

Cholla-MHD: An Exascale-Capable Magnetohydrodynamic Extension to the Cholla Astrophysical Simulation Code

ROBERT V. CADDY ¹ AND EVAN E. SCHNEIDER ¹

¹*University of Pittsburgh, Department of Physics & Astronomy
3941 O'Hara St
Pittsburgh, PA 15260, USA*

ABSTRACT

We present an extension of the massively parallel, GPU native, astrophysical hydrodynamics code Cholla to magnetohydrodynamics (MHD). Cholla solves the ideal MHD equations in their Eulerian form on a static Cartesian mesh utilizing the Van Leer + Constrained Transport integrator, the HLLD Riemann solver, and reconstruction methods at second and third order. Cholla's MHD module can perform ≈ 260 million cell updates per GPU-second on an NVIDIA A100 while using the HLLD Riemann solver and second order reconstruction. The inherently parallel nature of GPUs combined with increased memory in new hardware allows Cholla's MHD module to perform simulations with resolutions $\sim 500^3$ cells on a single high end GPU (e.g. an NVIDIA A100 with 80GB of memory). We employ GPU direct MPI to attain excellent weak scaling on the exascale supercomputer *Frontier*, while using 74,088 GPUs and simulating a total grid size of over 7.2 trillion cells. A suite of test problems highlights the accuracy of Cholla's MHD module and demonstrates that zero magnetic divergence in solutions is maintained to round off error. We also present new testing and continuous integration tools using GoogleTest, GitHub Actions, and Jenkins that have made development more robust and accurate and ensure reliability in the future.

1. INTRODUCTION

Over the past decade it has become increasingly clear that magnetohydrodynamics (MHD) plays a significant role in a variety of astrophysical phenomena (e.g. Charbonneau 2014; Beck 2016; Naab & Ostriker 2017; Han 2017; Kaspi & Beloborodov 2017; Blandford et al. 2019; Nakariakov & Kolotkov 2020; Davis & Tchekhovskoy 2020; Trujillo Bueno & del Pino Alemán 2022; Philippov & Kramer 2022; Brandenburg & Ntormousi 2023). Magnetic fields couple to gas both directly through plasma interactions with the magnetic field, and indirectly through cosmic ray transport (e.g. Pfrommer et al. 2017; Girichidis et al. 2019; Chan et al. 2019; Buck et al. 2020; Werhahn et al. 2021; Yoshida et al. 2021; Girichidis et al. 2022; Nuñez-Castiñeyra et al.

2022; Werhahn et al. 2023), anisotropic conduction (e.g. Yang et al. 2012; Hanasz et al. 2013; Simpson et al. 2016; Girichidis et al. 2016; Pakmor et al. 2016; Brüggén et al. 2023), and other physical effects.

Many studies have focused on the role of magnetic fields in galaxy evolution and have provided intriguing hints as to the possible effects of magnetic fields in galaxy dynamics (Shin et al. 2008; Banda-Barragán et al. 2016; Grand et al. 2017; Pakmor et al. 2017; Hopkins et al. 2020; Wibking & Krumholz 2021). Different simulations employing different numerical methods show varying impacts of magnetic fields, ranging from magnetic fields being largely irrelevant on large scales to magnetic fields being critically important in determining the evolution and structure of the interstellar medium (ISM), modifying galactic feedback, and influencing the structure of the circumgalactic medium (CGM) (Martin-Alvarez et al. 2020; Dobbs & Price 2007; Kim & Ostriker 2015; Hanasz et al. 2009; Pakmor & Springel 2013; Banda-Barragán et al. 2017). One factor in this uncertainty is the effect of numerical resolution on MHD simulations – galactic magnetic fields are likely amplified by a turbulent dynamo, which operates over a large dynamic range (Martin-Alvarez et al. 2018; Beck et al. 1996; Gent et al. 2021; Carteret et al. 2022; Brandenburg & Ntormousi 2023). The accuracy with which the dynamo is captured thus depends on both the numerical method that is employed as well as the resolution. Thus, a simple way to extend the dynamic range captured in an MHD simulation is by employing very high resolution MHD simulations – simulations that are now possible thanks to recent developments in hardware, numerical algorithms, and software.

Modern numerical methods for MHD are sophisticated and robust, but even with highly optimized codes, MHD simulations remain very computationally expensive (Stone et al. 2020). This computational expense is a result of the high number of floating point calculations required by modern finite-volume methods and the heavy memory bandwidth demands of MHD codes (Grete et al. 2021). In addition, MHD turbulent dynamos operate across a large dynamic range, from the full scale of a galaxy all the way down to a few parsecs or smaller, five or more orders of magnitude spatially (Pariev & Colgate 2007; Pariev et al. 2007; Ntormousi et al. 2020; Galishnikova et al. 2022). As a result, high resolution simulations are critical in order to accurately capture the effects of magnetic fields in astrophysical simulations.

This need has driven a push to develop MHD codes that can take advantage of modern computer architectures (e.g. Schive et al. 2018; Almgren et al. 2020; Zingale et al. 2020; Shankar et al. 2022; Liska et al. 2022; Bégué et al. 2023; Holmen et al. 2023; Grete et al. 2023). Very large, high resolution MHD simulations require supercomputers to run due to their high computational cost. The primary source of computational power in most new supercomputers is Graphics Processing Units (GPUs)¹. For example, of the top ten supercomputers named in the November 2023

¹ <https://www.top500.org/lists/top500/2023/11/>

Top500 list, only one - Fugaku - does not rely on GPUs for the majority of its performance.² The ubiquity of GPUs in modern supercomputers thus necessitates the development of GPU-based astrophysical MHD simulation codes.

The induction equation implies a simple constraint; it requires that the magnetic field be divergence free, i.e. the Universe does not contain magnetic monopoles. However, achieving this constraint numerically is not trivial, and several methods have been developed. Among these are the Powell 8 Wave scheme, which adds an additional source term to the induction equation and uses an 8 wave Riemann solver with the 8th wave corresponding to the magnetic divergence (Powell 1997); vector potential/projection methods, which project the magnetic field into the scalar and vector potential and then perform a cleaning step to reduce the divergence to zero by solving a Poisson equation (Brackbill & Barnes 1980; Ryu et al. 1995; Crockett et al. 2005; Torrilhon 2005); divergence cleaning methods, which operate by modifying the system of conserved equations with corrections that dissipate and propagate out the magnetic divergence (Dedner et al. 2002; Mignone & Tzeferacos 2010); and constrained transport (CT), which evolves face-centered values of the magnetic field and updates them using the electromotive forces (Evans & Hawley 1988). We have chosen to implement the CT method in Cholla due to its overall accuracy, because the algorithm pairs well with static structured grids, and because the computational efficiency of GPUs is a good match to the algebraic complexity of CT.

Constrained transport is formally divergence free, and when implemented numerically it typically results in divergence errors on the order of machine round off error (Evans & Hawley 1988; Gardiner & Stone 2005; Stone et al. 2008; Stone & Gardiner 2009; Zingale et al. 2020; Almgren et al. 2020). This is accomplished by tracking magnetic fields on a staggered, face centered grid rather than using cell-centered averages. These face centered values are used in conjunction with Riemann fluxes to calculate edge centered electric fields, and those electric fields are used to update the magnetic field. Thus, the trade-off for a divergence-free method is significant additional algorithmic complexity and associated computational expense, which can make the method more challenging to implement in a particle based scheme or when using unstructured meshes.


The Cholla code (Computational Hydrodynamics On paraLLeL Architectures) (Schneider & Robertson 2015) is a fixed grid, finite volume hydrodynamics code for astrophysics that was designed to run natively on GPU-based supercomputers. It employs an unsplit 3D hydrodynamics integrator based on the Van Leer predictor-corrector method (Falle 1991; Van Leer 2006) and was designed to be extended to MHD using constrained transport (Evans & Hawley 1988). This work presents the MHD extension of Cholla. Our MHD implementation largely follows the Van Leer + Constrained Transport (VL+CT) method presented in Stone & Gardiner (2009) with

² Fugaku employs custom CPUs that utilize vector processors similar to GPUs for much of its performance.

modifications for GPUs. It also uses an HLLD Riemann solver (Miyoshi & Kusano 2005) and includes second (Stone & Gardiner 2009) and third (Felker & Stone 2018) order reconstruction in the characteristic variables (Stone et al. 2008). We also highlight implementation choices that are particularly relevant to solving these equations on GPUs.

The extension of Cholla to include MHD allows the simulation of previously unreachable domains. The VL+CT integrator provides high accuracy results with divergences that are zero to round off error. Given current memory constraints, the code is fast enough that a $\sim 459^3$ cell MHD simulation can be run on a single AMD MI250X Graphics Compute Die (GCD) (logically a single GPU), allowing high resolution simulations to be run with only a small number of local resources. In addition, Cholla scales up to power of the largest available supercomputers, enabling simulations up to $19,278^3 \approx 7.2$ trillion cells on *Frontier*³, the world’s first exascale supercomputer. This will allow MHD simulations of entire galaxies with a constant resolution of a few parsecs per cell, turbulent box simulations with many trillions of cells, or many thousands of lower resolution simulations to be computed rapidly. For example, with approximately the same computing resources used to evolve a $19,278^3$ simulation one could run thousands of $2,000^3$ -cell simulations, enabling entire parameter studies with resolutions comparable to current cutting edge CPU-based simulations.

In addition to requiring complex algorithms to produce accurate results, modern community-developed simulation codes like Cholla require robust testing and software-development infrastructure to maintain their reliability. This work also presents the implementation of an automated testing/continuous integration (CI) pipeline for Cholla. CI tools have expanded rapidly in functionality and popularity over the last 20 years and their usefulness in scientific software is well established (Beck 1999; Wilson et al. 2014, 2017). Particularly in the last 5 years with the advent of GitHub Actions, Jenkins, and similar easily accessible and cheap (or even free) tools, CI pipelines have become much more straightforward to set up and run even for small groups and individuals. We present our implementation of testing and CI that is designed to be straightforward and scalable from a single GPU all the way up to an exascale machine.

The outline of this paper is as follows. In Section 2, we describe our implementation of the VL+CT algorithm in detail along with the modifications we made to efficiently run on GPUs. In Section 3, we demonstrate the correctness and accuracy of Cholla on a suite of MHD test problems. We also describe Cholla’s performance and weak scaling behavior on up to 74,088 GPUs using *Frontier*. In Section 4, we discuss the new continuous integration and automated testing framework. We conclude in Section 5. In figure captions the Zenodo icon, , links to the Zenodo repository

³ <https://www.olcf.ornl.gov/frontier/>

that contains the python scripts to generate each figure. This Zenodo repository has sufficient information to reproduce the figures shown in this paper.

2. METHODS

2.1. Magnetohydrodynamics

Cholla solves the ideal MHD equations in their conserved Eulerian form using a finite volume method (Godunov & Bohachevsky 1959). These equations neglect all dissipative processes, including finite viscosity, electrical resistivity, and thermal conductivity. These approximations are reasonable when simulating regions of with very high Reynolds numbers as is common in many astrophysical problems. We note that although we neglect these additional processes at present, the methods implemented here are fully compatible with future extensions to include additional physics that depends on magnetic fields, such as anisotropic conduction, cosmic ray transport, or non-ideal MHD effects.

The ideal MHD equations are:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (1)$$

$$\frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v} - \mathbf{B} \otimes \mathbf{B} + P_T \mathbf{I}) = 0 \quad (2)$$

$$\frac{\partial E}{\partial t} + \nabla \cdot ((E + P_T) \mathbf{v} + \mathbf{B}(\mathbf{B} \cdot \mathbf{v})) = 0 \quad (3)$$

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) = 0 \quad (4)$$

where ρ is density, $\mathbf{v} = (v_x, v_y, v_z)$ is the velocity vector, t is time, $\mathbf{B} = (B_x, B_y, B_z)$ is the magnetic field, \mathbf{I} is the identity tensor, $P_T \equiv P + \frac{1}{2}(\mathbf{B} \cdot \mathbf{B})$ is the total pressure, and E is the total energy per unit volume $E \equiv \epsilon + \frac{1}{2}\rho(\mathbf{v} \cdot \mathbf{v}) + \frac{1}{2}(\mathbf{B} \cdot \mathbf{B})$. We adopt units in which the magnetic permeability $\mu_0 = 1$.

Equation 1 describes the conservation of mass, equation 2 describes the conservation of momentum, equation 3 describes the conservation of energy, and equation 4 is the induction equation, which describes the divergence free condition. Cholla uses an ideal gas equation of state which is $P \equiv \epsilon(\gamma - 1)$ where γ is the adiabatic index and ϵ is the internal energy density.

In practice these equations are used in their vector form, where \mathbf{U} and \mathbf{W} are the conserved and primitive variables respectively:

$$\mathbf{U} = [\rho, \rho v_x, \rho v_y, \rho v_z, E, B_x, B_y, B_z] \quad (5)$$

$$\mathbf{W} = [\rho, v_x, v_y, v_z, P, B_x, B_y, B_z]. \quad (6)$$

The conservation equations, in Cartesian coordinates, can then be rewritten as

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}_x}{\partial x} + \frac{\partial \mathbf{F}_y}{\partial y} + \frac{\partial \mathbf{F}_z}{\partial z} = 0 \quad (7)$$

where \mathbf{F}_x , \mathbf{F}_y , and \mathbf{F}_z are the vectors of fluxes in the x , y , and z direction respectively and are given by

$$\mathbf{F}_x = \begin{bmatrix} \rho v_x \\ \rho v_x^2 + P_T - B_x^2 \\ \rho v_x v_y - B_x B_y \\ \rho v_x v_z - B_x B_z \\ v_x (E + P_T) - B_x (\mathbf{v} \cdot \mathbf{B}) \\ 0 \\ B_y v_x - B_x v_y \\ B_z v_x - B_x v_z \end{bmatrix} \quad (8)$$

$$\mathbf{F}_y = \begin{bmatrix} \rho v_y \\ \rho v_y v_x - B_y B_x \\ \rho v_y^2 + P_T - B_y^2 \\ \rho v_y v_z - B_y B_z \\ v_y (E + P_T) - B_y (\mathbf{v} \cdot \mathbf{B}) \\ B_x v_y - B_y v_x \\ 0 \\ B_z v_y - B_y v_z \end{bmatrix} \quad (9)$$

$$\mathbf{F}_z = \begin{bmatrix} \rho v_z \\ \rho v_z v_x - B_z B_x \\ \rho v_z v_y - B_z B_y \\ \rho v_z^2 + P_T - B_z^2 \\ v_z (E + P_T) - B_z (\mathbf{v} \cdot \mathbf{B}) \\ B_x v_z - B_z v_x \\ B_y v_z - B_z v_y \\ 0 \end{bmatrix}. \quad (10)$$

2.2. The VL+CT Integrator

To integrate these equations in Cholla, we implement the VL+CT (Van Leer plus Constrained Transport) integrator introduced in [Stone & Gardiner \(2009\)](#), along with the HLLD Riemann solver described in [Miyoshi & Kusano \(2005\)](#). Many of the equations were first described in the context of the constrained-transport extension to the corner-transport upwind (CTU) method ([Colella 1990](#)) described in 2D ([Gardiner](#)

& Stone 2005) and 3D (Gardiner & Stone 2008) for the Athena MHD code (Stone et al. 2008). Our specific implementation of the piecewise parabolic method follows Felker & Stone (2018), which is itself an extension of the original PPM method presented by Colella & Woodward (1984). The VL+CT integrator is very similar in structure to the MUSCL-Hancock integrator (Van Leer 2006; Falle 1991; Toro 2009), though with significant new additions for Constrained Transport (CT). We note that this is now the default integrator used in Cholla, as we have found it to be more robust than the CTU algorithm described in Schneider & Robertson (2015).

