Improved moving-puncture techniques for compact binary simulations

Zachariah B. Etienne[*]

Department of Physics, University of Idaho, Moscow, Idaho 83843, USA;
Department of Physics and Astronomy, West Virginia University, Morgantown,
West Virginia 26506, USA; and Center for Gravitational Waves and Cosmology,
West Virginia University, Chestnut Ridge Research Building, Morgantown,
West Virginia 26505, USA

To fully unlock the scientific potential of upcoming gravitational wave (GW) interferometers, numerical relativity (NR) simulation accuracy will need to be greatly enhanced. We present three infrastructure-agnostic improvements to the moving-puncture approach for binary black hole (BBH) simulations, aimed at greatly reducing constraint violation and improving GW predictions. Although these improvements were developed within the highly efficient NR code Blackholes@Home, we demonstrate their effectiveness in the widely adopted Einstein TOOLKIT/Carpet adaptive mesh refinement framework. Our improvements include a modified Kreiss-Oliger dissipation prescription, a Hamiltonian constraint damping adjustment to the Baumgarte-Shapiro-Shibata-Nakamura (BSSN) equations, and an extra term to the $1 + \log$ lapse evolution equation that slows the development of the sharp lapse feature, which dominates numerical errors in BBH simulations. With minimal increase in computational cost, these improvements greatly reduce GW noise, enabling the extraction of high-order GW modes previously obscured by numerical noise. They also improve convergence properties near and inside the convergent regime, reduce Hamiltonian (momentum) constraint violations in the strong-field region by roughly 2 (3) orders of magnitude, and in the GW-extraction zone by 5 (2) orders of magnitude. To promote community adoption, we have open-sourced the improved Einstein TOOLKIT thorn BaikalVacuum used in this work. Although our focus is on BBH evolutions and the BSSN formulation, these improvements may also benefit compact binary simulations involving matter and other formulations, a focus for future investigations.

DOI: 10.1103/PhysRevD.110.064045

I. INTRODUCTION

Numerical relativity (NR) simulations of compact binary mergers play a crucial role in gravitational wave (GW) data analysis, enabling reliable estimation of physical parameters from the late-inspiral, merger, and postmerger phases of GW signals. To date, detector noise in GW observations has generally surpassed the numerical errors intrinsic to NR simulations and GW approximants derived from them, so GW parameter estimation errors have generally been dominated by noise in the GW signal. However, improved sensitivities offered by current and forthcoming GW detectors threaten to change this dynamic. Third-generation (3G) detectors will necessitate a roughly tenfold improvement in NR simulation accuracy [1,2]. Therefore, enhancing the accuracy of NR codes is vital for unlocking the full scientific potential of GW astronomy moving forward.

Improved accuracy can be achieved through various means. One widely adopted approach involves building next-generation NR codes upon more scalable infrastructures.

Increased scalability enables simulations to be performed at higher resolutions and with greater physical realism. While running simulations at higher resolutions generally reduces numerical errors, it does not directly address one of the primary sources of error in existing NR codes: noise associated with the selection of numerical grids.

Cartesian adaptive mesh refinement (AMR) grids, the most popular gridding approach in NR, span the many decades in physical scales intrinsic to compact binaries with nested Cartesian grid patches. The highest resolution patches are placed where the spacetime fields are sharpest: at the compact objects themselves. These patches are embedded in increasingly coarser patches, extending to the outer boundary. Cartesian AMR grids are both numerically convenient and highly scalable with modern AMR algorithms.

When it comes to solving Einstein's equations in NR, the moving-puncture approach [3,4] is the most popular method. This approach generally combines the BSSN [5,6], CCZ4 [7], or Z4c [8] $3 + 1$ decompositions of Einstein's equations with the moving-puncture gauge conditions [3,4,9], enabling the evolution of black holes (BHs) that traverse the numerical grid without the need to excise

---

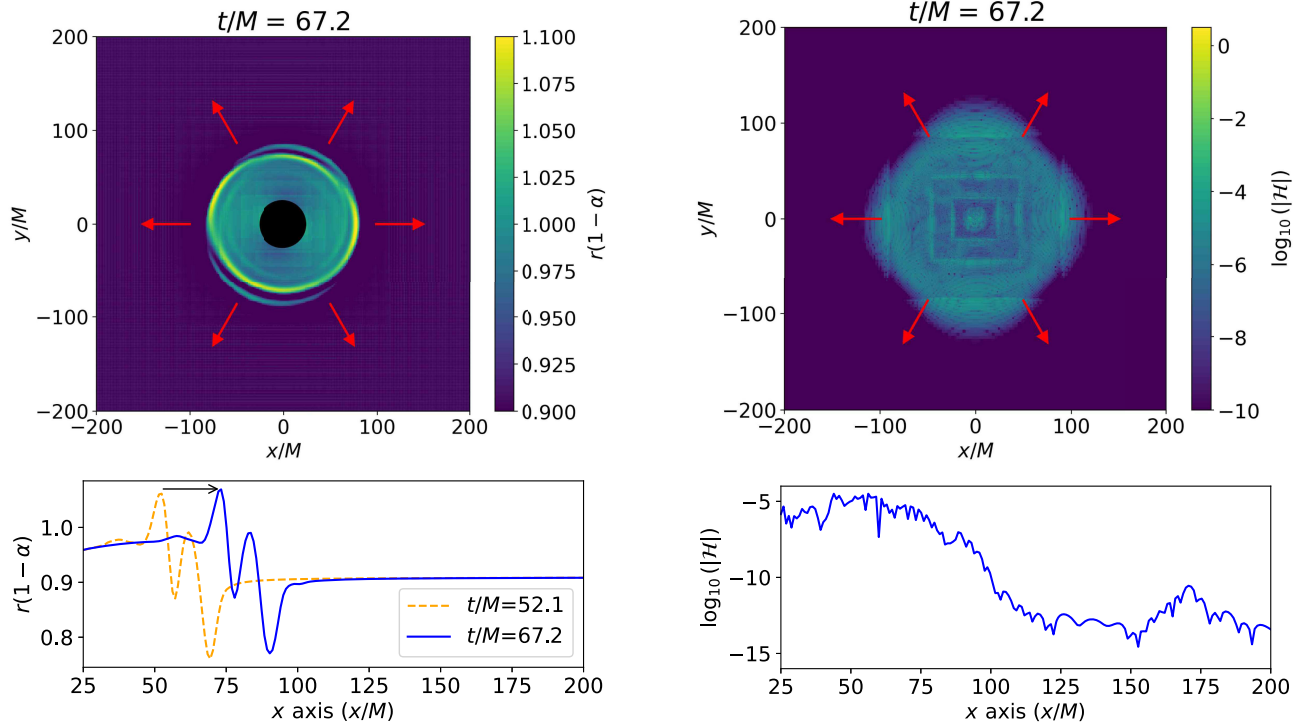[*]Contact author: zetienne@uidaho.edu

FIG. 1.   Development of sharp moving-puncture lapse feature and its aftermath. Top left: $r(1 - \alpha)$ color map on the orbital plane, for $r > 25M$; data for $r \leq 25M$ were excised (represented by a black dot). Bottom left: $r(1 - \alpha)$ data along the x axis, for $x > 25M$. Top right: $\mathcal{H}$ constraint violation, plotted as $\log_{10} |\mathcal{H}|$ on the orbital plane. The dashed orange curve corresponds to the data in the top-left plot, and the blue solid curve corresponds to the data at $t = 15.1M$ later. The black arrow denotes the propagation distance and direction of a wave propagating outward at $\sqrt{2}c$, from $t/M = 52.1$ to $t/M = 67.2$. Bottom right: $\log_{10} |\mathcal{H}|$ along the x axis at $t = 67.2M$, for $x > 25M$.

their interiors. However, mixing the moving-puncture technique with Cartesian AMR grids results in significant numerical noise that negatively impacts the accuracy of compact binary simulations.

As shown in Fig. 1, during the initial phase of a moving-puncture binary black hole (BBH) evolution, the lapse rapidly settles to new values in the strong-field region, releasing a sharp, superluminal gauge wave packet that radiates outward from the BHs (top panel) at a speed of $\sqrt{2}c$—the characteristic speed of the $1 + \log$ lapse [10] and the only superluminal characteristic speed [11]. Given our standard choice of initial ("precollapsed") lapse, in the weak field the lapse tends to a value of $1 - \alpha \propto r^{-1}$, both before and after the wave packet passes. Thus away from the wave packet, the $r(1 - \alpha)$ plotted in Fig. 1 tends to constant values in the weak field.

As the sharp lapse wave packet travels outward from the strong-field region, it crosses multiple AMR refinement boundaries. At each boundary, the numerical resolution drops discontinuously by a factor of two [12]. This crossing partially reflects the wave and generates substantial numerical noise in its wake.

Although this is a gauge wave, errors in its evolution directly contaminate the evolution of physical quantities, particularly the extrinsic curvature, which depends on

second-order derivatives of the lapse. As Einstein's equations are highly nonlinear, errors in the lapse's evolution contaminate all physical quantities, including the Einstein constraints. The right panel of Fig. 1 illustrates the roughly seven-decade jump in the Hamiltonian constraint violation $\mathcal{H}$ from its initial values following the passage of this wave. Elevated levels of $\mathcal{H}$ violation persist near refinement boundaries, evident as square shapes in the color map.

Meanwhile, in the strong-field region, $\mathcal{H}$ violation acts as an effective energy density, directly influencing the dynamics of the compact binary and, thus, one of the core outputs of the NR simulation: the GW prediction. Regarding GW predictions, noise appears as spurious high-frequency oscillations in the outgoing-radiation Weyl scalar ψ_4. ψ_4 is decomposed into spin-weight $s = -2$ spherical harmonics and integrated twice in time to compute the GW strain. While the noise is evident in the $(\ell = m = 2)$ dominant mode of ψ_4, it completely obscures modes with $\ell > 4$. Such high-order modes may be detectable by next-generation GW observatories, even in near-symmetric binaries [13]. Thus minimizing this noise is crucial for precise BBH waveform predictions for upcoming GW detectors.

We emphasize that this discussion has been limited to Cartesian AMR grids not set to refine at the sharp gauge features, as this approach is widely used in the field.

Improved refinement algorithms or grid structures may result in significantly less constraint violation and GW noise than what is reported here.

This paper introduces improvements to the moving-puncture approach, aimed at curtailing the development of sharp features in the lapse and mitigating their impact. Our first two improvements focus on decreasing the strength of low-pass filters near punctures—where sharp fields are desirable—and increasing the strength far away.

The first improvement multiplies the strength parameter of Kreiss-Oliger (KO) dissipation by $e^{-4\phi}$, where ϕ is the standard BSSN conformal factor, and increases its strength from typical values by 2 to 5 times.

The second improvement modifies a powerful but not widely used low-pass parabolic filter, which adds a term proportional to $\mathcal{H}$ to the right-hand side of the ϕ evolution equation. In the same vein as [14], we set the strength of the filter based on the maximum value allowed by the Courant-Friedrichs-Lewy (CFL) condition, according to the local numerical grid spacing. This enables us to increase its strength by more than an order of magnitude on coarser numerical grids in the region far from the punctures.

Our third improvement adds a term to the $1 + \log$ lapse evolution equation that slows the evolution of α in the area immediately outside the BH horizon. This arrests the formation of the sharp lapse wave packet, both reducing its magnitude and greatly smoothing it.

Modifying the moving-puncture approach in Cartesian-AMR BBH evolutions to mitigate the impact of the sharp lapse feature is not a new idea; [10] introduced several techniques to this end, achieving about a 40% reduction in GW noise (in ψ_4) and about 1.3 orders of magnitude reduction in $|\mathcal{H}|$ violations outside the horizons. While these improvements were effective, their recipes depend on careful tuning to the specific grid structure.

While we adopt the BSSN formalism (like [10]) and employ a constraint damping term, the moving-puncture approach is also widely used within the CCZ4 and Z4c formalisms, which contain constraint damping terms. [15] performed evolutions of puncture BBHs in the Z4c formalism, reporting 1–2 orders of magnitude reductions in Hamiltonian constraint violation outside the punctures, and GW phase and amplitude errors up to a factor of two smaller as compared to the standard BSSN moving-puncture approach.

