PHYSICAL REVIEW D **80,** 064019 (2009)

Use and abuse of the model waveform accuracy standards

Lee Lindblom

Theoretical Astrophysics 350-17, California Institute of Technology, Pasadena, California 91125, USA

(Received 2 July 2009; published 15 September 2009)

Accuracy standards have been developed to ensure that the waveforms used for gravitational-wave data analysis are good enough to serve their intended purposes. These standards place constraints on certain norms of the frequency-domain representations of the waveform errors. Examples are given here of possible misinterpretations and misapplications of these standards, whose effect could be to vitiate the quality control they were intended to enforce. Suggestions are given for ways to avoid these problems.

DOI: 10.1103/PhysRevD.80.064019 PACS numbers: 07.05.Kf, 04.25.D−, 04.25.dg, 04.30.−w

I. INTRODUCTION

Model waveforms are used in gravitational-wave data analysis in two different ways. A signal is first identified in a detector's noisy data stream when it is found to have a significantly large projection onto a model waveform. If the model waveforms were not accurate enough, then an unacceptably large fraction of real signals would fail to be detected in this way. The second use of model waveforms in gravitational-wave data analysis is to measure the physical properties of any signals identified in the first detection step. These measurements are performed by fine-tuning the model-waveform parameters (e.g., the masses and spins of the sources, the source's orientation, the times of arrival of the signals, etc.) to achieve the largest correlation with the data. If the model waveforms were not accurate enough, these measured parameters would fail to represent the true physical properties of the sources to the level of precision commensurate with the intrinsic quality of the data. So separate waveform-accuracy standards have been formulated to prevent these potential losses of scientific information in the detection and measurement phases of gravitational-wave data analysis [1,2].

These accuracy standards are expressed as limits on the waveform-modeling errors δh_m, i.e., the difference between a model waveform h_m and its exact counterpart h_e: $\delta h_m = h_m - h_e$. Model gravitational waveforms, and the errors associated with them, are most easily determined as functions of time $\delta h_m(t) = h_m(t) - h_e(t)$. In contrast, gravitational-wave data analysis and the accuracy standards for model waveforms are most conveniently formulated in terms of the frequency-domain representations of the waveforms and their errors, $\delta h_m(f)$. The time and frequency representations of waveform-modeling error are related to one another by the Fourier transform

$$\delta h_m(f) = \int_{-\infty}^{\infty} \delta h_m(t) e^{-2\pi i f t} dt. \tag{1}$$

(This paper follows the convention of the LIGO Scientific Collaboration [3] and the signal-processing community by using the phase factor $e^{-2\pi i f t}$ in these Fourier transforms. Most of the early gravitational-wave literature and essentially all other computational physics literature use $e^{2\pi i f t}$, but this choice does not affect any of the subsequent equations in this paper.)

The simplest way to express the standards needed to ensure the appropriate levels of accuracy for model gravitational waveforms is to write them in terms of a particular norm of the model-waveform errors $\langle \delta h_m | \delta h_m \rangle$. This norm, defined by

$$\langle \delta h_m | \delta h_m \rangle = 4 \int_0^{\infty} \frac{\delta h_m(f) \delta h_m^*(f)}{S_n(f)} df, \tag{2}$$

weights the different frequency components of the waveform error by the power spectral density of the detector noise $S_n(f)$. In terms of this norm, the accuracy requirement that ensures no loss of scientific information during the measurement process is

$$\langle \delta h_m | \delta h_m \rangle < 1, \tag{3}$$

cf. Eq. (5) of Ref. [1]. Similarly, the accuracy requirement that ensures no significant reduction in the rate of detections is

$$\langle \delta h_m | \delta h_m \rangle < 2\epsilon_{\max}\rho^2, \tag{4}$$

where ρ is the optimal signal-to-noise ratio of the detected signal, and $\epsilon_{\max}$ is a parameter that determines the fraction of detections lost due to waveform-modeling errors, cf. Eq. (14) of Ref. [1].

These basic accuracy requirements, Eqs. (3) and (4), assume the detector is ideal in the sense that any measurement errors made in calibrating the response function of the detector are negligible compared to the waveform-modeling errors. It is more realistic to expect that the detector will be calibrated only to the level of accuracy needed to ensure the calibration errors satisfy approximately the same accuracy requirements as the waveform-modeling errors. In that case the modeling-accuracy standards must be somewhat stricter than those for the ideal-detector case [2]. Assuming equal calibration and waveform-modeling errors, the modeling-accuracy requirement for measurement becomes

$$\langle \delta h_m | \delta h_m \rangle < \tfrac{1}{4}, \tag{5}$$

while the requirement for detection becomes

$$\langle\delta h_m|\delta h_m\rangle < \frac{\epsilon_{\max}}{2}\rho^2. \tag{6}$$

To keep things as simple as possible, this discussion uses these somewhat stronger accuracy requirements. However, none of the potential abuses nor any of the methods discussed here to avoid these problems depend critically on what the ratio of the waveform-modeling error to calibration error is ultimately chosen to be.

The discussion of various possible misapplications of the accuracy standards in the following section will be simplified by introducing a little more notation. It is useful to define the logarithmic amplitude χ and the phase Φ of the frequency-domain representation of a waveform in the following way: $h = e^{\chi+i\Phi}$. Consequently, the frequency-domain amplitude and phase errors $\delta\chi_m$ and $\delta\Phi_m$ of a model waveform are defined as

$$h_m = e^{\chi_e+\delta\chi_m+i\Phi_e+i\delta\Phi_m} \approx h_e(1+\delta\chi_m+i\delta\Phi_m). \tag{7}$$

It is also useful to define certain averages, $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$, of these waveform-modeling errors:

$$\overline{\delta\chi_m}^2 = \int_0^\infty (\delta\chi_m)^2\frac{4|h_e|^2}{\rho^2 S_n(f)}df, \tag{8}$$

and

$$\overline{\delta\Phi_m}^2 = \int_0^\infty (\delta\Phi_m)^2\frac{4|h_e|^2}{\rho^2 S_n(f)}df. \tag{9}$$