Constrained transport treats the magnetic field as a surface averaged quantity centered at cell interfaces rather than a volume averaged, cell centered quantity like the hydro variables; i.e. a staggered grid. Each face stores only the magnetic field perpendicular to that face, i.e. the $x, i + 1/2, j, k$ face stores the $B_{x,i+1/2,j,k}$ magnetic field. At each time step, the magnetic field is then updated using edge averaged electric fields computed from the magnetic flux returned by the Riemann solver. Updating the magnetic field with the electric field automatically fulfills the divergence free condition for magnetic fields, assuming that the initial conditions are also divergence free.

A brief overview of the algorithm for a single time step using the VL+CT integrator is as follows; more detailed discussion of each step is presented in the following subsections.

1. Compute the time step, Δt .
2. Reconstruct interface states using a piecewise constant method (PCM) approximation.
3. Solve the Riemann problem on every interface using the PCM interface states.
4. Compute the edge centered electric fields.
5. Update the $t = n$ conserved variables to $t = n + \Delta t/2$ using the computed fluxes and electric fields.
6. Reconstruct interface states using a piecewise linear method (PLM) or piecewise parabolic method (PPM) approximation.
7. Solve the Riemann problem on every interface using the higher order reconstructed interface states.
8. Recompute the edge centered electric fields.
9. Update the $t = n$ conserved variables to $t = n + \Delta t$ using the half time step fluxes and electric fields

2.2.1. Step 1: Compute the Time Step

The first step is to compute the time step. Cholla implements uniform time steps across the grid, so this is the minimum crossing time of any wave in any cell multiplied by a Courant factor to maintain the Courant–Friedrichs–Lewy condition [Courant et al. \(1967\)](#):

$$\Delta t = C_{CFL} \min \left(\frac{\Delta x}{|v_{x,i,j,k}^n| + c_{f,i,j,k}^n}, \frac{\Delta y}{|v_{y,i,j,k}^n| + c_{f,i,j,k}^n}, \frac{\Delta z}{|v_{z,i,j,k}^n| + c_{f,i,j,k}^n} \right), \quad (11)$$

where Δt is the time step, $C_{CFL} \leq 0.5$ is the CFL number, $|v_{l,i,j,k}^n|$ is the magnitude of the velocity in the l direction velocity in the i, j, k cell, and c_f^n is the fast magnetosonic wave-speed computed using *cell centered* values. The method is formally accurate at $C_{CFL} = 0.5$, though that is an estimate ([Stone & Gardiner 2009](#)). We have found empirically that the method is not stable at $C_{CFL} = 0.5$ but is stable at $C_{CFL} = 0.4^4$. The fast and slow magnetosonic speeds are

$$c_{f,s} = \sqrt{\frac{\gamma p + |\mathbf{B}|^2 \pm \sqrt{(\gamma p + |\mathbf{B}|^2)^2 - 4\gamma p B_x^2}}{2\rho}} \quad (12)$$

where the $+$ option corresponds to the fast magnetosonic speed, and $-$ to the slow.

Equation 11 computes the minimum crossing time of a wave in a specific cell. A global reduction is then performed to find the minimum in the entire grid; that minimum is used as the time step Δt . This reduction is done primarily on the GPU followed by an MPI ALL_REDUCE. Because GPU's programming methods are inherently parallel, GPU reductions are not trivial and their performance is sensitive to the method used. Details of our reduction method are discussed in section 2.3.

The cell-centered magnetic field is computed with a direct average of the face centered values:

$$\begin{aligned} B_{x,i,j,k}^n &= \frac{1}{2} (B_{x,i+1/2,j,k}^n + B_{x,i-1/2,j,k}^n) \\ B_{y,i,j,k}^n &= \frac{1}{2} (B_{y,i,j+1/2,k}^n + B_{y,i,j-1/2,k}^n) \\ B_{z,i,j,k}^n &= \frac{1}{2} (B_{z,i,j,k+1/2}^n + B_{z,i,j,k-1/2}^n) \end{aligned} \quad (13)$$

These cell-centered values for the magnetic field are used several times throughout the integrator. In CPU-based codes it is typically more efficient to compute these cell centered magnetic fields once and save them, since traditional CPU-based codes are typically compute limited. Cholla is generally limited instead by GPU memory, so we recompute the cell-centered values when they are needed rather than permanently allocating the associated memory.

⁴ Typically we use $C_{CFL} = 0.3$ to improve the accuracy of the operator splitting

2.2.2. Step 2: First Order Reconstruction

Reconstructing the interface states at first order, or the piecewise constant method (PCM), is done by setting the primitive interface states values to the same value as the cell:

$$\mathbf{W}_{L,i+1/2} = \mathbf{W}_{R,i-1/2} = \mathbf{W}_i \quad (14)$$

where $\mathbf{W}_{L/R,i\pm 1/2}$ is the state on the left or right side of the cell. Although Cholla evolves the conserved variables, primitive variables are typically used for interface reconstruction since they are required by the Riemann solver used to compute the fluxes. In higher order spatial reconstructions we also use the characteristic variables derived from the primitive variables.

Although PCM is too diffusive to be used on its own in both reconstruction steps, we note that that mode is excellent for debugging. Here, it offers a computationally-inexpensive way to calculate first-order fluxes, which will be used for the “predictor” step of this predictor-corrector algorithm (Van Leer 2006; Stone & Gardiner 2009). Most higher order methods revert to PCM near large discontinuities (LeVeque 2002).

2.2.3. Step 3: First Riemann Solve

The next step is to solve the Riemann problem with the first order interface states. To do this, we employ the HLLD Riemann solver introduced in Miyoshi & Kusano (2005). Because we have implemented the Riemann solver exactly as described in Miyoshi & Kusano (2005) we do not reproduce all the details here. The longitudinal magnetic field does not require reconstruction and can be used directly since it is stored at the face. The transverse fields (i.e. the fields parallel to the interface) are reconstructed from the cell centered average (computed as shown in section 2.2.1), then reconstructed using PCM identically to other fields.

The magnetic fluxes returned by the Riemann solver are the face centered electric fields (see section 5.3 of Stone et al. 2008). For clarity, we list in Table 1 the correlation between flux and electric field.

2.2.4. Step 4: Compute the Constrained Transport Electric Field

The next step is to calculate the constrained transport electric field. These fields are *line averaged* along each cell vertex. These line averaged fields are constructed by averaging the face centered electric fields from the previous step, and their slopes. Equation 15 is used to reconstruct these line averaged electric fields; only the equation for the z -direction is presented, the equations for the other directions can be found by cyclic permutation. The equations for computing the other directions are obtained by substituting out the z index with x or y and changing the derivatives appropriately.

Table 1. Magnetic Flux to Face Centered electric field

HLLD Solve Direction	Equation for Magnetic Flux	Eqn. as a Cross Product	electric field
X	$V_x B_y - B_x V_y$	$(V \times B)_z$	$-\varepsilon_z$
X	$V_x B_z - B_x V_z$	$-(V \times B)_y$	ε_y
Y	$V_x B_y - B_x V_y$	$(V \times B)_z$	$-\varepsilon_x$
Y	$V_x B_z - B_x V_z$	$-(V \times B)_y$	ε_z
Z	$V_x B_y - B_x V_y$	$(V \times B)_z$	$-\varepsilon_y$
Z	$V_x B_z - B_x V_z$	$-(V \times B)_y$	ε_x

NOTE—The directions used here are relative to the internal workings of the HLLD solver. Since the HLLD solver is inherently 1D we run it once for each of the faces of a cell. So in the case where the solver is running in the Y direction the solver’s Y field is actually the Z field and the solver’s Z field is actually the X field, cyclically extended for the Z direction.

$$\begin{aligned}
\mathcal{E}_{z,i-1/2,j-1/2,k} = & \frac{1}{4} \left(\mathcal{E}_{z,i-1/2,j,k} + \mathcal{E}_{z,i,j-1/2,k} + \mathcal{E}_{z,i-1/2,j-1,k} + \mathcal{E}_{z,i-1,j-1/2,k} \right) \\
& + \frac{\Delta y}{8} \left(\left(\frac{\partial \mathcal{E}_z}{\partial y} \right)_{i-1/2,j-1/4,k} + \left(\frac{\partial \mathcal{E}_z}{\partial y} \right)_{i-1/2,j-3/4,k} \right) \\
& + \frac{\Delta x}{8} \left(\left(\frac{\partial \mathcal{E}_z}{\partial x} \right)_{i-1/4,j-1/2,k} + \left(\frac{\partial \mathcal{E}_z}{\partial x} \right)_{i-3/4,j-1/2,k} \right)
\end{aligned} \tag{15}$$

$\mathcal{E}_{z,i-1/2,j-1/2,k}$ is the line averaged electric field; the first four terms on the right are the face averaged electric fields, and the four derivative terms are the derivatives of those fields in the direction towards the edge, details of which are in Equation 17. Figure 1 shows the spatial relationships between the derivatives and the relevant edge states and reference fields. Each edge requires 4 derivatives and they are computed as differences between a reference state and an edge state.

Technically, constrained transport uses the magnetic flux as the conserved variable and EMF (ElectroMotive Force) to update it. However, because those values only differ from the magnetic flux density (i.e. the magnetic field) and the electric field respectively by factors of unit length either can be treated as the conserved variable, the same way either density or mass can be evolved in the hydro fields (Stone et al. 2008). As such we use the magnetic field and electric field to evolve the grid. Electric fields have the proper units to evolve the magnetic flux density, B , whereas EMF has the proper units to evolve the magnetic flux.

On any face there are two non-zero electric fields; both transverse to the face. The component that is used to calculate the field along a given edge is the component that is parallel to that edge; i.e. if the edge points along the z -direction then the field pointing along the z -direction is used, not the field in the x or y direction.

The derivatives from Equation 15 are computed using the upwinded slope as follows

$$\begin{aligned}
312 \quad \left(\frac{\partial \mathcal{E}_z}{\partial y} \right)_{i-1/2, j-1/4, k} &= \begin{cases} \left(\frac{\partial \mathcal{E}_z}{\partial y} \right)_{i-1, j-1/4, k} & \text{for } v_{x, i-1/2} > 0 \\ \left(\frac{\partial \mathcal{E}_z}{\partial y} \right)_{i, j-1/4, k} & \text{for } v_{x, i-1/2} < 0 \\ \frac{1}{2} \left(\left(\frac{\partial \mathcal{E}_z}{\partial y} \right)_{i-1, j-1/4, k} + \left(\frac{\partial \mathcal{E}_z}{\partial y} \right)_{i, j-1/4, k} \right) & \text{otherwise} \end{cases} \quad (16)
\end{aligned}$$

313 where, for example, the derivatives are given by

$$314 \quad \left(\frac{\partial \mathcal{E}_z}{\partial y} \right)_{i, j-1/4, k} = 2 \left(\frac{\mathcal{E}_{z, i, j-1/2, k} - \mathcal{E}_{z, i, j, k}^{ref}}{\Delta y} \right). \quad (17)$$

315 $\mathcal{E}_{z, i, j, k}^{ref}$, is the cell centered reference field. This reference field is computed with the
316 following cross product

$$317 \quad \mathcal{E}_{i, j, k}^{ref, n} = -(\mathbf{v}_{i, j, k}^n \times \mathbf{B}_{i, j, k}^n). \quad (18)$$

318 2.2.5. Step 5. Perform the Half Time-step Update

319 We first update the density, momenta, and energy, but not the magnetic fields,
320 using the standard conservative update equation and the first-order fluxes from Step
321 3:

$$\begin{aligned}
322 \quad \mathbf{U}_{i, j, k}^{n+1/2} &= \mathbf{U}_{i, j, k}^n - \frac{\Delta t}{\Delta x} (\mathbf{F}_{x, i+1/2, j, k}^n - \mathbf{F}_{x, i-1/2, j, k}^n) \\
&\quad - \frac{\Delta t}{\Delta y} (\mathbf{F}_{y, i, j+1/2, k}^n - \mathbf{F}_{y, i, j-1/2, k}^n) \\
&\quad - \frac{\Delta t}{\Delta z} (\mathbf{F}_{z, i, j, k+1/2}^n - \mathbf{F}_{z, i, j, k-1/2}^n). \quad (19)
\end{aligned}$$

323 We then update the magnetic field using the electric fields computed in Step 4:

$$\begin{aligned}
324 \quad B_{x, i-1/2, j, k}^{n+1/2} &= B_{x, i-1/2, j, k}^n + \frac{\Delta t}{\Delta z} (\mathcal{E}_{y, i-1/2, j, k+1/2}^n - \mathcal{E}_{y, i-1/2, j, k-1/2}^n) \\
&\quad - \frac{\Delta t}{\Delta y} (\mathcal{E}_{z, i-1/2, j+1/2, k}^n - \mathcal{E}_{z, i-1/2, j-1/2, k}^n) \quad (20)
\end{aligned}$$

$$\begin{aligned}
325 \quad B_{y, i, j-1/2, k}^{n+1/2} &= B_{y, i, j-1/2, k}^n + \frac{\Delta t}{\Delta x} (\mathcal{E}_{z, i+1/2, j-1/2, k}^n - \mathcal{E}_{z, i-1/2, j-1/2, k}^n) \\
&\quad - \frac{\Delta t}{\Delta z} (\mathcal{E}_{x, i, j-1/2, k+1/2}^n - \mathcal{E}_{x, i, j-1/2, k-1/2}^n) \quad (21)
\end{aligned}$$

$$\begin{aligned}
326 \quad B_{z, i, j, k-1/2}^{n+1/2} &= B_{z, i-1/2, j, k}^n + \frac{\Delta t}{\Delta y} (\mathcal{E}_{x, i, j+1/2, k-1/2}^n - \mathcal{E}_{x, i, j-1/2, k-1/2}^n) \\
&\quad - \frac{\Delta t}{\Delta x} (\mathcal{E}_{y, i+1/2, j, k-1/2}^n - \mathcal{E}_{y, i-1/2, j, k-1/2}^n). \quad (22)
\end{aligned}$$

