Formulation improvements for critical collapse simulations

Daniela Cors,[1] Sarah Renkhoff,[1] Hannes R. Rüter,[2] David Hilditch,[2] and Bernd Brügmann[1]
[1]*Friedrich-Schiller-Universität Jena, 07743 Jena, Germany*
[2]*Centro de Astrofísica e Gravitação—CENTRA, Departamento de Física, Instituto Superior Técnico—IST, Universidade de Lisboa—UL, Avenida Rovisco Pais 1, 1049-001 Lisboa, Portugal*

(Received 7 August 2023; accepted 16 October 2023; published 7 December 2023)

The precise tuning required to observe critical phenomena in gravitational collapse poses a challenge for most numerical codes. First, threshold estimation searches may be obstructed by the appearance of coordinate singularities, indicating the need for a better gauge choice. Second, the constraint violations to which simulations are susceptible may be too large and force searches to terminate prematurely. This is a particularly serious issue for first-order formulations. We want our adaptive pseudospectral code BAMPS to be a robust tool for the study of critical phenomena so, having encountered both of these difficulties in work on the vacuum setting, we turn here to investigate these issues in the classic context of a spherically symmetric massless scalar field. We suggest two general improvements. We propose a necessary condition for a gauge choice to respect discrete self-similarity (DSS). The condition is not restricted to spherical symmetry and could be verified with any $3+1$ formulation. After evaluating common gauge choices against this condition, we suggest a DSS-compatible gauge source function in generalized harmonic gauge (GHG). To control constraint violations, we modify the constraint damping parameters of GHG, adapting them to collapse spacetimes. This allows us to improve our tuning of the critical amplitude for several families of initial data, even going from 6 up to 11 digits. This is the most precise tuning achieved with the first-order GHG formulation to date. Consequently, we are able to reproduce the well-known critical phenomena as well as competing formulations and methods, clearly observing up to three echoes.

DOI: 10.1103/PhysRevD.108.124021

I. INTRODUCTION

The threshold of gravitational collapse separates spacetimes at the verge of black hole (BH) formation from those in which a black hole forms. Spacetimes near this region in solution space have large, dynamical curvature and, depending on the specific model under consideration, infinitesimal black holes. This makes them extremely interesting, albeit hard to treat. In 1993, Choptuik [1] tackled the problem by evolving massless scalar fields minimally coupled to the Einstein field equations in spherical symmetry. He observed three features bearing a strong resemblance to those observed near critical points in other fields of physics. He noted that, near the threshold, the spacetime becomes discretely self-similar (DSS), a periodic fractal-like behavior also referred to as echoing. Second, as a result of scale invariance, scalars like the black hole mass M_{BH} on one side of the threshold and the Kretschmann scalar $I = R_{abcd}R^{abcd}$ on the other obey a power law of the form

$$M_{\mathrm{BH}} \sim |p - p_\star|^\lambda, \qquad {I_{\mathrm{max}}}^{-1/4} \sim |p - p_\star|^\lambda, \tag{1}$$

as a function to the parametric distance to the threshold, labeled here by the critical parameter $p_\star$. In the DSS case, this scaling behavior appears with a superimposed Δ-periodic wiggle [2,3]. Third, by evolving different families of data in spherical symmetry, Choptuik concluded that these features are universal. In particular, the scaling exponents λ in the power laws (1), as well as the DSS echoing periods Δ, were independent of the initial data. Notably, all the families of initial data, when evolved close enough to the threshold, share a common configuration, now referred to as the Choptuik spacetime. A spacetime with this specific symmetry has since been proven to exist [4]. These three features (self-similarity, power-law scaling, and universality) constitute what is now known as critical phenomena in gravitational collapse.

Critical phenomena in gravitational collapse have since then been observed in various matter models, mostly but not exclusively in spherical symmetry (see [5] for a detailed review). Families of initial data with a varying parameter p are evolved numerically such that for small values of p, the evolution eventually leads to flat spacetime, whereas large values lead to horizon formation. Following this end-state classification one can bisect towards a better estimation of the threshold parameter p_* delimiting the threshold of collapse. In this way one can measure the distance to the threshold by the precision with which p_* is tuned. The more digits are known, the closer to the threshold the

spacetime is and the more likely we are to observe critical phenomena.

Choptuik achieved an exquisite 13- (or more) digit tuning of p_* by using a maximally constrained formulation of general relativity (GR) with zero shift and areal radius in an adaptive mesh finite-differencing code. Subsequently, this setup was studied with a variety of formulations of GR tailored to spherical symmetry. For instance, Garfinkle and Duncan [6] used null coordinates, and Martín-García and Gundlach [7] constructed coordinate systems adapted to self-similar spherically symmetric spacetimes. There are a number of studies of the scalar-field model with aspherical perturbations. It was demonstrated numerically that at the linear level there is only one growing-mode solution [8]. In pioneering work [9], nonlinear axisymmetric perturbations of the Choptuik spacetime were studied, and their findings were confirmed much more recently by Baumgarte in [10] with a 13-digit tuning. In the latter, the BSSN formulation of the Einstein field equations was used with $1 + \log$ slicing and Γ-driver shift condition. Fully 3d nonlinear perturbations have been studied in [11,12], understandably with far less tuning to the threshold. Of these, Deppe *et al.*'s paper [12] is of particular interest to us, because they use a similar setup to ours. They studied the collapse of massless scalar fields using the generalized harmonic gauge (GHG) formulation of GR using a pseudospectral approximation. Coordinates that gradually zoom into the center of the computational domain and a new gauge source function were employed. They managed to properly reproduce Choptuik's results, for the first time with a pseudospectral code, tuning up to 6 digits.

Choptuik's results are relevant to our understanding of cosmic censorship, as his spacetime should contain a naked singularity, but generic physical data are not spherical, and so assessing the generality of his findings without spherical symmetry is fundamental. The 6-digit level of tuning of [12] gives echoing periods and scaling exponents consistent with those of spherical symmetry, whereas the finer-tuned data of [9,10] find that these scalars are dependent on the aspherical deviation of the initial data. To assess whether the different results are due to the difference in tuning, we need to improve the best level of tuning with GHG and a pseudospectral setup.

Although nonspherical models are generally more difficult to treat, an understanding is emerging. In vacuum [13–18], in the presence of electromagnetic waves [19,20] and for scalar fields as mentioned above, numerical evidence now consistently suggests a deviation from the spherical phenomenology. Confidence in some of these results, for instance whether exact DSS occurs beyond spherical symmetry, relies on the degree of fine-tuning to the threshold, as for instance a minimum number of periods needs to be observed to assess DSS. There is however room for improvement. For example, in the challenging case of gravitational waves, the best tuning [15,16] is around 6 digits. It should be noted that for one of the two common families evolved in these works, the oblate centered Brill wave, the tuning in [16] performed with first-order GHG was a digit further from the threshold than the one performed with BSSN in [15].

Here, we pave the way to improved studies of the threshold of collapse without symmetry, focusing on a subset of the obstructions to tuning we found in [13,16] for Brill waves, but retreating to the classic spherical scalar-field setup as a testbed. In those searches we had to stop tuning either because our code failed to stably evolve the fields until they dispersed or formed an horizon, or because our postprocessing apparent horizon finder AHloc3d [21] could not locate the presumably present, but strangely shaped, horizons that formed. Work concerning apparent horizon finders is ongoing and will be reported elsewhere. In this paper, we focus instead on solving the problems that prevented our code from reliably evolving extreme data.

Some families of data were prone to develop undesirable coordinate features. A shift with a large gradient would be the most common sign of this instance. Coordinate singularities have been reported to be the main obstacle of other studies of critical collapse whether performed with BSSN or GHG. For BSSN evolutions this motivated the use of "quasimaximal" slicing [14] and of the shock-avoiding coordinate conditions [17,22–24]. In GHG, the remaining gauge freedom lies in the choice of gauge source functions. The authors of [12] report that to avoid divergences, they needed to employ substantially modified gauge source functions as compared with the standard choice [25] for binary spacetimes. Our first goal here is therefore to look for a suitable gauge. Similar in spirit to [26], we look for a coordinate choice compatible with DSS. In particular we find and check a necessary condition for compatibility. We then construct a gauge source function that satisfies this condition.

Large constraint violations were present in the failed evolutions of most unclassified data. Even using adaptive mesh refinement to help control error, the closer the spacetimes were to the threshold, the more constraint violations we observed. For the examples considered here this is a particularly serious problem, probably since the first-order formulation we employ necessarily introduces a large number of constraints. Our second goal was therefore to reassess and adjust the constraint damping parameters of GHG, particularly focusing on the first-order reduction constraints in the context of collapsing spacetimes. These adjustments make a very significant improvement to our best tuning and indeed enable our code to display the well-known phenomena first discovered by Choptuik.

In Sec. II, we present the numerical and physical setup of our pseudospectral code BAMPS, which was used to carry out all the present simulations. In Sec. III we derive a DSS-compatibility condition, evaluate most gauge choices against it and suggest a compatible gauge source function.

In Sec. IV we present a mode analysis and the resulting improved constraint damping parameters for GHG. Finally, in Sec. V we apply these modifications to simulations of massless scalar fields in spherical symmetry. In Sec. VI we give a brief summary.

II. SETUP

A. Code BAMPS

All the simulations we present in this paper were carried out with the pseudospectral code BAMPS [27]. Its grid is divided into three main patches: a cubed box at the center of the simulation domain, a transition shell in the shape of a cubed sphere, and a spherical shell circling the domain up to the outer boundary (see Fig. 1). Each of these patches is itself divided into cells, inside of which the evolved fields are stored on Chebyshev-Gauss-Lobatto collocation points.

hp-adaptive mesh refinement (AMR) controls the resolution that every portion of the fields is granted by locally adapting the number of collocation points. It does so both by adding or removing points in a given cell (p refinement), and by subdividing or joining cells (h refinement). In order to evaluate the refinement needed, AMR can assess the error produced during the simulation and also the smoothness of the evolved fields. Details of the refinement scheme are given in [28].

Neighboring cells communicate with each other through characteristic fields using the penalty method. At the outer boundary, we use the constraint preserving boundary conditions described in [27,29]. Symmetries are handled by the cartoon method, using double cartoon for spherically symmetric runs.

In BAMPS we employ a standard free-evolution approach. The constraints are only explicitly solved for the initial data but not at any time step during the evolution. The scalar-field initial data are generated by an elliptic solve integrated within BAMPS, which uses the hyperbolic relaxation method [30].

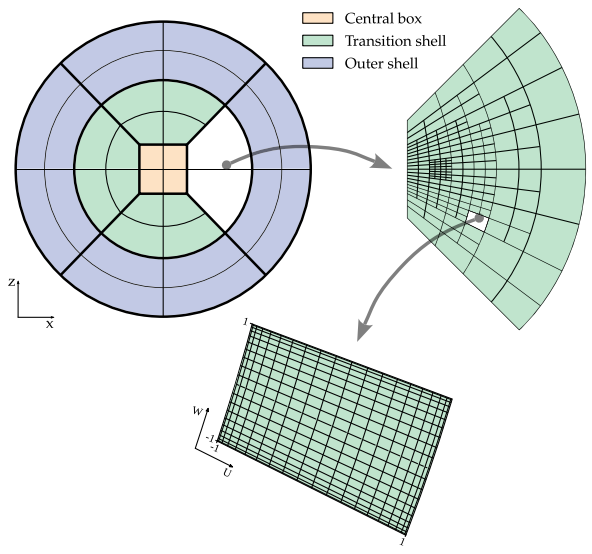

FIG. 1. The BAMPS grid structure illustrating h-refinement.

Time evolution is carried out using a fourth-order Runge-Kutta method. The time steps are adjusted during the evolution in order to preserve the Courant-Friedrich-Lewy condition.

A typical simulation takes place in the BAMPS grid where the central cube patch finishes at code unit 3, where the transition cubed sphere starts, until code unit 14. The outer boundary where the spherical shell ends is at code unit 30. The simulations use 21 to 31 points per grid, which can be refined up to 15 levels of 2-to-1 refinement, as explained in [28]. If the truncation error estimate for g_{ab} or ϕ is higher than 10^{-9}, or 10^{-10} for some runs, then h refinement will refine the grids, or/and p refinement will add more points to each grid. If it is lower than 10^{-12} then grids will be coarsened and/or points will be removed. The equivalent range when we use the smoothness refinement indicator for g_{ab} and ϕ is 0.001–0.0005. In all simulations we use the filtering as described in [27]. Simulations run until coordinate time 30, although the latest time of classification we have encountered occurred at around coordinate time 11.

B. Physical system

In geometric units the trace-reversed Einstein field equations read

$$R_{ab} = 8\pi\left(T_{ab} - \frac{1}{2}g_{ab}T\right), \tag{2}$$

where g_{ab} is the 4d metric, R_{ab} the Ricci tensor, T_{ab} the energy momentum tensor, and $T = g^{ab}T_{ab}$ its trace. Time evolutions are enabled by a $3+1$ splitting of the 4d metric g_{ab}. This gives the 3d spatial metric γ_{ij} and a normal unit vector $n^a = \alpha^{-1}(1, -\beta^i)$, where α is the lapse and β^i the shift. In these variables the line element becomes

$$ds^2 = -\alpha^2 dt^2 + \gamma_{ij}(\beta^i dt + dx^i)(\beta^j dt + dx^j). \tag{3}$$

We denote 4d-component indices with Latin letters starting from a and 3d spatial ones starting from i.

We consider a massless real scalar field φ minimally coupled to the Einstein field equations. The corresponding scalar-field energy momentum tensor is given by

$$T_{ab} = \nabla_a\varphi\nabla_b\varphi - \frac{1}{2}g_{ab}(\nabla^c\varphi\nabla_c\varphi). \tag{4}$$

C. Initial data

In this work we consider two families of initial data, one starting at a moment of time symmetry, the other with a predominantly incoming pulse. The moment of

time-symmetry data is constructed from a scalar field with a Gaussian profile of the form

$$\varphi = A(e^{-(r+R_0)^2} + e^{-(r-R_0)^2}), \tag{5}$$

and with a vanishing derivative along n^a:

$$n^a \nabla_a \varphi = 0. \tag{6}$$

For moment of time-symmetry initial data we use $R_0 = 0$ as in [10]. Observe that (5) implies that our A is a half of the equivalent parameter η used in [10].