The quantity ρ that appears in these integrals is the optimal signal-to-noise ratio, defined by

$$\rho^2 = \int_0^\infty \frac{4|h_e|^2}{S_n(f)}df. \tag{10}$$

The weight factor $4|h_e|^2/\rho^2 S_n$ that appears in Eqs. (8) and (9) has the effect of emphasizing those frequency components of the errors where the wave amplitude $|h_e|$ is large and the noise S_n is small. This weight factor has integral one; so these are true (signal and noise-weighted) averages of $\delta\chi_m$ and $\delta\Phi_m$, respectively. The waveform-accuracy standards of Eqs. (5) and (6) take simple and intuitive forms when expressed in terms of these averages:

$$\sqrt{\overline{\delta\chi_m}^2+\overline{\delta\Phi_m}^2} < \frac{1}{2\rho}, \tag{11}$$

for measurement and

$$\sqrt{\overline{\delta\chi_m}^2+\overline{\delta\Phi_m}^2} < \sqrt{\frac{\epsilon_{\max}}{2}}, \tag{12}$$

for detection. In this form, the accuracy standards merely state that the (amplitude and noise-weighted) averages of the combined amplitude and phase errors, $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$, must be less than $1/2\rho$ for measurement and $\sqrt{\epsilon_{\max}/2}$ for detection.

II. POTENTIAL ABUSES

Several possible misinterpretations of the waveform-modeling-accuracy standards, Eqs. (11) and (12), are discussed in this section. These fallacies can (although not necessarily always will) result in false conclusions about the suitability of model waveforms for gravitational-wave data analysis. Methods for avoiding these potential abuses are presented as part of the discussion of each fallacy.

A. Maximum error fallacy

The waveform-modeling-accuracy standards expressed in Eqs. (11) and (12) are easy to understood as requirements on the average values of the amplitude and phase errors of the model waveforms. When assessing the accuracy of the waveforms produced by numerical simulations, it is natural and reasonable to attempt to estimate their errors by producing graphs, such as those in Ref. [4], showing the time dependence of the differences in the amplitudes and phases of the model waveforms produced by simulations at different numerical resolutions, etc. It would be tempting to conclude that the waveforms in question are good enough whenever such graphs indicate that the maximum amplitude and phase errors of the model waveforms satisfy the inequalities of Eqs. (11) and (12). If the maximum errors satisfy the required inequalities, then it seems reasonable to expect that any average of these errors would satisfy the inequalities as well.

Unfortunately this would be wrong, for a long list of reasons. For the purposes of this discussion, let us put aside the issue of systematic errors (e.g., failure to impose purely outgoing, nonreflective, constraint preserving outer boundary conditions, or failure to extract the waveform in a gauge invariant way, etc.), which cannot be measured simply by performing convergence tests on a series of simulations. Rather let us focus on the narrow issue of whether it is sufficient to guarantee that the maximum errors in the time-domain representations of the waveforms satisfy the inequalities specified in the accuracy standards.

The fundamental misinterpretation that leads to this fallacy is the blurring of the distinction between time-domain and frequency-domain representations of the waveform errors. While fairly straightforward, the Fourier transform that connects these representations is nonlocal, and does not map the amplitudes and phases of one representation into the amplitudes and phases of the other in a simple way. In particular, it is fairly easy to construct examples that demonstrate that even when the maximum values of the time-domain amplitude and phase errors $\delta\mu_\chi = \max|\delta\chi_m(t)|$ and $\delta\mu_\Phi = \max|\delta\Phi_m(t)|$ satisfy the accuracy standards, Eqs. (11) and (12), there is no guarantee that the analogous frequency-domain averages $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$ do as well.

The time-domain representation of any exact waveform can be expressed in terms of an amplitude $A_e(t)$ and phase

$\Phi_e(t)$,

$$h_e(t) = A_e(t)\cos[\Phi_e(t)]. \tag{13}$$

Here and throughout the remainder of this paper the time-domain waveform $h_e(t)$ is taken to be real, consisting of the particular combination of $+$ and $\times$ polarizations observable by a particular detector. Similarly, the time-domain representation of the corresponding model waveform can be written as

$$h_m(t) = A_e(t)[1 + \delta\chi_m(t)]\cos[\Phi_e(t) + \delta\Phi_m(t)], \tag{14}$$

where $\delta\chi_m$ and $\delta\Phi_m$ represent (real) time-domain versions of the logarithmic amplitude and phase errors. It is useful to express these errors in the form

$$\delta\chi_m(t) = \delta\mu_\chi g_\chi(t), \tag{15}$$

$$\delta\Phi_m(t) = \delta\mu_\Phi g_\Phi(t), \tag{16}$$

where g_χ and g_Φ are taken to satisfy $|g_\chi(t)| \le 1$ and $|g_\Phi(t)| \le 1$, so the constants $\delta\mu_\chi$ and $\delta\mu_\Phi$ represent the maximum errors. The goal of this example is to find fairly "realistic" error functions $\delta\chi_m(t)$ and $\delta\Phi_m(t)$, having the property that $\delta\mu_\chi$ and $\delta\mu_\Phi$ are less than the corresponding averages $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$. Any such example would demonstrate that limiting the time-domain maxima, $\delta\mu_\chi$ and $\delta\mu_\Phi$, with the inequalities in the accuracy standards is insufficient to guarantee that the actual standards are satisfied.

To guide our selection of the error functions g_χ and g_Φ for this example, let us examine the estimates of $\delta\chi_m(t)$ and $\delta\Phi_m(t)$ from actual numerical simulations, e.g., Figs. 6–8 of Ref. [4]. It appears that in some simulations at least, the waveform errors can have fairly monochromatic oscillatory time dependence with frequencies a few times the basic gravitational-wave frequency. So let us consider error functions of the form

$$g_\chi(t) = g_\Phi(t) = \cos[\lambda\Phi_e(t)], \tag{17}$$

where the parameter λ sets the frequency of the errors relative to the basic gravitational-wave frequency. The waveform errors, Eq. (17) for this example, are combined with the "exact" waveform functions $A_e(t)$ and $\Phi_e(t)$ to produce $h_m(t)$ according to Eq. (14). For the purposes of this example, the exact waveform $h_e(t)$ is taken to be one obtained by patching a post-Newtonian waveform onto a numerical waveform from an equal-mass nonspinning binary black-hole simulation [4,5]. Once assembled this example waveform error $\delta h_m(t) = h_m(t) - h_e(t)$ is Fourier transformed numerically, and the result used to evaluate the average error quantities $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$ according to the prescriptions in Eqs. (8) and (9).

