A GPU-accelerated mixed-precision WENO method for extremal black hole and gravitational wave physics computations

Scott E. Field · Sigal Gottlieb · Zachary J. Grant · Leah F. Isherwood · Gaurav Khanna

1 Introduction

The field of computational relativity is undergoing a transformative renaissance of sorts, due to multiple new discoveries such as the direct detection of gravitational waves from a black hole binary system by LIGO in 2015 [1], which was awarded a Nobel Prize in 2017, the first-ever image of a black hole event horizon by the Event Horizon Telescope in 2019 [2], and others. Accurate computational models are critical for the success of such efforts since high-fidelity models enable improved...
data analysis efforts required for parameter estimation, tests of general relativity, and studying the formation channels of compact binaries [3, 4, 5].

Black hole perturbation theory is the standard framework for studying gravitational phenomena, such as gravitational waves and linear stability of black hole solutions. The theory of such perturbations is well developed, and starts with pioneering investigations by Regge and Wheeler [6] for Schwarzschild (nonspinning) black holes. The theory was later extended by Teukolsky to handle perturbations of the Kerr metric [7], which is an exact solution of the Einstein equation. The Kerr solution is parameterized by the black hole mass, \( M \), and the hole’s spin, \( a \), and small perturbations about this solution obey the Teukolsky master equation Eqn. (1).

For many realistic problems, we require highly-accurate numerical solutions to the Teukolsky equation over extremely long simulation times. Yet most state-of-the-art solvers are based on a finite-difference numerical evolution scheme stabilized with numerical dissipation [8, 9, 10, 11], which, as we will show, can significantly degrade the quality of the numerical solution.

In this paper, we develop and use a novel mixed-precision weighted essentially non-oscillatory (WENO) method for solving the Teukolsky equation. Our work is one of the first applications of WENO methods to the modeling of gravitational waves, and the first WENO-based Teukolsky solver. To showcase the improvements of this new solver, we compare a standard sixth-order centered differencing approach (with an eighth-order Kreiss-Oliger dissipation operator [12] for numerical stability) and third-order and fifth-order weighted essentially non-oscillatory (WENO) schemes. We show that the fifth order WENO of Jiang and Shu [13] performs well in the sense of stability and preservation of accuracy over long evolutions.

While our new numerical solver can be used to study a variety of interesting questions in gravitational physics, in this paper, we focus our attention on a specific computational problem in the area of black hole physics. The physical scenario one may imagine is that of an isolated rotating, Kerr black hole, spinning at the maximum possible rate, i.e. an extremal Kerr black hole. If such a black hole is perturbed because of any nearby material, or scalar or gravitational wave, analytic results suggest that the hole will undergo a rather unexpected evolution that would result in a unique signal which, if measured, would be a “signature” of an extremal black hole [14, 15]. In particular, Aretakis derived a mathematical quantity, often called “charge” or “horizon hair”, that is conserved (stays constant in time) in such a system and this quantity is measurable from a far away distance. Black holes that are not extremally spinning are unable to preserve the constancy of such a quantity [15].

Aretakis had derived those key results in the context of scalar waves, that are simpler to work with. However, the key ideas of Aretakis have been extended to the astrophysically realistic context of gravitational waves using a computational approach by Khanna and collaborators [17]. In this paper, we focus entirely on the realistic gravitational waves scenario and discuss the computational challenges involved with the relevant simulations and how to solve them. The main challenge that is involved is the accuracy of computed numerical solution and being able maintain that accuracy in long evolutions. As a simple example, one may compute the Aretakis conserved charge numerically and study how long the numerical simulation is able to maintain its constancy (within some tolerance). That is a
key test that we use in this work. There are also additional similar tests based on
the expected temporal behavior of the spatial derivatives of the solution that we
conduct in this work. The numerical results from our WENO simulations retain
the predicted behavior of the Aretakis conserved charge in fidelity with rigorous
mathematical results. The simulations are computationally costly, and take many
months of computational time. To alleviate this problem, we implement a GPU-
accelerated version of this code on a GPU-cluster, IBM/MIT’s Satori a 64-node
GPGPU system with 256 Nvidia V100 GPUs. Furthermore, we introduce and use
a mixed precision WENO algorithm which speeds up the computation by a factor
of between 3 – 4 without meaningfully reducing its accuracy.

2 Numerical Solution of the Teukolsky Equation

In this section we briefly describe the Teukolsky equation 1 and the context
in which it arises. We provide a description of the coordinate-systems used, the
relevant evolution equations, and a fully first-order reduction of the Teukolsky
equation 8 which is subsequently discretized using the WENO method presented
in Sec. 3.

The Kerr metric, which describes a rotating black hole, is both an astrophysi-
cally and theoretically important exact solution of the Einstein equation. Studying
small perturbations of the Kerr metric is an especially important line of inquiry
and can be used, for example, to understand the behavior of gravitational waves
in the spacetime of a rotating black hole. The coordinate system that is typi-
cally used to describe the spacetime of black holes is the Boyer-Lindquist system,
\((t, r, \theta, \phi)\), that has many similarities with spherical coordinates. However, because
they suffer from a coordinate-singularity at the horizon locations, Boyer-Lindquist
coordinates are not well for numerical computations. We instead make use of a
better suited coordinate system known as ingoing Kerr coordinates. These are a
Kerr spacetime generalization of the better-known Eddington-Finkelstein coordi-
nates that are able to smoothly “penetrate” the horizon of a black hole. In the
following subsection, we review the relationship between these different coordinate
systems and emphasize some of their important aspects.

The main evolution equation of interest in this work is the Teukolsky master
equation that describes scalar, vector and tensor field perturbations in the space-
time of a Kerr black hole 7 to linear order. We numerically solve this equation
for the gravitational waves case using a compactified form of the ingoing Kerr
coordinates. Using hyperboloidal compactification allows us to directly sample the
behavior of fields throughout the spacetime, including even (null) infinity. One
important aspect of this work is that we must evolve the fields for a long duration
because we are interested in the late-time, power-law decay behavior of these
fields. This behavior typically appears after the quasi-normal modes of the system
have exponentially decayed enough to become subdominant. This poses certain
challenges that are explained in some detail in the following sections.