For the incoming initial data we use the same initial scalar field as in [12]. Thus, we have

$$\varphi = A\frac{e^{-(r+R_0)^2} - e^{-(r-R_0)^2}}{r} \tag{7}$$

and

$$n^a \nabla_a \varphi = 2A\frac{(r+R_0)e^{-(r+R_0)^2} + (r-R_0)e^{-(r-R_0)^2}}{r}, \tag{8}$$

with $R_0 = 5$ and the appropriate limits taken at the origin.

To solve the Arnowitt Deser Misner (ADM) constraints we make a conformal decomposition of the metric and solve the extended conformal thin sandwich equations (XCTS) [31,32]. We choose the conformal metric to be flat, $\bar{\gamma}_{ij} = \delta_{ij}$, and take its time derivative to vanish, $\partial_t \bar{\gamma}_{ij} = 0$. Furthermore, we impose maximal slicing on the initial data, so that both the trace of the extrinsic curvature and its time derivative vanish, $K = \gamma^{ij} K_{ij} = 0$, $\partial_t K = 0$. After fixing these variables the XCTS equations become a set of coupled elliptic partial-differential equations for the conformal factor ψ, the shift β^i, and the lapse α.

At the outer boundary, $\partial\Omega$, we impose Robin boundary conditions compatible with a $1/r$ decay for ψ, β^i, and α. Concretely, the boundary conditions are given by

$$s^i \partial_i \psi|_{\partial\Omega} = \frac{1-\psi}{r}, \tag{9}$$

$$s^i \partial_i \alpha|_{\partial\Omega} = \frac{1-\alpha}{r}, \tag{10}$$

$$s^i \partial_i \beta^j|_{\partial\Omega} = \frac{\beta^j}{r}, \tag{11}$$

where $s^i = \bar{\gamma}^{ij}\partial_j r/L$ is the spatial normal to the outer boundary, with the normalization factor L such that $\bar{\gamma}^{ij} s_i s_j = 1$.

For incoming initial data we have to solve the full coupled set of the XCTS equations. For moment of time-symmetry initial data on the other hand, we only have to solve one equation for ψ. Thanks to the choice in Eq. (6) the solutions for shift and lapse are trivially given by $\beta^i = 0$ and $\alpha = 1$. The remaining equation for ψ has the form

$$0 = \delta^{ij}\partial_i\partial_j\psi + \pi\psi\delta^{ij}\partial_i\varphi\partial_j\varphi, \tag{12}$$

where π here is the mathematical constant not to be confused with the field π introduced later. We use the hyperbolic relaxation method [30] implemented in BAMPS to solve the XCTS equations or Eq. (12).

D. Evolution equations

1. Scalar-field evolution equations

The equation of motion of a massless scalar field is given by the Klein-Gordon equation, $\nabla^a\nabla_a\varphi = 0$. We work under a first-order reduction with π as the time reduction variable $n^a\partial_a\varphi$, and χ_i as the spatial reduction variable associated with the reduction constraint $S_i := \partial_i\varphi - \chi_i = 0$. This yields the first-order massless Klein-Gordon evolution system:

$$\partial_t\varphi = \alpha\pi + \beta^i\chi_i, \tag{13}$$

$$\begin{aligned}\partial_t\pi = {}& \beta^i\partial_i\pi + \gamma^{ij}(\chi_j\partial_i\alpha + \alpha\partial_i\chi_j - \alpha^{(3)}\Gamma^k{}_{ij}\chi_k) \\ & + \alpha\pi K + \sigma\beta^i S_i,\end{aligned} \tag{14}$$

$$\partial_t\chi_i = \pi\partial_i\alpha + \alpha\partial_i\pi + \chi_j\partial_i\beta^j + \beta^j\partial_j\chi_i + \sigma\alpha S_i, \tag{15}$$

where σ is a damping term and the spatial connection is denoted by ${}^{(3)}\Gamma^i{}_{jk}$.

2. GHG evolution equations

In BAMPS, the Einstein field equations are formulated with the first-order reduction of the GHG formulation [33]. The evolved variables are the metric components g_{ab}, the time reduction variable Π_{ab} corresponding to $-n^d\partial_d g_{ab}$, and the spatial reduction variable Φ_{iab} associated with the reduction constraint $C_{iab} := \partial_i g_{ab} - \Phi_{iab} = 0$. The evolution equations are

$$\partial_t g_{ab} = \beta^i\partial_i g_{ab} - \alpha\Pi_{ab} + \gamma_1\beta^i C_{iab}, \tag{16}$$

$$\begin{aligned}\partial_t\Pi_{ab} = {}& \beta^i\partial_i\Pi_{ab} - \alpha\gamma^{ij}\partial_i\Phi_{jab} + \gamma_1\gamma_2\beta^i C_{iab} \\ & + 2\alpha g^{cd}(\gamma^{ij}\Phi_{ica}\Phi_{jdb} - \Pi_{ca}\Pi_{db} - g^{ef}\Gamma_{ace}\Gamma_{bdf}) \\ & - 2\alpha\left(\nabla_{(a}H_{b)} + \gamma_4\Gamma^c{}_{ab}C_c - \frac{1}{2}\gamma_5 g_{ab}\Gamma^c C_c\right) \\ & - \frac{1}{2}\alpha n^c n^d\Pi_{cd}\Pi_{ab} - \alpha n^c\gamma^{ij}\Pi_{ci}\Phi_{jab} \\ & + \alpha\gamma_0(2\delta^c{}_{(a}n_{b)} - g_{ab}n^c)C_c \\ & - 16\pi\alpha\left(T_{ab} - \frac{1}{2}g_{ab}T^c{}_c\right),\end{aligned} \tag{17}$$

$$\partial_t \Phi_{iab} = \beta^j \partial_j \Phi_{iab} - \alpha \partial_i \Pi_{ab} + \gamma_2 \alpha C_{iab} + \frac{1}{2}\alpha n^c n^d \Phi_{icd}\Pi_{ab} + \alpha \gamma^{jk} n^c \Phi_{ijc}\Phi_{kab}. \quad (18)$$

The canonical constraint damping parameters are $\gamma_1 = -1$ and $\gamma_4 = \gamma_5 = 1/2$ as introduced in [27]. We use $\alpha\gamma_0 = 2$ for incoming data and $\alpha\gamma_0 = 4$ for the moment of time-symmetry data. In Sec. IV we examine the role of γ_2, directly analogous to the damping parameter of the scalar field σ, in relation to the spatial reduction constraint. The harmonic constraint corresponds to

$$C_a := H_a + \Gamma_a = 0, \quad (19)$$

where H_a is the gauge source function, which we are free to specify (see Sec. II E). As pointed out in [33], $\{C_a, \partial_t C_a\} = \{0, 0\}$ encodes the Hamiltonian and momentum constraints, thereby ensuring both that the coordinates are harmonic and that the Einstein field equations are satisfied. The system of equations (16)–(18) is symmetric hyperbolic, ensuring a well-posed initial boundary-value problem provided suitable initial and boundary values.

E. Gauge source functions

Harmonic coordinates satisfy

$$\Box x^a = -\Gamma^a = 0. \quad (20)$$

The GHG formulation generalizes this gauge to a less restrictive condition with the addition of a source term H_a to the right-hand side of (20). This generalization is satisfied as long as the harmonic constraint (19), $C_a = 0$, is satisfied. All the gauge freedom of GHG is contained in H_a. When the harmonic constraint is satisfied, the expressions of Γ^a in standard $3+1$ formalism give the following evolution equations for the lapse and shift:

$$\mathrm{d}_t \alpha = -\alpha^2 (n^a H_a + K), \quad (21)$$

$$\mathrm{d}_t \beta^i = \alpha^2 (H^i + {}^{(3)}\Gamma^i - \partial^i \ln \alpha), \quad (22)$$

where $\mathrm{d}_t = \alpha n^a \partial_a = \partial_t - \beta^k \partial_k$ and the 3d contracted spatial connection is ${}^{(3)}\Gamma^i = {}^{(3)}\Gamma^i{}_{jk}\gamma^{jk}$.

To guarantee symmetric hyperbolicity, the arbitrary function of spacetime H_a cannot contain derivatives of the fields. In that sense the gauge freedom of GHG is rather limited. We assess here in detail the choice of harmonic damped wave gauge (HDWG) as well as two modifications of it used in studies of critical collapse.

1. HDWG

As explained in [25], the harmonic condition (20) can lead to strong gauge dynamics when the physical degrees of freedom are very dynamical. An approach to combat this effect is to reduce those gauge dynamics, with the time and spatial coordinates subject to slightly different considerations. For instance, to suppress gauge dynamics associated with the spatial coordinates, the shift can directly be damped through the gauge source function, yielding a shift condition very similar to the Γ-driver one (57). The equations for harmonic slicing and a shift damped by a factor η in GHG are

$$H_a = -\eta \gamma_{ai} \beta^i \alpha^{-1}, \quad (23)$$

$$\mathrm{d}_t \alpha = -\alpha^2 K, \quad (24)$$

$$\mathrm{d}_t \beta^i = -\alpha \eta \beta^i + \alpha^2 ({}^{(3)}\Gamma^i - \partial^i \ln \alpha). \quad (25)$$

This choice of H_a only damps the spatial gauge dynamics, but an extra term is necessary in order to damp the temporal gauge dynamics. Besides, the authors of [25] report a concern about the growth of $\alpha^{-1}\sqrt{\gamma}$ in some simulations, where $\gamma = \det(\gamma_{ij})$ is the spatial volume element. Since $n^a H_a = n^a \partial_a \ln(\alpha^{-1}\sqrt{\gamma}) - \alpha^{-1}\partial_k \beta^k$, they suggest to fix $n^a H_a = -\eta_L \ln(\alpha^{-1}\sqrt{\gamma})$ in order to exponentially suppress $\alpha^{-1}\sqrt{\gamma}$. This is the well-known HDWG for GHG [see Eq. (A15) in [25]]:

$$H_a = \eta_L \ln(\alpha^{-1}\sqrt{\gamma}) n_a - \eta_S \gamma_{ai} \beta^i \alpha^{-1}, \quad (26)$$

$$\mathrm{d}_t \alpha = \alpha^2 (\eta_L \ln(\alpha^{-1}\sqrt{\gamma}) - K), \quad (27)$$

$$\mathrm{d}_t \beta^i = -\alpha \eta_S \beta^i + \alpha^2 ({}^{(3)}\Gamma^i - \partial^i \ln \alpha), \quad (28)$$

with a canonical choice of coefficients being $\eta_L = 1$ and $\eta_S = 2$. Although the HDWG choice of H_a was extremely successful, yielding black hole simulations for the authors of [34], in critical collapse studies we have found that it fell short to handle the spacetimes of interest. Some modifications of this function have shown to be more appropriate for critical collapse studies, as we present below.

2. HDWG-α^2 H_a

In [13], Eqs. (1) and (5), and more recently in [16], Eq. (4), we gave a variant of the HDWG gauge source function characterized by the powers of the lapse as

$$H_a = \eta_L \alpha^{-2} \ln(\alpha^{-1}\sqrt{\gamma}) n_a - \eta_S \gamma_{ai} \beta^i \alpha^{-2}, \quad (29)$$

$$\mathrm{d}_t \alpha = \eta_L \ln(\alpha^{-1}\sqrt{\gamma}) - \alpha^2 K, \quad (30)$$

$$\mathrm{d}_t \beta^i = -\eta_S \beta^i + \alpha^2 ({}^{(3)}\Gamma^i - \partial^i \ln \alpha). \quad (31)$$

These modifications take into account the collapse of the lapse in spacetimes of extreme curvature (see a description in Sec. IV) so that relevant terms survive despite a vanishing lapse. In the shift-evolution equation

one can see that the damping of the spatial gauge dynamics takes place whether or not the lapse collapses. This success inspired one of the modifications we present below, the adjusted damping parameter (see Sec. IV). Large values of η_S had to be used to deal with undesirable coordinate features for some families of data in [16]. Empirically we found $\eta_L = 1$ and $\eta_S = 2$ to be a reliable choice.

3. *HDWG-*$\ln(\alpha)$ H_a

In [12] the authors introduce a new version of the HDWG gauge source function in order to further enhance the suppression of $\alpha^{-1}\sqrt{\gamma}$ and to control the growth of α^{-1}.

$$H_a = R(t)W(x^i)(((\ln(\alpha^{-1}\sqrt{\gamma}))^5 + (\ln(\alpha^{-1}))^5)n_a - (\ln(\alpha^{-1}\sqrt{\gamma}))^4\gamma_{ai}\beta^i\alpha^{-1}), \tag{32}$$

$$\mathrm{d}_t\alpha = R(t)W(x^i)\alpha^2((\ln(\alpha^{-1}\sqrt{\gamma}))^5 + (\ln(\alpha^{-1}))^5 - K), \tag{33}$$

$$\mathrm{d}_t\beta^i = R(t)W(x^i)(-\alpha(\ln(\alpha^{-1}))^4\beta^i) + \alpha^2({}^{(3)}\Gamma^i - \partial^i \ln\alpha), \tag{34}$$

where $R(t)$ is a roll-on function and $W(x^i)$ is a spatial weight function. $R(t)$ allows to start simulations with maximal slicing and then move on smoothly to their modified HDWG while $W(x^i)$ enforces pure harmonic gauge at the boundary. As we are only interested in testing the impact of the modified HDWG, in our tests we consider $R(t) = W(x^i) = 1$. The authors of [12] also mention that for reliable evolutions of critical collapse, the powers in these logarithms are higher than the ones typically needed in binary black hole evolutions.

III. DSS-COMPATIBLE COORDINATES

As hinted in the previous subsection, choosing a gauge for critical collapse simulations is not trivial. In GHG, gauge source functions have been constructed by careful experimentation and used to control divergences and a collapsing lapse. Even then, code failures coinciding with undesirable coordinate features such as a shift with a large gradient, occur. Ultimately this may result in a coordinate singularity. In practice, since these coordinate features trigger large constraint violations they tend to result in a crash before unambiguously determining the presence of such a singularity. Nevertheless, our best hypothesis thus far is that at least a subset of our code crashes are due to these undesirable coordinate features. Therefore, we look for a gauge choice that keeps the lapse and shift "well behaved" during the simulation by reducing any additional coordinate features.

