$ rotation.

late ringdown phase relative to $t_{\text{common horizon}}$. This optimization is performed numerically using `scipy`'s `minimize` function via the `sxs` package [17, 18]. Both simulations are quasi-circular, non-precessing systems. As can be seen, by making each simulation 'shorter', one can decrease the mismatch by many orders of magnitude. Equation (14) predicts that on a log-log plot like figure 2, we would see a slope of roughly -2, and indeed using `scipy`'s `curve_fit` we find slopes in good agreement with this prediction.

At this point, one may worry that the results of figure 2 as well as the $\mathcal{M} \leqslant 10^{-6}$ mismatch requirement of LISA (i.e. an SNR of $\rho \sim 1000$ in equation (6)) seem to suggest that current NR codes are no where near accurate enough to simulate the $\mathcal{O}(10^3)$ orbit binaries that future detectors will observe. We stress, however, that this concern is partially resolved by utilizing PN hybridizations [19]. Specifically, provided one can run a simulation long enough such that, over its length it is accurate to $\mathcal{M} \leqslant 10^{-6}$ and over a comparison window the NR/PN hybrid is accurate to $\mathcal{M} \leqslant 10^{-6}$ and PN itself before the window is accurate to $\mathcal{M} \leqslant 10^{-6}$[14], then one need not actually simulate $\mathcal{O}(10^3)$ orbits. Thus far some work has been done to study how feasible this is. In particular, [19] has shown that many of the simulations produced by the SXS collaboration can obtain hybridization mismatches $\lesssim \mathcal{O}(10^{-6})$ for spin-aligned systems and $\lesssim O(10^{-2})$ for precessing systems, with the primary limitations being residual eccentricity and the lack of higher-order PN spin terms. Thus, with more accurate eccentricity control and higher-order PN terms, and perhaps even aid from the effective-one-body framework [20], it

---

[14] To measure this, one unfortunately needs to compare to a long NR simulation. But performing this comparison for one point in the more complicated region of parameter space is likely sufficient to claim that PN (at whatever PN order one has) is sufficiently accurate for modeling purposes.

is likely that PN hybridizations can ready the community for future detectors without the need to push NR to be able to accurately simulate the entirety of $\mathcal{O}(10^3)$ orbits[15].

3. Discussion and conclusions

In this Note, we highlight a commonly overlooked feature of the waveform mismatch: its dependence on the length of an NR simulation. In particular, by assuming that two waveforms differ by some small phase error—as is the case for waveforms from different resolution NR simulations—we showed analytically that the mismatch tends to increase with the square of the length of the simulation. Moreover, by examining simulations from the SXS collaboration, we showed that this increase in mismatch can be as large as a three orders of magnitude when increasing the length of a simulation by $\sim$100 000 M or $\sim$100 orbits. Consequently, demonstrating that a relatively short simulation is accurate to within a target mismatch is not sufficient to claim a NR code's accuracy for much longer simulations, i.e. those which are required for spanning the entire frequency band of future GW detectors.

This shows that NR groups have much to achieve to be prepared for the next era of GW science [22, 23]. There are now many independent NR simulation catalogs [16, 24–36], but most only provide a single resolution for each system. We encourage future catalogs to provide at least two, preferably three, resolutions so that their truncation errors can be quantified. We also point out that there are many important studies that one can pursue to help prepare us for next-generation sensitivities: better control of eccentricity in simulations, improving the accuracy of PN waveforms with more terms and effective-one-body calculations, and extending mass ratio coverage with gravitational self-force.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://data.black-holes.org/simulations/index.html.

Acknowledgments

---

[15] There is also a crucial need from the self-force community [21] to extend the parameter space coverage provided by NR simulations to higher mass ratios, as well as an effort from all communities to understand how to model eccentric-precessing systems more generally.