By contrast, the three improvements presented in this work, when combined, reduce Hamiltonian (momentum) constraint violations in the strong-field region by about 2 (3) orders of magnitude and in the GW extraction zone by 5 (2) orders of magnitude. Further, constraint violation convergence is generally improved, implying that these gaps widen with increased resolution. Coordinate eccentricity is slightly decreased, and numerical noise in the dominant $\ell = m = 2$ mode of ψ_4 is reduced by an average factor of 4.3 across resolutions. Fidelity of higher-order

modes is greatly improved, enabling the $\ell = m = 6$ mode, previously obscured by noise, to become visible.

Although our focus is primarily on the impact of these improvements on BBH evolutions with the BSSN formalism, we expect that they will also be beneficial in other compact binary evolutions and formalisms that use the moving-puncture technique. Determining whether our improved techniques significantly improve the fidelity of binary neutron star (BNS) simulations will be a focus of an upcoming paper. In addition, formulations like Z4c have proven quite effective in reducing constraint violations in BNS simulations [15]. Comparing BNS evolutions performed with our improved BSSN moving-puncture approach to Z4 techniques widely used across the community will also be a focus of future work.

The rest of the paper is structured as follows. Section II introduces the basic equations, including the BSSN evolution, moving-puncture gauge, and constraint equations. Building on this, Sec. III presents prescriptions for all three of our improved techniques. Next, Sec. IV describes our algorithmic approach for simulating the BBH mergers presented in this work. Results from these simulations, discussed in Sec. V, demonstrate the impacts of our improved techniques. Finally, we conclude and describe avenues for future work in Sec. VI.

II. BASIC EQUATIONS

We adopt $G = c = 1$ units, and set our mass scale $M = 1$ to be the sum of the individual puncture masses reported by our initial data solver TwoPunctures. In these units the total ADM mass is 0.989946. Further we adopt standard Einstein summation rules on repeated indices, and the convention that Greek and Latin indices represent four- and three-dimensional quantities, respectively.

A. BSSN evolution equations

In this paper, we investigate the numerical evolutions of BBHs using the covariant "Lagrangian" BSSN formulation in vacuum, as presented in [16]. This formulation has been adapted for spherical coordinate reference metrics by [17] and for more general reference metrics by [18].

Starting with the ADM line element,

$$ds^2 = -\alpha^2 dt^2 + \gamma_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt), \quad (1)$$

γ_{ij}, β^i, and α denote the spatial 3-metric, shift vector, and lapse, respectively.

We then define the conformal metric $\bar{\gamma}_{ij}$ as

$$\bar{\gamma}_{ij} = \left(\frac{\bar{\gamma}}{\gamma}\right)^{\frac{1}{3}}\gamma_{ij} = e^{-4\phi}\gamma_{ij}, \quad (2)$$

where ϕ represents the conformal factor. Instead of evolving ϕ, we follow previous works (e.g., [19,20]) and

evolve $W = e^{-2\phi}$, as W exhibits greater smoothness near puncture BHs, resulting in reduced truncation errors (also see discussion in [3]). To this end, derivatives of the conformal factor ϕ appearing in the BSSN evolution equations are calculated with respect to the evolved variable W via $\partial_\mu W = -2e^{-2\phi}\partial_\mu\phi \Rightarrow \partial_\mu\phi = -\partial_\mu W/(2W)$, and so forth for second derivatives.

We also introduce a reference metric $\hat{\gamma}_{ij}$, defined by

$$\bar{\gamma}_{ij} = \hat{\gamma}_{ij} + \varepsilon_{ij}. \tag{3}$$

In this work, the reference metric corresponds to the Cartesian metric $\hat{\gamma}_{ij} = \delta_{ij}$, which remains constant in both space and time. Thus, all derivatives of the conformal metric $\bar{\gamma}_{ij}$ are evaluated directly with respect to ε_{ij}, since $\partial_\mu\bar{\gamma}_{ij} = \partial_\mu\varepsilon_{ij}$. Although analytically equivalent, computing finite-difference derivatives of diagonal metric components with respect to ε_{ij} instead of $\bar{\gamma}_{ij} = 1 + \varepsilon_{ij}$ results in reduced roundoff error in the weak-field region.

Building on these definitions, we adopt the following time-evolution equations, replacing all occurrences of ϕ and its derivatives with equivalent expressions in terms of W:

$$\partial_t\varepsilon_{ij} = \mathcal{L}_\beta\varepsilon_{ij} + \frac{2}{3}\bar{\gamma}_{ij}(\alpha\bar{A}^k_k - \bar{D}_k\beta^k) - 2\alpha\bar{A}_{ij}, \tag{4}$$

$$\partial_t\bar{A}_{ij} = \mathcal{L}_\beta\bar{A}_{ij} - \frac{2}{3}\bar{A}_{ij}\bar{D}_k\beta^k - 2\alpha\bar{A}_{ik}\bar{A}^k{}_j + \alpha\bar{A}_{ij}K$$
$$+ e^{-4\phi}\{-2\alpha\bar{D}_i\bar{D}_j\phi + 4\alpha\bar{D}_i\phi\bar{D}_j\phi$$
$$+ 4\bar{D}_{(i}\alpha\bar{D}_{j)}\phi - \bar{D}_i\bar{D}_j\alpha + \alpha\bar{R}_{ij}\}^{\mathrm{TF}}, \tag{5}$$

$$\partial_t\phi = \mathcal{L}_\beta\phi + \frac{1}{6}(\bar{D}_k\beta^k - \alpha K), \tag{6}$$

$$\partial_t K = \mathcal{L}_\beta K + \frac{1}{3}\alpha K^2 + \alpha\bar{A}_{ij}\bar{A}^{ij} - e^{-4\phi}(\bar{D}_i\bar{D}^i\alpha + 2\bar{D}^i\alpha\bar{D}_i\phi), \tag{7}$$

$$\partial_t\bar{\Lambda}^i = \mathcal{L}_\beta\bar{\Lambda}^i + \bar{\gamma}^{jk}\hat{D}_j\hat{D}_k\beta^i + \frac{2}{3}\Delta^i\bar{D}_j\beta^j$$
$$+ \frac{1}{3}\bar{D}^i\bar{D}_j\beta^j - 2\bar{A}^{ij}(\partial_j\alpha - 6\partial_j\phi)$$
$$+ 2\bar{A}^{jk}\Delta^i_{jk} - \frac{4}{3}\alpha\bar{\gamma}^{ij}\partial_j K, \tag{8}$$

where the TF superscript indicates the trace-free part, and $\mathcal{L}_\beta$ denotes the Lie derivative along the shift vector β^i. The covariant derivatives with respect to the reference metric $\hat{\gamma}_{ij}$ and the barred spatial 3-metric $\bar{\gamma}_{ij}$ are represented by $\hat{D}_j$ and $\bar{D}_j$, respectively. The tensor Δ^i_{jk} is formed from the difference between the barred and hatted Christoffel symbols:

$$\Delta^i_{jk} = \bar{\Gamma}^i_{jk} - \hat{\Gamma}^i_{jk},$$

where Δ^i is defined $\Delta^i \equiv \bar{\gamma}^{jk}\Delta^i_{jk}$. The conformal, trace-free extrinsic curvature, $\bar{A}_{ij}$, is given by

$$\bar{A}_{ij} = e^{-4\phi}\left(K_{ij} - \frac{1}{3}\gamma_{ij}K\right),$$

where K represents the trace of the extrinsic curvature K_{ij}.

B. Moving-puncture gauge evolution equations

To close the system of time-evolution equations, we choose the moving-puncture gauge conditions [3,4,9], which include the $1 + \log$ slicing condition for the lapse and the Γ-driving (here, Λ takes the place of Γ) condition for the shift:

$$\partial_t\alpha = \beta^j\partial_j\alpha - 2\alpha K, \tag{9}$$

$$\partial_t\beta^i = \beta^j\partial_j\beta^i + B^i, \tag{10}$$

$$\partial_t B^i = \beta^j\partial_j B^i + \frac{3}{4}(\partial_t\bar{\Lambda}^i + \beta^j\partial_j\bar{\Lambda}^i) - \eta\beta^i, \tag{11}$$

where the Γ-driving shift damping parameter is set in this paper to $\eta = 1/M$.

Variations in moving-puncture gauge conditions, such as a first-order implementation of the shift condition, do exist. Although we did not explore these alternatives in our study, we direct the reader to van Meter *et al.* [11] for a comprehensive analysis of different implementations. Our choice of gauge conditions was primarily motivated by their finding that incorporating advection terms into the lapse and shift conditions leads to well-behaved characteristic speeds.

C. BSSN constraint equations

Analytically (as opposed to numerically), the BSSN Hamiltonian constraint $\mathcal{H} = 0$, so any expression for $\mathcal{H}$ can be multiplied by an arbitrary prefactor and still remain perfectly valid. As such, two expressions for the BSSN Hamiltonian constraint $\mathcal{H}$ are commonly found in the literature, differing only in their prefactors.

In this work, we adopt $\mathcal{H}$ with a negative prefactor in front of $\bar{D}^2\phi$, its principal part. To differentiate our definition of $\mathcal{H}$ from other choices, we will refer to it as $\mathcal{H}_-$ when necessary:

$$\mathcal{H}_- = \frac{2}{3}K^2 - \bar{A}_{ij}\bar{A}^{ij} + e^{-4\phi}(\bar{R} - 8\bar{D}^i\phi\bar{D}_i\phi - 8\bar{D}^2\phi) = \mathcal{H} = 0. \tag{12}$$

The second convention (e.g., [21]) multiplies both sides by $-e^{5\phi}/8$. This results in a sign flip on $\mathcal{H}$, and thus the prefactor on $\bar{D}^2\phi$ becomes positive. We therefore refer to this form as $\mathcal{H}_+$:

$$\mathcal{H}_+ = -\frac{e^{5\phi}}{8}\mathcal{H}_-. \tag{13}$$

Although the two expressions $\mathcal{H}_+ = \mathcal{H}_- = 0$ are equivalent, it is important to distinguish between them. As we will soon see, the specific implementation of Hamiltonian constraint damping (Sec. III B) will depend on the choice of this prefactor.

Regarding the physical implications of constraint violation, if we write $\mathcal{H}_-$ in its general form with the matter source term included,

$$\mathcal{H}_-^m = \mathcal{H}_- - 16\pi\rho = 0, \tag{14}$$

then we may conclude that, in vacuum evolutions, positive Hamiltonian constraint violations $\mathcal{H}_-$ behave as a positive *effective* energy density ρ_{eff}, and negative violations as a negative ρ_{eff}:

$$\mathcal{H}_- = 16\pi\rho_{\text{eff}}. \tag{15}$$

As for the momentum constraint, we adopt the following form:

$$\mathcal{M}^i = e^{-4\phi}\left(\hat{D}_j\bar{A}^{ij} + 2\bar{A}^{k(i}\Delta_{jk}^{j)} + 6\bar{A}^{ij}\partial_j\phi - \frac{2}{3}\bar{\gamma}^{ij}\partial_j K\right) = 0, \tag{16}$$

and define the scalar $\mathcal{M}^2 = \bar{\gamma}_{ij}\mathcal{M}^i\mathcal{M}^j$, which we use for diagnostics in this work.

As Hamiltonian and momentum constraint violations are defined at each of the $\gtrsim 10^7$ points in our computational domain, for convenience we will often compute them as a volume-averaged L_2 norm:

$$\|\mathcal{H}\| = \sqrt{\frac{\int_\mathcal{V} \mathcal{H}^2 d^3x}{\int_\mathcal{V} d^3x}}, \tag{17}$$

$$\sqrt{\|\mathcal{M}^2\|} = \sqrt[4]{\frac{\int_\mathcal{V}(\bar{\gamma}_{ij}\mathcal{M}^i\mathcal{M}^j)^2 d^3x}{\int_\mathcal{V} d^3x}}, \tag{18}$$

where the volume $\mathcal{V}$ will generally exclude small spheres around the BHs and be tailored to either a strong-field or weak-field region.

III. IMPROVED TECHNIQUES

We present three techniques that lower numerical errors in moving-puncture simulations, with minimal additional computational cost. These include curvature-adjusted Kreiss-Oliger (CAKO) dissipation, coarse-grid-adjusted Hamiltonian constraint damping (CA$\mathcal{H}$D), and the slow-start lapse (SSL) technique.