The following subsections offer additional details including the main expres-
sions for the quantities involved and also our computational methodology.
2.1 Teukolsky Equation

The Teukolsky master equation describes scalar, vector and tensor field perturbations in the space-time of Kerr black holes [7]. In Boyer-Lindquist coordinates, this equation takes the form

\[
\begin{align*}
- \left[ \frac{(r^2 + a^2)^2}{\Delta} - a^2 \sin^2 \theta \right] \partial_t \Psi & - \frac{4Ma}{\Delta} \partial_\phi \Psi \\
- 2s \left[ r - \frac{M(r^2 - a^2)}{\Delta} + ia \cos \theta \right] \partial_\theta \Psi & + \Delta^{-s} \partial_r \left( \Delta^{s+1} \partial_r \Psi \right) + \frac{1}{\sin \theta} \partial_\theta (\sin \theta \partial_\theta \Psi) + \\
\left[ \frac{1}{\sin^2 \theta} - \frac{a^2}{\Delta} \right] \partial_{\phi \phi} \Psi & + 2s \left[ \frac{a(r - M)}{\Delta} + \frac{i \cos \theta}{\sin^2 \theta} \right] \partial_\phi \Psi
\end{align*}
\]

where \( M \) is the mass of the black hole, \( a \) its angular momentum per unit mass, \( \Delta = r^2 - 2Mr + a^2 \) and \( s \) is the “spin weight” of the field. The \( s = \pm 2 \) versions of these equations describe the radiative degrees of freedom of the gravitational field, and thus are the equations of interest here. As mentioned previously, this equation is an example of linear, hyperbolic, homogeneous PDEs which are quite common in several areas of science and engineering. Finite-difference methods are by far the most common numerical schemes developed for the Teukolsky Equation.

2.2 Teukolsky Equation in Ingoing-Kerr Coordinates

We begin with an expression of the usual Boyer-Lindquist coordinate version of the Kerr spacetime metric and the associated Teukolsky equation [7]. It is clear that the equation exhibits pathological behavior at the horizon locations, i.e. when \( \Delta = 0 \). Note that this coordinate singularity can be easily removed by a suitable change of coordinates.

To remove the coordinate singularity at the horizon locations determined by \( \Delta = 0 \), we consider the above equations in the ingoing Kerr coordinate system \( (\tilde{t}, r, \theta, \tilde{\phi}) \), also called “horizon penetrating” coordinates. These coordinates are related to the Boyer-Lindquist coordinates through the transformations \( \tilde{\phi} = \phi + \int \frac{a}{\Delta} \, dr \) and \( \tilde{t} = t - r + r_* \), where the “tortoise” radial coordinate \( r_* = \int (r^2 + a^2)^{1/2} \, dr \). This system does not suffer from any pathologies at the horizon locations and is therefore well-suited for analyzing fields both in the exterior and interior spacetimes of a rotating black hole. It is also useful to note that at the horizon location at late times, the \( \tilde{t} \) variable is essentially the null variable \( \psi = t + r_* \).

The last ingredient that goes into the setup of our coordinate system is hyperboloidal compactification as developed by Zenginoğlu [18,19,20,21,22,23,24,25,26]. We define a compactified coordinate system \( (\tau, \rho, \theta, \tilde{\phi}) \) by

\[
\tau = \tilde{t} - r^2/(r + S) + 4 \ln[S/(r + S)]
\]
and

$$\rho = r/[1 + r/S]$$  \hspace{1cm} (3)

where a free parameter $S$ controls both the domain and also the foliation. Note that $\rho \in [0, S]$ maps $r \in [0, \infty)$ and is therefore a one-to-one compactifying coordinate. A Penrose diagram of the slices defined by these coordinates in the Kerr spacetime context can be found in Ref. [27]. We do not show the final form of the Teukolsky master equation in these compactified coordinates because of the lengthy nature of the expression and the fact that it is not particularly illuminating. We simply refer to it in the symbolic form,

$$\begin{align*}
A^{\tau\tau}\partial_{\tau}^2\Psi + A^{\tau\rho}\partial_\tau\partial_\rho\Psi + A^{\rho\rho}\partial_\rho^2\Psi + A^{\theta\theta}\partial_\theta^2\Psi + B^{\tau}\partial_\tau\Psi + B^{\rho}\partial_\rho\Psi + B^{\theta}\partial_\theta\Psi + C\Psi &= 0 \hspace{1cm} (4),
\end{align*}$$

and refer the reader to Refs. [28,29] wherein additional details may be found.

The computational grid is defined as a uniform grid over the compactified $\rho$ coordinate. As pointed out earlier, this allows us to access null infinity directly on the computational grid ($\rho = S$ maps to infinity). Moreover, the compactification offers a solution to the “outer boundary problem” in numerical relativity. Typical boundary conditions used in the research community lead to spurious wave reflections from the edge of the computational grid. However, with the approach of hyperboloidal compactification, one is able to extend the computational domain to infinity, making it possible to completely eliminate any such reflections [28].

In addition, the compactification allows us to employ a very dense computational grid (typically, $S \sim 20$) which results in highly accurate numerical results. Those details are provided in the next subsection.

2.3 Additional Implementation Remarks

After performing the transformations presented in the previous sections, we rewrite the vacuum equation in the form

$$\begin{align*}
\partial_{\tau}^2\Psi &= \tilde{A}^{\tau\rho}\partial_\tau\partial_\rho\Psi + \tilde{A}^{\rho\rho}\partial_\rho^2\Psi + \tilde{A}^{\theta\theta}\partial_\theta^2\Psi \\
&\quad + \tilde{B}^{\tau}\partial_\tau\Psi + \tilde{B}^{\rho}\partial_\rho\Psi + \tilde{B}^{\theta}\partial_\theta\Psi + \tilde{C}\Psi, \hspace{1cm} (5)
\end{align*}$$

where the coefficients with a tilde are obtained by dividing the coefficients of Teukolsky equation (4) by $-A^{\tau\tau}$. We put the equation (5) in first-order form (in $\rho$ and $\tau$) by defining two new field variables

$$\begin{align*}
\pi &\equiv \partial_\tau\Psi + b\partial_\rho\Psi, \hspace{1cm} (6) \\
b &\equiv -(\tilde{A}^{\tau\rho} + \sqrt{(\tilde{A}^{\tau\rho})^2 + 4\tilde{A}^{\rho\rho}})/2. \hspace{1cm} (7)
\end{align*}$$