The ratio R, defined by

$$R = \sqrt{\frac{\overline{\delta\chi_m}^2 + \overline{\delta\Phi_m}^2}{\delta\mu_\chi^2 + \delta\mu_\Phi^2}}, \tag{18}$$

measures how faithfully the maximum time-domain errors $\delta\mu_\chi$ and $\delta\mu_\Phi$ estimate the averages $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$. When $R > 1$ the maximum time-domain errors underestimate the frequency-domain averages of these errors, and could not be used therefore to verify the waveform-accuracy standards. The ratio R is (essentially) independent of $\delta\mu_\chi$ and $\delta\mu_\Phi$ in the limit of small $\delta\mu_\chi$ and $\delta\mu_\Phi$. Figure 1 illustrates R as a function of the total binary black-hole mass for several values of the parameter λ. The averages $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$ were computed here using an Advanced LIGO noise curve [6], and maximum time-domain errors $\delta\mu_\chi = \delta\mu_\Phi = 0.01$. It is not hard to imagine how model-waveform errors, having frequencies a few times the fundamental gravitational-wave frequency (e.g., from inadvertent excitations of the individual black holes), could enter even the best numerical simulations. This example shows that in such cases the frequency-domain error averages $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$ could exceed the simple time-domain maxima $\delta\mu_\chi$ and $\delta\mu_\Phi$ by a significant amount. It would be an abuse of the accuracy standards therefore to conclude that model waveforms are suitable for gravitational-wave data analysis simply by verifying that their maximum time-domain errors satisfy those conditions. The cure for this fallacy is straightforward: do not use the time-domain maxima as surrogates for the relevant signal and noise-weighted averages when enforcing the waveform-accuracy standards.

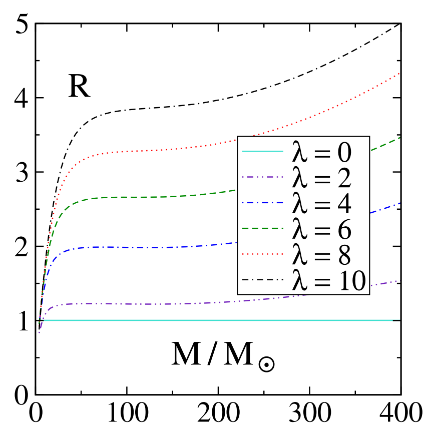

FIG. 1 (color online). Shown is R, the ratio of the proper frequency-domain waveform-error measure to the simple time-domain maximum measure, for a range of masses and values of the frequency parameter λ. When $R > 1$ the accuracy standards could be violated if the simple maximum error measure were used instead of the proper frequency-domain measure.

B. Error-envelope fallacy

The discussion of the maximum error fallacy in Sec. II A shows that a proper application of the gravitational-waveform-accuracy standards, Eqs. (11) and (12), requires more information about waveform errors than a simple knowledge of their time-domain maxima. What additional information is needed? Rather than tackle that difficult question directly, let us consider instead a more modest question: What additional information about waveform errors is likely to be available? A complete knowledge of the waveform error $\delta h_m(t)$ will obviously never be available to verify the accuracy standards. If it were, the exact waveform $h_e(t) = h_m(t) - \delta h_m(t)$ would also be known, and there would be no need for any waveform-accuracy standards. It seems likely that the most that will ever be available are reliable local-in-time bounds on the waveform errors. Careful convergence tests together with a detailed analysis of all the systematic errors, could plausibly produce error-envelope functions $\delta\chi_E(t)$ and $\delta\Phi_E(t)$, having the property that they strictly and tightly bound the waveform errors

$$\delta\chi_E(t) \geq |\delta\chi_m(t)|, \tag{19}$$

$$\delta\Phi_E(t) \geq |\delta\Phi_m(t)|. \tag{20}$$

Given such error-envelope functions, it would be straightforward to construct the full model-waveform-error function, $\delta h_E(t)$, based on them:

$$\begin{aligned}\delta h_E(t) &= A_m(t)[1 + \delta\chi_E(t)]\cos[\Phi_m(t) + \delta\Phi_E(t)] \\ &\quad - A_m(t)\cos[\Phi_m(t)], \\ &\approx A_m(\delta\chi_E\cos\Phi_m - \delta\Phi_E\sin\Phi_m),\end{aligned} \tag{21}$$

where A_m and Φ_m are the amplitude and phase of the model waveform. This waveform-error estimate, $\delta h_E(t)$ could then be Fourier transformed and the resulting frequency-domain error estimate used to construct the averages $\overline{\delta\chi_E}$ and $\overline{\delta\Phi_E}$ in a fairly straightforward way. It seems reasonable to expect that the resulting $\overline{\delta\chi_E}$ and $\overline{\delta\Phi_E}$ would be upper bounds on the actual waveform-error averages $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$ needed for the accuracy standards. So these error-envelope estimates should be exactly what are needed to enforce the accuracy standards in a rigorous and reliable way.

Unfortunately, this would be incorrect: error envelopes produce error averages $\overline{\delta\chi_E}$ and $\overline{\delta\Phi_E}$ that are not in general upper bounds on the needed waveform-error averages $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$. Consequently, they are not a useful test of whether the accuracy standards are actually satisfied or not. A very simple example of this failure can be seen in the example introduced in Sec. II A to illustrate the maximum error fallacy. The maximum time-domain errors $\delta\mu_\chi$ and $\delta\mu_\Phi$ can be used to construct very simple (constant in time) but crude error-envelope functions:

$$\delta\chi_E(t) = \delta\mu_\chi \geq |\delta\chi_m(t)|, \tag{22}$$

$$\delta\Phi_E(t) = \delta\mu_\Phi \geq |\delta\Phi_m(t)|. \tag{23}$$

The $\lambda = 0$ case in Eq. (17) corresponds to this choice of envelope function. It is obvious from Fig. 1 that the error-envelope case $\lambda = 0$ has smaller values of $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$ than any of the other cases. So this simple example illustrates that in general, the error-envelope estimates, $\overline{\delta\chi_E}$ and $\overline{\delta\Phi_E}$ do not provide upper limits on the needed waveform-error averages $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$.