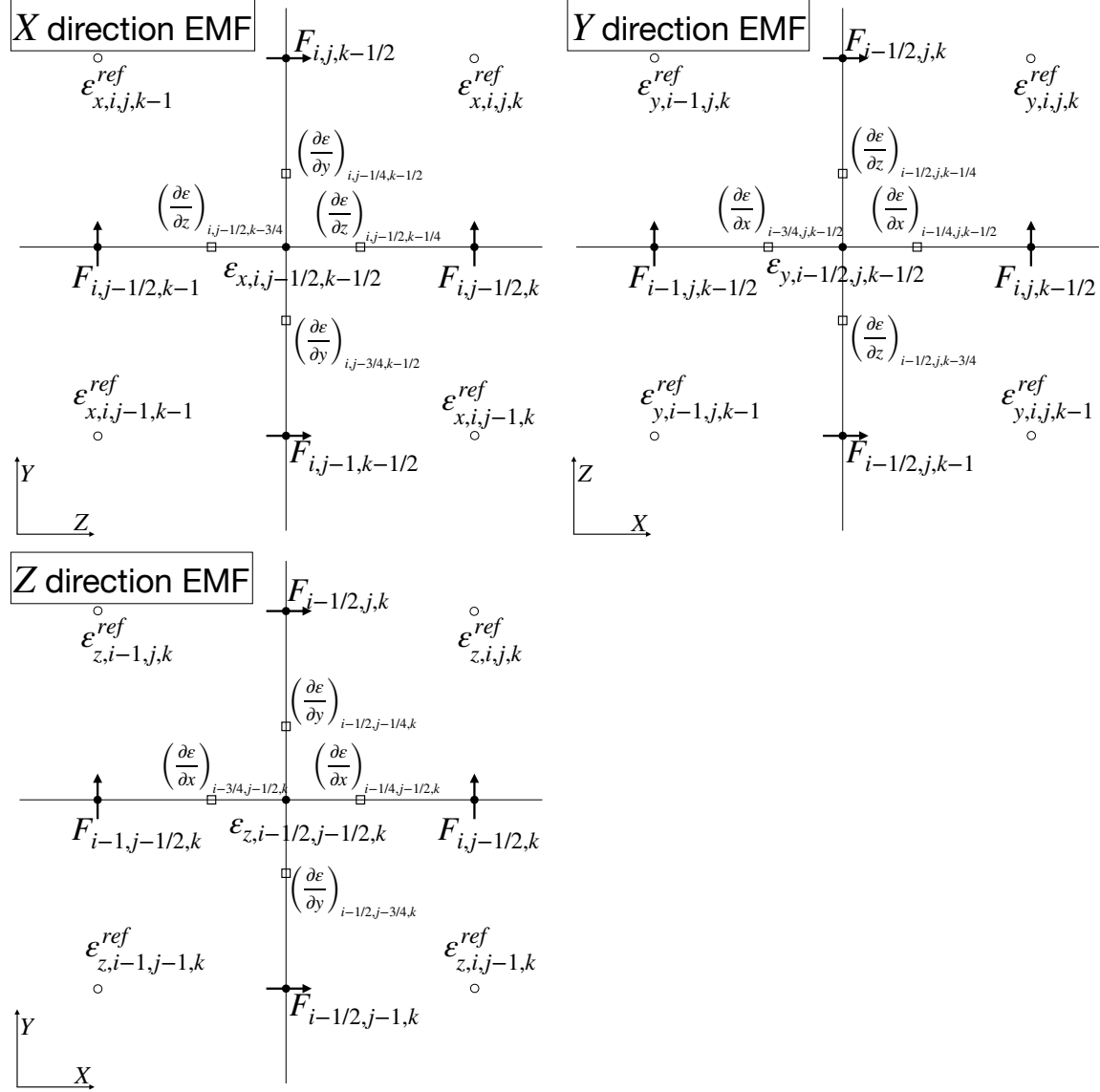


Figure 1. 2D slices in all three planes showing the location of the fluxes, edge electric fields, and derivatives. Based on Figure 5 of [Stone et al. \(2008\)](#).

2.2.6. Step 6. Half Time-step Second Order Reconstruction

Step 5 results in first order time-averaged values for the cell-centered conserved variables and face-centered magnetic fields. To make the integration second-order in time, we need to perform a “corrector” step. First, we perform a higher order interface reconstruction. The method shown here is for Piecewise Linear Method (PLM), reconstruction. Cholla currently implements piecewise constant, piecewise linear, and piecewise parabolic reconstruction, with limiting in the characteristic variables, for MHD. The piecewise parabolic method that Cholla utilizes is discussed in detail in [Felker & Stone \(2018\)](#). Using the third order piecewise parabolic method for spatial reconstruction does typically give slightly more accurate results at a given resolution compared to PLM, but since the method is formally second order it does

not improve the overall order of convergence. Note that at a given face only the transverse components of the electric field need to be reconstructed. The longitudinal component is already given at the face.

The steps of the PLM update are:

1. Compute the primitive variables from the conserved variables.
2. Compute the left, right, centered, and Van Leer differences in the primitive variables

$$\delta \mathbf{W}_{L,i} = \mathbf{W}_i - \mathbf{W}_{i-1} \quad (23)$$

$$\delta \mathbf{W}_{R,i} = \mathbf{W}_{i+1} - \mathbf{W}_i \quad (24)$$

$$\delta \mathbf{W}_{C,i} = \frac{\mathbf{W}_{i+1} - \mathbf{W}_{i-1}}{2} \quad (25)$$

$$\delta \mathbf{W}_{VL,i} = \begin{cases} \frac{2\mathbf{W}_{L,i}\mathbf{W}_{R,i}}{\mathbf{W}_{L,i} + \mathbf{W}_{R,i}}, & \text{if } \mathbf{W}_{L,i}\mathbf{W}_{R,i} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

3. Project the slopes into the characteristic variables, \mathbf{a} , using the eigenvectors listed in the appendix of [Stone et al. \(2008\)](#). Note that to maintain mathematical consistency we use the eigenvectors of the \mathbf{w}_i cell for all four slopes and the later projection back to primitive variables.
4. Apply monotonicity constraints to the characteristic differences to ensure that the reconstruction is total variation diminishing (TVD). We use the following limiter, given in [LeVeque \(2002\)](#):

$$\delta \mathbf{a}_{m,i} = \begin{cases} \text{sgn}(\mathbf{a}_{C,i}) \min(2|\mathbf{a}_{L,i}|, 2|\mathbf{a}_{R,i}|, |\mathbf{a}_{C,i}|, |\mathbf{a}_{VL,i}|), & \text{if } \mathbf{a}_{L,i}\mathbf{a}_{R,i} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (27)$$

5. Project the limited characteristic slopes, $\delta \mathbf{a}_m$, back into the primitive variables, $\delta \mathbf{W}_m$, using the eigenvectors.
6. Compute the interface states $\mathbf{W}_{L,i+1/2}$ and $\mathbf{W}_{R,i-1/2}$:

$$\begin{aligned} \mathbf{W}_{L,i+1/2} &= \mathbf{W}_i + \frac{\delta \mathbf{W}_{m,i}}{2} \\ \mathbf{W}_{R,i-1/2} &= \mathbf{W}_i - \frac{\delta \mathbf{W}_{m,i}}{2} \end{aligned} \quad (28)$$

where $\mathbf{W}_{L/R,i\pm 1/2}$ is the state on the left or right side of the cell and $\delta \mathbf{W}_{m,i}$ is the monotonically limited primitive slope. Limiting can be done in the primitive variables by skipping steps 3 and 5 and replacing the characteristic slopes in Equation 27 with their primitive counterparts.

2.2.7. Step 7. Second Riemann Solve

Solve the Riemann problem for each interface again using the higher order interface states computed in step 6.

2.2.8. Step 8. Compute the Constrained Transport Electric Fields

Repeat step 3, but using the fluxes from the second Riemann solve and the half time step MHD variables computed in Step 5.

2.2.9. Step 9. Perform the Full Time-step Update

Update the cell-centered hydro variables from $t = n$ to $t = n + \Delta t$ using the conserved update equation and the second order fluxes computed in Step 7:

$$\begin{aligned} \mathbf{U}_{i,j,k}^{n+1} = \mathbf{U}_{i,j,k}^n & - \frac{\Delta t}{\Delta x} \left(\mathbf{F}_{x,i+1/2,j,k}^{n+1/2} - \mathbf{F}_{x,i-1/2,j,k}^{n+1/2} \right) \\ & - \frac{\Delta t}{\Delta y} \left(\mathbf{F}_{y,i,j+1/2,k}^{n+1/2} - \mathbf{F}_{y,i,j-1/2,k}^{n+1/2} \right) \\ & - \frac{\Delta t}{\Delta z} \left(\mathbf{F}_{z,i,j,k+1/2}^{n+1/2} - \mathbf{F}_{z,i,j,k-1/2}^{n+1/2} \right). \end{aligned} \quad (29)$$

Update the face-centered magnetic field using the electric fields calculated in Step 8:

$$\begin{aligned} B_{x,i-1/2,j,k}^{n+1} = B_{x,i-1/2,j,k}^n & + \frac{\Delta t}{\Delta z} \left(\mathcal{E}_{y,i-1/2,j,k+1/2}^{n+1/2} - \mathcal{E}_{y,i-1/2,j,k-1/2}^{n+1/2} \right) \\ & - \frac{\Delta t}{\Delta y} \left(\mathcal{E}_{z,i-1/2,j+1/2,k}^{n+1/2} - \mathcal{E}_{z,i-1/2,j-1/2,k}^{n+1/2} \right) \end{aligned} \quad (30)$$

$$\begin{aligned} B_{y,i,j-1/2,k}^{n+1} = B_{y,i,j-1/2,k}^n & + \frac{\Delta t}{\Delta x} \left(\mathcal{E}_{z,i+1/2,j-1/2,k}^{n+1/2} - \mathcal{E}_{z,i-1/2,j-1/2,k}^{n+1/2} \right) \\ & - \frac{\Delta t}{\Delta z} \left(\mathcal{E}_{x,i,j-1/2,k+1/2}^{n+1/2} - \mathcal{E}_{x,i,j-1/2,k-1/2}^{n+1/2} \right) \end{aligned} \quad (31)$$

$$\begin{aligned} B_{z,i-1/2,j,k}^{n+1} = B_{z,i-1/2,j,k}^n & + \frac{\Delta t}{\Delta y} \left(\mathcal{E}_{x,i,j+1/2,k-1/2}^{n+1/2} - \mathcal{E}_{x,i,j-1/2,k-1/2}^{n+1/2} \right) \\ & - \frac{\Delta t}{\Delta x} \left(\mathcal{E}_{y,i+1/2,j,k-1/2}^{n+1/2} - \mathcal{E}_{y,i-1/2,j,k-1/2}^{n+1/2} \right). \end{aligned} \quad (32)$$

2.2.10. Step 10. Increment the Time by Δt

Increment the time by Δt . Additional physics modules are added in a first order operator-split fashion after this point (chemistry, radiation transport, etc.), after which all MPI communication is done to exchange the ghost/halo cells that surround each MPI rank's subdomain.

2.3. Implementation on GPUs

2.3.1. Memory bandwidth constraints

While the implementation of MHD on GPUs is similar to the implementation on CPUs, there are some crucial differences, especially regarding data handling and

movement. Compared to CPUs, GPUs have higher memory bandwidth and extremely high FLOPS, but limited memory capacity and limited functionality due to their fundamentally SIMD (Single Instruction Multiple Data) nature. Also, while GPU memory bandwidth is higher on the whole, due to their parallel nature GPUs can request many more values from memory at once, meaning GPU-based codes are often still memory bandwidth bound.

This leads to some implementation choices that may seem counter-intuitive. For example, while an optimized CPU-based code may compute the cell-centered magnetic fields once and then save them to memory, Cholla recomputes them in each function call where they are required. Similarly, Cholla does not store the primitive variables, only the conserved ones, and recomputes the primitive variables as needed. This approach reduces global memory usage by not storing an entire second grid. It also generally reduces memory bandwidth requirements, since often both the conserved and primitive variables are needed within a function, but only one set needs to be loaded.

Another potential source of memory bandwidth optimization is fusing functions together. By combining multiple GPU kernels into one, different functions can share data without a write/read cycle to global memory. For example, we have implemented this in Cholla by combining the first order reconstruction and Riemann solver kernels, since each interface that is reconstructed is only used in the next Riemann solve, and does not need to be used later in the integrator. Fusing the PCM “reconstruction” into the Riemann solver led to a $\sim 10\%$ improvement in overall runtime. However, fusing the PLMC reconstruction into the Riemann solver actually slowed the code down by $\sim 10 - 15\%$. This slowdown is caused by the increased register usage of the new, larger kernel, which in turn reduced overall GPU occupancy. Thus, the trade-off between reducing global memory accesses and increasing register usage is not always straightforward to determine, and we have found that it depends on compiler optimizations which can vary between platforms, as well.

Another major challenge in GPU computing is CPU-to-GPU bandwidth which is much lower than GPU memory bandwidth. To address this problem, Cholla now keeps as much data as possible in the GPU memory instead of moving it to and from main system memory, in contrast with previous versions of the code. Although it has always been a GPU-native code, historically Cholla used an extreme version of the “offload” model of GPU programming, keeping the “primary” data (like conserved variable arrays) in CPU memory, and copying that data to the GPU to carry out computations before copying it back each time step. Although this approach originally had the advantage of allowing larger grid sizes to be computed on a single GPU when GPU memory sizes were extremely limited, as GPUs have gotten dramatically faster and their memory capacity has increased over the last decade, these full-grid copies between CPU and GPU memory would now dominate the simulation run time. In addition, most new supercomputers, such as *Frontier*, have GPUs that are directly

connected to the network, so direct off-node GPU-to-GPU MPI communication is now possible. Between these two factors it is much more efficient to store all the simulation data on the GPU, and only move it back to the CPU for i/o operations. In the new version of Cholla described in this work, all computations are carried out on the GPU, and MPI communication proceeds directly from GPU memory buffers. The only functions that remain on the CPU are the setting of the initial conditions, which only occurs once, and reading and writing output files.⁵

2.3.2. Performance portability

Cholla is written in C++ and CUDA, the C++ extension developed by NVIDIA. However, CUDA code can only be compiled using NVIDIA’s compiler `nvcc` for NVIDIA GPUs. Thus, with the advent of AMD-based supercomputers like *Frontier*, the problem of code portability becomes relevant. We approached this problem in two ways. First, AMD has developed HIP, a cross platform equivalent to CUDA, which can be compiled using their `hipcc` to target either NVIDIA or AMD hardware. Although porting Cholla’s CUDA code to HIP did not prove challenging due to the extreme similarity between the two platforms, rewriting the entire code base in HIP has the drawback of making it impossible to compile on systems with NVIDIA GPUs that do not have HIP installed. Thus, we chose to employ a less invasive modification: we introduced a header file (`gpu.hpp`) which uses C++ preprocessor macros to convert between CUDA and HIP versions of functions at compile time. When building the code, the user merely has to specify whether the compilation should use `hipcc` or `nvcc`, and the code is compiled into HIP or CUDA accordingly. This method also has the advantage of continuing to allow us to maintain a single code-base written in CUDA.