A natural suggestion is to tailor coordinates to the symmetries of the studied system. In the case of DSS spacetimes, using DSS-adapted coordinates would have the key advantage of directly tackling the problem of coordinate singularities since, if one could compute through one complete period, subsequent periods ought to follow without extra gauge features, as long as numerical error were sufficiently well controlled. From the numerical point of view such adapted coordinates are therefore attractive because they should potentially reduce the computational cost to reach a desired error. In [26], the authors study coordinates based on the ADM equations and on the homothetic Killing vector describing the continuous self-similarity (CSS) or DSS present in critical collapse spacetimes. With a particular boundary prescription, they test two gauges adapted to self-similarity in spherical symmetry. They call these coordinates *symmetry seeking*. Unfortunately neither is a simple modification of popular dynamical gauge conditions now known to be well behaved when treating a range of generic $3+1$ dimensional data. Therefore, with the same general motivation, we take a different approach. Starting from the expression of the assumed symmetry—DSS in this case—on the evolved variables, we derive a necessary condition for gauge choices to be *compatible* with DSS (CSS would be similar). This condition is not restricted to spherical symmetry and can be straightforwardly used with any $3+1$ formulation of GR.

A. DSS-compatibility condition

1. *Periodic rescaling*

By definition [5], there exist coordinates (T, x^i) in which the metric of a DSS spacetime displays a periodic conformal rescaling of the form

$$g_{ab}(T, x^i) = e^{-2T}\tilde{g}_{ab}(T, x^i), \tag{35}$$

where the conformal metric $\tilde{g}_{ab}$ is periodic, with fixed period Δ,

$$\tilde{g}_{ab}(T + \Delta, x^i) = \tilde{g}_{ab}(T, x^i). \tag{36}$$

Such coordinates are referred to as *DSS-adapted* coordinates. Slow time is a time coordinate adapted to DSS. It can be thought of as the logarithm of an ever-decreasing spacetime scale. An example of slow time is the one computed as the logarithmic distance of proper time τ to the accumulation time τ_*:

$$T_p = -\ln|\tau_* - \tau|, \tag{37}$$

along a future directed timelike curve that terminates at the accumulation point. (In general the coordinate then needs to be suitably extended away from the observer.) In (T, x^i) coordinates, the characteristic timescale is Δ; therefore, it would be numerically beneficial to evolve the spacetime

using T, avoiding then the need to resolve ever-decreasing timescales.

Obviously, combining (35) and (36) we obtain that after a period Δ, the 4d metric is rescaled as

$$g_{ab}(T+\Delta, x^i) = e^{-2\Delta} g_{ab}(T, x^i). \tag{38}$$

This transformation of the metric can be translated to the $3+1$ variables in (3) as

$$\gamma_{ij}(T+\Delta, x^k) = e^{-2\Delta}\gamma_{ij}(T, x^k), \tag{39}$$

$$\alpha(T+\Delta, x^i) = e^{-\Delta}\alpha(T, x^i), \tag{40}$$

$$\beta^i(T+\Delta, x^j) = \beta^i(T, x^j). \tag{41}$$

The transformation after a period Δ of other relevant quantities can be derived from these, giving

$$\sqrt{\gamma(T+\Delta, x^i)} = e^{-3\Delta}\sqrt{\gamma(T, x^i)}, \tag{42}$$

$${}^{(3)}\Gamma^i(T+\Delta, x^j) = e^{2\Delta\,(3)}\Gamma^i(T, x^j), \tag{43}$$

$$K(T+\Delta, x^i) = e^{\Delta} K(T, x^i). \tag{44}$$

2. Coordinate choice

Let us briefly refer to (39)–(43) holding at all times as (i). A consequence of these, in particular of (40) and (41), is that (ii)

$$\mathrm{d}_T\alpha(T+\Delta, x^i) = e^{-\Delta}\mathrm{d}_T\alpha(T, x^i), \tag{45}$$

$$\mathrm{d}_T\beta^i(T+\Delta, x^j) = \mathrm{d}_T\beta^i(T, x^j), \tag{46}$$

hold at all times too. Thus (ii) is necessary for (i), and in turn for coordinates to be adapted to DSS. It is not a sufficient condition. (ii) does not imply (i) because the time integration required in that step is also affected by the remaining degrees of freedom of GR.

In the $3+1$ split, we choose coordinates by imposing evolution equations for the lapse and shift:

$$\mathrm{d}_t\alpha = F[g_{ab}], \tag{47}$$

$$\mathrm{d}_t\beta^i = G^i[g_{ab}], \tag{48}$$

where the right-hand sides are functionals permitted to depend also upon derivatives of the metric. Given this gauge choice, (ii) implies that (iii)

$$F[g_{ab}(T+\Delta, x^i)] = e^{-\Delta} F[g_{ab}(T, x^i)], \tag{49}$$

$$G^i[g_{ab}(T+\Delta, x^j)] = G^i[g_{ab}(T, x^j)], \tag{50}$$

holds, with the arguments assumed to respect such a symmetry according to (39)–(41), namely (i) holds. (iii) is a necessary condition for (ii), because without (iii) it cannot be the case that (ii) holds for the given gauge choice. By transitivity, (iii) is also a necessary condition of (i). We call (iii) the DSS-compatibility necessary condition for an arbitrary choice of gauge-evolution equations to render the associated coordinates DSS adapted.

Since (ii) is not sufficient to have (i) DSS-adapted coordinates, neither is (iii). Note that (iii) is not even a sufficient condition for (ii) to hold, as without additionally assuming (i), (iii) alone could not yield (ii):

$$(i) \Rightarrow (ii) \Rightarrow (iii),$$
$$(iii) \not\Rightarrow (ii) \not\Rightarrow (i).$$

One might hope that satisfying the DSS-compatible condition would at least remove undesirable gauge features emerging from the incompatibility between the gauge choice and the self-similarity. This in itself would also be a good step towards avoiding coordinate singularities, the best step being of course actually evolving in adapted coordinates as explained at the beginning of this section.

Evolving a spacetime in DSS-adapted coordinates is not needed to assess the presence of exact DSS, as this can be verified after the evolution through coordinate transformations from the evolved coordinates to constructed DSS-adapted ones (see for instance Sec. V C). However, it would not only be desirable from the numerical point of view (potential efficiency and coordinate singularity avoidance), but also because the dynamical construction of DSS-adapted coordinates would be the least ambiguous demonstration of such a symmetry.

In practice we can now use conditions (49) and (50) to assess whether the different gauges that are commonly used in critical collapse studies may be compatible with the symmetry we expect at the threshold of collapse. We refer to gauge choices satisfying the DSS-compatibility condition (49) and (50) as DSS compatible. If a gauge is not DSS compatible, the associated coordinates will not be DSS adapted.

B. Moving puncture gauge

1. Bona-Masso slicing condition

Most finite-difference simulations using the BSSN [35–37] or conformal Z4 formulations [38–40] choose the Bona-Masso [41] family of slicing conditions:

$$\mathrm{d}_t\alpha = -\alpha^2 f(\alpha) K. \tag{51}$$

Harmonic gauge corresponds to $f(\alpha) = 1$. The use of (40) and (44) gives

$$\mathrm{d}_t\alpha(T+\Delta,x^i) = -e^{-\Delta}\alpha^2(T,x^i)K(T,x^i) \tag{52}$$

$$= e^{-\Delta}\mathrm{d}_t\alpha(T,x^i), \tag{53}$$

which satisfies the DSS-compatibility condition (49). More interesting is to consider $1+\log$ slicing, which corresponds to $f(\alpha)=\frac{2}{\alpha}$, as it is the most popular choice. After a period Δ, we obtain

$$\mathrm{d}_t\alpha(T+\Delta,x^i) = -2\alpha(T,x^i)K(T,x^i) \neq e^{-\Delta}\mathrm{d}_t\alpha(T,x^i), \tag{54}$$

which does not satisfy the condition (49). Similarly, shock-avoiding slicing, corresponding to $f(\alpha)=1+\frac{\kappa}{\alpha^2}$ for a given constant κ, also fails to satisfy the DSS-compatibility condition:

$$\mathrm{d}_t\alpha(T+\Delta,x^i) = (-e^{-\Delta}\alpha^2(T,x^i)+e^{+\Delta}\kappa)K(T,x^i) \neq e^{-\Delta}\mathrm{d}_t\alpha(T,x^i). \tag{55}$$

2. Γ-driver shift condition

The conformal nature of the BSSN and conformal Z4 formulations plays a role in the choice of shift condition. These formalisms are expressed in terms of a conformal spatial metric $\bar{\gamma}_{ij}=\chi\gamma_{ij}$ where the conformal factor is proportional to the determinant of the spatial metric $\chi=\gamma^{-1/3}$. The Christoffel symbols associated with the conformal spatial metric $\bar{\Gamma}^i{}_{jk}$ can be contracted with this metric to give ${}^{(3)}\bar{\Gamma}^i=\bar{\gamma}^{jk}{}^{(3)}\bar{\Gamma}^i{}_{jk}$. Using (39)–(43) we obtain

$${}^{(3)}\bar{\Gamma}^i(T+\Delta,x^j) = {}^{(3)}\bar{\Gamma}^i(T,x^j). \tag{56}$$

In terms of this connection, the Γ-driver condition reads as

$$\mathrm{d}_t\beta^i = -\mu_S\,{}^{(3)}\bar{\Gamma}^i - \eta\beta^i, \tag{57}$$

with μ_S a constant. We find that this choice does satisfy the DSS-compatibility condition (50):

$$\mathrm{d}_t\beta^i(T+\Delta,x^i) = \mathrm{d}_t\beta^i(T,x^i). \tag{58}$$

We conclude that the most common gauge choices used in finite-differencing codes will not provide coordinates adapted to DSS due to the slicing condition, with the exception of pure harmonic gauge.

C. Gauge in GHG

Since in GHG the gauge freedom is fixed through the gauge source function H_a, the DSS-compatibility condition (49) and (50) can be rewritten in terms of H_a. Using (39)–(44) in (21) and (22) gives

$$\mathrm{d}_t\alpha(T+\Delta,x^i) = -e^{-\Delta}\alpha^2(T,x^i)(n^a(T,x^i)H_a(T+\Delta,x^i) + K(T,x^i)), \tag{59}$$

$$\mathrm{d}_t\beta^i(T+\Delta,x^j) = \alpha^2(T,x^j)\gamma^{ik}(T,x^j)(H_k(T+\Delta,x^j) + {}^{(3)}\Gamma_k(T,x^j) + \partial_k\ln(\alpha(T,x^j))). \tag{60}$$

So, requiring (49) and (50) is equivalent to requiring

$$H_a(T+\Delta,x^i) = H_a(T,x^i). \tag{61}$$

The HDWG gauge source function does not satisfy the DSS-compatibility condition (61), since

$$H_a^{\mathrm{HDWG}}(T+\Delta,x^i) = -2\Delta\eta_L e^{-\Delta}n_a(T,x^i) + e^{-\Delta}\eta_L\ln\left(\frac{\sqrt{\gamma(T,x^i)}}{\alpha(T,x^i)}\right)n_a(T,x^i) - e^{-\Delta}\eta_S\frac{\gamma_{ak}(T,x^i)\beta^k(T,x^i)}{\alpha(T,x^i)} \neq H_a^{\mathrm{HDWG}}(T,x^i). \tag{62}$$

Similarly, neither variant of HDWG, HDWG-α^2 (29), nor HDWG-$\ln(\alpha)$ (32), satisfies (61). We have instead that Eq. (29),

$$H_a^{\mathrm{HDWG}\text{-}\alpha^2}(T+\Delta,x^i) = -2\Delta\eta_L e^{\Delta}n_a(T,x^i) + e^{\Delta}\eta_L\ln\left(\frac{\sqrt{\gamma(T,x^i)}}{\alpha(T,x^i)}\right)\frac{n_a(T,x^i)}{\alpha^2(T,x^i)} - \eta_S\frac{\gamma_{ak}(T,x^i)\beta^k(T,x^i)}{\alpha^2(T,x^i)} \neq H_a^{\mathrm{HDWG}\text{-}\alpha^2}(T,x^i) \tag{63}$$

and that Eq. (32),

$$\begin{aligned} &H_a^{\mathrm{HDWG}\text{-}\ln(\alpha)}(T+\Delta,x^i) \\ &= e^{-\Delta}n_a(T,x^i)\left\{\left(-2\Delta+\ln\left(\frac{\sqrt{\gamma(T,x^i)}}{\alpha(T,x^i)}\right)\right)^5 + \left(-\Delta+\ln\left(\frac{1}{\alpha(T,x^i)}\right)\right)^5\right\} \\ &\quad - e^{-\Delta}\frac{\gamma_{ak}(T,x^i)\beta^k(T,x^i)}{\alpha(T,x^i)}\left(-2\Delta+\ln\left(\frac{\sqrt{\gamma(T,x^i)}}{\alpha(T,x^i)}\right)^4\right) \\ &\neq H_a^{\mathrm{HDWG}\text{-}\ln(\alpha)}(T,x^i). \end{aligned} \tag{64}$$

We conclude that none of these popular dynamical gauge conditions that have recently been used in studies of critical collapse satisfy the DSS-compatibility condition.

D. DSS-compatible gauge sources

Remaining in the context of GHG from now on, the DSS-compatibility condition (61) can be taken into account to construct gauge sources designed to satisfy it. As can be seen in (63), the variant HDWG-α^2 H_a only fails to satisfy (61) in the factor proportional to η_L, for $a = 0$. For $a = i$, the extra α^{-1}, which was initially added to counteract the collapse of the lapse in the shift-evolution equation of (26), cancels out the factors $e^{-2\Delta}$ that appear in γ_{ij} after a Δ period. However, the same extra α^{-1} added to the first term of (29) with the same intention but regarding this time the slicing condition (21) fails to satisfy the DSS-compatibility condition and needs to be adjusted.