The fundamental misconception leading to the error-envelope fallacy is the expectation that local-in-time bounds, e.g., $G(t) \geq |g(t)|$, lead to analogous bounds on the frequency-domain representations of those functions, i.e., $G(f) \geq |g(f)|$. It is easy to find examples that illustrate that this is not the case. The exemplar model-waveform error $g(t)$, shown as the dotted curve in Fig. 2, is bounded by the envelope function $G(t)$ shown as the solid curve in this figure. The Fourier transforms of these functions produce the frequency-domain representations $g(f)$ and $G(f)$ illustrated in Fig. 3. The basic problem is that the error-envelope function does not place any limit on the frequency content of the function it bounds. So in the frequency domain, the actual error represented by $g(f)$ can be much larger for some frequencies than the envelope function $G(f)$. This can lead to waveform-error averages $\overline{\delta\chi_m}$ and $\overline{\delta\Phi_m}$, which are much larger than those based on the envelope functions $\overline{\delta\chi_E}$ and $\overline{\delta\Phi_E}$. This can occur when the weight function used in Eqs. (8) and (9) is large in the frequency range where the waveform errors are large, and small in the frequency range where the error envelope is large.

In the discussion above, it was argued that the error-envelope functions $\delta\chi_E(t)$ and $\delta\Phi_E(t)$ are probably the most information about the actual waveform errors $\delta\chi_m(t)$ and $\delta\Phi_m(t)$ that will ever be available. So unfortunately

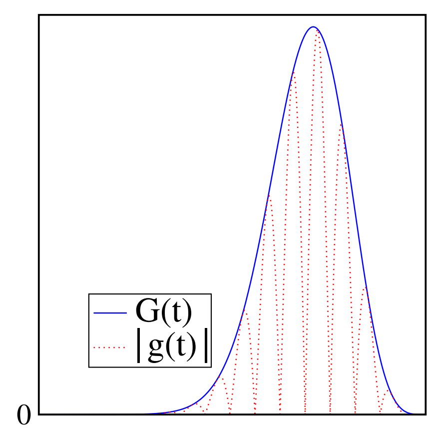

FIG. 2 (color online). Example of a time-domain waveform error $g(t)$ and an envelope function $G(t)$ that satisfies $G(t) \geq |g(t)|$.

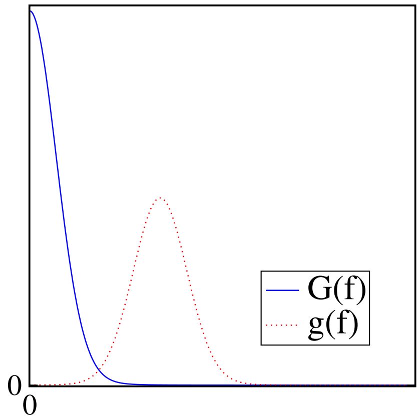

FIG. 3 (color online). Frequency-domain representations of the model-waveform error and envelope functions $g(f)$ and $G(f)$, whose time-domain representations appear in Fig. 2. The envelope function $G(f)$ obviously fails to provide a local bound on the frequency-domain error $g(f)$.

the error-envelope fallacy suggests that the waveform-accuracy standards, Eqs. (11) and (12), may never be enforceable.

Fortunately, this too would be incorrect: there is a way to enforce the accuracy standards rigorously with only an error-envelope estimate of the actual waveform-modeling error. Parseval's theorem provides a direct, exact, connection between the norms of time-domain functions and their frequency-domain counterparts. In particular, the L^2 norm of the time-domain waveform error

$$\|\delta h_m(t)\|^2 = \int_{-\infty}^{\infty} |\delta h_m(t)|^2 dt \tag{24}$$

is identical to the L^2 norm of its frequency-domain counterpart

$$\|\delta h_m(t)\|^2 = \|\delta h_m(f)\|^2 = \int_{-\infty}^{\infty} |\delta h_m(f)|^2 df. \tag{25}$$

An important feature of these L^2 norms is that local-in-time bounds also provide bounds on the frequency-domain L^2 norms. So if $G(t)$ is an envelope function for $g(t)$, then the L^2 norm $\|G(t)\|$ bounds the L^2 norms of $g(t)$ and $g(f)$: $\|G(t)\| \geq \|g(t)\| = \|g(f)\|$. In this way the error-envelope waveform $\delta h_E(t)$ provides an upper bound on the L^2 norm of the frequency-domain error $\delta h_m(f)$. Keeping terms in Eq. (14) to first order in $\delta\chi_m$ and $\delta\Phi_m$, it follows (using the triangle inequality) that

$$\begin{aligned}\|\delta h_m(t)\| \leq{}& \|A_m(t)\delta\chi_m(t)\cos[\Phi_m(t)]\| \\ &+ \|A_m(t)\delta\Phi_m(t)\sin[\Phi_m(t)]\|, \\ \leq{}& \|A_m(t)\delta\chi_E(t)\cos[\Phi_m(t)]\| \\ &+ \|A_m(t)\delta\Phi_E(t)\sin[\Phi_m(t)]\|.\end{aligned} \tag{26}$$

Therefore, error-envelope waveforms would be useful in applying the accuracy standards if they could be reformulated in terms of L^2 norms $\|\delta h_m\|$, rather than noise-weighed inner products $\langle \delta h_m | \delta h_m \rangle$.