2.3.3. GPU reductions

There are several places in Cholla where a grid wide reduction must be performed. The primary example is the calculation of the time step, which requires finding the minimum crossing time in the full simulation grid. While CPU and MPI based reductions are relatively simple to implement, often through library calls, the inherently parallel nature of the GPU programming model makes GPU based reductions somewhat more complex. For GPU reductions in Cholla, we use a method similar to that described by NVIDIA⁶ which describe the challenges of a GPU reduction well. This reduction method uses atomics for the final level of reduction. While this method is straightforward to implement in HIP, at the time of writing, CUDA does not support floating point `atomicMax`. To work around this we have adopted a method from the RAPIDS cuML library⁷ which, with some encoding, uses the integral `atomicMax`. Overall this method performs slightly faster than Cholla’s previous hybrid GPU+CPU

⁵ At the time of writing, the HDF5 library does not support writing files directly from GPU memory.

⁶ <https://developer.nvidia.com/blog/faster-parallel-reductions-kepler/>

⁷ <https://github.com/rapidsai/cuml/blob/dc14361ba11c41f7a4e1e6a3625bbadd0f52daf7/cpp/src/primstats/minmax.cuh>

reduction for the number of elements we typically encounter. Primarily though, it is much simpler to use and performs dramatically better as the number of elements increases.

3. MHD TESTS

In this Section, we show the results of a suite of test problems that demonstrate the accuracy, robustness and performance of Cholla MHD. We have included many problems that are common in the literature and specifically selected tests that are challenging for the integrator and, when possible, have quantifiable measures of correctness for comparison to other methods and codes. All of these problems are also included in our automated test suite as either accuracy and/or regression system tests (see Section 4 for details). Section 3.1 discusses tests for accuracy, while Section 3.2 discusses the performance and scaling of Cholla MHD.

3.1. Accuracy Tests

3.1.1. Linear Wave Convergence

The propagation of the four MHD linear waves provides an excellent quantitative measure of the accuracy of a numerical MHD method. Our linear wave tests use a periodic domain of $1.0 \times 1.0 \times 1.0$ and a resolution of $N \times 16 \times 16$ where N goes from 16 to 512 in powers of 2. The wave equation is

$$q = \bar{q} + AR_w \sin \frac{2\pi x}{\lambda}, \quad (33)$$

where q is the conserved variable, \bar{q} is the mean background state, $A = 10^{-6}$ is the amplitude of the wave, R_w is the right eigenvector in conserved variables for the wave mode w , x is the position, and $\lambda = 1$ is the wavelength of the wave. The adiabatic index γ is 5/3 and the background state is: $\bar{\rho} = 1.0$, $\bar{v}_x = \bar{v}_y = \bar{v}_z = 0$ (except for the contact wave where $\bar{v}_x = 1$), $\bar{P} = 1/\gamma$, $\bar{B}_x = 1$, $\bar{B}_y = 1.5$, and $\bar{B}_z = 0$. The right eigenvectors for this state are given in Appendix A of Gardiner & Stone (2008).

The wave propagates for one period, after which the error is computed between the initial and final state. First we compute the L1 norm, which is the absolute difference for each conserved variable between the initial and final state:

$$\delta q_s = \frac{1}{n_x n_y n_z} \sum_{i,j,k} |q_{i,j,k,s}^f - q_{i,j,k,s}^i|, \quad (34)$$

where q_s is a specific conserved variable. We then compute the L2 norm of this vector of L1 norms as

$$\|\delta q\| = \sqrt{\sum_s (\delta q_s)^2}. \quad (35)$$

These L2 errors are plotted in Figure 2 for both the PLM and PPM reconstructions. The results are comparable to the results in Stone & Gardiner (2009) and

demonstrate the expected second order convergence. Using PPM improves the accuracy of the solution by approximately an order of magnitude at any given resolution, but maintains the second order convergence due to the second order nature of the integrator. We have implemented these tests in all three directions with the waves moving in the positive or negative directions and find identical results.

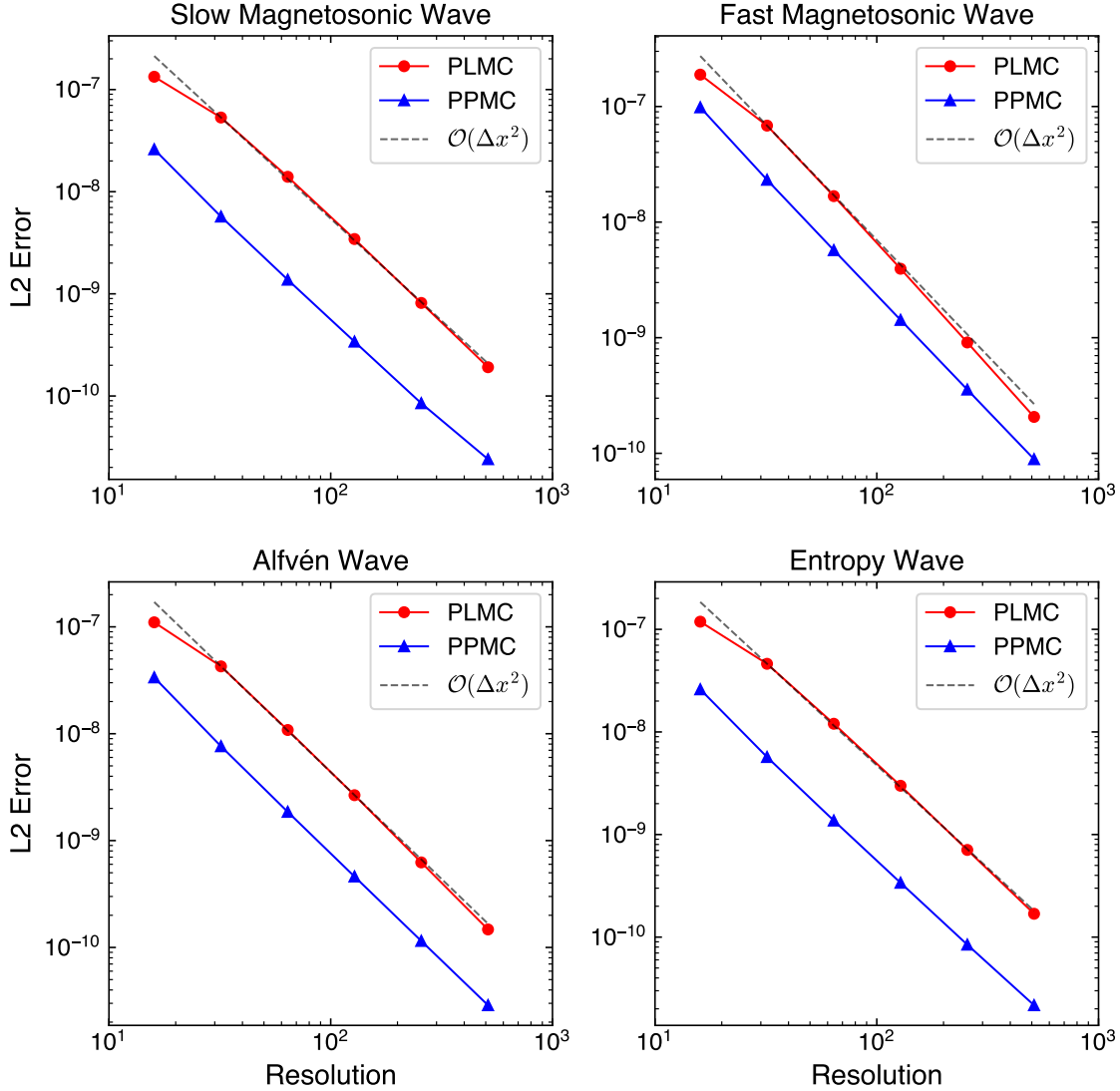


Figure 2. Linear Wave Convergence of all four MHD waves using PLM and PPM reconstruction. [arXiv](#)

3.1.2. Circularly Polarized Alfvén Wave

The circularly polarized Alfvén wave is a non-linear wave that tests a code’s accuracy in the non-linear regime with the quantitative benefits of a regular wave test (Tóth 1996). The tests use a periodic domain of $3.0 \times 1.5 \times 1.5$ and a resolution of $2N \times N \times N$

where N goes from 8 to 256 in powers of 2. The wave is initialized at an oblique angle the grid, making this a fully 3D test.

In a coordinate system aligned with the movement of the wave, the initial conditions are $\rho = 1.0$, $P = 0.1$, $v_x = (0, -1)$ for traveling or standing waves respectively, $v_y = A \sin \frac{2\pi x}{\lambda}$, $v_z = A \cos \frac{2\pi x}{\lambda}$, $B_x = 1.0$, $B_y = A \sin \frac{2\pi x}{\lambda}$, and $B_z = A \cos \frac{2\pi x}{\lambda}$, where the amplitude of the wave $A = 0.1$ and the wavelength $\lambda = 1.0$. These coordinates are then transformed with the rotation

$$\begin{aligned} x' &= x \cos \alpha \cos \beta - y \sin \beta - z \sin \alpha \cos \beta \\ y' &= x \cos \alpha \sin \beta + y \cos \beta - z \sin \alpha \sin \beta \\ z' &= x \sin \alpha + z \cos \alpha \end{aligned}$$

with $\sin \alpha = 2/3$ and $\sin \beta = 1/\sqrt{5}$. This ensures the domain is fully periodic through the boundaries and the wave can travel (or stand) indefinitely. The magnetic fields are initialized with the vector potential to ensure initial divergence is zero to round off. The waves are then run for a single period and the L2 norm of the L1 error vector is plotted in Figure 3 using the same method as in Section 3.1.1. It is interesting to note that the accuracy improvement of PPM versus PLM seen for the linear waves is absent in this non-linear test.

These Alfvén waves are subject to a parametric instability (Del Zanna et al. 2001), which should not be present for these initial conditions. However, the truncation error will result in small variations in the magnetic pressure which drives low amplitude compression waves (Stone et al. 2008).

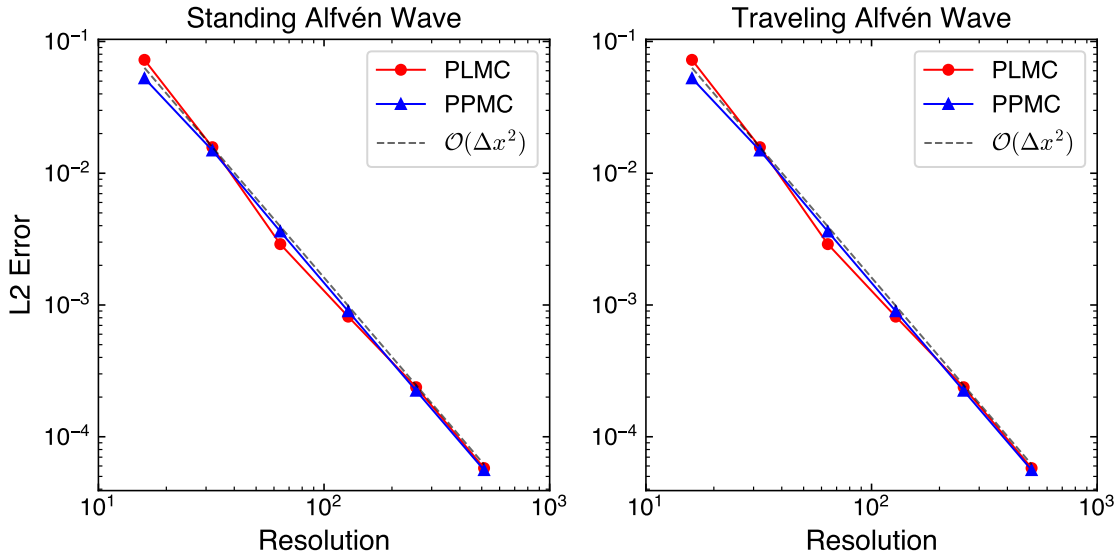


Figure 3. Circularly Polarized Alfvén Wave Convergence using PLM and PPM reconstruction. 

3.1.3. *Advecting Field Loop*

The advecting field loop test consists of a tilted spherical current loop which travels across the domain at an oblique angle to the grid. This test requires particularly accurate balancing of the non-zero components of the induction equation. It also has zero magnetic field outside the spherical current loop; as the current loops moves across the grid those cells that are no longer in the loop should return to zero to within round off error. It is also a good test of the dissipation of the magnetic field, as the magnetic pressure should remain constant.

The initial conditions for this test are most easily described using the magnetic vector potential, which is the vector quantity whose curl equals the magnetic field, $\mathbf{B} = \nabla \times \mathbf{A}$. The background state is $\rho = 1.0$, $P = 1.0$, $v_x = 1.0$, $v_y = 1.0$, $v_z = 2.0$, $B_x = 0$, $B_y = 0$, and $B_z = 0$.

In the central region the state is given by the following vector potential, which we have chosen such that $A_x = 0$:

$$A_y = A_z = \begin{cases} A(R - r), & \text{for } r < R \\ 0, & \text{otherwise} \end{cases} \quad (36)$$

where r is the Euclidean distance from the center of the domain, $R = 0.3$, and the amplitude $A = 10^{-3}$. Note that since the vector potential is along the vertices of the cells, A_y and A_z will never have the same value at the same position as they are not stored at identical positions. The test is conducted on a grid of $N \times N \times 2N$ cells for $N = (32, 64, 128, 256)$ with a periodic domain of $1.0 \times 1.0 \times 2.0$ centered at zero and evolved for two periods; $t_{max} = 2.0$.

Figure 4 shows the mean of cell centered B^2 , normalized to the initial value, in order to demonstrate the convergence of the dissipation rate. The dissipation rate is comparable to those found in the literature (Stone et al. 2008) and improves at approximately first order. Figure 4 also shows the maximum divergence in the domain as a function of time. Throughout the entire evolution it remains near round off and, after an initial rise, remains fairly constant. The zero magnetic field region outside of the current loop also remains near zero throughout the entire evolution of the problem. Figure 5 shows cross sections of the loop initial conditions and after one period with a resolution of $128 \times 128 \times 256$ cells. The shape is well maintained with minimal dissipation.

3.1.4. *MHD Riemann Problems*

MHD Riemann problems are standard tests for new MHD codes and methods due to their mix of different flow types and extreme conditions. There are many different MHD Riemann problems in the literature (Brio & Wu 1988; Einfeldt et al. 1991; Ryu & Jones 1995; Dai & Woodward 1998); we have chosen to present five that cover a variety of challenging cases.