These improvements are meant to be both easy to implement and agnostic to numerical grid structures. To this end, while we demonstrate their effectiveness in the widely used Einstein TOOLKIT [22,23] Carpet [24,25] AMR
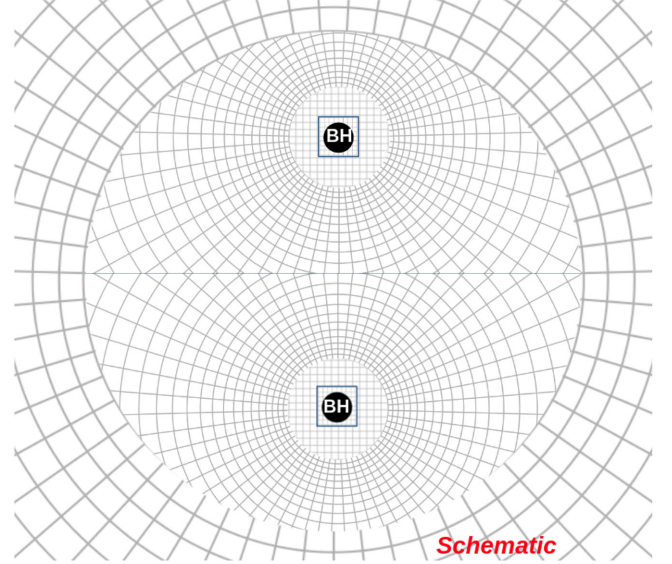

FIG. 2. Schematic of BBH inspiral numerical meshes in Blackholes@Home, the code used to develop and tune the improvements presented here.

infrastructure, the new techniques were originally devised and tuned during the development of a new, super-efficient NR code called Blackholes@Home.

Illustrated in Fig. 2, Blackholes@Home adopts a unique multipatch grid structure that combines grids in spherical-like and Cartesian-like coordinates. Einstein's equations are solved in the corresponding spherical-like and Cartesian-like bases, leveraging recent developments in covariant BSSN [16], BSSN with a reference metric [17,18], and NR code generation with NRPy [18,26].

With these new grid structures, Blackholes@Home performs full, $3+1$ NR BBH evolutions using about 3 GB of RAM or less, enabling its use on consumer-grade desktop computers. As such, it aims to generate the largest-ever full-NR BBH GW catalog through an upcoming volunteer-computing project of the same name.

Owing to Blackholes@Home's high efficiency, the improvements presented here were tuned over approximately 10 simulations of a single BBH scenario, and the tuned values were proven similarly effective across more than 100 different BBH scenarios simulated with Blackholes@Home. These scenarios spanned mass ratios from $q=1$ to $q=6$ and included arbitrarily directed spin vectors with dimensionless magnitudes up to 0.5.

Building on the success of these Blackholes@Home -tuned improvements, this work focuses on evaluating their performance on Cartesian AMR grids. To this end, we implement the improvements within our BaikalVacuum Einstein TOOLKIT thorn, and release the improved version of BaikalVacuum used in this paper's open source [27]. This updated version of BaikalVacuum and its sister (nonvacuum BSSN) code Baikal will be incorporated into a future Einstein TOOLKIT release.

We next describe each improvement in detail.

A. Curvature-adjusted Kreiss-Oliger dissipation

Kreiss-Oliger dissipation [28], a widely adopted low-pass filter in numerical relativity, reduces numerical noise. In Cartesian coordinates, the addition of the nth-order KO dissipation operator (where n is odd) to the right-hand side of each evolution equation is formulated as

$$D_{\mathrm{KO}}^n \vec{f} = \epsilon_{\mathrm{KO}} \frac{(-1)^{(n+3)/2}}{2^{n+1}} (\Delta x^i)^n \partial_i^{(n+1)} \vec{f}, \qquad (19)$$

where $\vec{f}$ denotes the vector of evolved functions on a numerical grid, ϵ_{KO} is the dimensionless KO strength factor, and $\partial_i^{(n+1)}$ represents the $(n+1)$st partial derivative in the x^i direction. This derivative is generally implemented by a centered finite-difference operator with the smallest possible stencil. For example, first-order KO dissipation includes the second derivative, which is represented by the standard 3-point stencil: $\partial_x^2 f = (f_{i-1} - 2f_i + f_{i+1})/(\Delta x)^2$. The $(\Delta x^i)^n$ prefactor ensures that the operator converges to zero in the continuum limit; KO dissipation is purely a numerical filter acting to damp high-frequency oscillations.

When solving a partial differential equation (PDE) with a dominant finite-difference truncation error of jth order (j even), the $n = j + 1$-order KO dissipation operator is most often chosen, with the KO operator stencil comprising $j + 3$ grid points. In this study, we use $j = 8$, and apply ninth-order KO dissipation to the central point on an 11-point stencil.

Larger KO strength factors ϵ_{KO} lead to stronger filtering; however, ϵ_{KO} is CFL limited to values of approximately one or less when typical CFL factors and numerical integration schemes are chosen for evolving Einstein's equations in the moving-puncture formalism [29]. This said, so long as the CFL-limited value is not exceeded, ϵ_{KO} need not be the same for all evolved grid functions.

In [30], we found that setting $\epsilon_{\mathrm{KO,gauge}} = 0.9$ for the gauge quantities and $\epsilon_{\mathrm{KO,other}} = 0.3$ for the other spacetime quantities enhances stability when evolving boosted and unboosted near-extremal-spin ($a/M = 0.99$) BHs. Several years later, in [10], we refined this approach in the context of a mildly spinning BBH evolution. Building on the idea that the sharp fields inside and around puncture BHs are natural and should not be smoothed, we made $\epsilon_{\mathrm{KO,gauge}}$ a function of time and the distance from the center-of-mass r, such that it increased gradually over time from 0.3 to 0.98 in the weak-field region. Additionally, we adjusted $\epsilon_{\mathrm{KO,gauge}}$ to drop to zero at punctures by multiplying by the conformal factor W raised to some fractional power.

While this refinement was quite effective in the context of the BBH evolution presented in that paper, $\epsilon_{\mathrm{KO,gauge}}(t, r, W)$ would need to be modified and potentially retuned for BBH scenarios at different initial separations and potentially different AMR grid structures.

Here we introduce a much simpler adjustment to ϵ_{KO}: curvature-adjusted KO dissipation. It multiplies the KO dissipation strength by the conformal factor $W = e^{-2\phi}$,

which ensures the KO strength is nearly zero near punctures, tending to $\epsilon_{\mathrm{KO,CA}}$ in the weak field:

$$\epsilon_{\mathrm{KO}}(W) = W \epsilon_{\mathrm{KO,CA}}. \qquad (20)$$

Similar to our past work, we set $\epsilon_{\mathrm{KO,CA}}$ to 0.99 (near the CFL-limited maximum) for gauge quantities and to 0.3 for all other BSSN evolved variables. But unlike our past work, this prescription has been validated using both BLACKHOLES@HOME grid structures across more than 100 unique scenarios across BBH parameter space, and (as shown in this work) AMR grid structures within the EINSTEIN TOOLKIT in the context of a GW150914-like BBH system at multiple numerical resolutions.

B. Coarse-grid-adjusted $\mathcal{H}$ damping (CA$\mathcal{H}$D)

In 2002, [31] found that analytically, the following modification to the right-hand side (RHS) of the conformal factor evolution equation causes one constraint violating mode to exponentially damp:

$$\partial_t \phi = \partial_t \phi_{\mathrm{orig}} + \kappa_{\phi\mathcal{H}} \alpha \mathcal{H}_-, \qquad (21)$$

where $\kappa_{\phi\mathcal{H}} \leq 0$ for stability.

As $\mathcal{H} = 0$ in the continuum limit, this modification drops to zero with increasing resolution. Further, for unit consistency $\kappa_{\phi\mathcal{H}}$ must have units of mass M (i.e., the same as units of time and space in $G = c = 1$ units).

When analyzing the principal part of the adjusted $\partial_t \phi$ equation in the weak-field limit, we find

$$\partial_t \phi = \mathcal{H}_{-,\mathrm{pp}} = -8\kappa_{\phi\mathcal{H}} \partial_i \partial^i \phi, \qquad (22)$$

which is a *parabolic* PDE. If this parabolic equation is solved explicitly with a 3D Laplacian [10,32] in flat space, then at a numerical grid point with $\Delta s = \min(\Delta x^i)$ across all three spatial dimensions i, the CFL-constrained time step Δt is given by

$$-8\kappa_{\phi\mathcal{H}} \frac{\Delta t}{(\Delta s)^2} \leq \frac{1}{6} \Rightarrow \Delta t \leq \frac{(\Delta s)^2}{-48\kappa_{\phi\mathcal{H}}}. \qquad (23)$$

The $(\Delta s)^2$ on the right-hand side is undesirable. For instance, suppose a simulation is performed at the maximum stable $\kappa_{\phi\mathcal{H}}$ at a given resolution Δs and corresponding CFL-limited $\Delta t \propto \Delta s$. If a higher resolution (smaller Δs) is found necessary for a convergence study, $\kappa_{\phi\mathcal{H}}$ would need to be adjusted downward in order for the CFL condition to be satisfied. This would result in an inconsistency in the strength parameter across different resolutions, potentially invalidating the convergence study.

To address this, [21] included the simulation time step in the prefactor:

$$\partial_t \phi = \partial_t \phi_{\mathrm{orig}} + K \Delta t \mathcal{H}_+, \qquad (24)$$

where K is dimensionless and $K \geq 0$ for stability. Numerically, the term $\Delta t \propto \Delta s$ decreases the strength of the parabolic operator in direct proportion to the CFL limit, making this modification stable at all choices of grid resolution provided it is stable at a given choice of K. From an analytical perspective, both the Δt prefactor and $\mathcal{H}$ drop to zero in the continuum limit. Choosing $K = 0.02$–0.06, they showed that this damping term reduced constraint violations by almost an order of magnitude in the context of BNS simulations.

More recently, [32] explored the effect of three different coefficient prescriptions in the context of BNS evolutions. These include Eq. (24) with K set to a constant across all AMR grids, Eq. (24) with K *decreased* on coarser grids according to $\Delta s_{\min}/\Delta s_n$ (where Δs_n is the Δs on the nth refinement level), and the third method,

$$\partial_t \phi = \partial_t \phi_{\text{orig}} + c_{\mathcal{H}} \mathcal{H}_+, \qquad (25)$$

where $c_{\mathcal{H}}$ is set to a fixed dimensionful constant (independent of Δt). They found the third approach preferable because it improved the convergence properties of the postmerger phase. Additionally, they attributed the decreasing effectiveness of the [21] approach at higher resolutions to its relatively poor convergence behavior. As discussed, setting the coefficient on the parabolic operator to a fixed value results in instability if the evolution is repeated with sufficiently high resolution, so this approach is best applied when the highest resolution can be known *a priori*.

Our improvement to the approach inverts the second prescription of [32]. Instead of decreasing the strength K on coarser grids, we increase the strength according to the maximum allowed time step on the refinement level, similar to the approach of [14] for accelerating hyperbolic relaxation waves across a curvilinear numerical grid. Specifically, we adjust Eq. (24) as follows:

$$\partial_t \phi = \partial_t \phi_{\text{orig}} - C \Delta t_n \left[\text{CFL}_0 \frac{\Delta s_n}{\Delta t_n} \right] \mathcal{H}_-, \qquad (26)$$

where the n subscript refers to the local value of a given quantity on grid n, and CFL_0 is the CFL factor on numerical grid $n = 0$, which contains the smallest time step. We find $C = 0.15$ results in stable evolutions across BBH parameter space explored by Blackholes@Home, including nonzero spins and mass ratios from $q = 1$ to $q = 6$, and we generally set CFL_0 to 90% of its highest possible value as a safety margin.

On the finest AMR level ($n = 0$), the bracketed term reduces to 1. This approach maximally benefits standard (2∶1 refined) AMR grids with a global time step $\Delta t_n = \Delta t_0 = \Delta t$, as it *exponentially* increases the damping strength with n; in this limit the bracketed term reduces to 2^n. For the Δt_n used in this work (Sec. IV), the bracketed

term increases the damping strength by a factor of $2^5 = 32$ on the coarsest grid.

C. Slow-start lapse

Our third and final improvement focuses on rapid changes in the lapse near puncture BHs at the start of a spacetime evolution. As illustrated in Fig. 1, these changes send a sharp pulse outward at $\sqrt{2}c$. To minimize the early dynamics, the initial lapse α_0 is typically set to a "precollapsed" value of $\alpha_0 = W_0$, where $W_0 = e^{-2\phi_0}$ is the conformal factor W at $t = 0$.