We chose these auxiliary variables because it has been discovered through extensive experimentation that the resulting first-order form,

$$\begin{align*}
\partial_\tau u + M\partial_\rho u + Lu + Au &= 0, \hspace{1cm} (8)
\end{align*}$$

Note that a different approach towards compactification was used therein – a hyperboloidal compactified layer was attached to the outer part of a Boyer-Lindquist coordinates based computational grid.
is ideally suited for long stable evolutions. Here

\[ \mathbf{u} \equiv \{ \Psi_R, \Psi_I, \pi_R, \pi_I \} \]  

(9)
is the solution vector and the subscripts \( R \) and \( I \) refer to the real and imaginary parts respectively. The matrices \( M, A \) and \( L \) are obtained from (5) as in [8,9]. Here it will suffice to simply indicate the final form taken by these matrices:

\[
M \equiv \begin{pmatrix}
    b & 0 & 0 & 0 \\
    0 & b & 0 & 0 \\
    m_{32} & m_{32} & -b & 0 \\
    -m_{32} & m_{32} & 0 & -b \\
\end{pmatrix}, 
\]

(10)

\[
A \equiv \begin{pmatrix}
    0 & 0 & -1 & 0 \\
    0 & 0 & 0 & -1 \\
    a_{31} & a_{32} & a_{33} & a_{34} \\
    -a_{32} & a_{31} & -a_{34} & a_{33} \\
\end{pmatrix}, 
\]

(11)

and

\[
L \equiv \begin{pmatrix}
    0 & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 \\
    l_{31} & 0 & 0 & 0 \\
    0 & l_{31} & 0 & 0 \\
\end{pmatrix}. 
\]

(12)
The angular derivatives are encoded in \( L \).

The main advantage of casting the equation (5) into the form (8) is that the system has advantageous properties in the variable \( \rho \). The matrix \( M \) has a complete set of linearly independent eigenvectors with real eigenvalues. This is not a rigorous statement on the hyperbolicity of the system because the matrix \( L \) contains second-order angular derivatives. Nevertheless, experiments show that the system is numerically well-behaved.

### 3 Computational Approach

The numerical approach used to solve the first-order Teukolsky equation (8) in the compactified ingoing Kerr coordinate system is very similar to the one presented in our earlier work [28]. We simply outline the main steps here and refer the reader to that reference for additional details. We begin by taking advantage of Kerr spacetime’s axisymmetry and separating out the \( \tilde{\varphi} \) dependence of the system using an \( \exp(\rho i \tilde{\varphi}) \) form for the gravitational field \( \Psi \). This transforms the original \((3+1)\)D equation into a system of \((2+1)\)D equations. In this work we restrict ourselves to axisymmetric fields only, and therefore we set \( m = 0 \) throughout. Next, we cast the equations into a first-order hyperbolic partial differential equation form, by defining a new “momentum” field that is related to the derivative of the field \( \Psi \).

In the next subsections we will describe the numerical methods used to discretize these equations. We implement a fifth-order WENO finite-difference and a third-order WENO finite-difference scheme, and compare it with a sixth-order finite-difference scheme with numerical dissipation required for stability. The time-stepping method used is the third-order Shu-Osher explicit Runge–Kutta scheme.

It is worth commenting on the fact that numerical computations are rather challenging in the context of studying the late-time tails. As remarked before, these
computations must be long duration because the observed field initially exhibits an exponentially decaying oscillatory behavior known as quasi-normal ringing. Only much later, once the exponential decay has made these modes subdominant, does the field transition to a power-law tail. Moreover, there are often intermediate tails [26], that do not necessarily have the true late-time asymptotic rates that we are interested in here. These intermediate tails decay faster than the asymptotic rate, but may have dominant amplitudes for a period of time. We must evolve longer than these intermediate tails last in order to obtain the tail solution with the true asymptotic decay rate.

In addition, each of the field’s spherical harmonic multipoles \( Y_{\ell m} \) has its own decay rate (that is proportional to \( \ell \)). Thus, at late times we obtain numerical data in which different multipoles may have widely ranging amplitudes (typically 15 – 20 orders of magnitude apart!). It is thus important for the numerical scheme to have high grid density in order to reduce the truncation errors to very low levels. In addition, due to the very large range of amplitudes involved, these computations also require high-precision floating-point numerics that allow us to reduce round-off error that can otherwise easily overwhelm the fast decaying multipoles. In particular, we satisfy this requirement by using quad-precision numerics (128-bit or \( \sim 30 \) decimal digits). This keeps the round-off error in our computations at acceptably low levels.

Finally, to complete these long duration, high-accuracy and high-precision computations in a reasonable time-frame we make extensive use GPGPU-based parallel computing. For additional details on implementation of such intensive computations on a parallel GPU architecture, we refer the reader to our earlier work on the subject [30].

3.1 Spatial Discretization

We compare two approaches in this work: a sixth-order finite difference method that is stabilized by an eighth-order Kreiss-Oliger dissipation operator [12] and the weighed essentially non-oscillatory methods of Jiang and Shu [13]. Note that the WENO discretization is only applied in the radial (\( \rho \)) direction, not the angular (\( \theta \)) direction. This is because the solution is expected to have a very smooth profile in the \( \theta \) direction. In that direction, standard centered-differencing is used for all methods under consideration in this work.