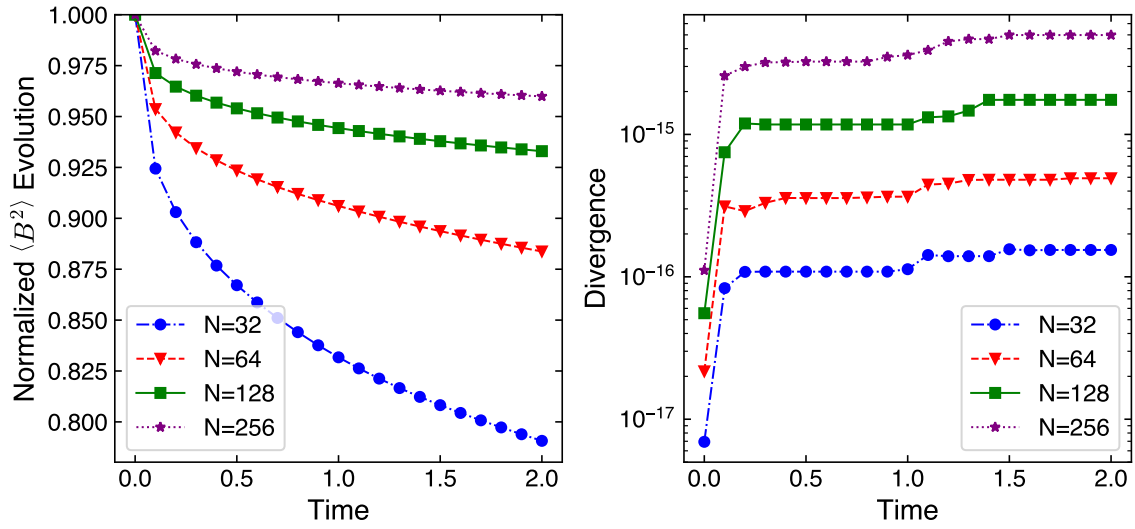


Figure 4. Evolution of tilted spherical magnetic field loop through two full periods using PPMC reconstruction

. Mean of B^2 normalized to the initial value as a function of time (left) and the maximum divergence in the domain as a function of time (right). [zenodo](#)

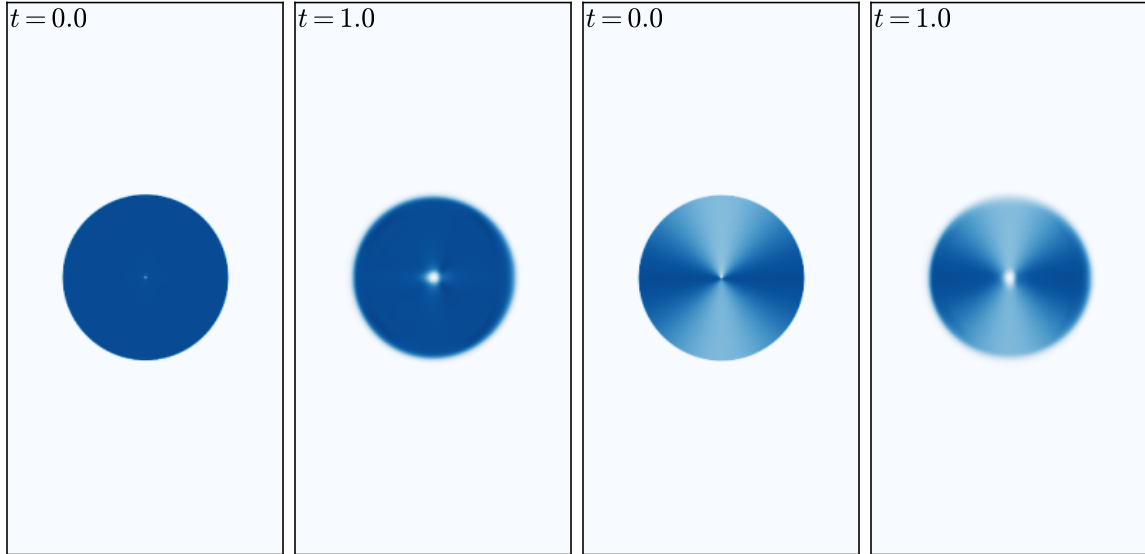


Figure 5. Cross sections of the spherical advecting field loop magnetic energy density at $t = 0.0$ and one period. The first and second panels show a slice centered on the loop through the plane of symmetry. The third and fourth panels show a slice along the $x - z$ plane. Note that these figures utilize PLMC reconstruction as PPMC introduced spurious oscillations in the direction of advection. [zenodo](#)

The typical Riemann problem setup uses a domain with a discontinuity at the midpoint between two different states. As the problem evolves in time, the waves propagate along characteristics such that the solution evolves self-similarly in time. All of Riemann problems shown in this section employ a domain of $1 \times 1 \times 1$ and reso-

Riemann Problem	γ	t_{max}	B_x	ρ_L	P_L	$v_{x,L}$	$v_{y,L}$	$v_{z,L}$	$B_{y,L}$	$B_{z,L}$
Brio & Wu	2	0.1	0.75	1	1	0	0	0	1	0
Dai & Woodward	$\frac{5}{3}$	0.2	$\frac{2}{\sqrt{4\pi}}$	1.08	0.95	1.2	0.01	0.5	$\frac{3.6}{\sqrt{4\pi}}$	$\frac{2}{\sqrt{4\pi}}$
Ryu & Jones 1a	$\frac{5}{3}$	0.08	$\frac{5}{\sqrt{4\pi}}$	1	20	10	0	0	$\frac{5}{\sqrt{4\pi}}$	0
Ryu & Jones 4d	$\frac{5}{3}$	0.16	0.7	1	1	0	0	0	0	0
Einfeldt Rarefaction	1.4	0.16	0	1	0.45	-2	0	0	0.5	0

Riemann Problem	ρ_R	P_R	$v_{x,R}$	$v_{y,R}$	$v_{z,R}$	$B_{y,R}$	$B_{z,R}$
Brio & Wu	0.125	0.1	0	0	0	-1	0
Dai & Woodward	1	1	0	0	0	$\frac{4}{\sqrt{4\pi}}$	$\frac{2}{\sqrt{4\pi}}$
Ryu & Jones 1a	1	1	-10	0	0	$\frac{5}{\sqrt{4\pi}}$	0
Ryu & Jones 4d	0.3	0.2	0	0	1	1	0
Einfeldt Rarefaction	1	0.45	2	0	0	0.5	0

Table 2. Riemann Problem Initial Conditions.

The L/R subscripts indicate that it is the left/right state. B_x is always the same in both states.

lution of $512 \times 16 \times 16$ and are run until the t_{max} which is specified for that particular problem. We use parabolic reconstruction with limiting in the characteristic variables unless otherwise noted. All have been run in all three spatial directions with both possible orientations of the two states and achieved identical results. The details of each left and right state are given in Table 2.

Brio & Wu Shock Tube—Figure 6 shows the Brio & Wu Shock Tube (Brio & Wu 1988) which is a staple test of MHD codes. This Riemann problem is essentially the Sod shock tube (Sod 1978) with a magnetic field. However, this shock tube is an excellent stress test for PPM reconstruction, as methods higher than second order tend to create large oscillations in the solution due to the slowly moving shock waves. We have implemented both the PPM reconstruction algorithm used in the VL+CT integrator from the method described in Stone et al. (2008) as well as the method described in Felker & Stone (2018), and find that the latter significantly reduces oscillations for this test; Figure 6 uses the method from Felker & Stone (2018), which is the version currently available in Cholla. With PPM reconstruction, we find oscillations in the solution when limiting in either the primitive or characteristic variables. No oscillations are present when using PLM reconstruction.

Dai & Woodward Shock Tube—Figure 7 shows the Dai & Woodward Shock Tube (also called Ryu & Jones 2a) (Dai & Woodward 1998; Ryu & Jones 1995) which produces all seven possible MHD waves. From left to right they are: fast shock, Alfvén wave, slow shock, contact discontinuity, slow shock, Alfvén wave, and fast shock. This makes it an excellent laboratory for checking that the full spread of wave modes are well resolved.

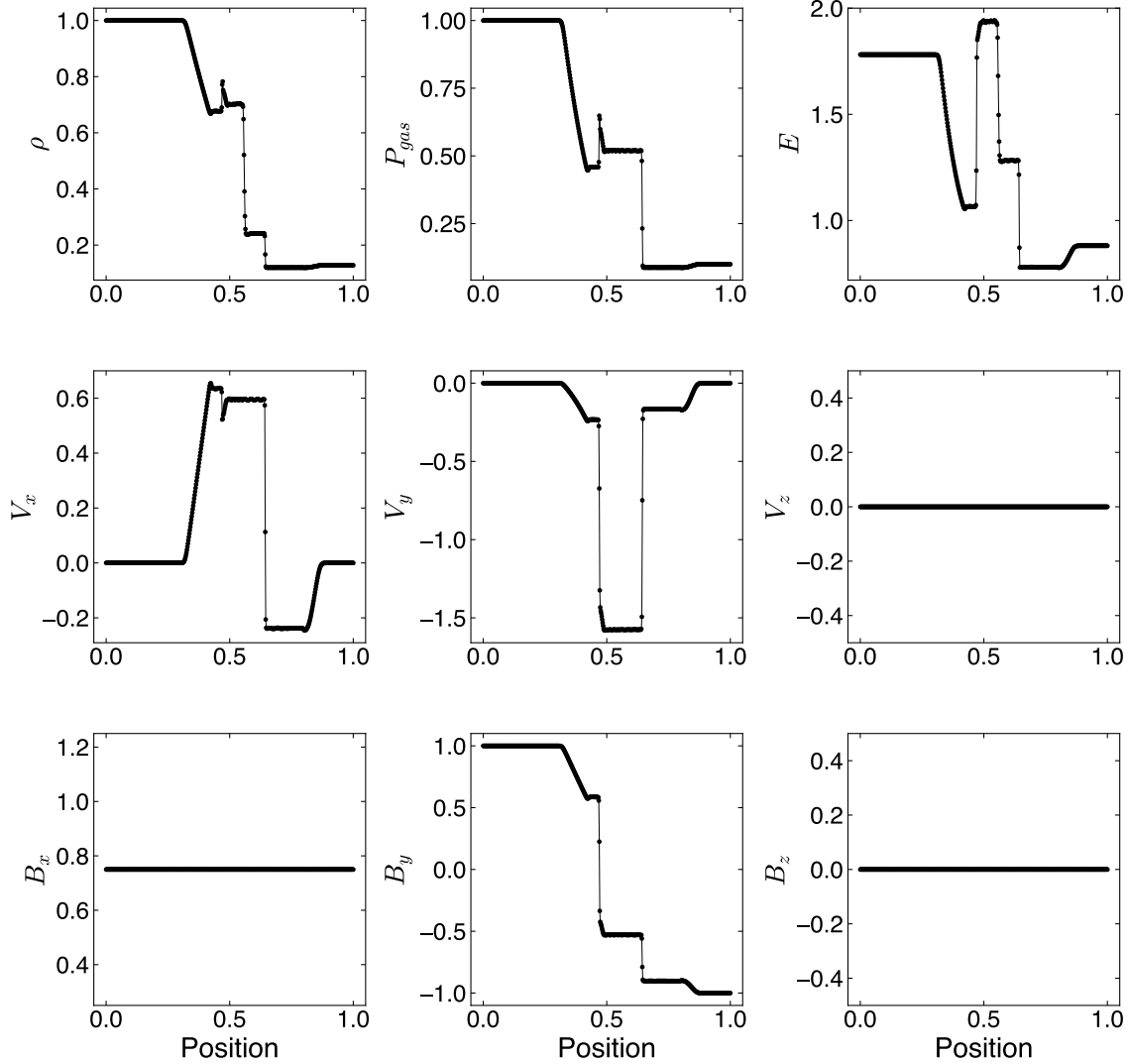


Figure 6. The Brio & Wu Shock Tube solution (Brio & Wu 1988). [arXiv](#)

Ryu & Jones 1a Shock Tube—Figure 8 shows the Ryu & Jones 1a Shock Tube solution (Ryu & Jones 1995) which is a less common test for MHD codes. However, in our experience, it is an excellent problem for debugging due to its relatively simple structure, which is easy to examine manually. The lack of any spikes and the presence of multiple types of strong shocks also make it a good diagnostic test for over/undershoot of the solution near discontinuities.

Ryu & Jones 4d Shock Tube—Figure 9 shows the Ryu & Jones 4d Shock Tube solution (Ryu & Jones 1995) which features a switch-on slow shock. Switch-on waves increase the strength of the transverse magnetic field while reducing the thermal pressure to maintain energy conservation. This is a simplified example of a type of magnetic field

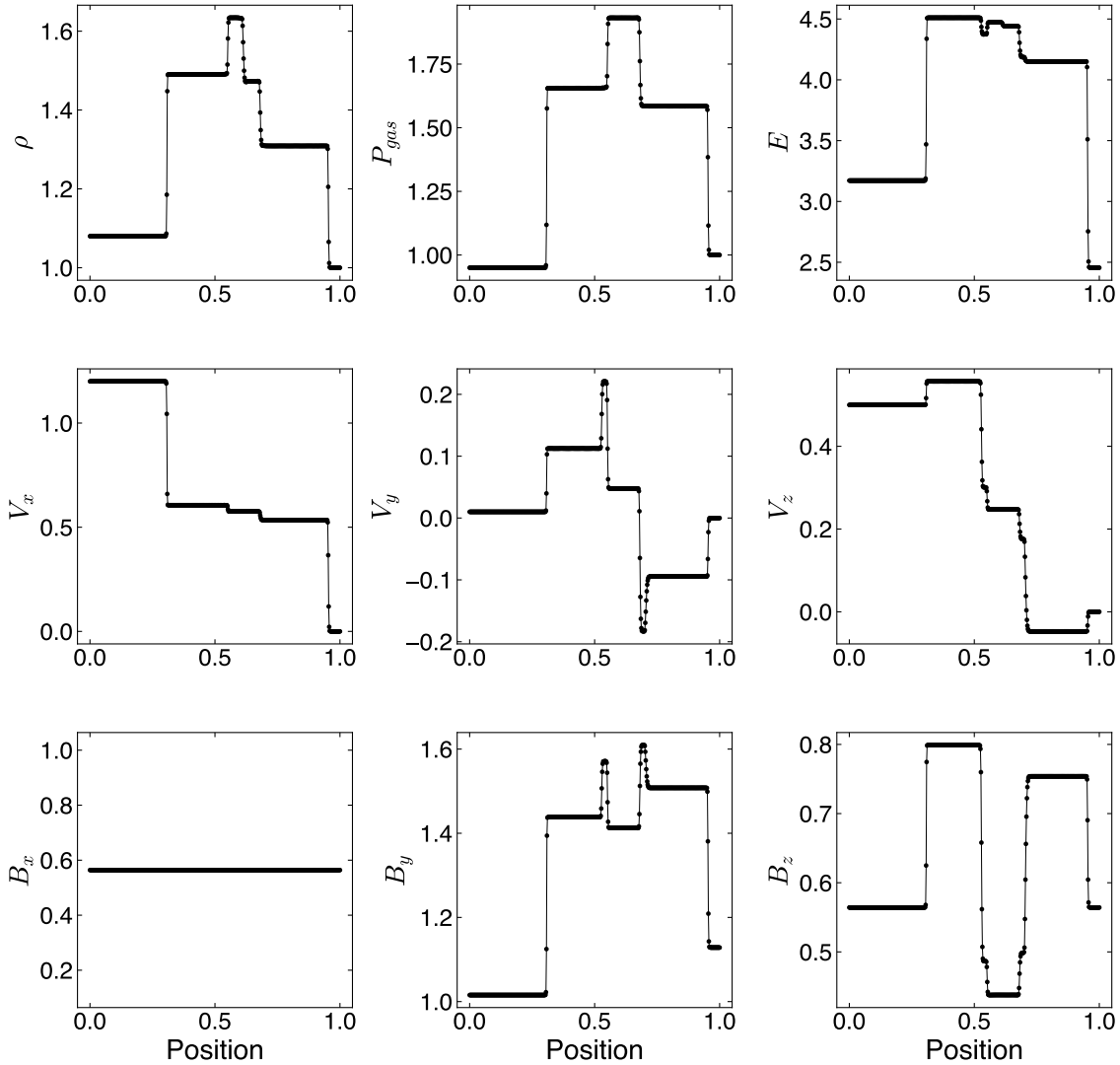



Figure 7. Dai & Woodward Shock Tube (also called Ryu & Jones 2a) solution (Dai & Woodward 1998; Ryu & Jones 1995). 

amplification and as such it is important to demonstrate that a code can replicate it accurately. A switch-off wave does the inverse.