Our slow-start lapse technique modifies the $1 + \log$ lapse evolution condition Eq. (9) as follows:

$$\partial_t \alpha = \partial_t \alpha_{\text{orig}} - W[he^{-t^2/(2\sigma^2)}](\alpha - W). \qquad (27)$$

SSL consists of three terms multiplied together. First, $-(\alpha - W)$ exponentially damps α to the (evolved) conformal factor W. By convention, at $t = 0$, α is set to W, rendering this term exactly zero at $t = 0$. Second, the Gaussian prefactor in square brackets sets the inverse damping timescale, such that the damping timescale is short and effective early in the evolution but grows to a very large and inert value later on. Third, to ensure that the wormhole-to-trumpet [33] transition inside the horizon is not affected, the damping strength is multiplied by an overall prefactor of W, which drops to zero at the puncture and smoothly increases away from the puncture.

In short, the SSL modification acts to quench rapid deviations of α from the quantity to which it is initially set, W, just outside the horizon, arresting the development of the sharp lapse feature in this region.

After many experiments across ~ 100 unique scenarios in BBH parameter space with Blackholes@Home, we find that a Gaussian height of $h = \frac{3}{5}M$ and a standard deviation of $\sigma = 20.0M$ are particularly effective in mitigating the initial lapse pulse. With these choices, SSL damps on a timescale of $e^{t^2/(2\sigma^2)}/h = \frac{5}{3}M$ at $t = 0$, $\sim 10^2 M$ at $t = 57M$, $\sim 10^5 M$ at $t = 94M$, and $\sim 10^{16}M$ at $t = 170M$. In this way, SSL impacts only the formation of the sharp lapse feature and is rendered completely ineffective shortly thereafter.

IV. ALGORITHMIC APPROACH

We present reproductions of the GW150914 gallery example [34] from the Einstein Toolkit [22,23], incorporating the aforementioned improvements in a systematic manner. This gallery example is notable, as it is perhaps the most comprehensively documented BBH NR simulation available; all NR codes, results, and plotting scripts required to reproduce the original simulation are openly accessible on Zenodo [34].

The physical parameters of the simulation are inspired by the first-ever directly detected GW event, GW150914 [35]. In particular, it uses the best-fit parameters for GW150914 [36]:

a mass ratio of $q = \frac{36}{29}$, with dimensionless spin parameters of $\vec{\chi}_M = \{\chi_M^x, \chi_M^y, \chi_M^z\} = \{0, 0, 0.31\}$ for the more massive BH and $\vec{\chi}_m = \{\chi_m^x, \chi_m^y, \chi_m^z\} = \{0, 0, -0.46\}$ for the less massive one. The initial separation is set to $10M$, which places the BHs approximately 5 orbits before common-horizon formation.

As in the gallery example, we use the Einstein TOOLKIT [22,23] and its Carpet [24,25] AMR infrastructure to perform these simulations. The numerical parameters for the evolution are largely the same: eighth-order finite differencing, fifth-order prolongation in space, second-order prolongation in time, RK4 time stepping. The TwoPunctures thorn is used to construct initial data [37], AHFinderDirect [38] for apparent-horizon diagnostics, QuasiLocalMeasures [39] for isolated horizon diagnostics, and the Kranc-generated [40] WeylScal4 for ψ_4 extraction. However, there are some notable differences.

First, instead of the ML_BSSN [40–42] thorn, the NRPy-generated [18] BaikalVacuum thorn, modified as needed with the improved techniques, is used for the spacetime evolution. BaikalVacuum and its sister code Baikal are BSSN evolution codes in the Einstein TOOLKIT: BaikalVacuum solves the BSSN equations without stress-energy source terms, and Baikal includes these source terms. The latter was used in [43] and [44] for the evolving the spacetime fields in BNS simulations.

Second, the TwoPunctures grid resolution was increased from its default values of $N_A \times N_B \times N_\phi = 30 \times 30 \times 16$ to $48 \times 48 \times 20$, as the default values result in initial data constraint violations dominating the early evolution. Also, the initial lapse was chosen to be $W = e^{-2\phi}$, instead of the "|TwoPunctures-averaged|" value, which sets the lapse to $\alpha(r) = \frac{1}{2} \sum_{i=1}^{2}(1 - m_i/(2r_i))$ for each puncture i of mass m_i at distance r from r_i. The former approach has the advantage of exhibiting the same asymptotic behavior at large radius and late times: $\alpha_\infty = 1 - M/r$ for total mass M. Meanwhile, the latter approach tends to $1 - M/(2r)$ in the weak field. While TwoPunctures has an option to set the initial lapse to any power of $\psi = e^\phi$, it chooses a value of ψ that does not include the correction term u [Eq. (30) in [37]]. To set the initial value of the lapse to the evolved quantity W, as needed for SSL, we modified TwoPunctures to include this option. Our modified version may be found in this GitHub repository [27].

Third, the Llama [45] cubed-spheres grid in the gravitational wave zone was replaced with box-in-box Cartesian adaptive mesh refinement grids of comparable or higher resolution. In addition, we choose a default resolution of all AMR grids 50% higher, so that, for example, the resolution at the punctures is $M/42$ instead of the gallery example value of $M/28$. As we will find in Sec. V, even $M/42$ lies slightly outside the convergent regime for modeling this physical scenario. For completeness, our AMR grid hierarchy comprises 11 levels of overlapping grid cubes, centered on and comoving with each puncture, with a

standard factor-of-2 refinement across adjacent AMR levels (i.e., there are 10 levels of refinement built upon the coarsest level). The AMR grid cubes have half-sidelengths of $\{\frac{2}{3}M, \frac{4}{3}M, \frac{8}{3}M, \frac{16}{3}M, \frac{64}{3}M, \frac{128}{3}M, \ldots, \frac{4096}{3}M\}$. As our default choice we choose grid resolutions of $\{\frac{M}{42}, \frac{M}{21}, \frac{M}{10.5}, \ldots, 24.38M\}$ with CFL factors of $\frac{1}{640}\{288, 288, 144, 72, 36, 18, 9, 9, 9, 9, 9\}$, respectively.

Finally, the imposed symmetry across the z axis was removed as BaikalVacuum does not currently support symmetries.

V. RESULTS

To investigate the individual and combined effectiveness of the improvements introduced in this work, we conducted five simulations based on the GW150914-inspired scenario detailed in Sec. IV at the default resolution. These simulations are denoted as Control, CAKO, CA$\mathcal{H}$D, SSL, and All3. Control serves as the no-improvement baseline, while the CAKO, CA$\mathcal{H}$D, and SSL each integrate a single enhancement: curvature-adjusted Kreiss-Oliger, $\mathcal{H}$-damping, and slow-start lapse, respectively. All3 combines all three enhancements. For convergence analyses, we executed the Control and All3 simulations at resolutions 1.25x, 1.5x, and 1.71x higher than the default resolution (1.0x) [46].

A. Effect of improvements on trajectories and lapse evolution

As the three improvements modify the evolution equations in the finite-resolution limit, as well as the $1 + \log$ lapse condition, it is reasonable to be concerned that All3 may influence the trajectories of the BHs, especially at lower resolutions, possibly inducing unwanted eccentricity. To investigate this, in Fig. 3 a cubic polynomial $P_3(t)$ is fit to $d_M(t)$—the distance of the more massive BH from the center of mass (the origin)—over the first three orbits ($0 \leq t \leq 600M$).

$d_M(t)$ and $P_3(t) - 0.1$ are plotted in the top panels of each plot. Comparing the top and bottom plots, the $P_3(t)$ fits—and thus the trajectories—appear to converge with increased resolution. In addition, the sharp nonconvergent oscillation in Control at $t \approx 20M$ is eliminated in All3. Comparing the effects of the individual improvements, we find that the SSL enhancement is responsible for the removal of this sharp feature. As SSL most significantly impacts the early gauge evolution, this is not surprising.

Regarding possible eccentricity enhancement in All3, the residual $d_M(t) - P_3(t)$ (bottom panels of Fig. 3) shows the *opposite* is the case: at both default and highest resolutions, the coordinate eccentricity over the first three orbits is slightly *lowered* in All3 as compared to Control. That is to say, the amplitude of the residual oscillation is reduced in All3 as compared to Control. Further analysis reveals that All3 reduces the average absolute value of the residual by 33% at low resolution, from 0.0088 in Control
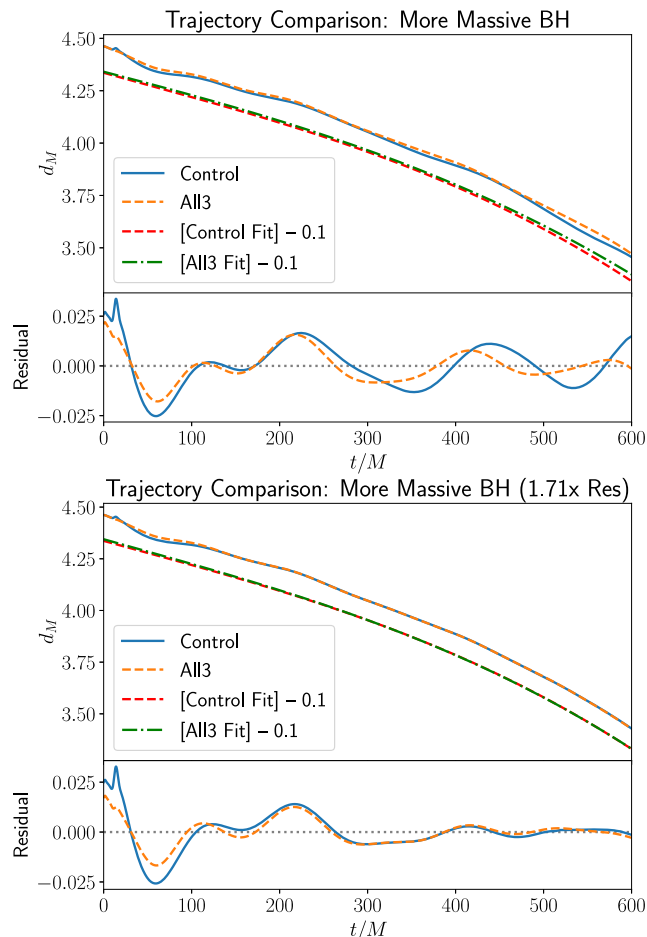
FIG. 3.   Trajectory comparison: Control vs All3. Top panels, upper curves: distance from the origin of the more massive BH d_M, comparing Control (blue solid) and All3 (orange dashed) over the first three orbits. Top panels, lower curves: least-squares cubic polynomial fit $P_3(t)$ to All3 (red dotted) and Control (green dotted), minus 0.1 to improve visibility. Bottom panels: difference between data and cubic polynomial fit.

FIG. 4.   Effects of SSL on early lapse evolution: $\alpha - W$ at various times during the early evolution along the x axis, in the vicinity of the more massive BH, comparing Control (blue solid) with SSL (orange dashed).

to 0.0059 in All3, and by 25% at the highest resolution, from 0.0059 to 0.0044. We stress that this is a measure of the smoothness of $d_M(t)$, not eccentricity.

We next turn our attention to further analyzing the effects of SSL. As discussed in Sec. III C, SSL slows the deviation of the lapse away from $\alpha = W$ outside the horizon at early times. Given that we adopt the standard "precollapsed" lapse choice $\alpha = W$ at $t = 0$, SSL has no impact on the evolution at $t = 0$.

Figure 4 illustrates how the evolution of the lapse deviates from the conformal factor shortly after the simulation begins, comparing $\alpha - W$ in Control to SSL at times $t/M = 1.37$, 5.49, and 10.97. For comparison, the first orbital period completes at $t \approx 222M$.

The more massive puncture is initially situated at $x = 4.46$, moving off the x axis shortly after the evolution begins. Thus the top panel shows the evolution very close to the puncture, and lower panels depict slices at larger
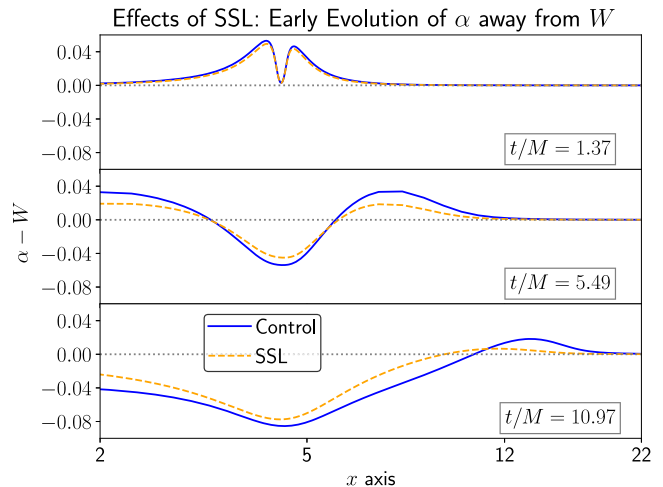
distances from the puncture and at later times. The conformal factor along the axis smooths significantly away from the puncture, so the oscillations in $\alpha - W$ at $t \geq 5.49M$ are primarily due to the lapse. While the effects of SSL are barely noticeable at $t = 1.37M$, they become quite apparent starting at $t = 5.49M$; SSL causes the oscillations in the lapse to be greatly reduced in amplitude, both at $t \approx 5.49M$ and $t \approx 10.97$.