3.1.1 Sixth Order Finite Difference Method

Standard sixth-order centered finite-difference stencils and an eighth-order Kreiss-Oliger dissipation operator are explicitly included below. The sixth-order derivative stencil we use takes this form

\[
\begin{align*}
    u'(\rho) = \frac{u_{j+3} - 9u_{j+2} + 45u_{j+1} - 45u_{j-1} + 9u_{j-2} - u_{j-3}}{60 \Delta \rho}.
\end{align*}
\]

The Kreiss-Oliger dissipation operator of the proper order for a sixth-order scheme is computed as [12]

\[
Q = \frac{\sigma h^7 D_+^4 D_-^4}{256}
\]

(14)
where $D_+ = (u_{j+1} - u_j)/\Delta \rho$ and $D_- = (u_j - u_{j-1})/\Delta \rho$ are the standard forward and backward differencing operators and $\sigma$ is parameter that is usually a value set between $(0, 1)$. Based on many numerical experiments, we found that a value of $\sigma = 0.01$ achieves stability for a large class of computations.

### 3.1.2 WENO Methods

To discretize the problem

$$u_t + f(u)_x = 0$$

in space using a WENO method we split the flux into its positive and negative parts

$$f(u) = f^+(u) + f^-(u),$$

such that

$$\frac{df^+(u)}{du} \geq 0, \quad \text{and} \quad \frac{df^-(u)}{du} \leq 0.$$

- The third order WENO method: For both the positive direction flux $\hat{f}^+$ and the negative direction flux $\hat{f}^-$ the smoothness measurements are:

$$IS_1 = (f^+_j - f^+_{j-1})^2, \quad IS_2 = (f^+_{j+1} - f^+_j)^2$$

and the weights of the candidate stencils are given by

$$\alpha_1 = \frac{1}{3} \left( \frac{1}{\epsilon + IS_1} \right)^2, \quad \alpha_2 = \frac{2}{3} \left( \frac{1}{\epsilon + IS_2} \right)^2,$$

$$\omega_1 = \frac{\alpha_1}{\alpha_1 + \alpha_2}, \quad \omega_2 = \frac{\alpha_2}{\alpha_1 + \alpha_2}.$$

The fluxes are:

$$\hat{f}^+_{j+\frac{1}{2}} = \omega_1 \left( \frac{3}{2} f^+_j - \frac{1}{2} f^+_{j-1} \right) + \omega_2 \left( \frac{1}{2} f^+_j + \frac{1}{2} f^+_{j+1} \right),$$

$$\hat{f}^-_{j+\frac{1}{2}} = \omega_1 \left( \frac{3}{2} f^-_j - \frac{1}{2} f^-_{j+1} \right) + \omega_2 \left( \frac{1}{2} f^-_j + \frac{1}{2} f^-_{j-1} \right).$$

- The fifth order WENO method: For the positive direction flux $\hat{f}^+$, the smoothness measurements are:

$$IS_0^+ = \frac{13}{12} \left( f^+_{j-2} - 2f^+_{j-1} + f^+_{j-1} \right)^2 + \frac{1}{4} \left( f^+_{j-2} - 4f^+_{j-1} + 3f^+_{j-1} \right)^2$$

$$IS_1^+ = \frac{13}{12} \left( f^+_{j-1} - 2f^+_{j+1} + f^+_{j+1} \right)^2 + \frac{1}{4} \left( f^+_{j-1} - f^+_{j+1} \right)^2$$

$$IS_2^+ = \frac{13}{12} \left( f^+_{j} - 2f^+_{j+2} + f^+_{j+2} \right)^2 + \frac{1}{4} \left( 3f^+_{j} - 4f^+_{j+1} + f^+_{j+2} \right)^2$$

For the negative direction flux, $\hat{f}^-$ the smoothness measurements are:

$$IS_0^- = \frac{13}{12} \left( f^-_{j-2} - 2f^-_{j-1} + f^-_{j-1} \right)^2 + \frac{1}{4} \left( 3f^-_{j-1} - 4f^-_{j+2} + f^-_{j+3} \right)^2$$

$$IS_1^- = \frac{13}{12} \left( f^-_{j} - 2f^-_{j+1} + f^-_{j+1} \right)^2 + \frac{1}{4} \left( f^-_{j} - f^-_{j+2} \right)^2$$

$$IS_2^- = \frac{13}{12} \left( f^-_{j-1} - 2f^-_{j+2} + f^-_{j+1} \right)^2 + \frac{1}{4} \left( f^-_{j-1} - 4f^-_{j} + 3f^-_{j+1} \right)^2$$
Now we discretize each flux as follows:

\[ \alpha_0^\pm = \frac{1}{10} \left( \frac{1}{\epsilon + IS_0^\pm} \right)^2 \quad \alpha_1^\pm = \frac{6}{10} \left( \frac{1}{\epsilon + IS_1^\pm} \right)^2 \quad \alpha_2^\pm = \frac{3}{10} \left( \frac{1}{\epsilon + IS_2^\pm} \right)^2 \]

and

\[ \omega_0^\pm = \frac{\alpha_0^\pm}{\alpha_0^\pm + \alpha_1^\pm + \alpha_2^\pm} \quad \omega_1^\pm = \frac{\alpha_1^\pm}{\alpha_0^\pm + \alpha_1^\pm + \alpha_2^\pm} \quad \omega_2^\pm = \frac{\alpha_2^\pm}{\alpha_0^\pm + \alpha_1^\pm + \alpha_2^\pm}. \]

Finally, we compute the fluxes, comprised of the candidate stencils with their weights:

\[ f_{j+\frac{1}{2}}^+ = \omega_0^+ \left( \frac{2}{6} f_{j-2}^+ - \frac{7}{6} f_{j-1}^+ + \frac{11}{6} f_j^+ \right) + \omega_1^+ \left( -\frac{1}{6} f_{j-1}^+ + \frac{5}{6} f_j^+ + \frac{2}{6} f_{j+1}^+ \right) \]
\[ + \omega_2^+ \left( \frac{2}{6} f_j^+ + \frac{5}{6} f_{j+1}^+ - \frac{1}{6} f_{j+2}^+ \right) \]

and

\[ f_{j+\frac{1}{2}}^- = \omega_0^- \left( -\frac{1}{6} f_{j-1}^- + \frac{5}{6} f_j^- + \frac{2}{6} f_{j+1}^- \right) + \omega_1^- \left( \frac{2}{6} f_j^- + \frac{5}{6} f_{j+1}^- - \frac{1}{6} f_{j+2}^- \right) \]
\[ + \omega_2^- \left( \frac{11}{6} f_{j+1}^- - \frac{7}{6} f_{j+2}^- + \frac{2}{6} f_{j+3}^- \right). \]