MHD Einfeldt Strong Rarefaction—Figure 10 shows the MHD Einfeldt Strong Rarefaction test (Einfeldt et al. 1991) which creates a strong outflow and central vacuum state. The diverging solution leads to an extremely strong and fast rarefaction where the energy is dominated by kinetic energy and as such can often reveal challenges for finite-volume methods with near-vacuum states, since some Riemann solvers will return unphysical solutions with negative density or negative internal energy. High values of the outflow velocity ($V_{out} \geq 3$) can also lead to spurious oscillations in the solution. $V_{out} = 2$ was chosen for this test (Miniati & Martin 2011). Cholla performs well on this test with no spurious oscillations or unphysical negative values.

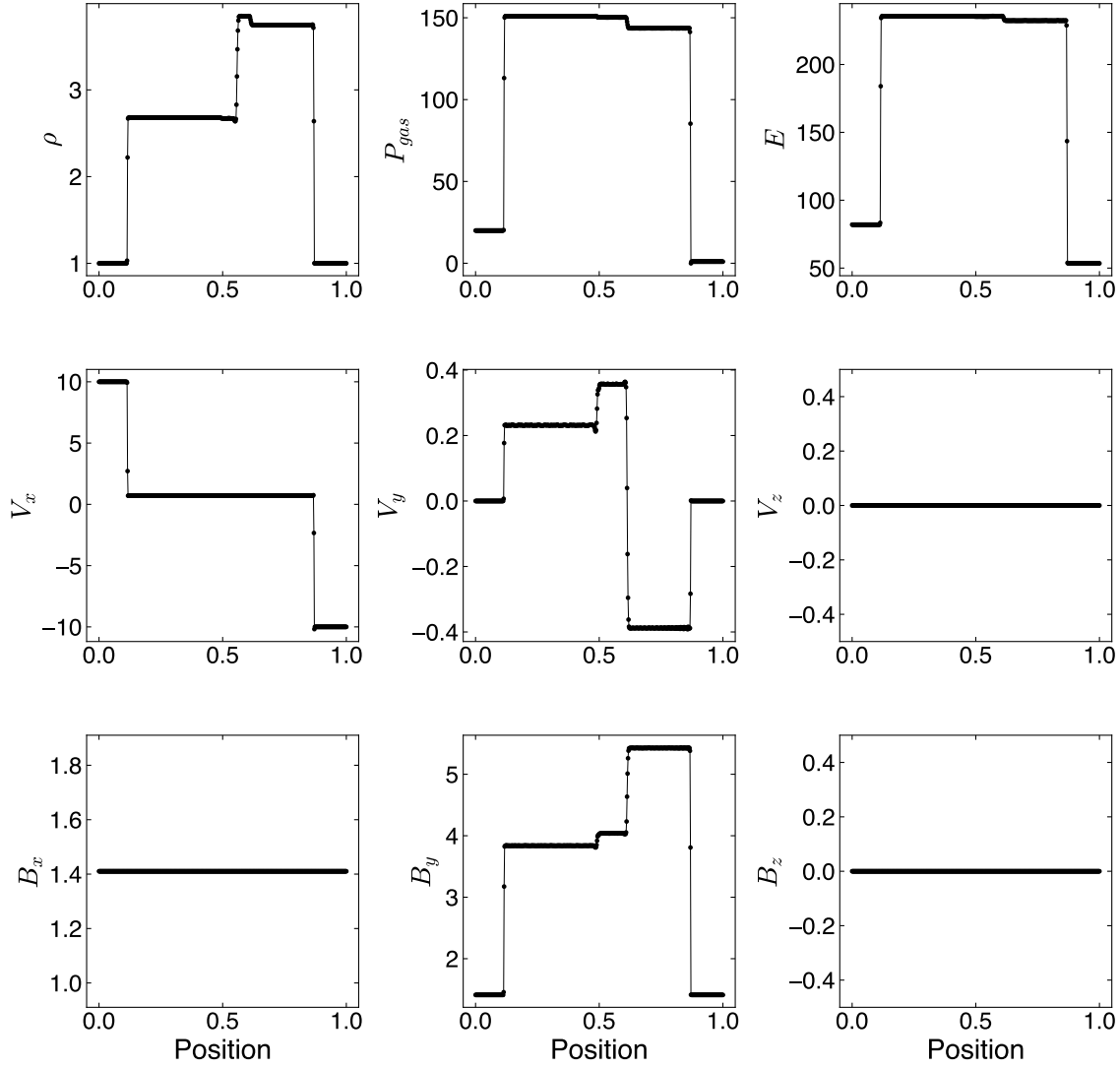



Figure 8. Ryu & Jones 1a Shock Tube solution (Ryu & Jones 1995). 

3.1.5. MHD Blast Wave in a Strongly Magnetized Medium

Blast waves in different forms are excellent tests for hydrodynamics and MHD codes. They combine strong shocked flows, smooth flows, and, in MHD, strong magnetic fields. The results are qualitative rather than quantitative, but thoroughly test the robustness of the algorithm and are excellent regression tests for automated testing (see Section 4). For this test we use $\beta = 0.2$; like Stone & Gardiner (2009), we find instabilities if β is decreased by a factor of 10.

The background state is $\rho = 1.0$, $P = 0.1$, $v_x = 0.0$, $v_y = 0.0$, $v_z = 0.0$, $B_x = 1/\sqrt{2}$, $B_y = 1/\sqrt{2}$, $B_z = 0.0$, and the over pressure region is a central sphere of size $R = 0.1$ which has $P = 10.0$. The test is then run on a domain of $1 \times 1.5 \times 1$ with a resolution of $200 \times 300 \times 200$ cells until $t = 0.2$. Figure 11 shows contours of the density

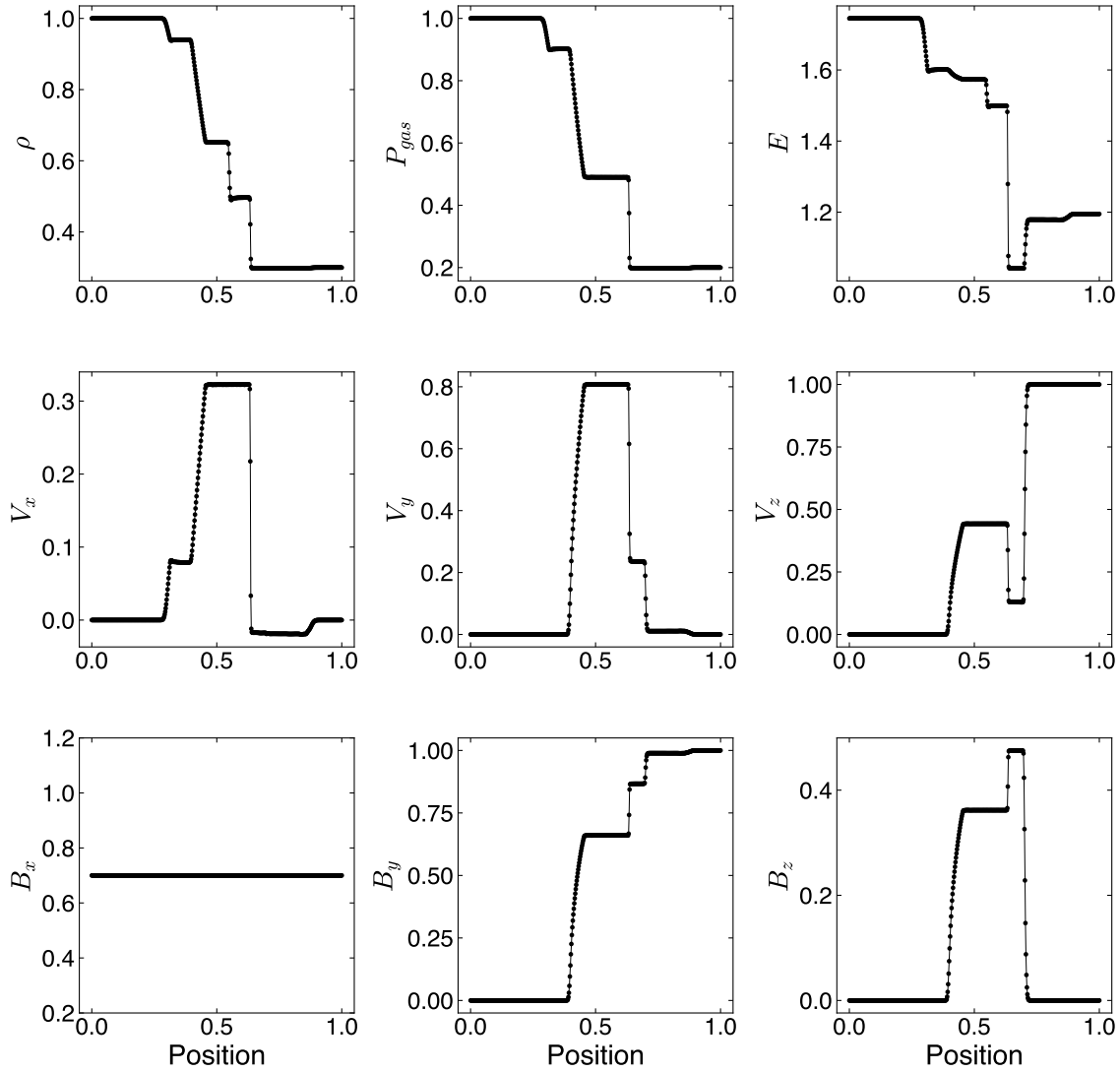



Figure 9. Ryu & Jones 4d Shock Tube solution (Ryu & Jones 1995). 

and magnetic energy fields in an $x - y$ slice through the center of the domain. The contours are smooth and symmetric and show clear elongation of the blast wave rarefaction parallel to the magnetic field. The blast wave propagates slowly parallel to the magnetic field but much more rapidly perpendicular to the magnetic field.

3.1.6. Orszag-Tang Vortex

The Orszag-Tang vortex is a standard 2D MHD test from Orszag & Tang (1979). While it does not provide a quantitative measure of accuracy like the linear wave tests or a test of the robustness of the method like the MHD blast wave, it does have a very complex flow that is sensitive to changes in the integrator, making it ideal for regression testing.

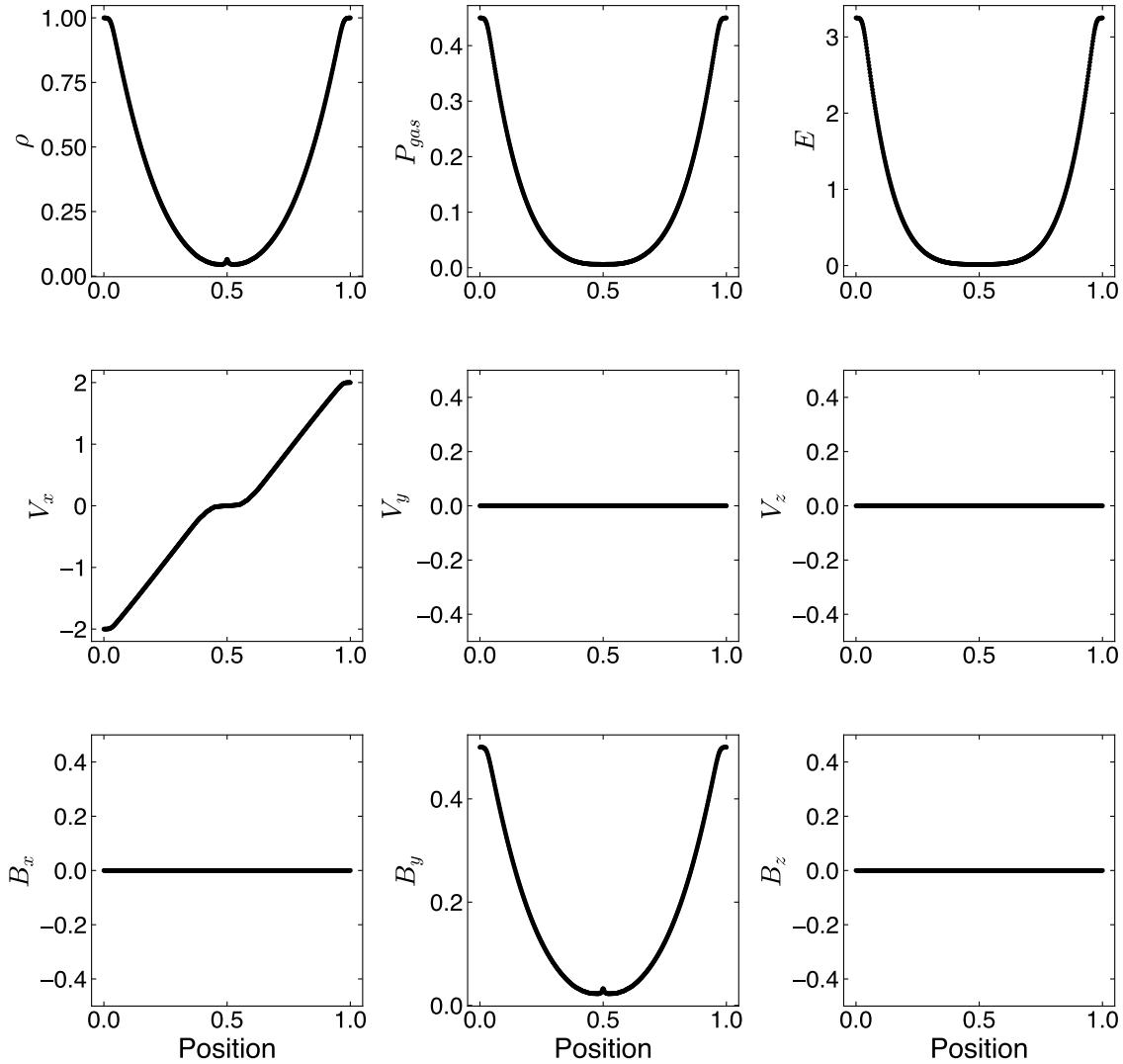



Figure 10. MHD Einfeldt Strong Rarefaction solution (Einfeldt et al. 1991). 

The test was conducted on a periodic domain of $1 \times 1 \times 1$ with a resolution of $192 \times 192 \times 192$ cells until $t = 0.5$ with the following initial conditions: $\rho = 25/(36\pi)$, $P = 5/(12\pi)$, $v_x = \sin 2\pi y$, $v_y = -\sin 2\pi x$, $v_z = 0.0$, $A_x = 0.0$, $A_y = 0.0$, $A_z = (B_0/4\pi)(\cos 4\pi x + 2\cos 2\pi y)$, with $B_0 = 1\sqrt{4\pi}$. The results, plotted in Figure 12, can be compared directly to Figure 22 in Stone et al. (2008) as a qualitative check for correctness of the flow structure.