The sharp lapse feature propagates outward and induces a great deal of constraint violation in its wake, as discussed in Sec. I. Figure 5 shows how each improvement impacts the sharp lapse feature evolution in the wave zone. As in Fig. 1, $r(1 - \alpha)$ is plotted, which should tend to a constant value in the limit $r \to \infty$.

At $t = 175.5M$, the wavefront in Control, CA$\mathcal{H}$D, and CAKO has propagated to roughly $x = 250M$, consistent with a wave propagating from near the origin at $t = 0$ at a speed of $\sqrt{2}c$. Notice this feature partially reflects off of the $x = 170.\bar{6}M$ AMR grid boundary, due to undersampling across this factor-of-two drop in resolution. Meanwhile, the delayed evolution of the lapse away from W early in the SSL and All3 evolution acts to delay the wavefront by approximately $t = 26M$; these wavefronts at $t = 201.6M$ overlap those of runs without SSL at $t = 175.5M$.

As the CA$\mathcal{H}$D and Control results are indistinguishable, we conclude that CA$\mathcal{H}$D has no significant impact on the evolution of the lapse itself. On the other hand, CAKO acts as a strong low-pass filter in this weak-field region, significantly damping oscillations in the sharp lapse feature and its reflections over the standard KO prescription adopted in Control and CA$\mathcal{H}$D.

Most notably, the SSL modification to the $1 + \log$ lapse condition significantly smooths the outgoing lapse pulse, resulting in virtually no reflection across the $x = 170.\bar{6}M$
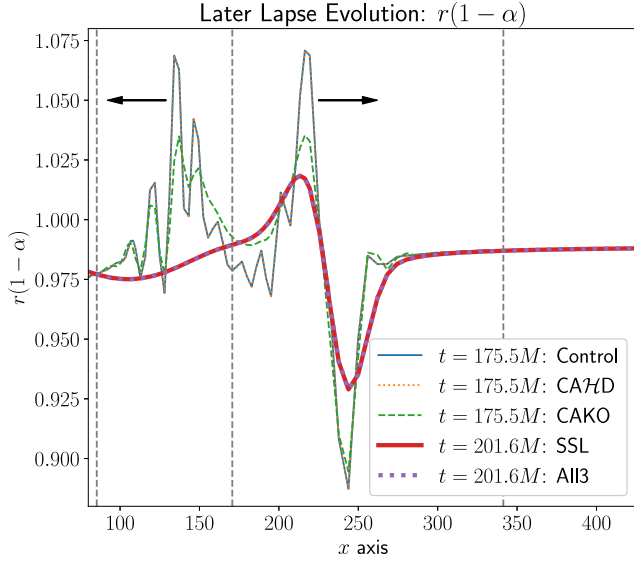
FIG. 5. Impact of improvements on the sharp lapse feature: $r(1 - \alpha)$ is plotted at $t/M = 175.5$ along the x axis for Control (thin solid), CA$\mathcal{H}$D (thin dotted; overlaps Control), and CAKO (thin dashed), and at $t/M = 201.6$ for SSL (thick solid) and All3 (thick dotted; overlaps SSL). Dashed vertical lines demarcate AMR boundaries along the x axis at $x = 85.\bar{3}$, $170.\bar{6}$, and $341.\bar{3}$. The transmitted and reflected wave pulse propagation directions in Control, CA$\mathcal{H}$D, and CAKO across the $170.\bar{6}$ AMR grid boundary are denoted by the horizontal arrows.

grid boundary. While this grid boundary generates minimal noise in SSL and All3 (which perfectly overlap in this plot), partial reflection is observed in SSL and All3 when the wavefront crosses the $x = 341.\bar{3}M$ grid boundary. After partially reflecting, All3 exhibits less noise than SSL due to the enhanced KO dissipation implemented in CAKO.

B. Strong-field region: Effect of improvements on constraint violations

As discussed in [10] and illustrated in Fig. 1, the outward-propagating sharp lapse pulse fills the computational domain with significant constraint violations in its wake. The top plot in Fig. 6 compares how our improvements, both individually and collectively, reduce the volume-averaged L_2 norm of the Hamiltonian constraint violation $\|\mathcal{H}\|$ in the strong-field region.

CA$\mathcal{H}$D has the most significant influence on $\|\mathcal{H}\|$, reducing it by between 2–3 orders of magnitude after $t = 10M$. This is consistent with its intended effect of causing one Hamiltonian constraint violating mode to exponentially damp. Further, All3 reduces $\|\mathcal{H}\|$ in the strong field by an additional 0.2–1.2 orders of magnitude. The spike in constraint violation observed at $t \approx 930M$ is coincident with common horizon formation. While the same spike is only observed in CA$\mathcal{H}$D and All3, the Hamiltonian constraint violation in these cases never exceeds Control, CAKO, or SSL.
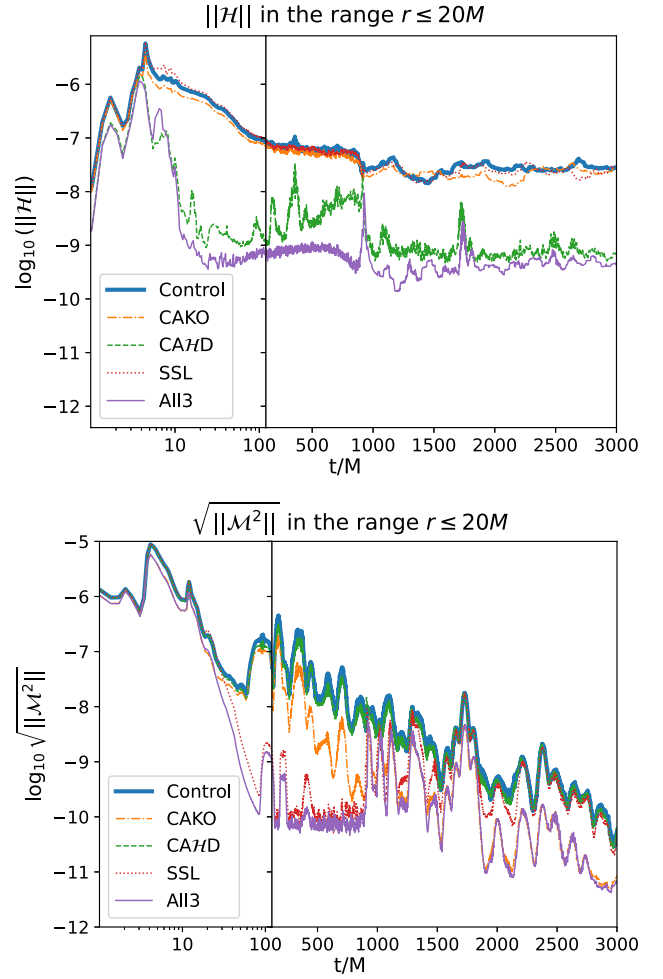

FIG. 6. Effect of improvements: Constraint violation in strong-field region: Volume-averaged L_2 norm of $\|\mathcal{H}\|$ (top plot) and $\sqrt{\|\mathcal{M}^2\|}$ (bottom plot) inside spherical region of radius $20M$, centered at the origin (center of mass), excluding two spheres of radius $1M$ centered at each BH. Results from five runs are compared: Control (thick solid blue), CAKO (dash-dotted orange), CA$\mathcal{H}$D (dashed green), SSL (dotted red), and All3 (thin solid purple). Volume excludes spheres of radius $r = 1M$ centered at each BH.

Meanwhile, as shown in the bottom plot of Fig. 6, at most times SSL greatly reduces momentum constraint violation $\sqrt{\|\mathcal{M}^2\|}$ in the strong field at early times, decreasing it by between 0–3.5 orders of magnitude. After the merger, CAKO proves to be most effective at reducing momentum constraint violation. Once more, the cumulative impact of all improvements (All3) surpasses that of any individual improvement.

Next we analyze the convergence behavior of the constraint violations. Assuming we are in the convergent regime (i.e., the region where truncation, and not roundoff or undersampling error dominates), we would expect the constraints to converge to zero between second and eighth order in our mixed-order scheme; second order corresponds

to the order at which data are interpolated (prolongated) in time at AMR grid boundaries, and eighth order corresponds to the finite-differencing order for spatial derivatives. Witnessing inconsistent convergence order in such a complex system typically implies that some physical length scale is not being properly resolved (undersampling error).

Given that constraint violations converge to zero in the continuum limit, in the truncation-error-dominated regime, we should find that the constraint violation C_i at resolution Δx_i and convergence order n should approximately obey the equation

$$C_i = A(\Delta x_i)^n, \tag{28}$$

where A is related to the $n+1$th derivative of the term in C_i that dominates the truncation error. So long as C_i is known at least two resolutions, both n and A can be computed.

Figure 7 plots the base-10 logarithm L_2 norms of Hamiltonian ($\|\mathcal{H}\|$) and momentum ($\sqrt{\|\mathcal{M}^2\|}$) constraint violations in the strong-field region, through merger and ringdown, up to $t = 1000M$. After this, the key payload from the simulation—the GW signal—has largely propagated away and no longer depends on data in this region. Data from Control and All3 at the default (1.0x) resolution, as well as 1.5x and 1.71x higher resolutions are shifted on this logarithmic scale such that the difference with the corresponding 1.25x case is minimized; from this difference an implied convergence order n can be immediately computed using Eq. (28):

$$n = \frac{\log_{10} C_i - \log_{10} C_{1.25x}}{\log_{10} \Delta x_i - \log_{10} \Delta x_{1.25x}}. \tag{29}$$

The top plot of Fig. 7 indicates that through most of the evolution, Hamiltonian constraint violation in All3 remains about *100 times* lower than Control at 1.25x resolution. More strikingly, the implied convergence orders at higher resolutions are consistently higher in All3 (4.6–6.4) compared to Control (3.5–4.2), indicating that this gap only widens at higher resolutions.

All3 is even more effective at reducing momentum constraint violations in the strong field (bottom plot of Fig. 7), dropping them by about 3–4 orders of magnitude as compared to Control at 1.25x resolution. Unlike the implied convergence order from Hamiltonian constraint violations, the momentum constraint convergence order is nearly *doubled* at higher resolutions in All3 (5.9–7.1) compared to Control (3.0–3.2), reflecting that at higher resolutions, the 3–4 decade gap between Control and All3 also widens. Finally, 1–2 orders-of-magnitude bumps in $\sqrt{\|\mathcal{M}^2\|}$ appear periodically as reflections from distant AMR boundaries enter this region. These reflections strongly impact All3's convergence around $t/M = 200$ and 400, but not at other times.
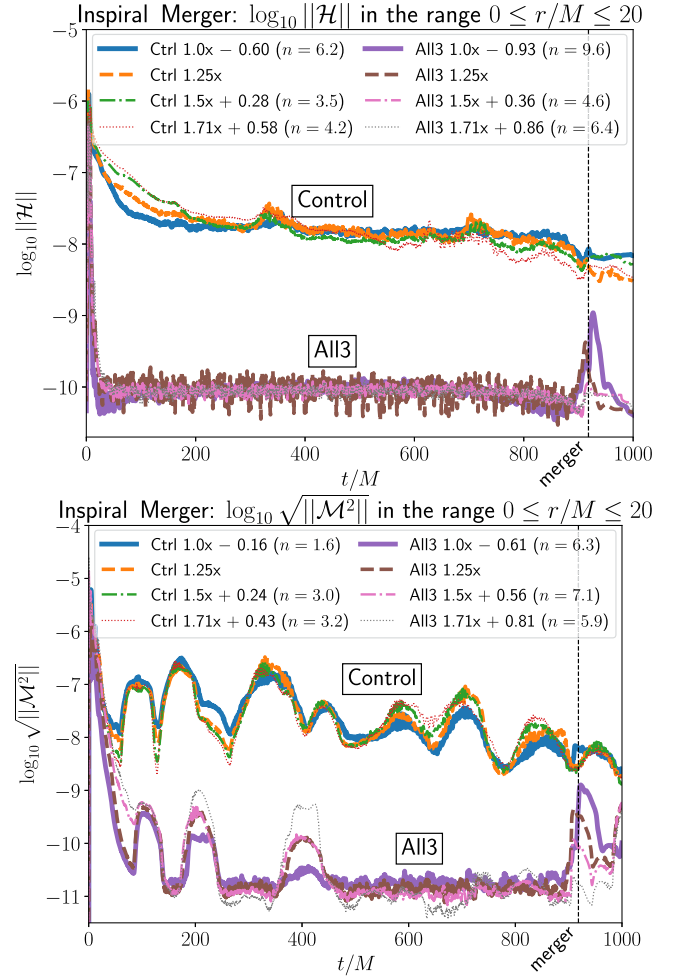

FIG. 7. Convergence of constraint violation in strong-field region: Convergence to zero of volume-averaged L_2 norms of Hamiltonian (top plot) and momentum (bottom plot) with increased resolution in a spherical region of radius $20M$, centered at the origin (center of mass), excluding two spheres of radius $1M$ centered at each BH. Control and All3 results are compared. The lowest resolution, as well as 1.5x and 1.71x higher resolution results have been shifted up or down to maximally overlap the 1.25x higher resolution results. Shifts and implied convergence orders (n) are reported in the legend for each case.