Now we discretize each flux as follows:

\[ f^+(u) = \frac{1}{\Delta \rho} \left( f_{j+\frac{1}{2}}^+ - f_{j-\frac{1}{2}}^- \right), \quad \text{and} \quad f^-(u) = \frac{1}{\Delta \rho} \left( f_{j+\frac{1}{2}}^- - f_{j-\frac{1}{2}}^- \right). \]

For the problem we are solving, we need to evaluate the term \( M \partial \rho u \). The structure of \( M \partial \rho u \) suggests that the first two elements in \( u \) have wavespeed \( b \), while the second two elements have wavespeed \( -b \). This is not completely correct, as the terms \( m_{31} \) and \( m_{22} \) have an impact as well; however, these terms are very small compared to the wavespeed for most of the computational domain. For simplicity, then, we simply approximate each element of \( \partial \rho u \) with a WENO discretization corresponding to the sign of the wavespeed. This convenient shortcut makes the numerical computation significantly faster and easier to code; however, the argument above is not rigorous and we do not have a complete numerical analysis that guarantees convergence of this process.

Proceeding with this approach, we begin by computing the smoothness measurements for each stencil. Note that each of the four elements of \( u \) is in fact a two-dimensional array, with the \( \rho \) and \( \theta \) directions. For convenience, we do not explicitly denote the \( \theta \) direction.

The determination of whether to use an upwind or downwind flux depends on whether the corresponding diagonal term in \( M \) is \( b \) or \( -b \). We treat the first two elements of \( u \) (which correspond to the value of \( b > 0 \) in the diagonal of \( M \)) as we would treat \( f^+(u) \), and the last two elements of \( u \) (which correspond to the value of \( -b < 0 \) in the diagonal of \( M \)) as we would treat \( f^-(u) \). For the first two elements (arrays) of \( u \), we define each of them in turn as an array \( f^+ \), where the values \( f^+_j \) refer to the \( \rho \) values. The WENO derivative is approximated for each value of \( \theta \), but the index refers to the grid-value of \( \rho \). For the last two elements of
we define each of them in turn as $\hat{f}^-$. The WENO process gives us the arrays $\hat{f}_{j+\frac{1}{2}}^\pm$ for all gridpoints $\rho_j$, and then we compute the first two arrays of $\partial_\rho u$ by

$$\frac{1}{\Delta \rho} \left( \hat{f}_{j+\frac{1}{2}}^+ - \hat{f}_{j-\frac{1}{2}}^+ \right)$$

and the last two element of $\partial_\rho u$ are computed by

$$\frac{1}{\Delta \rho} \left( \hat{f}_{j+\frac{1}{2}}^- - \hat{f}_{j-\frac{1}{2}}^- \right).$$

3.1.3 A mixed precision implementation of WENO methods

The WENO algorithm can be seen as having two parts: the inexpensive computation of a finite difference approximation to the derivative on several candidate stencils, and a costly nonlinear computation of the stencil weights. The computational bottleneck is in the computations of the stencil weights used to combine the stencil-based approximations. However, while the differentiation requires highly accurate computation, the calculation of the stencil weights does not require high precision. As long as the weights add up to 1.0 in high precision, the weights may not need to be high precision in the smooth regions, especially if the region where high precision is needed is near the location of the horizon. With this in mind we modified our quad-precision code to carry out a double-precision computation of the weights, then promoting them to quad-precision before finally assembling the WENO fluxes. This strategy is expected to speed up the computation significantly.

3.2 Time-discretization

When WENO is used to semi-discretize a problem of the form

$$u_t + f(u)_\rho = 0$$

we obtain a system of ODEs of the form

$$u_t = F(u).$$

The WENO method is designed to have an essentially non-oscillatory property when coupled with the forward Euler method,

$$u^{n+1} = u^n + \Delta t F(u^n)$$

under some stability condition $\Delta t \leq \Delta t_{FE}$. To preserve this property, we use a higher order strong stability preserving time discretization, which can be written as convex combinations of forward Euler schemes. Such time-stepping methods will preserve the properties of the spatial discretization coupled with forward Euler, under the modified time-step restriction

$$\Delta t \leq C \Delta t_{FE}.$$ 

If $C > 0$ we call the method Strong Stability Preserving (SSP). While the time-step depends on both the spatial and temporal discretizations, we isolate the contribution of the temporal discretization to the time-step restriction by considering the
ratio $C$ of the allowable time-step of the high order method to the forward Euler time-step. This ratio is called the strong stability preserving coefficient. Using this approach, we view the time-step restriction as a combination of two factors: the forward Euler time-step $\Delta t_{FE}$ that comes from the spatial discretization, and the SSP coefficient $C$ is a property only of the time-discretization. Among methods of similar types, a more relevant quantity is the effective SSP coefficient $C_{eff} = C/s$, which takes into account the computational cost of the method at each iteration, defined by the number of stages $s$ (typically also the number of function evaluations per time-step). In this work we consider two such methods [31].

The three stage, third order strong stability preserving Runge–Kutta method SSP-RK(3,3):

$$
\begin{align*}
    u^{(1)} &= u^n + \Delta t F(u^n) \\
    u^{(2)} &= \frac{3}{4}u^n + \frac{1}{4}(u^{(1)} + \Delta t F(u^{(1)})) \\
    u^{n+1} &= \frac{1}{3}u^n + \frac{2}{3}(u^{(2)} + \Delta t F(u^{(2)})).
\end{align*}
$$

This method has SSP coefficient $C = 1$ and effective SSP coefficient $C_{eff} = 1/3$.