3.2. MHD Performance Tests

Given that Cholla is a massively parallel code, its scaling properties warrant discussion. Our primary focus is on weak scaling rather than strong scaling, since good weak scaling enables much larger problems to be simulated, while strong scaling can lead to reducing the number of cells per GPU to the point where the whole GPU

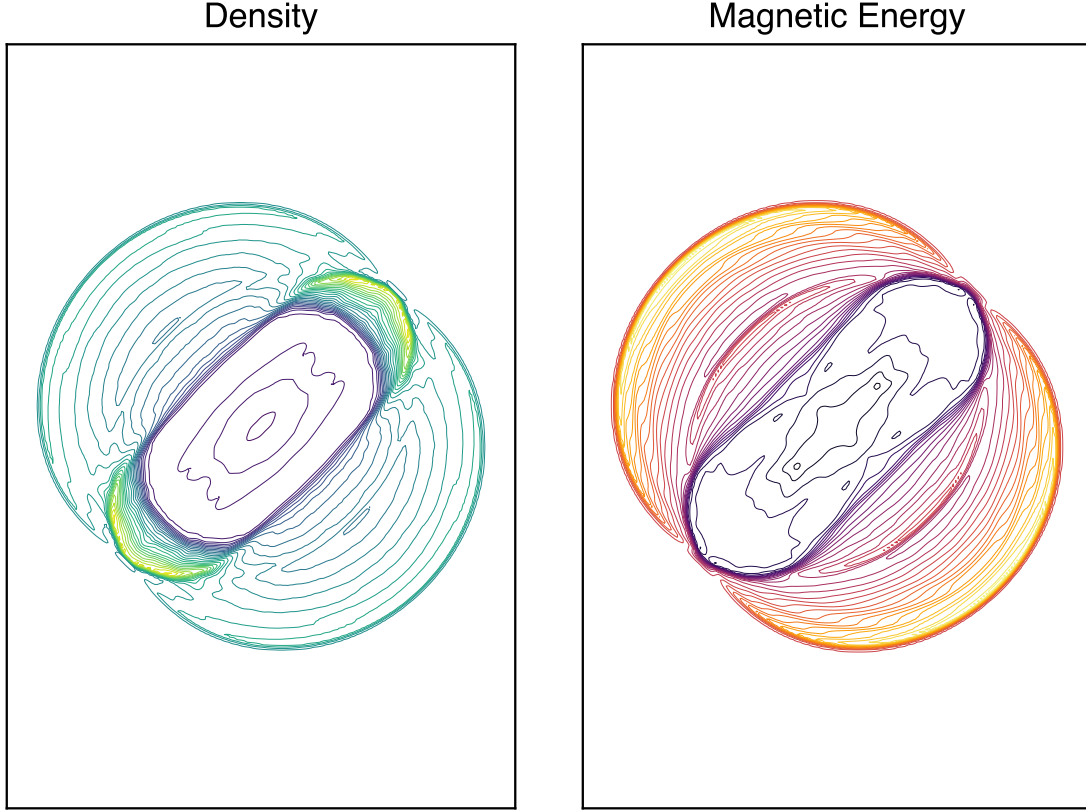


Figure 11. Contour plot of the MHD blast wave test at $t = 0.2$. 30 evenly spaced contours are shown in an $x - y$ slice through the center of the domain. [arXiv](#)

cannot be utilized, which will significantly impact performance. Results of our weak scaling tests are shown in Figure 13, while strong scaling is shown in Figure 14.

All of the weak scaling tests shown were performed with a slow magnetosonic wave perturbation (described in Section 3.1.1), periodic boundary conditions, in double precision, and with 459^3 cells per MPI rank; each rank is assigned one GPU. We employ the second order piecewise linear reconstruction method with limiting in the characteristic variables. The wave is evolved for 100 time steps (a wall-clock time of ~ 45 seconds) and the resulting time per step is averaged over the total number of time steps, excluding setup and tear down time. Strong scaling tests are performed with the same problem, though the wave is run through one full period, ~ 6100 time steps to ensure that it runs for a non negligible wall-clock time when large numbers of GPUs are utilized. The strong scaling test uses 459^3 cells, the most we can fit on a single MI250X GCD.

Our scaling tests were performed on the *Frontier* Supercomputer at the Oak Ridge Leadership Computing Facility. *Frontier* utilizes AMD MI250X GPUs, each of which contains two Graphics Compute Dies (GCDs) that largely function as separate GPUs and can be treated as such in software. Thus, for the sake of clear comparison to other systems, we will refer to each GCD as a single GPU for the remainder of this paper.

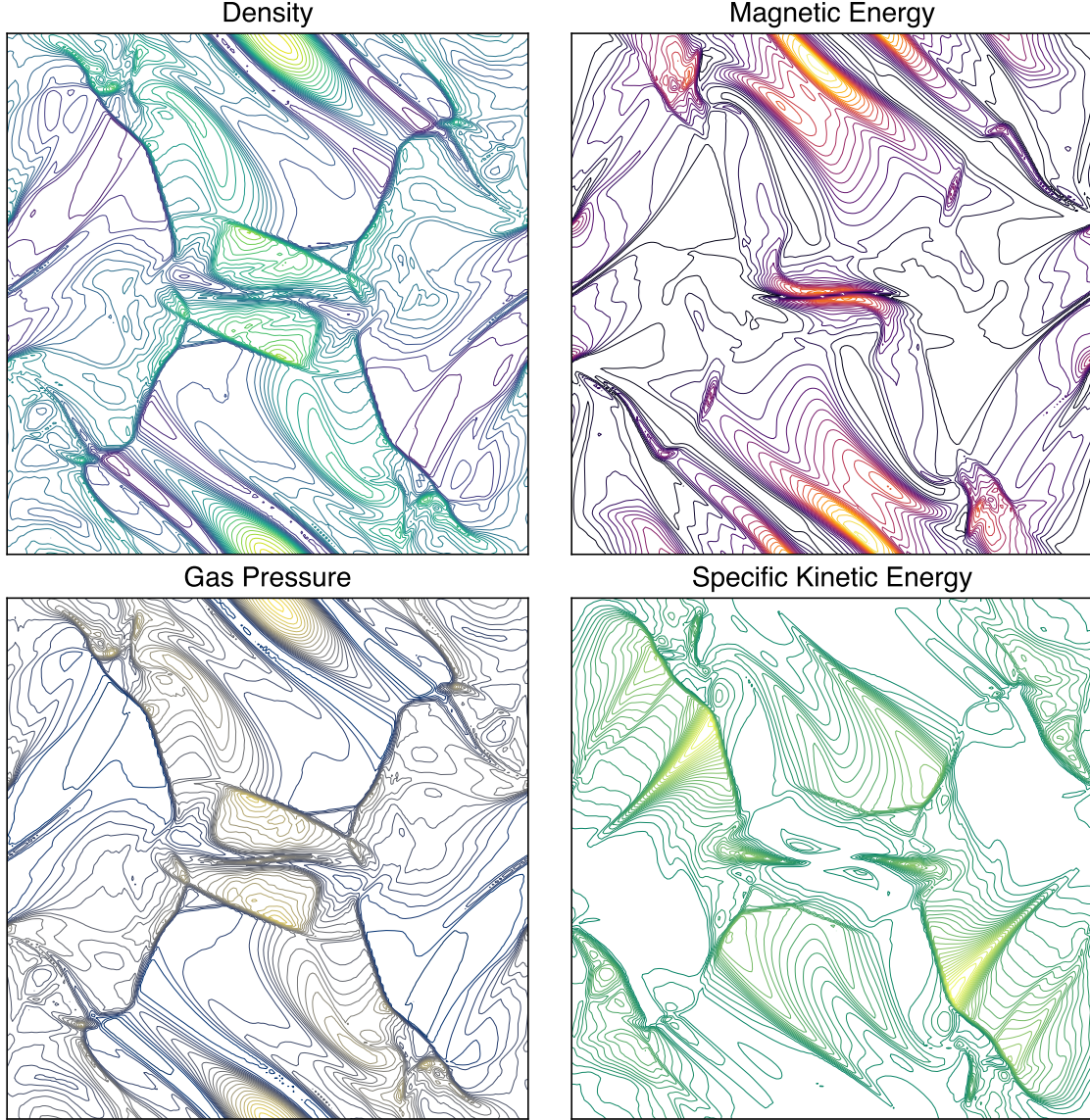


Figure 12. Contour plot of the Orszag-Tang Vortex at $t = 0.5$. Thirty evenly spaced contours are shown for each plot in an $x - y$ slice through the center of the domain. 2500

On *Frontier* Cholla updates 2.36×10^8 cells per second per GPU when running with a single GPU. The single-GPU performance is comparable on NVIDIA hardware: an NVIDIA V100 GPU achieved 160 million cell updates per second, and an A100 achieved 259 million cell updates per second. On 74,088 GPUs, nearly the entirety of *Frontier*, Cholla performs 1.89×10^8 cell updates per second per GPU, with a weak scaling efficiency of 82.2%. The 74,088 GPU run updated a total of 1.40×10^{13} cells per second on a total grid size of $19,278^3$ cells. This performance is comparable to other similar optimized GPU codes (see e.g. Figures 9 and 10 in Grete et al. (2023)).⁸

⁸ Note that the weak scaling plots in Grete et al. (2023) are normalized to single node performance, not single GPU performance, and they report performance for a 2nd order hydrodynamic solver, not MHD.

With a single MPI rank there is negligible communication overhead as no halo cells need to be exchanged between GPUs; the only boundary update required is to copy halo cell values from one location in GPU memory to another. This update is very fast, and accounts for $\sim 0.6\%$ of total runtime on a single rank. As the number of ranks grows the MPI overhead quickly stabilizes at around 15ms with a moderate increase when running on the full size of *Frontier*. Perhaps most importantly, the VL+CT integrator scales almost perfectly and takes up most of the time on each time step, dominating over the MPI communication time by nearly a factor of 10. The very slightly imperfect weak scaling of the integrator, which has no MPI communication, is due to GPU-to-GPU variance; as the number of GPUs increases the standard deviation in runtime for a single time step increases. Since all processes must wait for the slowest one to complete before beginning the next boundary exchange, this naturally leads to a few percent decrease in efficiency on these scales. Overall these tests demonstrate that Cholla has excellent weak scaling up to the full size of *Frontier*, over 74,000 GPUs.

Strong scaling is also an important metric in determining the optimally sized problem per GPU. Cholla's strong scaling performance (Figure 14) is close to ideal up to 32 GPUs (80% strong scaling efficiency) and does not drop below 50% until 256 GPUs. Strong scaling plateaus at around 512 GPUs, where the problem size per GPU is 58^3 cells. Running with a typical number of cells ($128^3 - 256^3$) to balance wall-time versus efficiency, we suffer at most a $\approx 55\%$ drop in efficiency. (In this strong scaling plot the closest problem size to 128^3 is the 64 GPU test, which has 115^3 cells per GPU and a strong scaling efficiency of 67.2%)

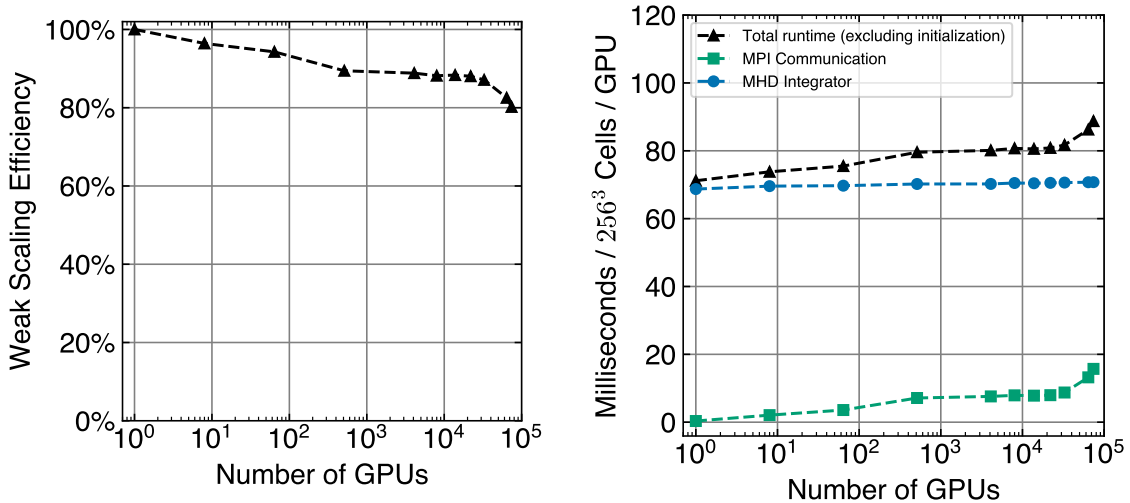



Figure 13. Weak scaling performance of Cholla MHD. When running on a single GPU Cholla updates 2.04×10^8 cells per second per GPU; the largest run with 74,088 GPUs updates 1.67×10^8 cells per second per GPU, a weak scaling efficiency of 82.2%. The 74,088 GPU run updates a total of 1.24×10^{13} cells per second. 

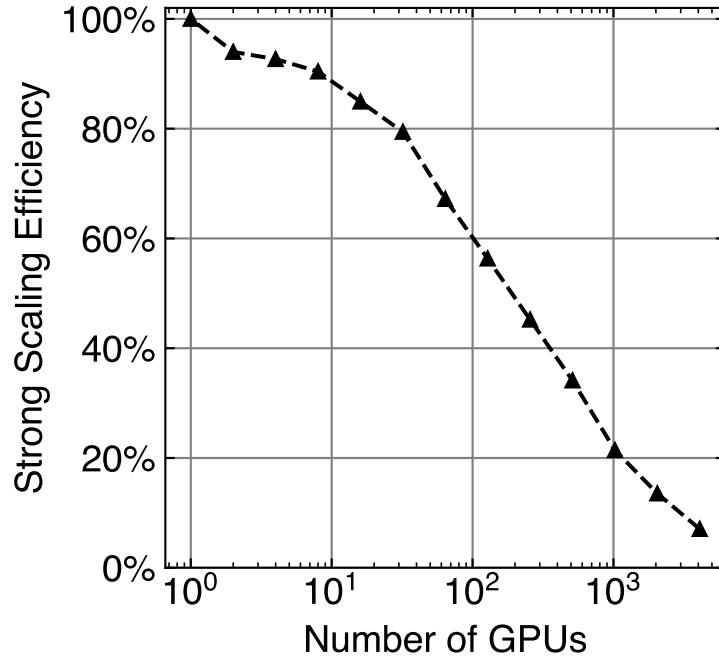


Figure 14. Strong scaling performance of Cholla MHD with a problem size of 459^3 cells. Note that a typical number of cells per GPU is $> 128^3$ which corresponds to 64 GPUs in this test. [zenodo](#)

4. AUTOMATED TESTING & CONTINUOUS INTEGRATION

As Cholla has continued to grow in complexity, and with the continued addition of large new physics modules like MHD that require changes to much of the code base, the need for a more robust, automated testing system has become increasingly apparent. Several challenges exist in implementing such a system for Cholla - not only must the testing system accommodate GPU hardware, but it must also work on large parallel systems like *Frontier*. We have addressed this need in three primary steps: 1) Choosing a testing framework for unit tests, 2) Writing the software tools needed to extend that testing framework for Cholla’s needs, and 3) adding automated testing as part of a continuous integration (CI) pipeline that runs whenever anyone submits a pull request to the Cholla GitHub repository. This Section describes the implementation of our automated testing and continuous integration framework.