Appendix. Other mismatch caveats

The GW literature contains varying degrees of parameter minimization when computing waveform mismatches. Above, we only (implicitly) considered time and $SO(3)$ rotations of two fixed waveforms. At a different extreme, in the context of GW detectability, one only cares if *any* template has a high overlap with a signal, so one might minimize mismatch over both extrinsic parameters (e.g. time translations, rotations of the source, and GW polarization angle (rotations about the line of sight to the source)) and intrinsic parameters (e.g. masses, spins, eccentricity), regardless of how close the template parameters are to the true source parameters.

A more subtle question is how to use mismatch to quantify the similarity of waveform models themselves, apart from detectors that live at one point on the source's celestial sphere. A waveform model is effectively a map $f\colon X \to \mathcal{F}(\mathcal{I}^+)$ where X is our parameter space (e.g. the 9-dimensional space of BBH mergers), $\mathcal{I}^+$ is future null infinity, which topologically is $S^2 \times \mathbb{R}$; and $\mathcal{F}(\mathcal{I}^+)$ is some appropriate function space (e.g. twice differentiable for the Weyl curvature scalars to exist; with square-integrable news for the memory to be finite). The overlap integral can be taken over the two-sphere, providing a metric on the function space (modulo the minor convergence issue related to memory effects mentioned in footnote 7).

Now we can ask how similar are two models f, g. Unfortunately this is not as simple as $\sup_{x \in X} \mathcal{M}(f(x), g(x))$ where sup is the supremum, because the two models f and g may have different meanings for the parameters; see the discussion of figure 1 in [19]. We can allow for different parameter definitions as follows. For some fixed waveform $f(x)$ we can find the 'shortest distance' from the image of g by finding $\mathcal{M}(f(x), g) = \inf_{x' \in X} \mathcal{M}(f(x), g(x'))$, where inf is the infimum. Doing this for all $x \in X$, we can find the point in the image of f whose shortest distance is greatest, $\sup_{x \in X} \mathcal{M}(f(x), g)$. Similarly, reversing the roles of f and g, we can find the point in the image of g whose shortest distance to the image of f is greatest—this need not be the same pair of points. Finally we take the larger of the two mismatches. This is the *Hausdorff distance* between the images of f and g. Unfortunately, this is only practical on a discrete subset $\Lambda \subset X$, but is still a useful measure of the similarity of f and g.

Finally, one might consider waveforms on $\mathcal{I}^+$ to be in equivalence classes of future null infinity's asymptotic symmetry group, i.e. the Bondi–van der Burg–Metzner–Sachs (BMS) group [37–42]. Some of these BMS transformations are clearly physically meaningful across the set of all binary merger observations—for example, each merger happens at a specific time, so it can not be translated to a different time. However in the model, we have a Universe with only a single GW source. The action of BMS on the function space $\mathcal{F}(\mathcal{I}^+)$ gives gauge-equivalent orbits, so a measurement of model similarity should be a Hausdorff distance on the quotient space $\mathcal{F}(\mathcal{I}^+)/\mathrm{BMS}$. Note that this is partially achieved when performing the optimizations described earlier, i.e. optimizing over time translations, source rotations, and GW polarization angle, but one should also include the Lorentz boost and supertranslation transformations.

ORCID iDs

Keefe Mitman https://orcid.org/0000-0003-0276-3856
Leo C Stein https://orcid.org/0000-0001-7559-9597
Michael Boyle https://orcid.org/0000-0002-5075-5116
Nils Deppe https://orcid.org/0000-0003-4557-4115
Lawrence E Kidder https://orcid.org/0000-0001-5392-7342