The low storage, ten stage, fourth order strong stability preserving Runge–Kutta method SSP-RK(10,4):

$$
\begin{align*}
    u^{(1)} &= u^n + \frac{1}{6}\Delta t F(u^n), \\
    u^{(i)} &= u^{(i-1)} + \frac{1}{6}\Delta t F\left(u^{(i-1)}\right), \text{ for } i=2:4 \\
    u^{(5)} &= \frac{3}{5}u^n + \frac{2}{5}u^{(4)} + \frac{1}{15}\Delta t F(u^{(4)}), \\
    u^{(i)} &= u^{(i-1)} + \frac{1}{6}\Delta t F\left(u^{(i-1)}\right), \text{ for } i=6:9 \\
    u^{n+1} &= \frac{1}{25}u^n + \frac{9}{25}u^{(4)} + \frac{3}{5}u^{(9)} + \frac{3}{50}\Delta t F(u^{(4)}) + \frac{1}{10}\Delta t F(u^{(9)}).
\end{align*}
$$

This method has SSP coefficient $C = 6$ and effective SSP coefficient $C_{eff} = 0.6$.

For the problems considered in this paper, we found no significant difference between the performance of the two.

4 Results

Aretakis’ rigorous results apply to fairly generic situations, as long as the black hole is extremal. In particular, the initial wave that perturbs an otherwise isolated extremal black hole could be fairly generic as long as it has support on the horizon. For the following results we chose a narrow Gaussian radial profile i.e., a “wave-packet” centered at $\rho = 1.0$ with a width of 0.22. The angular profile of this initial perturbation is chosen to be the $(\ell, m) = (2, 0)$ spherical harmonic. And we present results for the gravitational perturbation case with $s = -2$. The numerical simulations presented here used a grid size of $16384(\rho) \times 64(\theta)$ and the SSP-RK(3,3) time-stepper.

We wrote the Teukolsky equation for a Kerr black hole with parameters $M, a$ for the variable $\Phi$ which is related to the Teukolsky function $\Psi$ in the Kinnersley tetrad.
and Boyer-Lindquist coordinates via $\Phi = (r/\Delta^2) \Psi$, where $\Delta = r^2 - 2Mr + a^2$.

One may relate this function to the better-known Weyl curvature in the following manner. The Weyl curvature scalar $\psi_{\text{HH}}^4$ in the Hartle-Hawking tetrad is related to its Kinnersley tetrad counterpart, $\psi_{\text{K}}^4$, via a type-III transformation, or $\psi_{\text{HH}}^4 = \frac{4}{r^2 + a^2} \Delta^{-2} \psi_{\text{K}}^4$ [32] and that $\Psi = (r - ia \cos \theta)^4 \psi_{\text{K}}^4$ [7] we find that

$$\Phi = \frac{r (r - ia \cos \theta)^4}{4(r^2 + a^2)^2} \psi_{\text{HH}}^4,$$  \hspace{1cm} (15)

and use $\Phi$ with $\ell = 2, m = 0$ and $a = M$.

Previous work by Aretakis and others [14,15,17] show that the radial derivatives of the physical field on the horizon would be a conserved quantity, a so-called “charge”. In addition, the same analysis also leads to an expectation of the time-dependent behavior for the physical field itself on the horizon (proportional to inverse time) and for higher radial derivatives (proportional to an increasing positive power of time: second-derivative $\propto \tau$, third-derivative $\propto \tau^2$ and so on). To summarize:

$$\phi^{(p+1)} \propto \tau^p,$$  \hspace{1cm} (16)

when evaluated at the horizon an extremal black hole. It is this rigorous result that we will compare our numerical solutions with, thus offering a precise assessment of the quality of our computational results. We will show that the WENO(5,3) method performs the best amongst the different methods we tested.

We begin with an example of some high-quality results from the WENO(5,3) method. We plot $\Phi$ for a fixed $\theta$ as a function of $\rho$ for a set of $\tau$ values in Fig. 1. These radial snapshots of the solution $\Phi$ at different moments of time suggest how the solution evolves forming a sharper-and-sharper feature at the horizon. This plot is highly suggestive of the fact that the higher radial derivatives of the field on the horizon will not decay, even though the field itself decays everywhere.

To observe that more clearly, in Fig. 2 we show the time-dependence of the numerical solution $\Phi$, and its radial-derivative $\Phi'$ together. As pointed out above, the physical field $\phi \propto \tau^{-1}$ while the radial-derivative $\phi'$ should be a constant. The plot clearly shows that both WENO(3,1) and WENO(5,3) perform significantly better even when compared to a higher-order standard (6th-order) finite-difference scheme. The 6th-order standard scheme requires the addition of a small amount of dissipation (a standard Kreiss-Oliger 8th-order filter operator [12]) to suppress high-frequency instabilities. However, even a small amount of dissipation is enough to significantly degrade the quality of the numerical solution.

Next, in Fig. 3 we show the numerically computed local power-law index calculated according to the formula $p = \tau \phi' / \phi$. Here we can clearly see that the WENO(5,3) performs the best i.e. maintains a value of $p$ close to $-1$ better than WENO(3,1) and also the standard 6th-order finite-difference scheme with dissipation. The same can be seen in the $\Phi'$ data as well, as shown in Fig. 4 and for higher-order derivatives.

In Fig. 5 we depict the relative difference between two WENO solutions – one using full quadruple-precision numerics and the other, using a mixed precision approach i.e. only the WENO “weights” are computed using double-precision floating-point operations. The mixed precision approach offers a 3.3 fold speedup while having no impact on the results in the regime of interest i.e. near the horizon. Recall that the computed solution $\Phi$ is complex valued. It is interesting to
Fig. 1 The gravitational field spatial profile snapshots in the background of an extremal Kerr black hole: The solution’s radial dependence at different moments of time. It is clear that as the evolution advances, the solution at the horizon (left-end of the computational domain) develops an increasingly sharp profile with the higher gradients growing unboundedly.

note that even though we begin the evolution using purely real initial data, the system’s evolution introduces a physical phase shift due to the spin of the black hole resulting in a non-zero imaginary part. The mixed precision related error in the imaginary part is considerably lower than in the real part.

4.1 Parallel Scaling Results

In this section we present the parallel scaling performance results of our WENO(5,3) code on a GPGPU-cluster with multiple Nvidia V100 GPGPUs.