As mentioned in Sec. II E, the authors of [12,25] point out the importance of including natural logarithms of $\sqrt{\gamma}\alpha^{-1}$ in the slicing condition to prevent divergences. Therefore, we keep the logarithms, but suggest to balance out the $e^{\pm\Delta}$ factors inside the logarithm, in order to satisfy (61). This leads to

$$H_a = \eta_L \alpha^{-1} \ln(\alpha^{-3}\sqrt{\gamma}) n_a - \eta_S \gamma_{ai}\beta^i \alpha^{-2}, \tag{65}$$

$$\mathrm{d}_t\alpha = \eta_L \alpha \ln(\alpha^{-3}\sqrt{\gamma}) - \alpha^2 K, \tag{66}$$

$$\mathrm{d}_t\beta^i = -\eta_S \beta^i + \alpha^2({}^{(3)}\Gamma^i - \partial^i \ln\alpha), \tag{67}$$

such that

$$\begin{aligned} H_a^{\text{DSS-comp}}(T+\Delta, x^i) &= \eta_L \ln\left(\frac{\sqrt{\gamma(T,x^i)}}{\alpha^3(T,x^i)}\right)\frac{n_a(T,x^i)}{\alpha(T,x^i)} \\ &\quad - \eta_S \frac{\gamma_{ai}(T,x^i)\beta^i(T,x^i)}{\alpha^2(T,x^i)} \\ &= H_a^{\text{DSS-comp}}(T,x^i). \end{aligned} \tag{68}$$

This is far from a unique choice to fulfill the compatibility condition. One way to assess whether this choice is reasonable is to compare it with existing working gauge source functions. Figure 2 shows, for specific values of the trace of the extrinsic curvature K and determinant of the spatial metric γ, the profiles of $\mathrm{d}_t\alpha$ as functions of the lapse α itself. In this figure we can see the different ways in which variations of HDWG deal with a near-vanishing lapse. For HDWG one might get the impression that $\mathrm{d}_t\alpha$ is always negative, but in fact it becomes positive, in a small range near $\alpha = 0$ and the maximal value is also vanishingly small compared to the other gauges. That means that the HDWG gauge is counteracting the collapse of the lapse less strongly and also only at smaller α. The $\mathrm{d}_t\alpha$ corresponding to HDWG-ln(α) has a larger maximum and becomes positive at a larger α, intuitively resulting in pushing α towards larger values in an evolution, typically to the region in which $\mathrm{d}_t\alpha$ changes sign. In the unlikely case of the lapse somehow dropping to zero exactly it would also stay there.

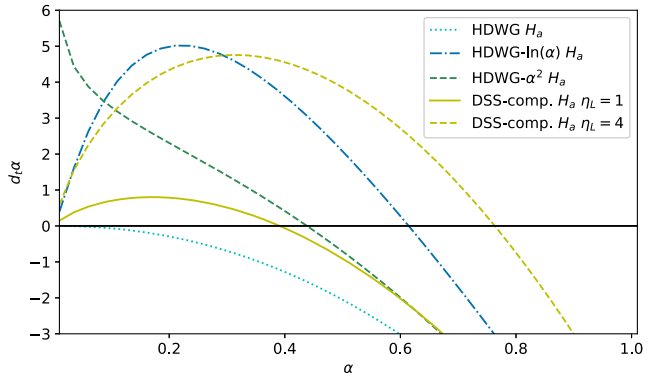

FIG. 2. Time-evolution equations of the lapse as functions of the lapse itself, for the various gauge source functions in Eqs. (27), (30), (33), and (66) with fixed $K = 10$ and $\gamma = 9$. The coefficients η_L for HDWG and HDWG-α^2 are all taken to be 1 because it is the canonical choice and it provides a fair comparison with HDWG-ln(α). We show the slicing condition corresponding to the DSS-compatible gauge source function with its η_L being 1 for the comparison, and also 4, a reliable value.

Very close to $\alpha = 0$ it would be pushed away from zero, but initially only very slowly. The $\mathrm{d}_t\alpha$ corresponding to HDWG-α^2 instead diverges as α approaches zero, an effect of the lack of α^2 in (30). This gauge will therefore push very small α more aggressively to the equilibrium point with $\mathrm{d}_t\alpha = 0$. Both modifications of HDWG successfully fight a vanishing lapse at least until it becomes too small to be dealt with. The DSS-compatible suggestion (66) keeps a factor α in the first term so it results in a very similar profile to the $\mathrm{d}_t\alpha$ from HDWG-ln(α). That can be better seen in Fig. 2 with the curve corresponding to the DSS-compatible gauge and $\eta_L = 4$. Coincidentally, that is also the value we found provided successful simulations.

In Sec. V we present numerical critical collapse evolutions using this gauge with the choice of $\eta_L = 4$ and $\eta_S = 6$. Assessing how well these coordinates adapt to DSS can only be done once the self-similar phase is successfully reached. To ensure we can stably evolve our simulations up to that point, we next treat a second major obstruction, namely constraint violations.

IV. THE REDUCTION CONSTRAINT DAMPING SCHEME

Ensuring that the constraints of the system are well satisfied throughout the evolution guarantees that the numerically evolved data do truly represent the mathematical and physical problem we set out to evolve. Since numerical error cannot be completely avoided, however, constraint violations will still occur. What is needed is a strategy to prevent them from dominating a simulation and potentially even causing the code to crash.

As suggested in [42], some systems of equations admit a damping scheme that reduces these violations by making

the constraint surface (in solution space) an attractor. This was done explicitly for the Z4 and GHG systems in [33,43] by adding to the evolution-equation terms containing multiples of the constraints and freely specifiable parameters. In (16)–(18) these are the terms with γ_0, γ_1, and γ_2.

As mentioned in Sec. I, we are primarily concerned with violations of the reduction constraints, as these are not present in second-order formulations, which yield a finer tuning of the threshold parameter and so are a point of suspicion. Consequently, we restrain our analysis to a linear model for GHG whose only nonprincipal terms are those containing the reduction constraint, plus linear terms.

A. Mode analysis

Taking the first-order GHG evolution equations (16)–(18) in vacuum with harmonic gauge, linearizing about a constraint satisfying background with perturbations that satisfy the harmonic constraint (19), working in the constant coefficient approximation, and finally assuming that $\Pi_{ab} = \Phi_{iab} = 0$ in the background gives

$$\partial_t g_{ab} = \beta^i \partial_i g_{ab} - \alpha \Pi_{ab} + \gamma_1 \beta^i C_{iab}, \tag{69}$$

$$\partial_t \Pi_{ab} = \beta^i \partial_i \Pi_{ab} - \alpha \gamma^{ij} \partial_i \Phi_{jab} + \gamma_1 \gamma_2 \beta^i C_{iab}, \tag{70}$$

$$\partial_t \Phi_{iab} = \beta^j \partial_j \Phi_{iab} - \alpha \partial_i \Pi_{ab} + \gamma_2 \alpha C_{iab}, \tag{71}$$

where here, and for the rest of this section, we overload the notation so that g_{ab}, and so forth, stand for the metric perturbation such that (69)–(71) hold up to first order in the perturbation. The remaining coefficients α, β^i, and γ^{ij} are the constant lapse, shift, and spatial metric in the background. These simplifying assumptions suppress any coupling between different components, making the following analysis algebraically tractable. Here, the analysis deviates from that done in [43] as the lapse and the shift are not taken to be 1 and 0, respectively, but instead to be constant in the background. It differs from the analysis provided in [33] as we here take into account nonprincipal terms. Although it would be interesting to do so, dropping any of the remaining assumptions would make the computation very substantially more complicated and, since we are only trying to motivate a simple change to our constraint damping scheme, does not seem worthwhile here.

A mode analysis of the evolution system can tell us the rate at which the constraint violations grow in this approximation. Writing $\mathbf{u} = (g_{ab}, \Pi_{ab}, \Phi_{iab})^T$, the system of equations (69)–(71) can be written as

$$\partial_t \mathbf{u} = \mathbf{A}^k \partial_k \mathbf{u} + \mathbf{B}\mathbf{u}. \tag{72}$$

Here, $\mathbf{A}^k$ corresponds to the principal part matrices of GHG, whereas $\mathbf{B}\mathbf{u}$ corresponds to the subset of nonprincipal terms that survive linearization and our simplifying assumptions. Observe that the 4d indices ab can be omitted for the analysis.

Moving to the frequency domain, we make a mode ansatz $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}_0 e^{st + i\omega_k x^k}$, where $\tilde{\mathbf{u}}$ is the Laplace-Fourier transform of $\mathbf{u}$, $\omega_k = |\omega|\hat{\omega}_k$ is an arbitrary 3d-vector in the frequency domain, and we write $|\omega| = \sqrt{\gamma^{ij}\omega_i\omega_j}$. The system (72) then takes the form of the following eigenvalue problem:

$$s\tilde{\mathbf{u}} = \mathbf{M}\tilde{\mathbf{u}}, \tag{73}$$

where $\mathbf{M} = i\mathbf{A}^k\omega_k + \mathbf{B}$.

Using the unit frequency vector $\hat{\omega}_i$ ($\hat{\omega}_i\hat{\omega}^i = 1$) and its orthogonal 2d projection operator q_{ij} ($\hat{\omega}^i q_{ij} = 0$), we can further simplify the system with a $2+1$ decomposition:

$$\gamma_{ij} = \hat{\omega}_i\hat{\omega}_j + q_{ij}, \qquad \beta^i = \hat{\omega}^i\beta^{\hat{\omega}} + q^i{}_A\beta^A, \tag{74}$$

where capital Latin letters A, B denote a projection by q_{ij}, running over the 2d plane orthogonal to $\hat{\omega}_k$. Our $2+1$ notation denotes components in the direction of $\hat{\omega}_i$ as $\beta^{\hat{\omega}} = \hat{\omega}_k\beta^k$ and 2d projected ones as $\beta^A = q^A{}_k\beta^k$. Any other vector or covector is decomposed using an analogous notation. In this notation the mode variables in the eigenvalue problem (73) become $\tilde{\mathbf{u}} = (\tilde{g}, \tilde{\Pi}, \tilde{\Phi}_{\hat{\omega}}, \tilde{\Phi}_A)^T$. The matrix $\mathbf{M}$ becomes

$$\begin{pmatrix} i(1+\gamma_1)|\omega|\beta^{\hat{\omega}} & -\alpha & -\gamma_1\beta^{\hat{\omega}} & -\gamma_1\beta^A \\ i\gamma_1\gamma_2|\omega|\beta^{\hat{\omega}} & i|\omega|\beta^{\hat{\omega}} & -i\alpha|\omega| - \gamma_1\gamma_2\beta^{\hat{\omega}} & -\gamma_1\gamma_2\beta^A \\ i\alpha\gamma_2|\omega| & -i\alpha|\omega| & i|\omega|\beta^{\hat{\omega}} - \alpha\gamma_2 & 0 \\ 0 & 0 & 0 & i|\omega|\beta^{\hat{\omega}} - \alpha\gamma_2 \end{pmatrix}. \tag{75}$$

The eigenvalues of $\mathbf{M}$ are

$$\begin{aligned} &s_1 = i|\omega|(\beta^{\hat{\omega}} - \alpha), \quad s_2 = i|\omega|(\beta^{\hat{\omega}} + \alpha), \\ &s_{3,4} = i|\omega|\beta^{\hat{\omega}} - \alpha\gamma_2, \quad s_5 = i|\omega|\beta^{\hat{\omega}}(1+\gamma_1) - \alpha\gamma_2. \end{aligned} \tag{76}$$

The imaginary parts of these eigenvalues represent the speeds of the system's propagation modes and the real parts capture exponential decay or growth. The expected speed of light is captured by s_1 and s_2 for $|\omega| = 1$. More interesting to our analysis is the role of the parameters γ_1 and γ_2 in the evolution of the constraints.

We construct the propagation modes $v = \mathbf{l}\cdot\tilde{\mathbf{u}}$ from the left eigenvectors $\mathbf{l}$ of $\mathbf{M}$ (see Appendix A). In the generic case $\beta^{\hat{\omega}} \neq 0$, the propagation modes corresponding to the eigenvalues that contain damping terms, are

$$v_{s_{3,4}} = \tilde{\Phi}_{Aab} = q_A{}^i\tilde{\Phi}_{iab} = -\tilde{C}_{Aab}, \tag{77}$$

$$v_{s_5} = \beta^A\tilde{\Phi}_{Aab} + \beta^{\hat{\omega}}(\tilde{\Phi}_{\hat{\omega}ab} - i|\omega|\tilde{g}_{ab}) = -\beta^i\tilde{C}_{iab}, \tag{78}$$

where for the last equalities we exploited the orthogonality between ${q_A}^i$ and $\hat{\omega}_i$ in $|\omega|\hat{\omega}_i\tilde{g}_{ab}$. This tells us that some modes of the reduction constraints propagate as $v_{s_{3,4}}$ with speed $s_{3,4}$, whereas some propagate as v_{s_5} with speed s_5. We can then confirm by direct computation that $\tilde{C}_{iab}$ is damped directly by a factor $-\alpha\gamma_2$. Similar results hold in the special cases $\beta^{\hat{\omega}} = 0$ and $\gamma_1 = 0$.

Despite the simplistic nature of this treatment it helps us to understand that there is a clear relation between moments where the lapse collapses and the growth of the constraints, as we explain in the next section.

B. Adjusted damping

1. Collapsing lapse

The mode analysis shows that violations of the reduction constraints, C_{iab}, are damped at a rate $-\alpha\gamma_2$. In the limit where the lapse tends to zero, constraint violations are no longer damped and can start to grow due to numerical round-off. It is an empirical fact that with many popular gauge conditions, the lapse α collapses to zero in the presence of an apparent horizon. Although this is a coordinate-dependent feature, it is so consistent that a collapsing lapse is often used to detect the presence of a black hole. To the best of our knowledge, there is no mathematical proof in which this is demonstrated for GHG, but we do observe such behavior numerically. Intuitively, this means that spatial slices stretch and wrap around a singularity, causing a rapid growth of the radial metric components while proper time gradually freezes inside the horizon, but advances on the outside. This time-freezing effect near a singularity can even be desirable for certain simulation purposes, for instance to steer clear of the singularity.

Although pure harmonic slicing is only "marginally singularity avoiding" in the terminology of Alcubierre [23], our experience with GHG in BAMPS is that every time we have confirmed the detection of an apparent horizon with either harmonic or one of the gauge source functions discussed above, the lapse had collapsed almost to zero, as Fig. 3 shows. Interestingly, we observe the same tendency when the curvature is extreme, even in the absence of an apparent horizon. For instance, in evolutions of subcritical data close to the threshold of gravitational collapse, we observe a nearly vanishing lapse at the time where the curvature invariant peaks. A crucial example of this are our studies of critical phenomena in gravitational collapse.