Harald P Pfeiffer https://orcid.org/0000-0001-9288-519X

Mark A Scheel https://orcid.org/0000-0001-6656-9134

References

[1] Chen Y *et al* 2024 *Phys. Rev. D* **110** 064049
[2] Finn L S and Chernoff D F 1993 *Phys. Rev. D* **47** 2198–219
[3] Cutler C and Flanagan E E 1994 *Phys. Rev. D* **49** 2658–97
[4] Flanagan E E and Hughes S A 1998 *Phys. Rev. D* **57** 4566–87
[5] Flanagan E E and Hughes S A 1998 *Phys. Rev. D* **57** 4535–65
[6] Chatziioannou K, Dent T, Fishbach M, Ohme F Pürrer M Raymond V Veitch J 2024 arXiv:2409.02037
[7] Lindblom L, Owen B J and Brown D A 2008 *Phys. Rev. D* **78** 124020
[8] Chatziioannou K, Klein A, Yunes N and Cornish N 2017 *Phys. Rev. D* **95** 104004
[9] Toubiana A and Gair J R 2024 arXiv:0801.4134
[10] Kinoshita H, Yoshida H and Nakai H 1990 *Celest. Mech. Dyn. Astron.* **50** 59–71
[11] Wisdom J and Holman M 1991 *Astron. J.* **102** 1528–38
[12] Yoshida H 1993 *Celest. Mech. Dyn. Astron.* **56** 27–43
[13] Levison H F and Duncan M J 1994 *Icarus* **108** 18–36
[14] Chambers J E 1999 *Mon. Not. R. Astron. Soc.* **304** 793–9
[15] Scuro S R and Chin S A 2005 *Phys. Rev. E* **71** 056703
[16] Boyle M *et al* 2019 *Class. Quantum Grav.* **36** 195006
[17] Virtanen P *et al* SciPy 10 Contributors 2020 *Nat. Methods* **17** 261–72
[18] Boyle M and Scheel M 2025 The sxs package
[19] Sun D, Boyle M, Mitman K, Scheel M A, Stein L C, Teukolsky S A and Varma V 2024 *Phys. Rev. D* **110** 104076
[20] Pompili L *et al* 2023 *Phys. Rev. D* **108** 124035
[21] Pound A and Wardell B 2021 arXiv:2101.04592
[22] Pürrer M and Haster C J 2020 *Phys. Rev. Res.* **2** 023151
[23] Chandramouli R S, Prokup K, Berti E and Yunes N 2025 *Phys. Rev. D* **111** 044026
[24] Aylott B *et al* 2009 *Class. Quantum Grav.* **26** 114008
[25] Ajith P *et al* 2012 *Class. Quantum Grav.* **29** 124001
[26] Mroue A H *et al* 2013 *Phys. Rev. Lett.* **111** 241104
[27] Hinder I *et al* 2014 *Class. Quantum Grav.* **31** 025012
[28] Jani K, Healy J, Clark J A, London L, Laguna P and Shoemaker D 2016 *Class. Quantum Grav.* **33** 204001
[29] Healy J, Lousto C O, Zlochower Y and Campanelli M 2017 *Class. Quantum Grav.* **34** 224001
[30] Healy J, Lousto C O, Lange J, O'Shaughnessy R, Zlochower Y and Campanelli M 2019 *Phys. Rev. D* **100** 024021
[31] Huerta E A *et al* 2019 *Phys. Rev. D* **100** 064003
[32] Healy J and Lousto C O 2020 *Phys. Rev. D* **102** 104018
[33] Healy J and Lousto C O 2022 *Phys. Rev. D* **105** 124010
[34] Ferguson D *et al* 2023 arXiv:2309.00262
[35] Hamilton E *et al* 2024 *Phys. Rev. D* **109** 044032
[36] Rashti A Gamba R Chandra K Radice D Daszuta B Cook W and Bernuzzi S 2024 arXiv:2303.05419
[37] Bondi H 1960 *Nature* **186** 535–535
[38] Sachs R 1961 *Proc. R. Soc. A* **264** 309–38
[39] Bondi H, der Burg M G J V and Metzner A W K 1962 *Proc. R. Soc. A* **269** 21–52
[40] Sachs R 1962 *Phys. Rev.* **128** 2851–64
[41] Sachs R 1962 *Proc. R. Soc. A* **270** 103–26
[42] Mitman K *et al* 2024 *Class. Quantum Grav.* **41** 223001