First, we offer some details on both the GPGPU many-core processor architecture and implementation details relevant to our code. Nvidia’s CUDA framework[2] is a set of software layers that allow for GPU devices to become more accessible to the average computational scientist. At a high level, there is a CUDA Runtime API; while at a low level, there is the CUDA Driver API. Each function call of the Runtime API is broken into simpler instructions and managed by the Driver API. Through CUDA, the GPU (called device) is accessible to the CPU (called host) as a co-processor with its own memory. The device executes a function (usually

[2] https://developer.nvidia.com/about-cuda
referred to as a *kernel* in a “data parallel” model, which means that a number of threads run the same program on different data.

A data-parallel model is straightforward to implement in a code like ours. We simply perform a domain-decomposition of our finite-difference numerical grid and allocate the different parts of the grid to different GPU cores. Each thread computes a time-step for a single pair of \( \rho \) and \( \theta \) grid values. Note that all these calculations are independent, i.e. no communication is necessary between the GPU threads. However, it is necessary to establish the appropriate data communication between the GPU-cores and the remaining code that is executing on the CPU respectively. Of course, this fine-grain decomposition on the GPU is performed after a standard coarse-grain level domain-decomposition using MPI over the many GPUs that are part of a tightly-coupled cluster.

It is worth pointing out that this CUDA implementation of our WENO code is fairly straightforward. It should also be mentioned that we do not attempt to hand-tune the codes to tailor them for each architecture, in order to obtain maximal performance. Instead, we rely on the mature compiler suites to perform all low-level optimizations (such as vectorization) automatically.
The gravitational field $\Phi$ power law tail on the horizon of an extremal Kerr black hole: The expected behavior is $\Phi \propto \tau^{-1}$ i.e. a power-law of $-1$. It is clear that the WENO methods perform significantly better when compared to an even higher-order (dissipative) scheme. WENO(5,3) performs better than WENO(3,1).

The performance results on MIT’s Satori GPGPU supercomputer are presented in Fig. 6. This system is an IBM machine with 64-nodes; each node offers two Power9 CPUs (32-cores each) and four (4) Nvidia V100 GPGPUs connected via NVLink2. We ran the exact same computation using different numbers of GPUs: 1, 2, 4, 8, 16, 32 each time. Note that the 32-GPU computation used 8-nodes of this system i.e. a 12.5% of the entire system size. The results are clearly indicative of nearly perfect scaling performance. As demonstrated by one of the authors in Ref. [30] the GPU-acceleration alone speeds up the computation by nearly a factor of 50 over a CPU-only computation.

5 Conclusion

In this work we applied a number of GPU-accelerated numerical methods, including a novel mixed-precision fifth order WENO method to solve the Teukolsky equation – the master equation of black hole perturbation theory. Our results are in three distinct areas: numerical analysis, efficient computing, and gravitational wave astrophysics.

First, we showed that the WENO finite-difference methods of Jiang and Shu [13] out-performed the sixth order centered difference method with a standard Kreiss-Oliger filter. The dissipation added to the sixth order centered difference is needed to stabilize the method; however, it adversely impacts the quality of
Fig. 4 The gravitational field $\Phi'$ power law tail on the horizon of an extremal Kerr black hole: The expected behavior is $\Phi' \propto \tau^0$, i.e., a power-law of 0. It is clear that the WENO methods perform significantly better when compared to an even higher-order (dissipative) scheme. WENO(5,3) performs better than WENO(3,1).

the solution so that the expected behavior is not observed near the horizon. The WENO methods correct this issue, as they achieve stability by clever automated stencil choosing and no additional dissipation. For this reason, they attain the correct behavior near the horizon. As expected, the low-order WENO is not sufficiently accurate for this problem, but it still out-performs the sixth order centered difference method with the filter. The fifth order (third order near discontinuities) WENO method we employ confirms the predicted behavior of the system and proves to be appropriate for solving the problem of interest i.e., computing the so-called Aretakis charge, which is an important conserved physical quantity.

Next, we considered two computational approaches to speed up this code. Regardless of the algorithms used, this is a computationally intensive problem, which requires a very high precision computation (quad precision) to accurately resolve the long-time integration. This problem requires GPU-acceleration to complete in reasonable time. We show that our GPU-accelerated code scales very well as more GPU devices are used. Additionally, to further speed up the computation, we considered a novel mixed-precision approach to the WENO algorithm, by computing the WENO weights in double precision while the rest of the code is computed in quad precision. The mixed precision WENO approach results in a speed-up of a factor of 3.3. We investigated the errors introduced by this approach, and found that the relative error in the solution is very small ($< 10^{-4}$) near the horizon, which
Fig. 5 Relative differences between two numerical solutions: spatial profile snapshot of the relative differences at the last time-step i.e. $\tau = 2000$ between two WENO solutions. One solution was generated using full quadruple-precision floating-point computations; the other using a mixed precision approach i.e. the WENO “weights” are computed in double-precision. While the relative error becomes significant in the middle portion of the computational domain; it stays very low near the horizon while offering a 3.3 fold speedup.

is the regime of interest. This minor loss in accuracy is a small and worthwhile price to pay for the dramatic speedup in the code.

Finally, such high-accuracy simulations may enable significant discoveries in the field of computational and observational gravitational-wave physics. For example, the gravitational wave analog of the (scalar) Aretakis charge is in principle a measurable quantity that future observatories, like LIGO/Virgo, could detect. Observational evidence for the Aretakis charge would reveal the existence of extremal or near-extremal black holes, settling a long-standing open problem in the field of black hole physics.