2. Adjusted damping parameters

Given that the rate of damping of C_{iab} is suppressed as the lapse collapses, and that in critical phenomena simulations, close to the threshold, the lapse consistently collapses, we hypothesize that this vanishing lapse results in the large constraint violations we observe close to the threshold. This may even result in code failures before field dispersion or trapped surface formation, thereby

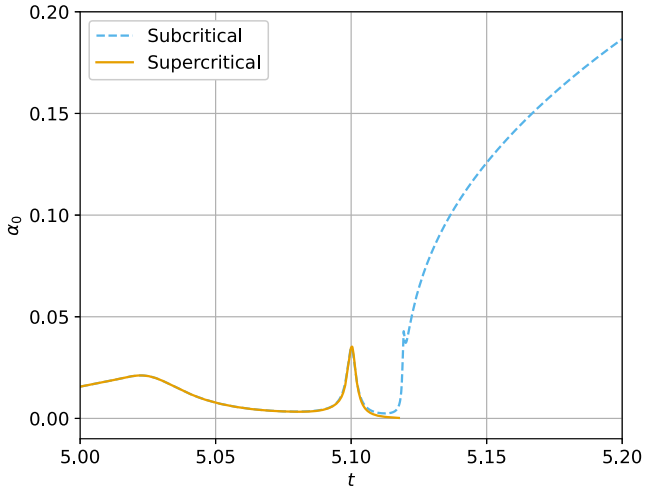

FIG. 3. The value of the lapse at the origin is plotted as a function of time in simulations of moment of time symmetry scalar field initial data. The orange line corresponds to a supercritical simulation that stopped and an apparent horizon was found at the time we observe the lapse reaches 0. The blue dashed line corresponds to a subcritical simulation where, despite a large peak of curvature exactly when the lapse approaches 0, the fields manage to eventually disperse.

obstructing the threshold parameter search. This could account for at least a subset of the shortcomings of first-order GHG in the tuning of the threshold parameter, as there is no reduction constraint to deal with in second-order systems. To test our hypothesis we adjusted the parameter γ_2 in the GHG equations (16)–(18) such that the damping terms in $s_{3,4}$ and s_5 decouple from the lapse:

$$\gamma_2 = c_{\gamma_2} \rightarrow \gamma_2 = \frac{c_{\gamma_2}}{\alpha}, \tag{79}$$

with typically $c_{\gamma_2} = 2$ or 4.

In this way, the damping of the reduction constraint violations can resist a collapsing lapse:

$$s_{3,4} = i|\omega|\beta^{\hat{\omega}} - c_{\gamma_2}, \tag{80}$$

$$s_5 = i|\omega|\beta^{\hat{\omega}}(1+\gamma_1) - c_{\gamma_2}. \tag{81}$$

Essentially with this change the damping timescale is given in terms of coordinate, rather than proper time.

3. Damped constraints

As shown in Fig. 4, using the adjusted damping reduces the constraint violation of both subcritical and supercritical evolutions as desired. Crucially, we can see that it holds the constraints low enough for a previously unclassified spacetime to be classified as subcritical, enabling the continuation of the parameter search. In Sec. V we now demonstrate that this improvement helps in the observation of critical phenomena using first-order GHG.

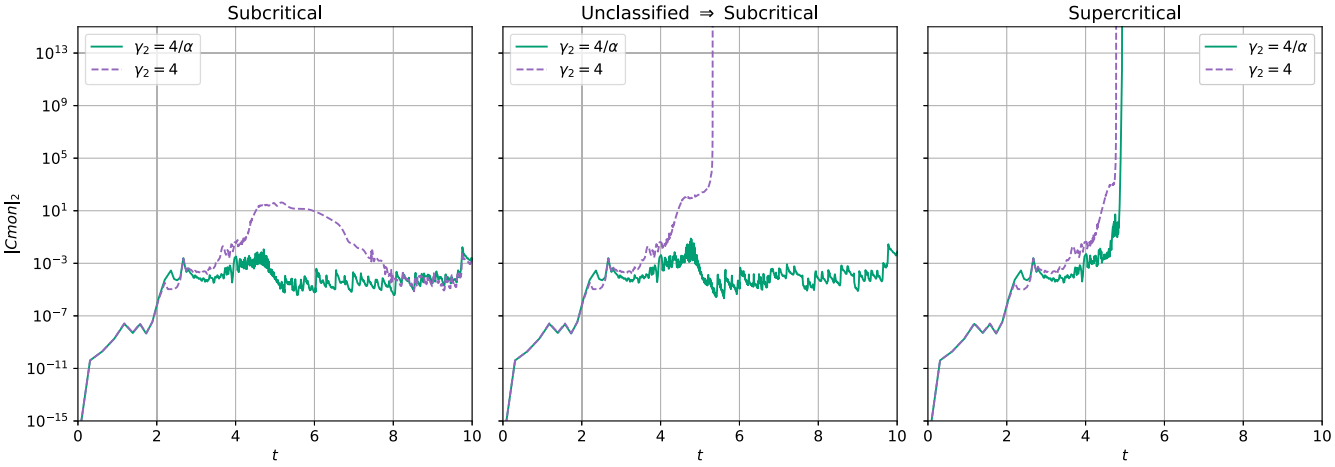

FIG. 4. The L_2 norm of the sum of all constraints (the "constraint monitor") with and without the adjusted damping parameter. Left: for a subcritical amplitude; Middle: for an amplitude that was not classified without the adjusted damping of Eq. (79), but could be classified as subcritical with the adjusted damping; and a supercritical amplitude. The data used in these plots are for configurations 4 and 3 in Table I.

V. CRITICAL COLLAPSE RESULTS

We now present critical collapse results for two one-parameter families of spherical initial data for the massless scalar field.

We treat these spacetimes as a (very challenging) testbed for the code, and as such assume some familiarity with the standard notions of critical collapse. We refer the reader to [5] for a comprehensive overview. The first family is the moment of time-symmetry data, Eqs. (5) and (6), the second incoming initial data, Eqs. (7) and (8). All bisection searches to the critical amplitude follow the same procedure. We classify data as subcritical if they evolve until the fields are fully dispersed. Although our code supports a robust method for spherical black hole excision [44], we are presently only interested in knowing whether or not an apparent horizon forms, so do not treat the subsequent black hole evolution carefully. The code therefore eventually fails for all supercritical data. In spherical symmetry, the presence of apparent horizons can be determined by zero crossings of the expansion, which is algebraic in our evolved variables [45]. We only classify failed evolutions as supercritical if they display such a negative expansion. Starting with a sufficiently wide initial window, we proceed to bisect evolving data whose amplitude is in the middle of the regime. If the evolution is subcritical we update the lower bound and it if its supercritical we update the upper bound. We proceed with the bisection until we reach data that cannot be classified. In all plots the slow time T_p, defined in (37), is computed from the proper time at the origin as

$$\tau(t') = \int_0^{t'} \alpha(t,0)dt, \tag{82}$$

and the accumulation time τ_* following Eq. (23) in [10]. We computed the period Δ using Eq. (24) in [10]. Similarly, we computed the proper length x_p from the x coordinate as

$$x_p(x') = \int_0^{x'} \sqrt{g_{xx}(t,x)}dx. \tag{83}$$

A. Constraint violations

Critical phenomena of a massless scalar field minimally coupled to GR in spherical symmetry have been very well studied and reproduced. It was therefore surprising to us that our initial automated bisections of moment of time-symmetry data with the standard first-order GHG constraint damping setup with HDWG-ln(α) H_a (32), apparently tuning up to 15 digits, displayed at most one and a half echoes. Closely looking at the data, large constraint violations were causing crashes, and noisy data caused incidental negative values of the expansion. The 15-digit tuning was not trustworthy, and the solution space had drifted towards the subcritical side by classifying as supercritical what was simply noisy data without horizons. The last reasonable estimation would only be after a 6-digit tuning, already giving indeed the same echo and a half we observed with the initially wrongly tuned 15 digits.

We obtained comparably poor results with the incoming data, only reliably tuning 7 digits. Those were our best results after increasing the damping parameters and forcing more refinement (to no avail). For this reason the tables and figures mention different refinement indicators and values of γ_2 for the two different families of data. With our initial gauge source function HDWG-α^2, Eq. (29), we obtained similarly large constraint violations and unreliable results far from the three echoes that [10] found.

TABLE I. Bisection intervals for various configurations. The critical amplitude A_* is located in the interval $[A_{\rm sub}, A_{\rm sup}]$. The identifier serves as reference for the configurations in the text. γ_2 is the constraint damping factor for the reduction constraints; see Eqs. (16)–(18). H_a indicates the choice of gauge source functions, which are introduced in Secs. II E and III D.

	Identifier	γ_2	Indicator for h refinement	H_a	$A_{\rm sub}$	$A_{\rm sup}$
Moment of time-symmetry data	1	$4/\alpha$	Smoothness	HDWG	0.15(0000000000000)	0.15(6250000000000)
	2	$4/\alpha$	Smoothness	HDWG-α^2	0.151675332247(121)	0.151675332247(212)
	3	4	Error	HDWG-ln(α)	0.151673(126220703)	0.151673(889160156)
	4	$4/\alpha$	Error	HDWG-ln(α)	0.15167533190(5423)	0.15167533190(6879)
	5	$4/\alpha$	Smoothness	HDWG-ln(α)	0.151675332(244484)	0.151675332(337617)
	6	$4/\alpha$	Smoothness	DSS-comp.	0.1516753322(44484)	0.1516753322(91050)
Incoming data	7	$2/\alpha$	Smoothness	HDWG	0.1(00000000000000)	0.1(18750000000000)
	8	$2/\alpha$	Smoothness	HDWG-α^2	0.10893328143(4289)	0.10893328143(6472)
	9	2	Smoothness	HDWG-ln(α)	0.1089332(75938033)	0.1089332(84878729)
	10	$2/\alpha$	Smoothness	HDWG-ln(α)	0.108933281(395000)	0.108933281(403731)
	11	$2/\alpha$	Smoothness	DSS-comp.	0.1089332815(08505)	0.1089332815(25968)

By instead choosing our adjusted damping parameter of Eq. (79), immune to the collapsing lapse that occurs as curvature approaches the supercritical side, we were able to control those constraint violations. As Fig. 4 shows, close to the time of highest curvature ($t \approx 5$ for those simulations), the constraints are damped much more effectively when using the adjusted damping factor.

This improvement of the constraints near criticality has allowed us to confidently classify data on both sides of the threshold. From the subcritical side, it avoids crashes caused by large constraint violations, allowing the simulation to last until the fields disperse, as the central plot in Fig. 4 shows. From the supercritical side, we observe clear and smooth indications of apparent horizons in the evolution of the expansion.

It should be stressed that the adjusted parameter does not provide an improvement of the constraints throughout the entire evolution, as can be seen at $t \approx 2.5$ in Fig. 4 or at later times. Instead, it makes a clear improvement in the key case where the lapse collapses and the spacetime starts displaying critical phenomena, at $t \approx 5$ in this case.

Table I provides an overview over all configurations discussed in the text. The table shows that the 7-digit estimation of the critical amplitude for the incoming data, and the 6-digit estimation for the moment of time-symmetry data, improved up to 9 and 11 digits, respectively, with the adjusted constraint damping parameter (compare configurations 9 and 3 with 10 and 4, respectively). As seen in Fig. 5, this has allowed us to observe up to three echoes.

Figure 6 shows three echoes for both the moment of time-symmetry data and the incoming data. The former become approximately self-similar slightly earlier, as their first echo is stronger than the incoming data's first echo, while their last one is incomplete compared to the incoming

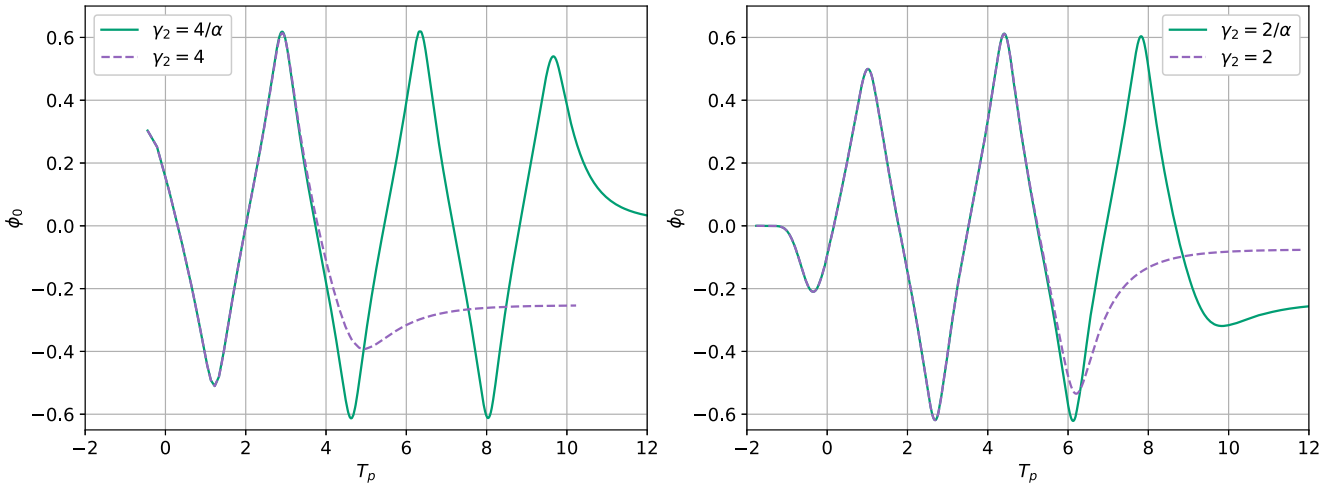

FIG. 5. Left: we show echoes of the highest subcritical amplitudes of two moment of time-symmetry data bisections: one with the original damping parameter (configuration 3) and one with the adjusted damping parameter of Eq. (79) (configuration 4). The lack of reliable tuning beyond 6 digits is an obstacle to see critical phenomena with the original damping scheme. Right: we show the same for the incoming initial data family (configurations 10 and 9).