Fortunately again, the accuracy standards have been rewritten in terms of the L^2 norms of the waveform errors [1]. In particular, the accuracy standard for measurement, corresponding to Eq. (5), becomes

$$\frac{\|\delta h_m(t)\|}{C\|h_m(t)\|} < \frac{1}{2\rho}, \tag{27}$$

and the standard for detection, corresponding to Eq. (6), becomes

$$\frac{\|\delta h_m(t)\|}{C\|h_m(t)\|} < \sqrt{\frac{\epsilon_{\max}}{2}}. \tag{28}$$

The quantity C that appears in these inequalities is the ratio of the standard signal-to-noise measure ρ to a nonstandard signal-to-noise measure based on L^2 norms

$$C^2 = \rho^2\left(\frac{2\|h_m(t)\|^2}{\min S_n(f)}\right)^{-1}. \tag{29}$$

This quantity is dimensionless, and independent of the absolute scale (i.e., the distance to the gravitational-wave source) of the waveform. Figure 4 illustrates C as a function of the total mass for nonspinning equal-mass binary black-hole waveforms constructed by patching together the waveform produced by a numerical simulation with a post-Newtonian waveform [1,4,5]. The quantity C can be evaluated in a straightforward way whenever model waveforms are computed. Using Eq. (26), it follows that the accuracy standards can also be written as conditions on the error-envelope functions $\delta\chi_E(t)$ and $\delta\Phi_E(t)$. These conditions are

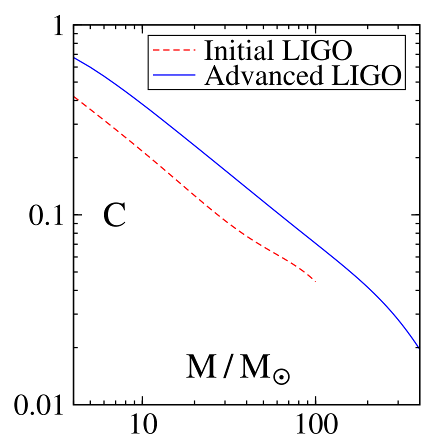

FIG. 4 (color online). Curves illustrate C, the ratio of the standard signal-to-noise measure ρ to a nonstandard measure defined in Eq. (29), as a function of the total mass for nonspinning equal-mass binary black-hole waveforms. Dashed curve is based on the initial LIGO noise spectrum [7]; solid curve is based on an advanced LIGO noise curve [6].

$$\frac{\|A_m(t)\delta\chi_E(t)\cos[\Phi_m(t)]\|}{C\|A_m(t)\cos[\Phi_m(t)]\|} + \frac{\|A_m(t)\delta\Phi_E(t)\sin[\Phi_m(t)]\|}{C\|A_m(t)\cos[\Phi_m(t)]\|} < \frac{1}{2\rho}, \tag{30}$$

for measurement, and

$$\frac{\|A_m(t)\delta\chi_E(t)\cos[\Phi_m(t)]\|}{C\|A_m(t)\cos[\Phi_m(t)]\|} + \frac{\|A_m(t)\delta\Phi_E(t)\sin[\Phi_m(t)]\|}{C\|A_m(t)\cos[\Phi_m(t)]\|} < \sqrt{\frac{\epsilon_{\max}}{2}}, \tag{31}$$

for detection.

This discussion shows that using error-envelope estimates of the waveform-modeling errors in a naive application of the gravitational-waveform-accuracy standards, Eqs. (11) and (12), would be an abuse of those standards. Fortunately, error-envelope estimates can be used to enforce the accuracy standards rigorously when they are reformulated in terms of the L^2 norms of the errors. The use of the L^2 norms rather than the noise-weighted norms prevents the suppression of any frequency components of the envelope functions. This ensures that the envelope functions provide real bounds on the norms of the actual errors, Eq. (26), thus making them useful for enforcing the accuracy standards. So the cure for the error-envelope fallacy is straightforward: use the time-domain L^2 norm versions of the accuracy standards, Eqs. (30) and (31), instead of the frequency-domain noise-weighted norm versions, Eqs. (11) and (12).

C. Universality fallacy

The discussion of the error-envelope fallacy in Sec. II B shows that it can be avoided by adopting versions of the accuracy standards based on L^2 norms, Eqs. (30) and (31), rather than noise-weighted norms. The L^2 norm versions of the standards are complicated however by the appearance of the quantity C, defined in Eq. (29). This quantity depends on the waveform and the detector noise spectrum, and can be evaluated in a straightforward way whenever a model waveform is computed. For example, Fig. 4 shows C as a function of mass for equal-mass nonspinning binary black-hole waveforms, cf. Fig. 4 of Ref. [1].

It has been argued that the quantity C is universal: the same for all gravitational waveforms of a given type [1]. Thus, it was believed that Fig. 4 represents all equal-mass nonspinning binary black-hole waveforms. This however is clearly wrong. The norm $\|h_m\|$, which appears in the expression for C in Eq. (29), depends on the length of the waveform: having different numerical values when computed for model waveforms with different numbers of wave cycles. Consequently, C too will depend on the length of the waveform, in addition to the waveform's other physical properties. This nonuniversality is illustrated in Fig. 5, where C is shown for equal-mass nonspinning

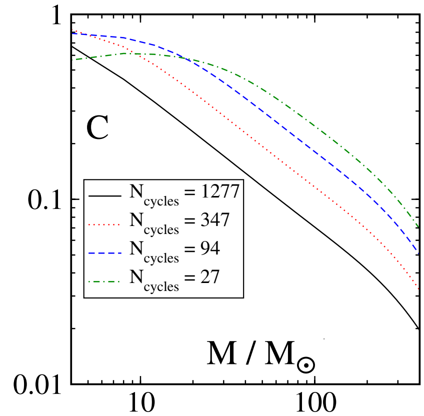

FIG. 5 (color online). Curves illustrate C, the ratio of the standard signal-to-noise measure ρ to a nonstandard measure defined in Eq. (29), as a function of the total mass for nonspinning equal-mass binary black-hole waveforms. Various curves use gravitational waveforms containing different numbers of gravitational-wave cycles, indicated by the parameter $N_{\rm cycles}$.

binary black-hole waveforms (constructed by patching a waveform from a numerical simulation onto a post-Newtonian waveform [4,5]) containing different numbers of gravitational-wave cycles. Each of the waveforms used to produce these curves includes the merger and ringdown of the final black hole, but different numbers of wave cycles from the inspiral portion of the binary evolution. The length (in time) of each successively shorter waveform used in Fig. 5 is one-eighth that of its predecessor; the number of wave cycles is indicated for each waveform by the parameter $N_{\rm cycles}$.