6 Acknowledgements

Many of the computations were performed on the MIT/IBM Satori GPU supercomputer supported by the Massachusetts Green High Performance Computing Center (MGHPCC). The authors acknowledge support of NSF Grants No. PHY-2010685 (G.K) and No. DMS-1912716 (S.F, S.G, and G.K), AFOSR Grant No. FA9550-18-1-0383 (S.G) and Office of Naval Research/Defense University Research
Fig. 6 Parallel scaling performance of the fifth order WENO method on a GPU cluster. We report the total wall-clock time for the same computation using different numbers of GPU resources (up to 32). The parallel framework used by the code is CUDA+MPI. The simulations were performed on MIT’s Satori supercomputer that has an IBM Power9 CPU + quad Nvidia V100 GPUs node architecture. For reference, perfect scaling data is indicated on the same graph. It is clear that the code exhibits near-ideal scaling performance. We suspect the good scaling performance is partly due to the quad-precision operations, which in turn increases the computational load of our algorithm relative to the data movement cost expected for wider stencils.
7 Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. B. P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. Phys. Rev. Lett., 116(6):061102, 2016.
2. Kazunori Akiyama, Antxon Alberdi, Walter Alef, Keiichi Asada, Rebecca Azulay, Anne-Kathrin Bacsko, David Ball, Mislav Baloković, John Barrett, Dan Bintley, et al. First m87 event horizon telescope results. iv. imaging the central supermassive black hole. The Astrophysical Journal Letters, 875(1):L4, 2019.
3. Daniel Wysocki, Jacob Lange, and Richard O'Shaughnessy. Reconstructing phenomenological distributions of compact binaries via gravitational wave observations. Phys. Rev., D100(4):043012, 2019.
4. B. P. Abbott et al. Tests of general relativity with GW150914. Phys. Rev. Lett., 116(22):221101, 2016.
5. B. P. Abbott et al. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. Phys. Rev., X9(3):031040, 2019.
6. Tullio Regge and John A. Wheeler. Stability of a schwarzschild singularity. Phys. Rev., 108:1063–1069, Nov 1957.
7. Saul A Teukolsky. Perturbations of a rotating black hole. 1. fundamental equations for gravitational electromagnetic and neutrino field perturbations. Astrophys. J., 185:635–647, 1973.
8. Pranesh A Sundararajan, Gaurav Khanna, and Scott A Hughes. Towards adiabatic waveforms for inspiral into kerr black holes: A new model of the source for the time domain perturbation equation. Physical Review D, 76(10):104005, 2007.
9. Pranesh A Sundararajan, Gaurav Khanna, Scott A Hughes, and Steve Drasco. Towards adiabatic waveforms for inspiral into kerr black holes. ii. dynamical sources and generic orbits. Physical Review D, 78(2):024022, 2008.
10. Pranesh A Sundararajan, Gaurav Khanna, and Scott A Hughes. Binary black hole merger gravitational waves and recoil in the large mass ratio limit. Physical Review D, 81(10):104009, 2010.
11. Anıl Zenginoglu and Gaurav Khanna. Null infinity waveforms from extreme-mass-ratio inspirals in kerr spacetime. Physical Review X, 1(2):021017, 2011.
12. Heinz Otto Kreiss and Joseph Oliger. Methods for the approximate solution of time dependent problems. 1973.
13. Guangle-Shan Jiang and Chi-Wang Shu. Efficient implementation of weighted eno schemes. Journal of computational physics, 126(1):202–228, 1996.
14. Stefanos Aretakis. A note on instabilities of extremal black holes under scalar perturbations from afar. Classical and Quantum Gravity, 30(9):095010, apr 2013.
15. Y. Angelopoulos, Stefanos Aretakis, and Dejan Gajic. Horizon hair of extremal black holes and measurements at null infinity. Phys. Rev. Lett., 121(13):131102, 2018.
16. Lior M. Burko, Gaurav Khanna, and Subir Sarkharwal. (Transient) Scalar Hair for (Nearly) Extreme Black Holes. Phys. Rev. Research., 1:033106, 2019.
17. Lior M Burko, Gaurav Khanna, and Subir Sarkharwal. Scalar and gravitational hair for extreme kerr black holes. arXiv preprint arXiv:2005.07294, 2020.
18. Anıl Zenginoglu and Manuel Tiglio. Spacelike matching to null infinity. Phys. Rev. D, 80:024044, 2009.
19. Anıl Zenginoglu. A Geometric framework for black hole perturbations. Phys. Rev. D, 83:127502, 2011.
20. Anıl Zenginoglu. Hyperboloidal layers for hyperbolic equations on unbounded domains. J. Comput. Phys., 230:2286–2302, 2011.
21. Anıl Zenginoglu. Asymptotics of black hole perturbations. Class. Quant. Grav., 27:045015, 2010.
22. Anil Zenginoğlu, Dario Nunez, and Sascha Husa. Gravitational perturbations of Schwarzschild spacetime at null infinity and the hyperboloidal initial value problem. *Class. Quant. Grav.*, 26:035009, 2009.
23. Anil Zenginoğlu. Hyperboloidal evolution with the Einstein equations. *Class. Quant. Grav.*, 25:195025, 2008.
24. Anil Zenginoğlu. A hyperboloidal study of tail decay rates for scalar and Yang-Mills fields. *Class. Quant. Grav.*, 25:175013, 2008.
25. Anil Zenginoğlu. Hyperboloidal foliations and scri-fixing. *Class. Quant. Grav.*, 25:145002, 2008.
26. Anil Zenginoğlu, Gaurav Khanna, and Lior M. Burko. Intermediate behavior of Kerr tails. *Gen. Rel. Grav.*, 46:1672, 2014.
27. Lior M. Burko, Gaurav Khanna, and Anil Zenginoğlu. Cauchy-horizon singularity inside perturbed Kerr black holes. *Phys. Rev. D*, 93(4):041501, 2016. [Erratum: Phys.Rev.D 96, 129903 (2017)].
28. Anil Zenginoğlu and Gaurav Khanna. Null infinity waveforms from extreme-mass-ratio inspirals in Kerr spacetime. *Phys. Rev. X*, 1:021017, 2011.
29. Enno Harms, Sebastiano Bernuzzi, Alessandro Nagar, and An Zenginoğlu. A new gravitational wave generation algorithm for particle perturbations of the Kerr spacetime. *Class. Quant. Grav.*, 31(24):245004, 2014.
30. Gaurav Khanna. High-precision numerical simulations on a cuda gpu: Kerr black hole tails. *Journal of Scientific Computing*, 56(2):366–380, 2013.
31. Sigal Gottlieb, David I Ketcheson, and Chi-Wang Shu. *Strong stability preserving Runge-Kutta and multistep time discretizations*. World Scientific, 2011.
32. Eric Poisson. Absorption of mass and angular momentum by a black hole: Time-domain formalisms for gravitational perturbations, and the small-hole or slow-motion approximation. *Physical Review D*, 70(8):084044, 2004.