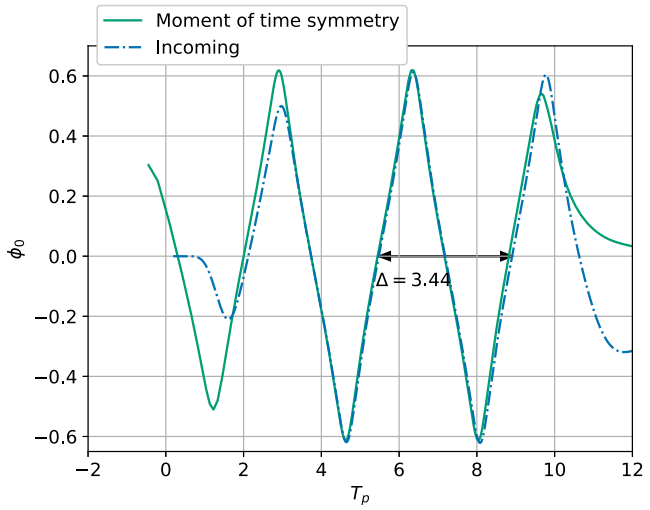

FIG. 6. Echoes of the highest subcritical amplitudes of incoming initial data (configuration 10) and moment of time-symmetry data (configuration 4) and the adjusted damping of Eq. (79). The incoming data have been shifted to the right to enable a direct comparison with the moment of time-symmetry data.

data. In this plot the incoming data have been shifted to the right to verify that indeed, different initial data display the same echoing near criticality with a period of $\Delta \simeq 3.44$ between the second and third echo. We also measure Δ from the zero crossings of the first and second echo and find that the second echo is 0.9% smaller for the moment of time-symmetry data and 1% smaller for the incoming data. This error is comparable with the one reported in [10]. We conclude that the adjusted damping was essential to see critical phenomena with GHG.

B. Coordinate behavior

Both physical and gauge dynamics become stronger closer to the threshold, which is where the latter can become an obstruction to the bisections. Once our bisections reach spacetimes that are close enough to criticality, we confront these gauge issues. Moreover, since all the interrupted bisections suspected to contain coordinate singularities also displayed large constraint violations, we only address coordinate problems once the constraint violations are effectively damped. In addition to this, since our strategy is to exploit the benefits of DSS, we can only evaluate the utility of the DSS-compatible gauge source function by evolving close to the threshold, where the DSS phase is reached for several periods.

We perform bisections of the same data using different gauge source functions in the evolutions, namely HDWG (26), HDWG-α^2 (29), HDWG-ln(α) (32), and finally our initial suggestion of a DSS-compatible gauge source function (65).

For this comparison we employ the same refinement strategy for both sets of data using the smoothness indicator

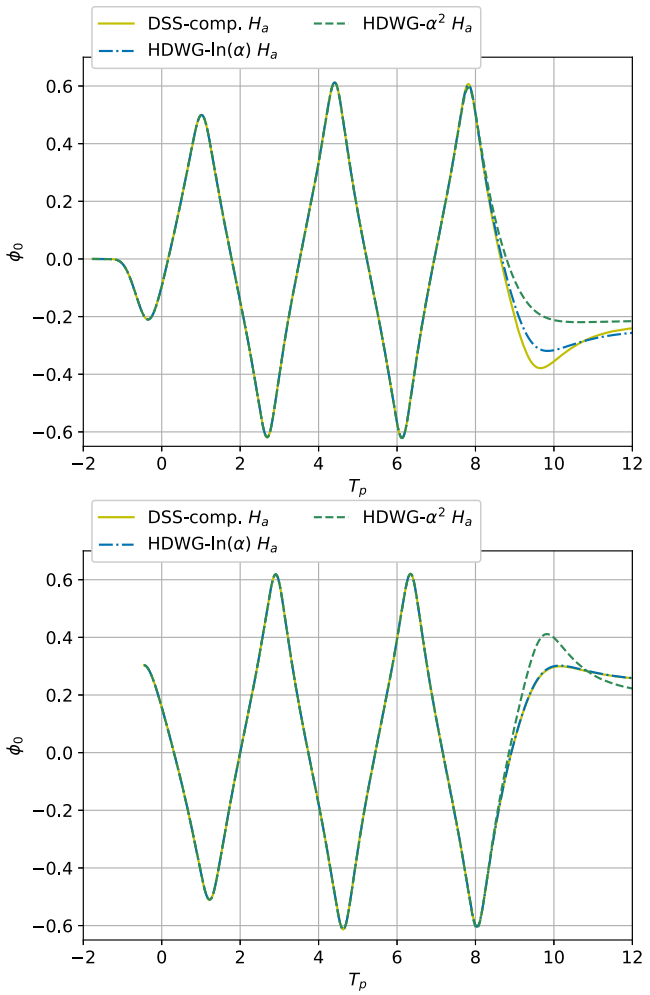

FIG. 7. Echoes of the highest subcritical amplitudes of incoming data bisections performed with the adjusted damping of Eq. (79) and all the mentioned gauge source functions. The top panel shows incoming data (configurations 11, 10, and 8) while the lower panel shows moment of time-symmetry data (configurations 6, 5, and 2). The DSS-compatible gauge source gives results comparable to the ones obtained with HDWG variants, being therefore also suited for critical collapse simulations.

for h refinement (see [28] for details). This provides low constraint violations with all gauge combinations but does actually display a lower peak in the third echo for the moment of time-symmetry data (compare the lower panel of Fig. 7 with the straight line in Fig. 6). The self-similar phase is well approached for all combinations once the adjusted damping parameter is used. As can be seen by comparing configuration 4 and 5 in Table I, the values of the critical regime do vary within a family with different h-refinement strategies after the 8th digit.

Table I shows the intervals that contain the critical amplitude A from the bisections. The values of $A_{\rm sub}$ correspond to the curves in Fig. 7. These runs have been checked and are not always the raw output of the automated bisections. In particular, like in the case discussed in the

previous subsection, the automated bisection of the moment of time-symmetry data evolved with HDWG-α^2 gauge source function (configuration 2) gave a misleading 14-digit tuning that actually showed no improvement in the echoing and large constraint violations after the 12th digit. We therefore trust the bisection up to the 12th digit only.

It is expected that the use of different coordinates not only affects how close we can evolve to the threshold but it also can drift the solution space slightly because of different numerical error. For example in Table I the critical intervals obtained with HDWG-α^2 (configuration 8) and the DSS-compatible gauge (configuration 10) differ in their 10th digit but none is in principle more real. The true value could perhaps be more accurately obtained with careful convergence testing, but since round-off error may already contribute with this level of tuning, and we are interested here primarily in capturing the correct phenomenology, we make no attempt to do so.

We can use again Fig. 7 to assess from which point this difference in tuning stops being meaningful with respect to the observed critical phenomena. For both sets of data, the low level of tuning obtained with vanilla HDWG (of only 1 or 2 digits) had a clear impact on our capacity to observe critical phenomena as it does not give any echoes for either data type. The difference in tuning obtained with vanilla HDWG compared to the other gauge source functions is therefore a severe shortcoming. Moving on to the other three gauge source functions, the lower panel of Fig. 7 displaying moment of time-symmetry data shows that the better tuning obtained with the HDWG-α^2 variant does indeed give the highest last echo. There, both the data evolved with HDWG-ln(α) and the DSS-compatible version behave in exactly the same way, so the difference in the tuning they provide, although not unphysical, is not significant. In the upper panel displaying incoming data, the best-tuned data, also obtained with HDWG-α^2 H_a, this time does not provide the best last echo. It comes third after the one obtained with HDWG-ln(α) H_a and the DSS-compatible version. The improved tuning obtained with HDWG-α^2 is therefore not as meaningful as that obtained with the DSS-compatible gauge for this family of data. Again, the constraint violations are all comparable so all critical intervals are valid, but it does seem that gauge differences affect how easily, after how many digits, we can evolve near criticality.

A sharp profile in the shift was particularly prominent when evolving vanilla HDWG with both sets of data. In order to examine coordinate behaviour more closely, we compare the last common-amplitude run in bisections of all three different gauge source functions that approach criticality. For the incoming data this corresponds to evolutions of $A = 0.108933281(456118)$. The evolution of these data with the DSS-compatible gauge shows a clear uninterrupted dispersion of the fields allowing to classify the spacetime as subcritical. In contrast, both adjustments to HDWG crash, classifying it as supercritical. A common feature to both of the HDWG-variant evolutions is the presence of what resembles a step function in the profile of the shift vector in space. The constraint violations are comparable enough to trust both classifications, but we see that these sharp features in the shift seem to appear later in the bisection for this family of data with the DSS-compatible gauge source function. This advantage does not seem to be the case for the last common amplitude in the moment of time-symmetry data, $A = 151675332(291050)$, which shows very similar profiles of the shift, none as smooth as would be desirable, for bisections with all three gauge source functions. This spacetime was classified as subcritical with HDWG-ln(α) H_a, whereas a horizon was found with the other two choices.

To summarize, the two HDWG variants considered and the DSS-compatible gauge agree very well to 9 digits for both sets of data with comparable critical results and constraint violations. With this level of tuning our results will be affected by round-off, so to tune further we may need to work with higher-precision arithmetic. More work will be needed to completely avoid the gauge features that affect the bisections near criticality, but we nevertheless conclude that the DSS-compatible gauge source function maintains well-behaved coordinates far into the critical regime.

C. DSS and gauge sources

As positive as it is to tackle the issues that were stopping our critical collapse bisections, the question of how well any of our gauges dynamically adapt to the approximate self-similarity of the evolved spacetimes remains. We compare directly with the HDWG-ln(α) non-DSS-compatible gauge source function, because we built the DSS-compatible suggestion so as to approximately agree with it, as can be seen in Fig. 2, but also because they provide very comparable results for critical phenomena, as demonstrated in Fig. 7.

First, we examine the presence of DSS in the fields evolved. After the evolution we transform the evolved coordinates into constructed canonical DSS-adapted coordinates to verify the symmetry. These coordinates are slow time computed from proper time T_p at the center, as defined in Eq. (37), and its spatial counterpart

$$X^i_p = \frac{x^i_p}{|\tau_* - \tau|}, \tag{84}$$

where x^i_p are proper lengths. In Fig. 8 we plot the scalar field of a near-critical spacetime evolved with HDWG-ln(α) and the DSS-compatible gauge sources, with respect to (T_p, X^i_p). This figure helps us also assess if any unwanted gauge features actually disrupt the self-similarity phase of the spacetime. This is luckily not the case for

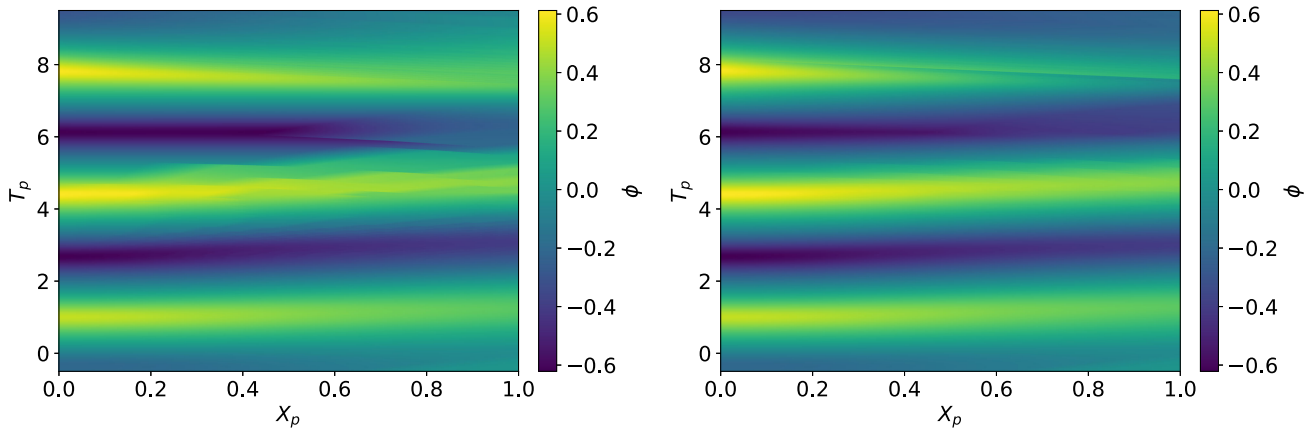

FIG. 8. The scalar field is plotted as a function of slow time and the similarity coordinate in the x direction X_p as defined in (84). These correspond to the highest subcritical evolutions of incoming data with the adjusted damping of Eq. (79), and on the left with HDWG-$\ln(\alpha)$ (configuration 10) and on the right with the DSS-compatible gauge source function (configuration 11).

either gauge choice. It seems that any undesirable gauge features reported in the previous Sec. V B are not sufficient to prevent reliable bisections, and not enough to disrupt the spacetimes that survive. The periodicity observed in these plots also confirms that we are close to the threshold, the only case where the similarity would be exact.

We proceed to examine whether our evolved coordinates are adapted to DSS. If our evolved time coordinate t was exactly T_p, as defined in Eq. (37), then Eq. (40) would hold, which can be rewritten in the same way as Eq. (35) as

$$\alpha(T_p, X_p^i) = e^{-T_p}\tilde{\alpha}(T_p, X_p^i), \tag{85}$$

$$\ln(\alpha(T_p, X_p^i)) + T_p = \ln(\tilde{\alpha}(T_p, X_p^i)), \tag{86}$$

where $\tilde{\alpha}$ is Δ periodic in T_p at fixed spatial similarity coordinate. In Fig. 9 we plot the left-hand side of Eq. (86) to check if indeed we find a periodic function of constant average as $\ln(\tilde{\alpha}(T_p, X_p^i))$ is. The periodicity is certainly noticeable for all gauge source functions, but the non-vanishing slope indicates that Eq. (86) is not satisfied for either gauge choice. This means that $t \neq T_p$.