It would be a mistake for the waveform simulation community to attempt to validate the accuracy of a model waveform by combining an error-envelope estimate of $\|\delta h_m\|/\|h_m\|$ from one model waveform with the C from another. In this case the mismatched version of the measure $\|\delta h_m\|/C\|h_m\|$, which appears on the left in Eqs. (27) and (28), would not be the appropriate one needed to enforce those standards. The cure for the universality fallacy is fortunately straightforward: simply determine the quantity C afresh whenever new waveform models are computed, and then evaluate the entire error measure $\|\delta h_m\|/C\|h_m\|$ from the same model waveform.

An analogous abuse of universality could potentially also occur with the quantity, $\tilde{C}$, used as part of the combined calibration and waveform-modeling-accuracy standards for gravitational-wave detectors [2]. This quantity is defined by

$$\tilde{C}^4 = \rho^4\left(\int_0^\infty \frac{4|h_m|^4}{\bar{n}^2 S_n}df\right)^{-1}, \tag{32}$$

where the total detector noise $\bar{n}$ is

$$\frac{1}{\bar{n}^2} = \int_0^\infty \frac{4}{S_n}df. \tag{33}$$

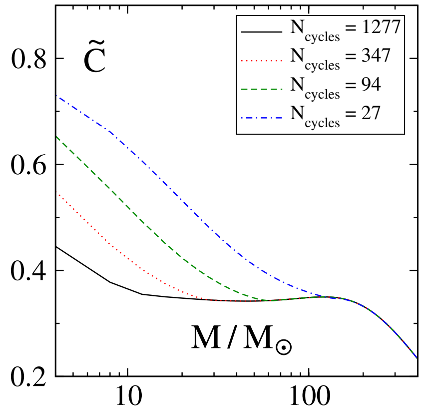

FIG. 6 (color online). Curves illustrate $\tilde{C}$, the ratio of the standard signal-to-noise measure ρ to another nonstandard measure defined in Eq. (32), as a function of the total mass for several nonspinning equal-mass binary black-hole waveforms. This quantity is used to enforce calibration accuracy standards for gravitational-wave detectors.

The quantity $\tilde{C}$ also depends on the length of the gravitational waveform. Illustrated in Fig. 6 are several curves showing $\tilde{C}$ as a function of mass, computed for the same waveforms of varying length used to compute the curves in Fig. 5. It would be a mistake to attempt to enforce the combined calibration and waveform-modeling-accuracy standards of Ref. [2] using C and $\tilde{C}$ computed from different waveform models.

III. DISCUSSION

This paper has identified several possible misinterpretations and misapplications of the waveform-modeling-accuracy standards. These potential abuses could result in the use of substandard model waveforms for gravitational-wave data analysis with a consequent loss of scientific information. This paper also outlines a series of steps that should be taken to avoid these problems:

(i) Use the specified waveform-error norms when applying the accuracy standards, not surrogates based on estimates of the maximum time-domain amplitude and phase errors, for example.

(ii) Construct careful error-envelope estimates of the time-domain amplitude and phase errors, and use these in the L^2 norm versions of the accuracy standards given in Eqs. (30) and (31).

(iii) Use the same model waveforms to construct the error-envelope estimates of the waveform-error norms, $\|\delta h_m\|/\|h_m\|$, and the quantities C and $\tilde{C}$ that appear in the combined waveform-modeling and calibration accuracy standards.

The L^2 norm versions of the accuracy standards, e.g., Eqs. (27) and (28), place limits on the allowed values of a particular measure of the waveform error $\|\delta h_m\|/C\|h_m\|$. A disturbing feature of this error measure is the fact that the quantity C can be rather small, resulting in very stringent requirements on the L^2 norm measure of the waveform error, $\|\delta h_m\|/\|h_m\|$. Figure 4 illustrates, for example, that C can be as small as 0.02 for $400M_\odot$ black-hole binary systems and an Advanced LIGO noise curve. This implies that the L^2 norm measure of the waveform errors, $\|\delta h_m\|/\|h_m\|$, must be about 50 times smaller than the requirement on the noise-weighted inner-product measure, $\sqrt{\langle\delta h_m|\delta h_m\rangle}/\sqrt{\langle h_m|h_m\rangle}$. The waveform-accuracy standard for detection in LIGO would require $\|\delta h_m\|/\|h_m\| \lesssim 0.001$ instead of $\sqrt{\langle\delta h_m|\delta h_m\rangle}/\sqrt{\langle h_m|h_m\rangle} \lesssim 0.05$, for example. This disparity in the level of accuracy required for these different error measures seems unreasonable. It is not clear (because no one has actually checked yet) whether any of the currently available model waveforms come close to satisfying the preferred L^2 norm accuracy standard, even the relatively weak standard for detection. Is this right? What is going on?

The L^2 norm versions of the accuracy standards, e.g., Eqs. (27) and (8), were derived by constructing a sequence of rigorous mathematical inequalities starting from the original noise-weighted inner-product version of the standards, Eqs. (5) and (6). Perhaps these inequalities are significantly weaker than optimal, forcing the final L^2 norm accuracy standards to be far more restrictive than they really have to be. Having waveform-accuracy standards that are closer to optimal would be quite desirable, if someone could find them. However, the chain of inequalities leading to Eqs. (30) and (31) appears to be fairly tight. It is possible that a factor of 2 or so has been lost due to suboptimal inequalities, but the loss of a factor of 50 does not seem plausible. It seems likely that the main cause of the disparity in the accuracy requirements lies elsewhere.