Analyzing implied convergence orders in Fig. 7, two results stand out: All3's Hamiltonian constraint violation exhibits $n = 9.6$ (1.0x vs 1.25x) and Control's momentum constraint violation exhibits $n = 1.6$ (1.0x vs 1.25x). These implied convergence orders lie outside the expected range for our mixed-order scheme of $n = 2$ to $n = 8$, indicating that our chosen default resolution may be outside the convergent regime.

Unusual implied convergence orders warrant further investigation. In Fig. 8, we examine Hamiltonian constraint violation in the region where the binary orbits, along the x axis at various times when the BHs are close to the y axis. Control is characterized by sharp, wavelike features about
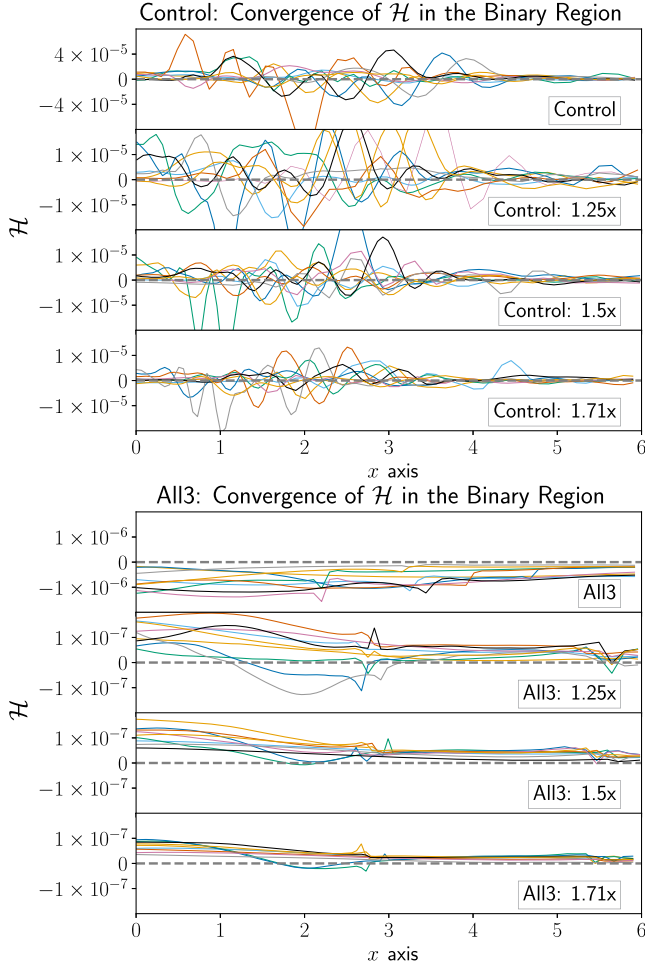
FIG. 8. Hamiltonian constraint violation in the binary region during inspiral: Convergence of $\mathcal{H}$ (thin solid curves) in Control with increased resolution (top plot) and All3 (bottom plot) along the positive x axis at times when the BHs are very close to the y axis, starting at approximately 0.75 orbits: $t/M \approx 168$, 272, 371, 464, 549, 628, 701, 765, and 828. A thick, gray, dashed line denotes $y = 0$. Note that both plots zoom out the top panel to better fit the data. The BHs are initially placed at $x = 4.46$ and $x = 5.54$, and the origin is the center of mass. Merger occurs at roughly $t/M = 930$.

$\mathcal{H} = 0$, while All3 exhibits a purely negative (at the lowest resolution) or mostly positive (at higher resolutions) systematic bias, albeit with constraint violation amplitudes roughly 1/100th that of Control (consistent with Fig. 7). For scale, All3 data would appear nearly indistinguishable from the horizontal dashed lines marking $\mathcal{H} = 0$ in the top plot.

While the amplitude of constraint violation indeed drops with increased resolution in both Control and All3, there is a clear zero crossing in All3 when comparing lowest and 1.25x higher resolution cases, consistent with the lowest resolution not yet being in the convergent regime. Further, the 1.25x resolution does not exhibit a clear bias in All3,
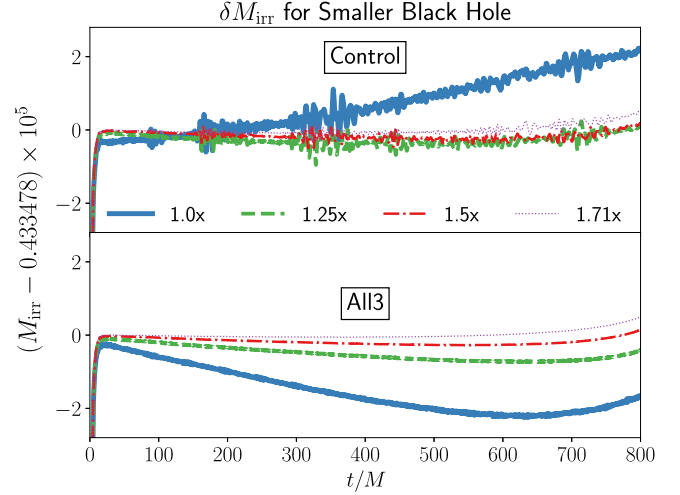

FIG. 9. Convergence behavior of M_{irr}, for smaller, faster-spinning BH: Comparison of irreducible mass for Control (top panel) and All3 (bottom panel) cases at all four resolutions. Thicker lines imply coarser resolution. Data beyond $t = 800M$ are removed, as this is close to the time of common horizon formation; merger occurs at $t \approx 910M$.

possibly implying that it lies very close to the boundary between the convergent and nonconvergent regimes.

As indicated by Eq. (15) and surrounding discussion, $\mathcal{H}$ violations act as an effective energy density ρ_{eff}; thus, the consistent $\rho_{\mathrm{eff}} < 0$ in 1.0x All3 implies that Hamiltonian constraint violation acts to continually push the BHs away from one another. If this were the dominant error, we would expect a delayed merger. Following the same reasoning, the consistent $\rho_{\mathrm{eff}} > 0$ in 1.5x and 1.71x resolutions of All3 imply an early merger, and the 1.25x resolution somewhere in between. As will be discussed in our analysis of the extracted waveforms (Sec. V C), this is exactly what we observe. Such a clean analysis is impossible in Control due to the extremely high level of noise, and implies that clean waveform convergence in Control might depend on near-exact cancellation of $\mathcal{H}$ waves.

We conclude that examining $\mathcal{H}$ in the binary region serves some use in identifying whether BBH evolutions are in the convergent regime, at least in the All3 cases. Regarding Control we found a low implied convergence order of $||\mathcal{M}^2||$ when comparing 1.0x and 1.25x resolutions. Further analysis of the irreducible mass (Fig. 9) of the less massive, faster-spinning BH lends additional insight.

Figure 9 shows that, after the early settling of M_{irr}, the irreducible mass remains constant to within a couple of parts in 10^5 across all cases. Comparing 1.0x with 1.25x and 1.5x resolutions, Control exhibits a *negative* implied convergence order beyond $t \approx 150M$, providing further evidence that the 1.0x resolution data in the strong field are not in the convergent regime. Measuring n directly for Control is quite challenging due to numerical noise, though given the stochastic behavior observed in $\mathcal{H}$ near the

punctures (Fig. 8), this is unsurprising. That said, comparing 1.25x, 1.5x, and 1.71x resolution data yields an implied convergence order of roughly $n = 1$ at $t = 400M$, but the 1.25x and 1.5x data completely overlap at $t \approx 600M$ (implying $n = 0$). Similar convergence behavior was observed in this quantity in [10].

Meanwhile, All3 exhibits far less noise and cleaner convergence in this diagnostic, with an implied convergence order of $n \approx 4.7$ in the range $200 < t/M < 600$ when analyzing the 1.0x, 1.25x, and 1.5x data, and $n \approx 2.0$ when analyzing the 1.25x, 1.5x, and 1.71x data. Although the All3 data have implied convergence orders consistent with the range expected from our mixed-order scheme ($2 \leq n \leq 8$), Fig. 8 has established that constraint violations in this region in the 1.0x All3 simulation are not in the convergent regime.

C. Weak-field region

While numerical noise in Control significantly influences the constraints and M_{irr} in the strong field, it is most prominent in the weak field region where GW content—typically the core payload of BBH simulations—is extracted. As our improvements act to mitigate this noise, they are particularly effective in the weak-field region. Figure 10 demonstrates that throughout much of the inspiral ($t \lesssim 1000M$), Hamiltonian constraint violations in the GW extraction region (the "wave zone") are reduced by roughly *4 to 6* orders of magnitude in All3, largely a consequence of SSL. This contrasts with the strong-field region (Fig. 6), where 2–3 orders of magnitude reductions are observed. Meanwhile, during the inspiral, All3 reduces momentum constraint violations by between 2–3 orders of magnitude, similar to the strong-field region.

At the merger, a spike in Hamiltonian and momentum constraint violations propagates from the strong-field region into the wave zone. In Control, this spike is observed only in momentum constraint violations, likely because noise levels in this region are far higher. Postmerger, Hamiltonian constraint violation reductions in All3 equilibrate to values roughly 3–4 orders of magnitude below Control, with near-equal contributions from SSL and CA$\mathcal{H}$D.

Careful analysis of momentum constraint violations in this region after merger shows that CA$\mathcal{H}$D and SSL overlap Control, making CAKO the dominant contributor to reducing momentum constraint violations. While All3 always falls below Control, they are quite comparable as constraint violations associated with common horizon formation propagate outward. Later, the gap between Control and All3 slowly widens to about 1 order of magnitude by the end of the simulation.