The DSS-adapted coordinates (T_p, X_p^i) are not unique. As explained in Appendix B, any coordinates (T, X^i) related to (T_p, X_p^i) by

$$T = T_p + \tilde{f}^T(T_p, X_p^i), \tag{87}$$

$$X^i = \tilde{f}^i(T_p, X_p^i), \tag{88}$$

where $\tilde{f}^a(T_p, X_p^i)$ is a function Δ periodic in T_p, will be themselves DSS adapted. The existence of such a wide range of DSS-adapted coordinates could make us hope that even if our coordinate t does not exactly coincide with T_p, it could still coincide with T in (87) and hence be DSS adapted. However, a short calculation also provided in

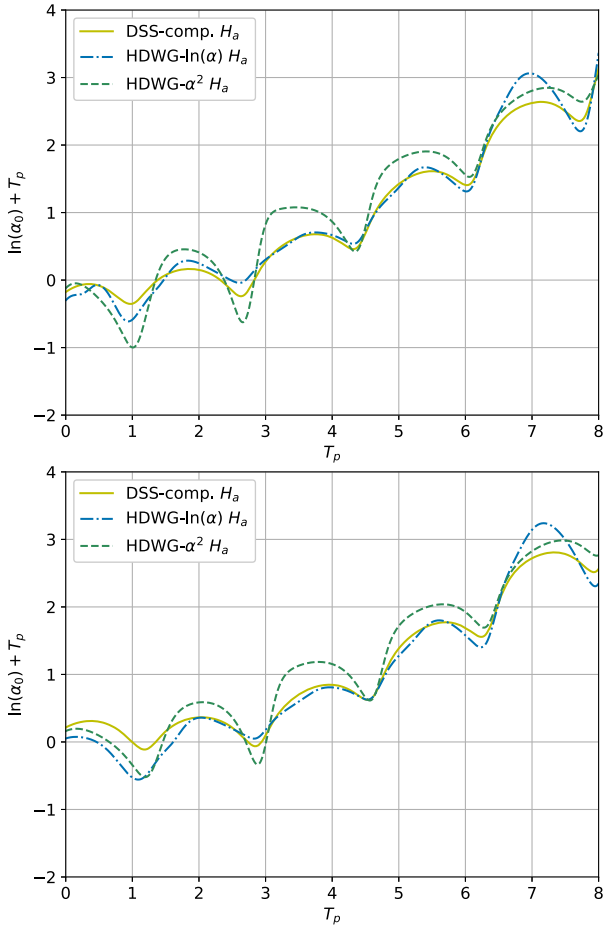

FIG. 9. Left-hand side of Eq. (86) for the highest amplitude of bisections of incoming data in the top panel and moment of time-symmetry data in the lower panel both with adjusted damping of Eq. (79) and both HDWG-variants H_a (configurations 8 and 10 in the top panel, 2 and 5 in the lower panel) and the DSS-compatible H_a, Eq. (65) (configuration 11 in the top panel, and 6 in the lower panel). Despite the different data used, each curve seems identical in both panels, mostly determined by the choice of H_a.

Appendix B shows that if t was DSS adapted it would satisfy the equation

$$\ln(\alpha(t,x^i)) + T_p = f, \tag{89}$$

where f must be a function Δ periodic in T_p. Figure 9 shows that $\ln(\alpha) + T_p$ has some oscillations with a characteristic frequency, but it is not periodic in T_p, which implies that t is not a DSS-adapted time coordinate.

We can see more explicitly that t is not DSS adapted at all. For our coordinate t to be DSS adapted we should have that

$$\ln(\alpha(t,x^i)) + t = \ln(\tilde{\alpha}(t,x^i)). \tag{90}$$

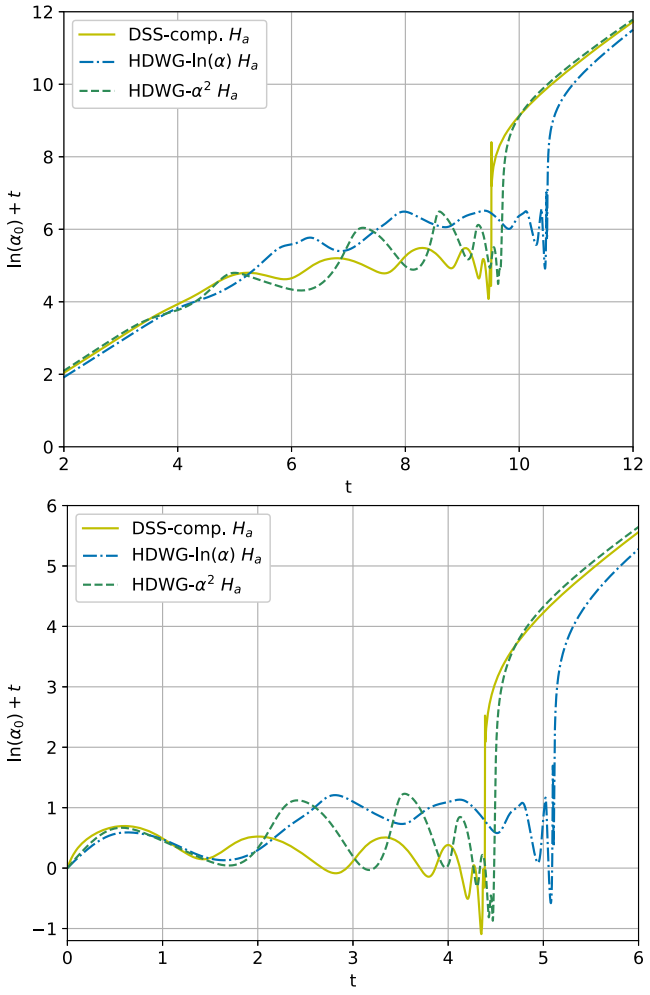

FIG. 10. Left-hand side of Eq. (90) for the highest amplitude of bisections of incoming data in the top panel and moment of time-symmetry data in the lower panel both with adjusted damping of Eq. (79) and both HDWG-variants H_a (configurations 10 and 8 in the top panel, 2 and 5 in the lower panel) and the DSS-compatible H_a, Eq. (65) (configuration 11 in the top panel, and 6 in the lower panel). The top panel is zoomed out to show how the DSS phase stands out in the evolution.

In Fig. 10 we plot the left-hand side of Eq. (90) to check again if we see a periodic function of constant average. We surprisingly observe opposite results to those of Fig. 9: although the average through a constant value during the self-similar phase is much more noticeable than in Fig. 9, there is not such a clear periodicity. We can now conclude that the dynamical coordinates associated with neither gauge are adapted to DSS. Following the calculation in Sec. III, we already knew that any non-DSS-compatible gauge choice could not yield DSS-adapted coordinates, and that even a DSS-compatible one could not guarantee them. The gauge choice Eq. (65) is unfortunately not the needle in the haystack.

The choice we made for a DSS-compatible gauge source function is only a first attempt. Many other gauge source functions could be constructed that satisfy the DSS-compatibility condition. As we already saw in Fig. 2, the DSS-compatible gauge source function we suggested is qualitatively closest to HDWG-ln(α) gauge source function, which may suggest the reason that this variant performs the best of all the non-DSS-compatible choices.

VI. SUMMARY AND DISCUSSION

In this paper, using the classic example of scalar-field collapse as a testbed [1], we have overcome two major obstacles that were stopping our critical collapse simulations of Brill waves in [16], namely constraint violations and undesirable coordinate features.

Taking into account the collapse of the lapse in near-critical spacetimes, we modified the reduction constraint damping parameter of GHG, γ_2, such that it keeps damping constraint violations near criticality. This can be thought of as changing the damping timescale from proper to coordinate time, and has allowed us to improve our critical parameter estimation and thus to clearly observe critical phenomena with two different families of real massless scalar fields in spherical symmetry minimally coupled to GR. Since pseudospectral codes typically employ first-order systems of equations, GHG in our case, we suspected that the inefficient damping of violations of the reduction constraint of GHG in spacetimes where the lapse collapses was responsible for the poorer tuning provided by pseudospectral searches of critical collapse in comparison to those performed with finite-difference codes and second order in space formulations of GR. With the improved damping scheme we can now confidently tune to a level competitive with bespoke spherical methods. It may be that the bottleneck in tuning is now the effect of round-off error, but this deserves further study.

Aiming to exploit the structure of a DSS spacetime to avoid coordinate singularities, we derived a necessary condition for a gauge choice to be compatible with DSS. The condition is appropriate also for CSS spacetimes. Using a specific gauge source function for GHG that satisfies the DSS-compatibility condition, we have reproduced

the critical phenomena of a massless scalar field at the level of commonly used gauge choices. The spatial profiles of the fields suggests that this gauge source function respects DSS slightly better than the other gauge sources, particularly vanilla HDWG, which was not designed with critical collapse in mind (see [46] for discussion). However, at the level of tuning that we can currently achieve, we do not see much improvement in the observation of critical phenomena, constraint violation reduction, nor in the avoidance of coordinate singularities with the DSS-compatible choice we made. This is more a practical than a principle issue. Every time we extend from one scale echo to the next, a larger class of coordinate singularities may occur if we do not satisfy the compatibility condition. Further work is however needed to come up with gauge source functions that give truly DSS-adapted coordinates.

The natural continuation of the present work is simply to apply our improvements to the more interesting setup of axisymmetry, both in the context of scalar fields and gravitational waves. We expect that an improved apparent horizon finder will be needed to classify spacetimes of the latter. By systematically treating each of the obstructions we encounter, we expect to be able to further investigate nonspherical spacetimes more accurately, with the ultimate aim of gaining a comprehensive understanding of the threshold of black hole formation in complete generality.

ACKNOWLEDGMENTS

We would like to thank Florian Atteneder for all his help throughout this project, and Isabel Suárez Fernández and Krinio Marouda for helpful discussions and feedback. This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant No. 406116891 within the Research Training Group RTG 2522/1. H. R. R. acknowledges support from the Fundação para a Ciência e Tecnologia (FCT) within the Projects No. UID/04564/2021, No. UIDB/04564/2020, No. UIDP/04564/2020, and No. EXPL/FIS-AST/0735/2021, and The H2020 ERC Advanced Grant "Black holes: gravitational engines of discovery" Grant Agreement No. Gravitas–101052587. This work was supported in part by FCT (Portugal) Project No. UIDB/00099/2020.

APPENDIX A: EIGENVALUE PROBLEM

The left eigenvectors $\mathbf{l}_s$ of $\mathbf{M}$ in (75) are

$$\mathbf{l}_{s_1}{}^T = \begin{pmatrix} \frac{-|\omega|(\alpha+\beta^{\hat{\omega}}\gamma_1)(|\omega|+i\gamma_2)}{\gamma_1\gamma_2} \\ \frac{(|\omega|(\alpha+\beta^{\hat{\omega}}\gamma_1)+i\alpha\gamma_2)(|\omega|+i\gamma_2)}{\gamma_1{\gamma_2}^2} \\ \frac{(|\omega|(\alpha+\beta^{\hat{\omega}}\gamma_1)-i\beta^{\hat{\omega}}\gamma_1\gamma_2)(|\omega|+i\gamma_2)}{\gamma_1{\gamma_2}^2} \\ \beta^A \end{pmatrix}, \tag{A1}$$

$$\mathbf{l}_{s_2}{}^T = \begin{pmatrix} \frac{-|\omega|(\alpha-\beta^{\hat{\omega}}\gamma_1)(|\omega|-i\gamma_2)}{\gamma_1\gamma_2} \\ \frac{(|\omega|(\alpha-\beta^{\hat{\omega}}\gamma_1)-i\alpha\gamma_2)(|\omega|-i\gamma_2)}{\gamma_1{\gamma_2}^2} \\ \frac{(|\omega|(-\alpha+\beta^{\hat{\omega}}\gamma_1)+i\beta^{\hat{\omega}}\gamma_1\gamma_2)(|\omega|-i\gamma_2)}{\gamma_1{\gamma_2}^2} \\ \beta^A \end{pmatrix}, \tag{A2}$$

$$\mathbf{l}_{s_3,s_4}{}^T = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \delta^B{}_A \end{pmatrix}, \qquad \mathbf{l}_{s_5}{}^T = \begin{pmatrix} -i|\omega|\beta^{\hat{\omega}} \\ 0 \\ \beta^{\hat{\omega}} \\ \beta^A \end{pmatrix}. \tag{A3}$$

The characteristic-like variables are

$$\begin{aligned} v_{s_1} = {} & \frac{(|\omega|+i\gamma_2)(|\omega|(\alpha+\beta^{\hat{\omega}}\gamma_1)+i\alpha\gamma_2)\Pi}{\gamma_2} \\ & + \gamma_1\gamma_2\beta^A\Phi_A \\ & + \left(i|\omega|\alpha + \frac{|\omega|^2(\alpha+\beta^{\hat{\omega}}\gamma_1)}{\gamma_2} + \gamma_1\gamma_2\beta^{\hat{\omega}}\right)\Phi_{\hat{\omega}} \\ & - |\omega|(\alpha+\gamma_1\beta^{\hat{\omega}})(|\omega|-i\gamma_2)g, \end{aligned} \tag{A4}$$

$$\begin{aligned} v_{s_2} = {} & \frac{(|\omega|-i\gamma_2)(|\omega|(\alpha-\beta^{\hat{\omega}}\gamma_1)-i\alpha\gamma_2)\Pi}{\gamma_2} \\ & + \gamma_1\gamma_2\beta^A\Phi_A \\ & + \left(i|\omega|\alpha + \frac{|\omega|^2(-\alpha+\beta^{\hat{\omega}}\gamma_1)}{\gamma_2} + \gamma_1\gamma_2\beta^{\hat{\omega}}\right)\Phi_{\hat{\omega}} \\ & + |\omega|(-\alpha+\gamma_1\beta^{\hat{\omega}})(|\omega|-i\gamma_2)g, \end{aligned} \tag{A5}$$

$$v_{s_4,s_3} = \Phi_A, \tag{A6}$$

$$v_{s_5} = \beta^A\Phi_A + \beta^{\hat{\omega}}(\Phi_{\hat{\omega}} - i|\omega|g). \tag{A7}$$

APPENDIX B: DSS-ADAPTED TIME COORDINATES

Consider a set of coordinates $x^a = (t, x^i)$ that are not adapted to DSS. Also consider two different foliations corresponding to two sets of DSS-adapted coordinates, $x^{\bar{a}} = (\bar{t}, x^{\bar{i}})$ and $x^{\hat{a}} = (\hat{t}, x^{\hat{i}})$.