What explains then the significantly stricter requirement on the L^2 norm measure of accuracy, $\|\delta h_m\|/\|h_m\|$? It is easy to show that the $\|h_m\|$ term that appears in the denominator of the expression for C, Eq. (29), is primarily responsible for its very small values in the waveforms of large mass binary black-hole systems. The norm $\|h_m\|$ scales with mass as $M^{3/2}$, while the standard signal-to-noise ratio ρ for binary black-hole signals scales approximately as $M^{0.8}$. Thus, C scales approximately as $M^{-0.7}$, as shown in Fig. 4. It is inevitable then that C becomes very small for large values of M. Figure 5 also reveals that C becomes smaller as the length of the model waveform becomes longer. In fact, C would approach zero as the length of the model waveform becomes infinite. This seems very odd.

Why does the accuracy standard become stricter for larger values of the mass and for model waveforms of greater length? Recall that the error measure that appears on the left in the accuracy standards, Eqs. (27) and (28), is $\|\delta h_m\|/C\|h_m\|$ not $\|\delta h_m\|/\|h_m\|$. The $\|h_m\|$ that appears in the denominator of C exactly cancels the $\|h_m\|$ that appears explicitly in $\|\delta h_m\|/C\|h_m\|$. So the size of $\|h_m\|$ (and consequently C) is basically irrelevant to the size of the real error measure $\|\delta h_m\|/C\|h_m\|$. The requirement on the L^2 norm error measure $\|\delta h_m\|/\|h_m\|$ only becomes excessively small when the reference norm $\|h_m\|$ becomes excessively large. Two questions arise immediately: Why does it make sense to introduce C at all then? Is there a real discrepancy between the waveform-accuracy requirement expressed in terms of the L^2 norm error measure $\|\delta h_m\|/C\|h_m\|$, and the noise-weighted measure $\sqrt{\langle\delta h_m|\delta h_m\rangle}/\sqrt{\langle h_m|h_m\rangle}$?

Consider first the question of why it makes sense to include the quantity C in the statements of the accuracy standards. The accuracy standard for measurement, Eq. (27), could be rewritten by replacing C with its definition from Eq. (29):

$$\|\delta h_m\| \leq \frac{\sqrt{\min S_n(f)}}{2\sqrt{2}}. \tag{34}$$

This expression is quite simple, but it has the disadvantage that it only applies when the model waveforms and their errors are scaled correctly: by the distance to the waveform's source. The model-waveform simulation community generally computes only the scaled waveform rh_m/M and its scaled error $r\delta h_m/M$. What distance r should be used when determining whether a given model waveform satisfies the standards?

It would be more convenient to write the accuracy standards in a way that can be applied to model waveforms with any scaling. A natural way to do this is to introduce as a natural scale, the L^2 norm of the model waveform itself, $\|h_m\|$. In this case the basic error standard, Eq. (34), can be rewritten as

$$\frac{\|\delta h_m\|}{\|h_m\|} \leq \frac{\sqrt{\min S_n(f)}}{2\sqrt{2}\|h_m\|}. \tag{35}$$

The left side of Eq. (35) is therefore the natural scale-invariant L^2 measure of the waveform error. Unfortunately, this trick merely pushes the scale problem to the right side of Eq. (35), where the $\|h_m\|$ term must still be scaled properly with the distance to the source. The quantity $\sqrt{2}\|h_m\|/\sqrt{\min S_n(f)}$ measures the signal-to-noise ratio of the waveform, so specifying its value is equivalent to fixing the distance and setting the waveform scale. Since this is not the standard signal-to-noise measure ρ used by the gravitational-wave data analysis community, it is natural to introduce the quantity C, the ratio of ρ to this nonstandard measure. This allows the accuracy standards to be written in terms of the standard signal-to-noise measure, as in Eqs. (27) and (28). Despite the confusing features of these representations of the standards (as pointed out above), their advantage is that they conveniently depend on the scale of the model waveforms only through the standard signal-to-noise ratio ρ.

Finally, consider the question of whether there is a real discrepancy between the waveform-accuracy requirement expressed in terms of the L^2 norm error measure $\|\delta h_m\|/C\|h_m\|$, and the noise-weighted measure $\sqrt{\langle\delta h_m|\delta h_m\rangle}/\sqrt{\langle h_m|h_m\rangle}$. This question is rather difficult, because the actual model-waveform error δh_m will never be known exactly for real numerically simulated waveforms, and so the noise-weighted measure can never be known for the reasons described in Sec. II B. The best that can be done are explorations of the differences using hypothetical waveform errors, and gaining experience by applying the proper L^2 norm error measures to real model waveforms. At the present time it appears that none of the waveform simulation groups have actually used these new methods to analyze the accuracy of their waveforms. Consequently, there is no direct experience yet on just how restrictive the L^2 norm accuracy requirement actually is, or whether the simulations currently being performed produce waveforms accurate enough according to this measure for LIGO's data analysis needs.

Until the L^2 norm accuracy measures are explored directly by many (most, all) simulation groups using their model waveforms, the best that can be done are explorations of how restrictive the $\|\delta h_m\|/C\|h_m\|$ error measure is in a more hypothetical context. The graphs showing estimates of the waveform errors, $\delta\chi_m(t)$ and $\delta\Phi_m(t)$, in Ref. [4], suggest that the errors in binary black-hole waveforms are largest during the merger phase when the amplitude of the wave $A_m(t)$ is largest. Thus, it seems plausible that carefully constructed error-envelope functions $\delta\chi_E(t)$ and $\delta\Phi_E(t)$ for these simulations will be similarly peaked near the maximum of the waveform amplitude $\max A_m(t)$. Consider hypothetical error-envelope functions of the form

$$\delta\chi_E(t) = \delta\mu_\chi\left[\frac{A_m(t)}{\max A_m(t)}\right]^p, \tag{36}$$

$$\delta\Phi_E(t) = \delta\mu_\Phi\left[\frac{A_m(t)}{\max A_m(t)}\right]^p, \tag{37}$$

where $p > 0$ is a constant that determines how narrowly peaked in time the errors are, and $\delta\mu_\chi$ and $\delta\mu_\Phi$ are the maximum time-domain waveform errors. Using these hypothetical error-envelope functions, it is straightforward to evaluate the error measure $\|\delta h_m\|/C\|h_m\|$, or more precisely the error-envelope version of this measure from the right sides of Eqs. (30) and (31). It is not possible, or at least it is not relevant (as shown in Sec. II B), to compare this error measure with the noise-weighted norm measure $\sqrt{\langle\delta h_m|\delta h_m\rangle}/\sqrt{\langle h_m|h_m\rangle}$. Instead compare this measure with the more familiar and easy to evaluate (yet also not strictly relevant) maximum time-domain errors by defining the ratio