Turning now to constraint convergence behavior in the wave zone at higher resolutions, Fig. 11 plots the constraint violation at all four resolutions, using the 1.25x resolution as the basis of comparison. In Control, neither Hamiltonian nor momentum constraint violations converge to zero in the
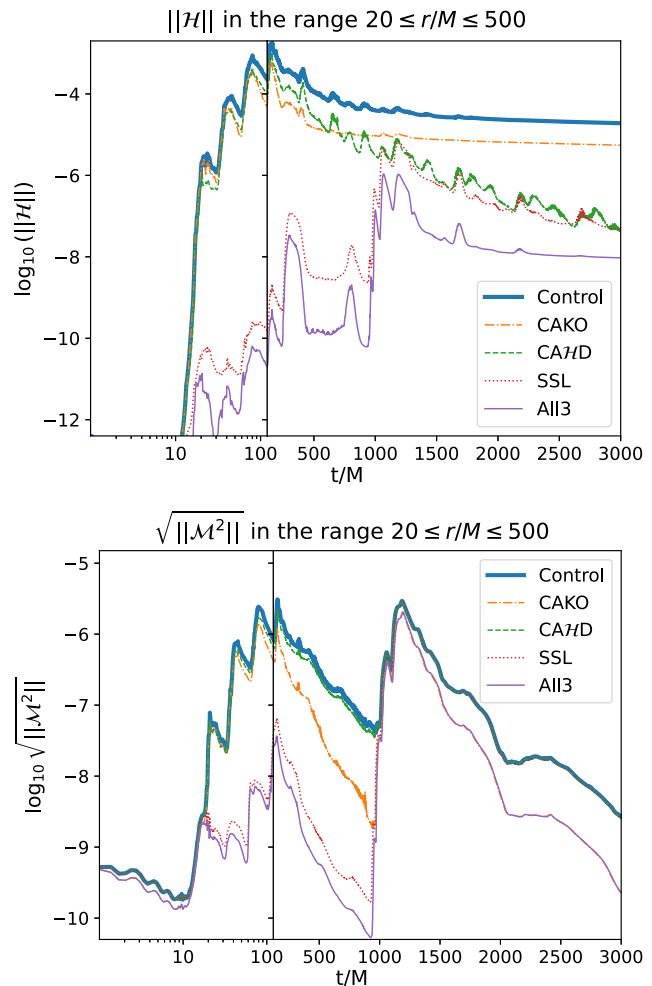

FIG. 10. Effect of improvements: Constraint violation in gravitational wave zone: Volume-averaged L_2 norm of $\|\mathcal{H}\|$ (left) and $\sqrt{\|\mathcal{M}^2\|}$ (right) within a thick spherical shell of $20 \leq r/M \leq 500$, centered at the origin (center of mass). Results from five runs are compared: Control (thick solid blue), CAKO (dash-dotted orange), CA$\mathcal{H}$D (dashed green), SSL (dotted red), and All3 (thin solid purple).

expected range of $2 \leq n \leq 8$, while, remarkably, All3 exhibits implied convergence orders of 3.5–3.8 in the Hamiltonian constraint violation. However, All3 falls outside the expected convergence-order range in momentum constraint violation, and in fact is slightly anticonvergent. While this is rather troubling for both Control and All3, upon closer analysis, we find that the level of violation in All3 remains lower than Control at all resolutions, just as in Fig. 7. We also find that if the integral is extended to the outer boundary as in [10], we find a consistent result: both Control and All3 exhibit convergence orders in the expected range $2 \leq n \leq 8$, implying convergent behavior outside of the wave zone.

Considering the GWs themselves, while GW detectors measure variation in GW strain, NR simulations typically compute the Weyl scalar ψ_4. As ψ_4 is the second time
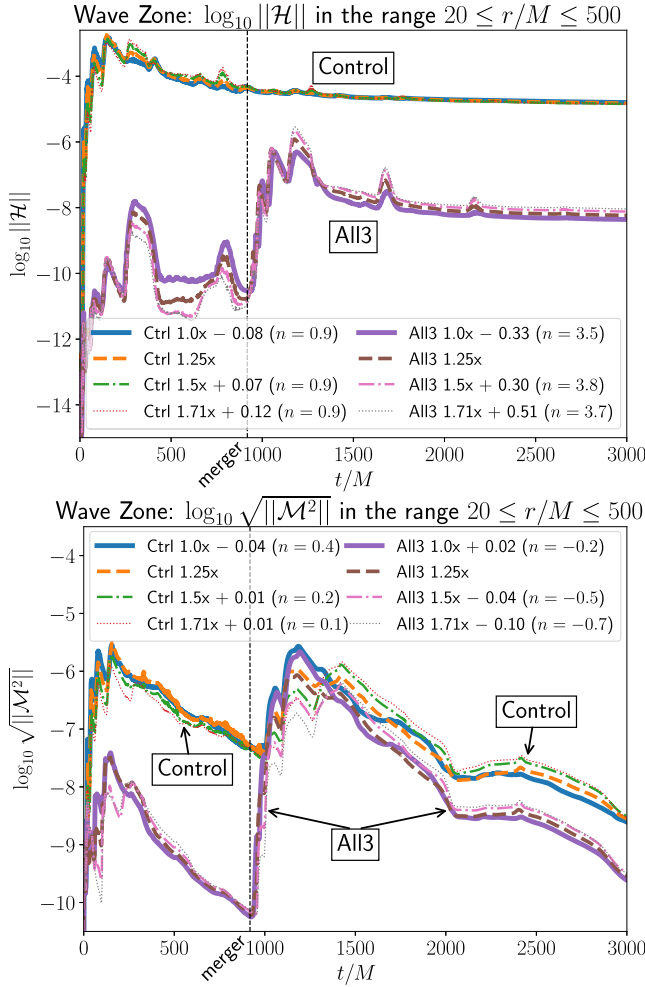
FIG. 11. Convergence of constraint violation in gravitational wave zone: Convergence to zero of volume-averaged L_2 norms of Hamiltonian (top plot) and momentum (bottom plot) with increased resolution in a thick spherical shell of inner radius $20M$ and outer radius $500M$, centered at the origin (center of mass). Lowest resolution as well as 1.5x and 1.71x higher resolution results have been shifted up or down to maximally overlap the 1.25x higher resolution results. Shifts and implied convergence orders (n) are reported for each case.

derivative of the strain in the weak-field limit—dependent on second spatial derivatives of the metric—it amplifies intrinsic numerical noise and thus provides an excellent noise diagnostic.

Figure 12 illustrates our procedure for quantifying noise in the dominant, $\ell = m = 2$, spin-weight $s = -2$ mode of ψ_4. In the top panel, we plot the amplitude of this mode extracted from the simulation at a radius $R_{\text{ext}} = 100M$, for both Control and All3 at 1.25x higher resolution than the lowest resolution. The vertical dashed lines denote the region where a quartic polynomial was fit through the amplitude data. The middle panel plots the absolute difference between the data and the quartic polynomial fit (i.e., the "noise"), and the bottom panel the noise integrated
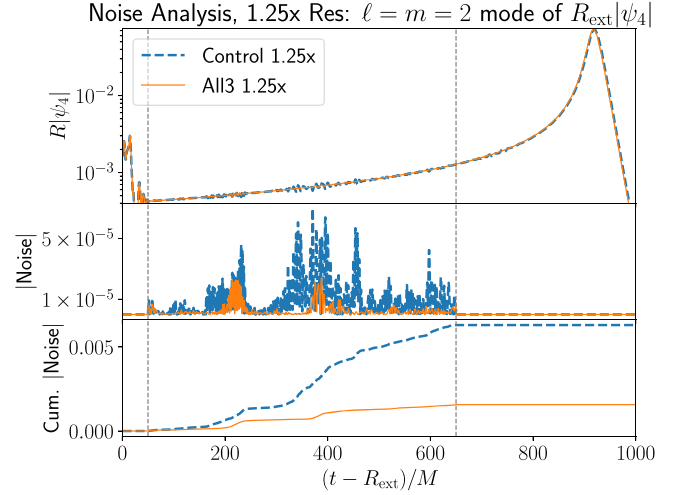

FIG. 12. GW noise analysis: dominant mode. Top: amplitude of dominant ($\ell = m = 2$) spin-weight $s = -2$ mode of ψ_4 versus retarded time. Vertical dashed lines: region where a quartic polynomial is fit to amplitude data. Middle: absolute difference between data and fit. Bottom: cumulative absolute difference between data and fit.

over time. Averaged over the time interval of the fit, our combined improvements reduce numerical noise in the $\ell = m = 2$ mode of ψ_4 by a factor of 4.0 at 1.25x resolution. Applying the same technique to other resolutions, we find noise reduction factors of 6.0 (1.0x), 4.3 (1.25x), and 3.1 (1.71x). This underscores the fact that this noise converges away extremely slowly at higher resolutions.

Such significant noise reductions in the dominant ψ_4 mode are quite important, as GWs are usually the core payload of NR BBH simulations. Looking ahead, the improved sensitivities of current and future GW observatories will require significantly improved accuracy of NR GW predictions [1,2], including higher-order modes. In anticipation of this, Fig. 13 shows All3's impact on reducing noise in subdominant, even $\ell = m$ modes at the highest (1.71x) resolution. The noise at $\ell = m = 4$ is greater in both All3 and Control than at $\ell = m = 2$, yet noise is still significantly reduced in All3. Turning to the $\ell = m = 6$ subdominant mode, we find a truly exciting result. This mode is completely obscured by noise through the entire inspiral in Control, yet is mostly visible in All3, enabling the strain prediction to be extracted.

Still, there are limits to All3's effectiveness: the $\ell = m = 8$ mode cannot be extracted from either All3 or Control. This underscores the need for further research into improved techniques and numerical grid structures such as those that Llama or Blackholes@Home provide.

We next turn our attention to the convergence behavior of ψ_4, splitting our analysis into the phase and amplitude of the dominant $\ell = m = 2$ mode. The phase in Control (top two panels of Fig. 14) is too noisy to measure n until a retarded time of about $780M$, when monotonic
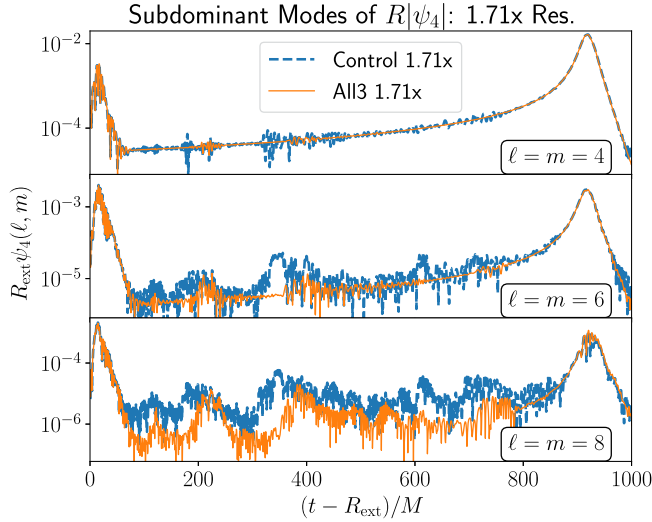
FIG. 13. Subdominant modes: Amplitude of spin-weight $s = -2$ subdominant even $\ell = m$ modes of ψ_4 up to 8, versus retarded time, at highest resolution. Top: $\ell = m = 4$ mode; middle: $\ell = m = 6$ mode; bottom: $\ell = m = 8$ mode.
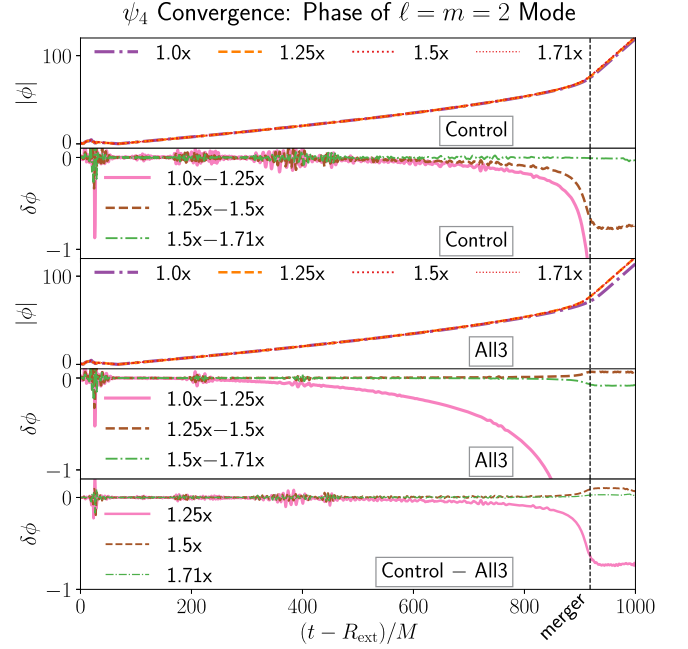

FIG. 14. GW phase convergence analysis: $\ell = m = 2$. Top: accumulated phase $|\phi|$ of dominant ($\ell = m = 2$) mode of ψ_4 at all four resolutions in Control. Second from top: accumulated phase difference across neighboring resolutions in Control. Third from top: same as top, but for All3. Fourth from top: same as second panel, but for All3. Bottom: accumulated phase difference between Control and All3 at the three highest resolutions. All data were extracted from the simulation at extraction radius $R_{\text{ext}} = 100M$.

convergence behavior starts to emerge. However, although the implied convergence order of 2.3 between 1.0x, 1.25x, and 1.5x resolutions falls within the expected range of $2 \leq n \leq 8$, when comparing differences in 1.25x, 1.5x, and 1.71x resolutions, n is roughly 12 before merger and 23 after merger, reflecting serious problems with convergence. We attribute this unusual behavior to the high-amplitude, stochastic oscillations in $\mathcal{H}$ between the BHs illustrated in Fig. 8. As discussed in Sec. II C, these oscillations equate to spurious positive and negative effective energy densities, which, when they do not cancel, will impose an effective attraction or repulsion between the BHs, respectively. We fear that noncancellations may directly impact the GW phase in potentially unpredictable ways.