Let a spacetime enter the DSS phase and call echo 1 the event where the scalar field first reaches 0.6 and echo 2 the subsequent event where the field reaches again 0.6. In general, we have $(t_2, x_2^i) = (t_1 + \delta t, x_1^i + \delta x^i)$ with arbitrary δt and δx^i while being in DSS-adapted coordinates implies $\delta t = \Delta$ and $\delta x^i = 0$.

Assuming that there is a functional relationship between the coordinates such that $x^{\bar{a}} = h^{\bar{a}}(x^a)$ and $x^{\bar{a}} = f^{\bar{a}}(x^{\hat{a}})$, then

$$\bar{t}_2 = h^{\bar{i}}(t_2, x_2^i) = h^{\bar{i}}(t_1 + \delta t, x_1^i) = \bar{t}_1 + \Delta = h^{\bar{i}}(t_1, x_1^i) + \Delta, \tag{B1}$$

$$\bar{t}_2 = f^{\bar{i}}(\hat{t}_2, x_2^{\hat{i}}) = f^{\bar{i}}(\hat{t}_1 + \Delta, x_1^{\hat{i}}) = \bar{t}_1 + \Delta = f^{\bar{i}}(\hat{t}_1, x_1^{\hat{i}}) + \Delta. \tag{B2}$$

From Eq. (B2) we can conclude that $f^{\bar{i}}(\hat{t}_1 + \Delta, x_1^{\hat{i}}) = f^{\bar{i}}(\hat{t}_1, x_1^{\hat{i}}) + \Delta$ and hence that the function

$$\tilde{f}^{\bar{i}}(\hat{t}, x^{\hat{i}}) \coloneqq f^{\bar{i}}(\hat{t}, x^{\hat{i}}) - \hat{t} = \bar{t} - \hat{t} \tag{B3}$$

must be Δ periodic in $\hat{t}$. Note that this is only the case because the DSS period is the same for both DSS-adapted coordinates, which is why it does not apply to the not-DSS-adapted coordinates as Eq. (B1) shows. Similarly, we have that $x^{\bar{i}} = \tilde{f}^{\bar{i}}(\hat{t}, x^{\hat{i}})$ is also Δ periodic in $\hat{t}$. From Eq. (B3) we have that

$$\nabla_{\bar{a}}\bar{t} = \frac{\partial x^{\hat{a}}}{\partial x^{\bar{a}}}[\nabla_{\hat{a}}\hat{t} + \nabla_{\hat{a}}\tilde{f}^{\bar{i}}(\hat{t}, x^{\hat{i}})] = \frac{\partial x^{\hat{a}}}{\partial x^{\bar{a}}}\tilde{C}_{\hat{a}}(x^{\hat{b}}), \tag{B4}$$

where everything inside the square brackets, now referred to as $\tilde{C}_{\hat{a}}(x^{\hat{b}})$, is Δ periodic in $\hat{t}$.

As in Eq. (85), in these DSS-adapted coordinates we have

$$\bar{\alpha}(\bar{t}, x^{\bar{i}}) = e^{-\bar{t}}\tilde{\bar{\alpha}}(\bar{t}, x^{\bar{i}}), \tag{B5}$$

$$\hat{\alpha}(\hat{t}, x^{\hat{i}}) = e^{-\hat{t}}\tilde{\hat{\alpha}}(\hat{t}, x^{\hat{i}}), \tag{B6}$$

where $\tilde{\bar{\alpha}}(\bar{t}, x^{\bar{i}})$ and $\tilde{\hat{\alpha}}(\hat{t}, x^{\hat{i}})$ are Δ periodic in $\bar{t}$ and $\hat{t}$, respectively. In terms of the conformal metrics $\tilde{g}^{\bar{a}\,\bar{b}}$ and $\tilde{g}^{\hat{a}\,\hat{b}}$, Δ periodic in $\bar{t}$ and $\hat{t}$, respectively, we have that Eq. (B4) implies

$$\begin{aligned}\tilde{\bar{\alpha}}(\bar{t}, x^{\bar{i}}) &= (-\tilde{g}^{\bar{a}\,\bar{b}}\nabla_{\bar{a}}\bar{t}\nabla_{\bar{b}}\bar{t})^{-1/2} \\ &= \left(-\tilde{g}^{\bar{a}\,\bar{b}}\left(\frac{\partial x^{\hat{a}}}{\partial x^{\bar{a}}}\tilde{C}_{\hat{a}}(x^{\hat{c}})\right)\left(\frac{\partial x^{\hat{b}}}{\partial x^{\bar{b}}}\tilde{C}_{\hat{b}}(x^{\hat{d}})\right)\right)^{-1/2} \\ &= (-\tilde{g}^{\hat{a}\,\hat{b}}\tilde{C}_{\hat{a}}(x^{\hat{c}})\tilde{C}_{\hat{b}}(x^{\hat{d}}))^{-1/2},\end{aligned} \tag{B7}$$

such that $\tilde{\bar{\alpha}}(\bar{t}, x^{\bar{i}})$ is also Δ periodic in $\hat{t}$.

Lastly, taking the logarithm of Eq. (B5) we have that

$$\ln(\bar{\alpha}(\bar{t}, x^{\bar{i}})) + \bar{t} = \ln(\tilde{\bar{\alpha}}(\bar{t}, x^{\bar{i}})), \tag{B8}$$

$$\ln(\bar{\alpha}(\bar{t}, x^{\bar{i}})) + \hat{t} = \ln(\tilde{\bar{\alpha}}(\bar{t}, x^{\bar{i}})) - \tilde{f}^{\bar{i}}(\hat{t}, x^{\hat{i}}), \tag{B9}$$

where the right-hand side should be Δ periodic in $\hat{t}$ as Eq. (B7) shows.

For the coordinates in the main text, we computed slow time from proper time T_p which corresponds to $\hat{t}$ and we would like to check whether our evolution coordinate t corresponds to another DSS-adapted coordinate here denoted as $\bar{t}$. The function f in Eq. (89) corresponds to the right-hand side of Eq. (B9).

[1] M. W. Choptuik, Universality and scaling in gravitational collapse of a massless scalar field, Phys. Rev. Lett. **70**, 9 (1993).
[2] C. Gundlach, Understanding critical collapse of a scalar field, Phys. Rev. D **55**, 695 (1997).
[3] S. Hod and T. Piran, Fine structure of Choptuik's mass scaling relation, Phys. Rev. D **55**, R440(R) (1997).
[4] M. Reiterer and E. Trubowitz, Choptuik's critical spacetime exists, Commun. Math. Phys. **368**, 143 (2019).
[5] C. Gundlach and J. M. Martín-García, Critical phenomena in gravitational collapse, Living Rev. Relativity **10**, 5 (2007).
[6] D. Garfinkle and G. C. Duncan, Scaling of curvature in subcritical gravitational collapse, Phys. Rev. D **58**, 064024 (1998).
[7] J. M. Martín-García and C. Gundlach, Global structure of Choptuik's critical solution in scalar field collapse, Phys. Rev. D **68**, 024011 (2003).
[8] J. M. Martin-Garcia and C. Gundlach, All nonspherical perturbations of the Choptuik spacetime decay, Phys. Rev. D **59**, 064031 (1999).
[9] M. W. Choptuik, E. W. Hirschmann, S. L. Liebling, and F. Pretorius, Critical collapse of the massless scalar field in axisymmetry, Phys. Rev. D **68**, 044007 (2003).
[10] T. W. Baumgarte, Aspherical deformations of the Choptuik spacetime, Phys. Rev. D **98**, 084012 (2018).
[11] J. Healy and P. Laguna, Critical collapse of scalar fields beyond axisymmetry, Gen. Relativ. Gravit. **46**, 1722 (2014).
[12] N. Deppe, L. E. Kidder, M. A. Scheel, and S. A. Teukolsky, Critical behavior in 3d gravitational collapse of massless scalar fields, Phys. Rev. D **99**, 024018 (2019).
[13] D. Hilditch, A. Weyhausen, and B. Brügmann, Evolutions of centered Brill waves with a pseudospectral method, Phys. Rev. D **96**, 104051 (2017).
[14] A. Khirnov and T. Ledvinka, Slicing conditions for axisymmetric gravitational collapse of Brill waves, Classical Quantum Gravity **35**, 215003 (2018).
[15] T. Ledvinka and A. Khirnov, Universality of curvature invariants in critical vacuum gravitational collapse, Phys. Rev. Lett. **127**, 011104 (2021).
[16] I. Suárez Fernández, S. Renkhoff, D. Cors, B. Brügmann, and D. Hilditch, Evolution of Brill waves with an adaptive

pseudospectral method, Phys. Rev. D **106**, 024036 (2022).
[17] T. W. Baumgarte, C. Gundlach, and D. Hilditch, Critical phenomena in the collapse of quadrupolar and hexadecapolar gravitational waves, Phys. Rev. D **107**, 084012 (2023).
[18] T. W. Baumgarte, B. Brügmann, D. Cors, C. Gundlach, D. Hilditch, A. Khirnov, T. Ledvinka, S. Renkhoff, and I. Suárez Fernández, Critical phenomena in the collapse of gravitational waves, Phys. Rev. Lett. **131**, 181401 (2023).
[19] T. W. Baumgarte, C. Gundlach, and D. Hilditch, Critical phenomena in the gravitational collapse of electromagnetic waves, Phys. Rev. Lett. **123**, 171103 (2019).
[20] M. F. Perez Mendoza and T. W. Baumgarte, Critical phenomena in the gravitational collapse of electromagnetic dipole and quadrupole waves, Phys. Rev. D **103**, 124048 (2021).
[21] S. Renkhoff, AHloc3d, https://git.tpi.uni-jena.de/srenkhoff/ahloc3d.
[22] M. Alcubierre, Appearance of coordinate shocks in hyperbolic formalisms of general relativity, Phys. Rev. D **55**, 5981 (1997).
[23] M. Alcubierre, Hyperbolic slicings of spacetime: Singularity avoidance and gauge shocks, Classical Quantum Gravity **20**, 607 (2003).
[24] T. W. Baumgarte and D. Hilditch, Shock-avoiding slicing conditions: Tests and calibrations, Phys. Rev. D **106**, 044014 (2022).
[25] L. Lindblom and B. Szilágyi, Improved gauge driver for the generalized harmonic Einstein system, Phys. Rev. D **80**, 084019 (2009).
[26] D. Garfinkle and C. Gundlach, Symmetry-seeking spacetime coordinates, Classical Quantum Gravity **16**, 4111 (1999).
[27] D. Hilditch, A. Weyhausen, and B. Brügmann, Pseudospectral method for gravitational wave collapse, Phys. Rev. D **93**, 063006 (2016).
[28] S. Renkhoff, D. Cors, D. Hilditch, and B. Brügmann, Adaptive hp refinement for spectral elements in numerical relativity, Phys. Rev. D **107**, 104043 (2023).
[29] O. Rinne, Stable radiation-controlling boundary conditions for the generalized harmonic Einstein equations, Classical Quantum Gravity **23**, 6275 (2006).
[30] H. R. Rüter, D. Hilditch, M. Bugner, and B. Brügmann, Hyperbolic relaxation method for elliptic equations, Phys. Rev. D **98**, 084044 (2018).
[31] T. W. Baumgarte and S. L. Shapiro, *Numerical Relativity: Solving Einstein's Equations on the Computer* (Cambridge University Press, Cambridge, England, 2010).
[32] W. Tichy, The initial value problem as it relates to numerical relativity, Rep. Prog. Phys. **80**, 026901 (2017).
[33] L. Lindblom, M. A. Scheel, L. E. Kidder, R. Owen, and O. Rinne, A new generalized harmonic evolution system, Classical Quantum Gravity **23**, S447 (2006).
[34] B. Szilagyi, L. Lindblom, and M. A. Scheel, Simulations of binary black hole mergers using spectral methods, Phys. Rev. D **80**, 124010 (2009).
[35] T. W. Baumgarte and S. L. Shapiro, On the numerical integration of Einstein's field equations, Phys. Rev. D **59**, 024007 (1998).
[36] M. Shibata and T. Nakamura, Evolution of three-dimensional gravitational waves: Harmonic slicing case, Phys. Rev. D **52**, 5428 (1995).
[37] T. Nakamura, K. Oohara, and Y. Kojima, General relativistic collapse to black holes and gravitational waves from black holes, Prog. Theor. Phys. Suppl. **90**, 1 (1987).
[38] S. Bernuzzi and D. Hilditch, Constraint violation in free evolution schemes: Comparing BSSNOK with a conformal decomposition of Z4, Phys. Rev. D **81**, 084003 (2010).
[39] D. Alic, C. Bona-Casas, C. Bona, L. Rezzolla, and C. Palenzuela, Conformal and covariant formulation of the Z4 system with constraint-violation damping, Phys. Rev. D **85**, 064040 (2012).
[40] D. Hilditch, S. Bernuzzi, M. Thierfelder, Z. Cao, W. Tichy, and B. Brügmann, Compact binary evolutions with the Z4c formulation, Phys. Rev. D **88**, 084057 (2013).
[41] C. Bona, J. Massó, E. Seidel, and J. Stela, New formalism for numerical relativity, Phys. Rev. Lett. **75**, 600 (1995).
[42] O. Brodbeck, S. Frittelli, P. Hübner, and O. A. Reula, Einstein's equations with asymptotically stable constraint propogation, J. Math. Phys. (N.Y.) **40**, 909 (1999).
[43] C. Gundlach, J. M. Martin-Garcia, G. Calabrese, and I. Hinder, Constraint damping in the Z4 formulation and harmonic gauge, Classical Quantum Gravity **22**, 3767 (2005).
[44] M. K. Bhattacharyya, D. Hilditch, K. Rajesh Nayak, S. Renkhoff, H. R. Rüter, and B. Brügmann, Implementation of the dual foliation generalized harmonic gauge formulation with application to spherical black hole excision, Phys. Rev. D **103**, 064072 (2021).
[45] M. Alcubierre, *Introduction to 3+1 Numerical Relativity* (Oxford University Press, Oxford, 2008).
[46] D. Cors, Towards finer estimations of the threshold of black hole formation, Ph.D. thesis, Friedrich-Schiller-Universität Jena, Jena, Germany (to be published).