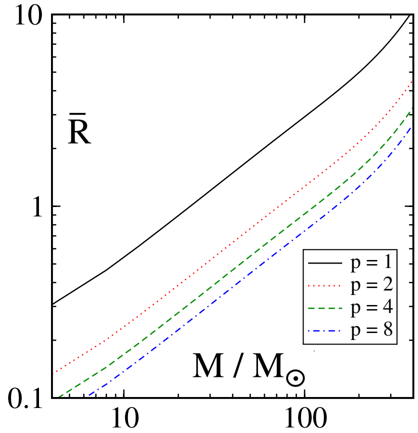

FIG. 7 (color online). Curves illustrate $\bar{R}$, as defined in Eq. (38) for waveform errors of the form $\delta\chi_E = \delta\Phi_E = \delta\mu(A_m/\max A_m)^p$. The quantity $\bar{R}$ measures the ratio between the time-domain L^2 norm measure of the waveform error $\|\delta h\|/C\|h\|$, and the maximum time-domain errors $\sqrt{\delta\mu_\chi^2 + \delta\mu_\Phi^2}$.

$$\bar{R} = \frac{\|A_m\delta\chi_E\cos\Phi_m\| + \|A_m\delta\Phi_E\sin\Phi_m\|}{C\|A_m\cos\Phi_m\|\sqrt{\delta\mu_\chi^2 + \delta\mu_\Phi^2}}. \tag{38}$$

Figure 7 illustrates $\bar{R}$ computed using the waveform-modeling error-envelope functions defined in Eqs. (36) and (37), with several values of the parameter p. This shows that the proper L^2 norm error measures are comparable to the maximum time-domain errors, for waveform-error functions that are narrowly peaked around the time of the maximum waveform amplitude. This example makes it plausible that the accuracy standards based on the proper L^2 norm error measures are not impossibly difficult to achieve for realistic numerical waveform simulations. Many of the currently available waveforms based on numerical simulations may well satisfy the LIGO detection standard. The waveform simulation community needs to explore this further by applying the recently developed waveform-accuracy standards [1,2] to the model waveforms being produced by their groups, and doing this in a way that avoids the misuses of those standards outlined here.

The L^2 norm based accuracy standards, Eqs. (30) and (31), are sufficient to guarantee the waveform accuracy needed for LIGO data analysis, and these standards are probably achievable for realistic waveform errors produced by currently available codes. It may be possible however to improve these standards somewhat in certain cases. For large mass black-hole binaries, only the last few orbits contribute significantly to the waveform in the frequency range where the detector is sensitive. Yet the L^2 norm includes errors from the full length of whatever waveform is tested. The L^2 norm based standards are overly restrictive therefore when unnecessarily long waveforms are tested, and this will be most pronounced for large mass systems. Thus, the L^2 norm based accuracy standards could be made more optimal by limiting their use to waveforms having the appropriate length.

What is the optimal length for a gravitational waveform? If the waveforms used for gravitational-wave data analysis are too short, their frequency-domain counterparts would not be accurate enough to describe the complete waveform in the full range of frequencies accessible to the gravitational-wave detector. If the waveforms were too long, their L^2 norm error estimates would include unnecessary contributions from the early parts of the waveform that can have no measurable influence on the detector. The optimal length is therefore the shortest waveform whose frequency-domain counterpart is accurate enough in the frequency range where the detector is most sensitive. The problem of turning these basic principles into useful specifications for optimal waveform length has yet to be studied properly. These optimal lengths will depend in complicated ways on the total mass of the binary system (which sets the frequency scale of the waveform), on the noise characteristics of the particular detector (which sets the relevant range of frequencies), on the method used to cut off the early part of the waveform (which determines the amount of Gibbs phenomenon produced in the frequency domain), and on the numerical accuracy requirement on the waveform. These issues will be addressed in a future study of this problem.

One final recommendation: The waveform simulation community should compute and publish graphs of the quantities C and $\tilde{C}$ whenever new waveforms are published. The quantity $\tilde{C}$, defined in Eq. (32), is needed to facilitate the decisions the LIGO experimental and data analysis communities must make about setting the appropriate calibration error levels for the detector [2].

ACKNOWLEDGMENTS

I thank B. Owen for helpful conversations concerning this work, and M. Scheel and U. Sperhake for helpful comments and suggestions on an earlier draft of this paper. This research was supported in part by a grant from the Sherman Fairchild Foundation, by NSF Grant Nos. DMS-0553302, PHY-0601459, and PHY-0652995, and by NASA Grant No. NNX09AF97G.

[1] L. Lindblom, B. J. Owen, and D. A. Brown, Phys. Rev. D **78**, 124020 (2008).
[2] L. Lindblom, Phys. Rev. D **80**, 042005 (2009).
[3] S. Anderson *et al.*, LIGO Report No. LIGO-T010095-00-Z, LIGO Project 2001, http://www.ligo.caltech.edu/docs/T/T010095-00.pdf.
[4] M. A. Scheel, M. Boyle, T. Chu, L. E. Kidder, K. D. Matthews, and H. P. Pfeiffer, Phys. Rev. D **79**, 024003 (2009).
[5] M. Boyle, D. A. Brown, and L. Pekowsky, Classical Quantum Gravity **26**, 114006 (2009).
[6] GWINC: Gravitational Wave Interferometer Noise Calculator, v1 default parameters, http://lhocds.ligo-wa.caltech.edu:8000/advligo/GWINC.
[7] A. Lazzarini and R. Weiss, LIGO Report No. LIGO-E950018-02-E 1995; See also, http://www.ligo.caltech.edu/jzweizig/distribution/LSC_Data/.