Returning to Fig. 8, $\mathcal{H}$ is roughly 100x lower between the BHs in All3 as compared to Control. In spite of this, the persistent negative effective energy density at 1.0x resolution likely causes a significantly delayed merger in All3 (middle panel, Fig. 14). While this may seem troubling for All3, it is even more troubling for Control, as it implies that the stochastic behavior in $\mathcal{H}$ between the BHs would need to cancel to better than one part in 100 over the inspiral in order to exhibit similar or lower phase errors. Given the small number of waves observed in the binary region, such cancellation cannot be guaranteed, and is likely the cause of the inconsistent convergence in Control.

Considering All3 at 1.25x and higher resolutions, Fig. 8 indicates that All3 exhibits slight *negative* oscillation about $\mathcal{H} = 0$ in 1.25x, but a *consistently positive* bias at 1.5x and 1.71x. Given inconsistent convergence observed in Control due to oscillations, as well as the clear negative bias of $\mathcal{H}$ in 1.0x All3 results, caution is clearly warranted when considering All3 1.25x ψ_4 phase results. The fourth panel

from the top of Fig. 14 indicates that while the difference between 1.25x and 1.5x phase results is smaller in All3 than Control (second panel from the top), the difference flips sign close to merger, as compared to all other resolution pairs (All3 or Control).

This likely implies that 1.25x resolution in All3 and Control are slightly outside the convergent regime. While this analysis of $\mathcal{H}$ in the binary region (Fig. 8) to explain nonconvergent behavior has been useful, it remains to be seen whether it will be predictive of convergence in other BBH scenarios.

Next, we assess whether Control and All3 waveforms converge to the same result. To this end, in the bottom panel of Fig. 14, we fix the resolution and simply plot the phase differences between Control and All3. Notice that phase differences between Control and All3 drop to tiny fractions of a radian with increasing resolution, indeed consistent with them converging to the same result.

For completeness, we present the same analysis for the rescaled amplitude of the extracted GW signal: $R_{\text{ext}}|\psi_4|$, for the dominant mode in Fig. 15. Perhaps unsurprisingly, the conclusions drawn from our phase analysis hold for the amplitude as well, though the significantly delayed merger in the 1.0x resolution All3 case is much more evident when considering the amplitude.
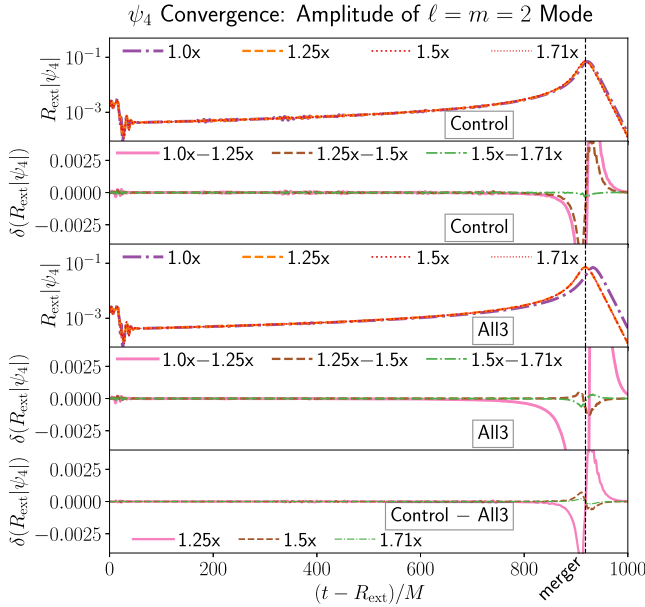
FIG. 15. GW amplitude convergence analysis. Same as Fig. 14, except instead of phase, we consider the rescaled amplitude $R_{ext}|\psi_4|$ of the dominant ($\ell = m = 2$) mode of ψ_4.

We conclude that improvements do not appear to negatively impact the dominant mode of the GW prediction; both amplitude and phase appear to converge to the same values as Control. That said, 1.25x resolution All3 amplitude and phase results are much closer to the converged solution than in 1.25x Control, implying smaller errors at resolutions bordering the nonconvergent regime. Perhaps most importantly, we find that All3's enormous reduction in Hamiltonian constraint noise in the binary region and the associated purely positive or negative biases, may enable us to predict for other BBH scenarios whether the actual merger time should be earlier or later than the given resolution. If so, Hamiltonian constraint information in the binary region, combined with All3, may be used to improve GW predictions from moving-puncture BBH merger simulations in the future.

VI. CONCLUSIONS AND FUTURE WORK

We have introduced three improvements to the moving-puncture technique, aimed at improving the accuracy of BBH evolutions. Our results show that the proposed improvements effectively reduce numerical errors and noise in different parts of the computational domain, from the strong-field area near the black holes to the weak-field region where GW predictions are extracted.

When combining all three improvements (All3), performance is significantly better than any single enhancement, cutting down Hamiltonian (momentum) constraint violations in the strong-field area by roughly 2 (3) orders of magnitude

and in the GW extraction zone by 5 (2) orders of magnitude. Further, the convergence of constraint violations improves near and inside the convergent regime, suggesting that these improvements become even more pronounced at higher resolutions.

The improvements also slightly reduce coordinate eccentricity and lower numerical noise in the dominant $\ell = m = 2$ mode of ψ_4 by an average factor of 4.3 across different resolutions. Additionally, the fidelity of higher-order modes increases significantly, making the $\ell = m = 6$ mode, which was previously hidden by noise, detectable. This is particularly encouraging given the demands of next-generation GW observatories.

Looking forward, we plan to explore the impacts of these improvements on BNS simulations and compare their efficacy to Z4-based formalisms. However, there are limitations to the wider applicability of these improvements: As the development of CA$\mathcal{H}$D was built upon characteristic analysis of BSSN, it is unlikely to be useful in Z4-based formalisms, which exhibit a different characteristic structure and include damping terms already. Further, the initial gauge pulse is likely significantly less in BNS evolutions, meaning that SSL may not be particularly helpful in this case.

Moving beyond Cartesian AMR-based grids, it will be interesting to learn the impacts of these improvements on mixed AMR/cubed-spheres grid structures like Llama [45], as they may act to eliminate reflections and have already proven quite effective in Blackholes@Home. Meanwhile, work will continue to improve upon the moving-puncture technique further, with particular emphasis on addressing spikes in constraint violation associated with merger.

[1] M. Pürrer and C.-J. Haster, Phys. Rev. Res. **2,** 023151 (2020).

[2] D. Ferguson, K. Jani, P. Laguna, and D. Shoemaker, Phys. Rev. D **104,** 044037 (2021).

[3] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, Phys. Rev. Lett. **96,** 111101 (2006).

[4] J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, Phys. Rev. Lett. **96,** 111102 (2006).

[5] T. W. Baumgarte and S. L. Shapiro, Phys. Rev. D **59,** 024007 (1998).

[6] M. Shibata and T. Nakamura, Phys. Rev. D **52,** 5428 (1995).

[7] D. Alic, C. Bona-Casas, C. Bona, L. Rezzolla, and C. Palenzuela, Phys. Rev. D **85,** 064040 (2012).

[8] S. Bernuzzi and D. Hilditch, Phys. Rev. D **81,** 084003 (2010).

[9] M. Alcubierre, B. Bruegmann, P. Diener, M. Koppitz, D. Pollney, E. Seidel, and R. Takahashi, Phys. Rev. D **67,** 084023 (2003).

[10] Z. B. Etienne, J. G. Baker, V. Paschalidis, B. J. Kelly, and S. L. Shapiro, Phys. Rev. D **90,** 064032 (2014).

[11] J. R. van Meter, J. G. Baker, M. Koppitz, and D.-I. Choi, Phys. Rev. D **73,** 124011 (2006).

[12] In virtually all AMR implementations, and even for those that offer refinement factors other than two, two is generally used without exception.

[13] P. Amaro-Seoane *et al.* (LISA Collaboration), arXiv:1702.00786.

[14] T. Assumpcao, L. R. Werneck, T. P. Jacques, and Z. B. Etienne, Phys. Rev. D **105,** 104037 (2022).

[15] D. Hilditch, S. Bernuzzi, M. Thierfelder, Z. Cao, W. Tichy, and B. Bruegmann, Phys. Rev. D **88,** 084057 (2013).

[16] J. D. Brown, Phys. Rev. D **79,** 104029 (2009).

[17] T. W. Baumgarte, P. J. Montero, I. Cordero-Carrion, and E. Müller, Phys. Rev. D **87,** 044026:1 (2013).

[18] I. Ruchlin, Z. B. Etienne, and T. W. Baumgarte, Phys. Rev. D **97,** 064036 (2018).

[19] W. Tichy and P. Marronetti, Phys. Rev. D **76,** 061502 (2007).

[20] P. Marronetti, W. Tichy, B. Bruegmann, J. Gonzalez, and U. Sperhake, Phys. Rev. D **77,** 064010 (2008).

[21] M. D. Duez, P. Marronetti, S. L. Shapiro, and T. W. Baumgarte, Phys. Rev. D **67,** 024004 (2003).

[22] F. Löffler, J. Faber, E. Bentivegna, T. Bode, P. Diener, R. Haas, I. Hinder, B. C. Mundim, C. D. Ott, E. Schnetter, G. Allen, M. Campanelli, and P. Laguna, Classical Quantum Gravity **29,** 115001 (2011).

[23] Einstein TOOLKIT Consortium Homepage (2024), http://einsteintoolkit.org/.

[24] E. Schnetter, S. H. Hawley, and I. Hawke, Classical Quantum Gravity **21,** 1465 (2004).

[25] Carpet: Adaptive mesh refinement for the Cactus framework, http://www.carpetcode.org/ (2024).

[26] NRPy webpage, https://github.com/nrpy/nrpy (2024).

[27] Z. Etienne, Updated BaikalVacuum with all improvements implemented, available at https://github.com/zachetienne/baikalvacimproved (2024).

[28] H.-O. Kreiss and J. Oliger, *Methods for the Approximate Solution of Time Dependent Problems*, GARP Publications Series No. 10 (World Meteorological Organization and International Council of Scientific Unions, Geneva, 1973).

[29] G. Bozzola and V. Paschalidis, Phys. Rev. D **104,** 044004 (2021).

[30] Y. T. Liu, Z. B. Etienne, and S. L. Shapiro, Phys. Rev. D **80,** 121503 (2009).

[31] G. Yoneda and H.-a. Shinkai, Phys. Rev. D **66,** 124003 (2002).

[32] C. A. Raithel and V. Paschalidis, Phys. Rev. D **106,** 023015 (2022).

[33] J. D. Brown, Phys. Rev. D **80,** 084042 (2009).

[34] B. Wardell, I. Hinder, and E. Bentivegna, Simulation of GW150914 binary black hole merger using the Einstein TOOLKIT, 10.5281/zenodo.155394 (2016).

[35] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116,** 061102 (2016).

[36] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116,** 241102 (2016).

[37] M. Ansorg, B. Brügmann, and W. Tichy, Phys. Rev. D **70,** 064011 (2004).

[38] J. Thornburg, Classical Quantum Gravity **21,** 743 (2004).

[39] O. Dreyer, B. Krishnan, D. Shoemaker, and E. Schnetter, Phys. Rev. D **67,** 024018 (2003).

[40] S. Husa, I. Hinder, and C. Lechner, Comput. Phys. Commun. **174,** 983 (2006).

[41] J. D. Brown, P. Diener, O. Sarbach, E. Schnetter, and M. Tiglio, Phys. Rev. D **79,** 044023 (2009).

[42] McLachlan, McLachlan, a public BSSN code, http://www.cct.lsu.edu/~eschnett/McLachlan/ (2008).

[43] F. G. L. Armengol *et al.*, Phys. Rev. D **106,** 083015 (2022).

[44] L. R. Werneck *et al.*, Phys. Rev. D **107,** 044037 (2023).

[45] D. Pollney, C. Reisswig, E. Schnetter, N. Dorband, and P. Diener, Phys. Rev. D **83,** 044045 (2011).

[46] AMR grid structures and resolutions were chosen such that each resolution maintains consistent physical boundaries for AMR grids in our moving-box AMR algorithm.

[47] Idaho C3 + 3 Collaboration, Falcon: High performance supercomputer, 10.7923/falcon.id (2022).