Taken together, the ability to add tests for any part of the code and their automated running on every pull request has meant that new features are faster and easier to add to Cholla with confidence. Errors or changes that may otherwise break older code are often caught by the tests before they ever make it into the code base.

4.1. Unit Testing Framework

Three main kinds of tests are needed for scientific code bases: unit tests, integration tests, and system tests (also called “end-to-end” tests). Unit tests check a single

“unit” of code, i.e. a single function, class method, data structure, etc. Integration tests check how units of code operate together; for example, a test of the HLLD solver is an integration test because the solver internally calls many different functions and uses multiple different data structures. System tests check the entire code base, often with a simple test problem like a Riemann problem or linear wave, and verify that the entire program produces the correct output.

We chose GoogleTest⁹ for the unit testing framework due to its large number of features, general popularity, and relative ease of use. Another primary requirement was a testing framework that supports “death tests”, i.e. tests that check if the internals of the test crash/segfault/etc. Since Cholla, like many high performance computing (HPC) codes, handles errors by reporting those errors and then exiting, a testing framework that can handle code crashes without crashing the tests as well is critical. GoogleTest is available already built on many HPC systems and, if it is not available, can be built as an optional part of the test running script.

4.2. Extensions for Cholla

Two primary extensions were required for fully testing Cholla: a robust method for comparing floating point numbers and a way to run system tests.

4.2.1. Floating Point Comparisons

In order to run either unit tests or system tests, the code must have a robust method for comparing floating point numbers for equality. Comparing floating point numbers for equality is notoriously challenging and generally the exact comparison method that should be used varies by application (Goldberg 1991; Muller et al. 2018)¹⁰. Absolute comparisons ($|a - b| < X$) work well for small numbers, but with larger numbers the difference between two successive floats can be much larger than a typical value for X and therefore a different comparison method is required. We chose a hybrid method of both an absolute comparison and a Units in Last Place (ULP) comparison. A ULP comparison determines how many representable floating point numbers there are between any two floats. By default the hybrid method we employ first performs an absolute check, $|a - b| < 10^{-14}$. This number was chosen as we found that typical differences in the Sod Shock Tube solution when comparing results with different hardware and compilers resulted in differences of $\sim 5 \times 10^{-15}$. After the absolute check a ULP check is performed with a maximum allowed error of 4. If either check passes then the numbers are deemed to be “equal”. Having a robust method to compare floating point numbers is critical since the output of a code is not guaranteed to be bitwise identical when the code is compiled with different compilers, run on different systems, etc.

4.2.2. System Tests

⁹ <https://github.com/google/googletest>

¹⁰ <https://randomascii.wordpress.com/2012/02/25/comparing-floating-point-numbers-2012-edition/>

GoogleTest provides most of the tools required to run unit and integration tests, but validating the results of an entire system test requires additional infrastructure. Running a system test requires launching the program to test with correct initial conditions, checking that the program did not crash, loading both the generated data to test and the fiducial data, then comparing those two data sets. In order to perform system tests with Cholla, we added a class that performs all of the required tasks, which include launching Cholla with any number of MPI ranks as well as comparing the results against fiducial data. To facilitate running across a wide range of MPI ranks and on clusters with queue systems, the class is designed to allow system tests to be run in different modes: one can either launch Cholla and save the results, compare already existing test data to fiducial data, or do both. This enables the user to run Cholla on many thousands of ranks then later launch a separate job to make the actual comparison. We have found that on up to 10,000 ranks, with small simulation grids (typically $< 64^3$ cells) per rank the latter comparison only takes a few minutes on a single CPU core. Most of the time, however, both steps can be run within the same job, since large tests with many ranks are not required for most development work.

Two primary methods of comparison are used to determine the success of a system test. These are either a direct cell-by-cell comparison of the results for each field using the floating point comparison tools described above (Section 4.2.1), or a calculation of the L2 norm of the L1 error vector as described in Section 3.1.1. The cell-by-cell comparison is quite accurate, but can be fragile on some complex tests if a small number of cells have errors that are slightly larger than typical, which can lead to false failures when comparing results between systems or compilers. The L2 norm method is less fragile to small errors in a handful of cells, but is generally less sensitive, so we only use it on the tests where it is required, namely the MHD blast wave (subsubsection 3.1.5) and advecting field loop (subsubsection 3.1.3).

4.3. Automated Testing

To ensure that these tests are run regularly and all new code is tested we have made the existing tests as easy to run as possible, and we require that they are all run automatically on each pull request. To facilitate this, Cholla's build directory includes a script which performs all required setup, installs GoogleTest (if requested), builds Cholla, and then runs all the tests. The script also includes a function that combines all of these into a single function call for ease of use, for example, if a user is running tests manually (say, prior to submitting a PR).

Implementing automated testing for HPC codes is not always an easy task. Automated tests are an aspect of Continuous Integration (CI) – the practice of automating the addition of code changes and additions from a team of developers into a software project. Cholla is currently designed to run on CUDA or HIP capable GPUs, so GPU hardware is required in order to incorporate continuous integration (CI), something

that few current CI services offer at a reasonable cost for academic users. Many of Cholla’s physics modules are turned on and off at compile time, which presents another challenge, since not all code will be tested for every compilation configurations. Similarly, we need to be able to ensure that Cholla works properly on both AMD and NVIDIA GPUs, which require different compilation targets. This means that there is no single binary file that contains all of Cholla. Instead we have multiple “builds” that each require testing. While this does result in a high performance executable that only contains the necessary code, it makes testing the code much more complex due to the number of possible build configurations.

Our solution to these issues was to use a mix of two different systems. When a pull request is submitted to the Cholla GitHub repository several jobs are launched: A GitHub Actions¹¹ job to check code formatting, a GitHub Actions matrix job to build all the HIP/AMD builds using a Docker container, and a Jenkins¹² matrix job running on local hardware that runs the CUDA/NVIDIA builds, tests, and static analyzers. Thus, every common configuration of Cholla is built with both CUDA and HIP on every pull request and the CUDA builds are also tested to ensure that no existing or new tests fail.

5. SUMMARY

We have presented the MHD extension to Cholla, a massively parallel, GPU native, astrophysical simulation code. MHD in Cholla uses the Van Leer plus Constrained Transport MHD integrator (VL+CT) (Stone & Gardiner 2009), the HLLD Riemann solver (Miyoshi & Kusano 2005), and includes multiple high order reconstruction methods to model numerical solutions to the Eulerian ideal MHD equations on a static mesh.

We showed the modifications required to implement MHD on GPUs compared to CPUs and discussed challenges working within the limits of GPUs in Section 2.3. One major challenge was moving computational work and data storage from the CPU to the GPU. Previous versions of Cholla did some of the computation on the CPU and used CPU memory to effectively expand GPU memory. This required regularly copying data between the CPU and the GPU which became a performance bottleneck. The current version of Cholla maintains all the data, and the vast majority of the work, on the GPU which has led to a considerable speedup. As demonstrated in Section 3.2, these optimizations combined with the highly parallel nature of GPUs make MHD in Cholla extremely fast, with 259 million cell updates per GPU-second when running on a single NVIDIA A100. Cholla also demonstrates excellent weak scaling, and achieves a weak scaling efficiency of 80% when scaled up to 74,088 GPUs on *Frontier* which utilizes AMD MI250X GPUs; with a total of 14.0 trillion cell updates per second (see Figure 13).

¹¹ <https://github.com/features/actions>

¹² <https://www.jenkins.io>

We have also presented a suite of canonical MHD tests in Section 3. These tests demonstrate the accuracy of Cholla across a broad range of problems. They also demonstrate that the VL+CT MHD algorithm does an excellent job of maintaining the divergence free condition even in highly challenging settings.

To accommodate the increasing complexity of Cholla and facilitate multiple simultaneous development efforts, we have also added a structured testing framework, described in Section 4. This framework is based on GoogleTest, augmented with custom testing tools. This approach enables running tests that range from single function unit tests to massively parallel system tests across the scale of an entire cluster within the same framework. We have further integrated these tests with GitHub Actions to run formatting, static analysis, and builds with various physics configurations along with Jenkins running on local resources.

Cholla is free and open source software available at <https://github.com/cholla-hydro/cholla>. In general, Cholla is designed to be flexible and modular, and can be run with or without the new MHD module. In addition, the Cholla framework can easily accommodate additional physics modules in the future, some of which are in progress. We welcome new development efforts, and hope that this work will be a resource for the broader astrophysics simulation community.

6. ACKNOWLEDGEMENTS

RVC thanks Alwin Mao, Helena Richie, Orlando Warren, Kyle Felker, and Seth Cook for many helpful discussions. Special thanks to Daniel Perrefort at the Center for Research Computing for his help in setting up and maintaining the Jenkins environment for our automated testing. This research was supported in part by the University of Pittsburgh Center for Research Computing, RRID:SCR_022735, through the resources provided. Specifically, this work used the H2P cluster, which is supported by NSF award number OAC-2117681. This research also used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725, using Frontier CAAR allocations CSC380 and AST181. E.E.S. acknowledges support from NASA TCAN grant 80NSSC21K0271, NASA ATP grant 80NSSC22K0720, StScI grant HST-AR-16633.001-A, and the David and Lucile Packard Foundation (grant no. 2022-74680).

Software: Cholla (Schneider & Robertson 2015; Caddy et al. 2024), numpy (Van Der Walt et al. 2011), matplotlib (Hunter 2007), HDF5 (The HDF Group. Hierarchical Data Format, version 5.), GoogleTest (<https://github.com/google/googletest>)

REFERENCES

- | | |
|---|--|
| <p>Almgren, A., Sazo, M., Bell, J., et al.
2020, Journal of Open Source Software,
5, 2513, doi: 10.21105/joss.02513</p> | <p>Banda-Barragán, W., Federrath, C.,
Crocker, R., & Bicknell, G. 2017,
MNRAS, 473, 3454,
doi: 10.1093/mnras/stx2541</p> |
|---|--|

- 874 Banda-Barragán, W. E., Parkin, E. R.,
875 Federrath, C., Crocker, R. M., &
876 Bicknel, G. V. 2016, Monthly Notices of
877 the Royal Astronomical Society, 455,
878 1309, doi: [10.1093/mnras/stv2405](https://doi.org/10.1093/mnras/stv2405)
- 879 Beck, K. 1999, Computer, 32, 70,
880 doi: [10.1109/2.796139](https://doi.org/10.1109/2.796139)
- 881 Beck, R. 2016, The Astronomy and
882 Astrophysics Review, 24, 4,
883 doi: [10.1007/s00159-015-0084-4](https://doi.org/10.1007/s00159-015-0084-4)
- 884 Beck, R., Brandenburg, A., Moss, D.,
885 Shukurov, A., & Sokoloff, D. 1996,
886 Annual Review of Astronomy and
887 Astrophysics, 34, 155,
888 doi: [10.1146/annurev.astro.34.1.155](https://doi.org/10.1146/annurev.astro.34.1.155)
- 889 Blandford, R., Meier, D., & Readhead, A.
890 2019, Annual Review of Astronomy and
891 Astrophysics, 57, 467, doi: [10.1146/annurev-astro-081817-051948](https://doi.org/10.1146/annurev-astro-081817-051948)
- 892 Brackbill, J. U., & Barnes, D. C. 1980,
893 Journal of Computational Physics, 35,
894 426
- 895 Brandenburg, A., & Ntormousi, E. 2023,
896 Annual Review of Astronomy and
897 Astrophysics, 61, 561, doi: [10.1146/annurev-astro-071221-052807](https://doi.org/10.1146/annurev-astro-071221-052807)
- 898 Brio, M., & Wu, C. C. 1988, JOURNAL
899 OF COMPUTATIONAL PHYSICS, 75,
900 40. <https://ui.adsabs.harvard.edu/abs/1988JCoPh..75..400B/abstract>
- 901 Brüggen, M., Scannapieco, E., & Grete, P.
902 2023, The Astrophysical Journal, 951,
903 113, doi: [10.3847/1538-4357/acd63e](https://doi.org/10.3847/1538-4357/acd63e)
- 904 Buck, T., Pfrommer, C., Pakmor, R.,
905 Grand, R. J. J., & Springel, V. 2020,
906 Monthly Notices of the Royal
907 Astronomical Society, 497, 1712,
908 doi: [10.1093/mnras/staa1960](https://doi.org/10.1093/mnras/staa1960)
- 909 Bégué, D., Pe'er, A., Zhang, G.-Q.,
910 Zhang, B.-B., & Pevzner, B. 2023, The
911 Astrophysical Journal Supplement
912 Series, 264, 32,
913 doi: [10.3847/1538-4365/aca276](https://doi.org/10.3847/1538-4365/aca276)
- 914 Caddy, R., Schneider, E., Villasenor, B.,
915 et al. 2024, bcaddy/cholla: Caddy &
916 Schneider 2024, v3.x, Zenodo,
917 doi: [10.5281/zenodo.10927183](https://doi.org/10.5281/zenodo.10927183)
- 918 Carteret, Y., Bendre, A. B., & Schober, J.
919 2022, Observational Signatures of
920 Galactic Turbulent Dynamos, arXiv.
921 <http://arxiv.org/abs/2208.14178>
- 922 Chan, T. K., Kereš, D., Hopkins, P. F.,
923 et al. 2019, Monthly Notices of the
924 Royal Astronomical Society, 488, 3716,
925 doi: [10.1093/mnras/stz1895](https://doi.org/10.1093/mnras/stz1895)
- 926 Charbonneau, P. 2014, Annual Review of
927 Astronomy and Astrophysics, 52, 251,
928 doi: [10.1146/annurev-astro-081913-040012](https://doi.org/10.1146/annurev-astro-081913-040012)
- 929 Colella, P. 1990, Journal of
930 Computational Physics, 87, 171,
931 doi: [10.1016/0021-9991\(90\)90233-Q](https://doi.org/10.1016/0021-9991(90)90233-Q)
- 932 Colella, P., & Woodward, P. R. 1984,
933 Journal of Computational Physics, 54,
934 174, doi: [10.1016/0021-9991\(84\)90143-8](https://doi.org/10.1016/0021-9991(84)90143-8)
- 935 Courant, R., Friedrichs, K., & Lewyt, H.
936 1967, IBM Journal
- 937 Crockett, R. K., Colella, P., Fisher, R. T.,
938 Klein, R. I., & McKee, C. F. 2005,
939 Journal of Computational Physics, 203,
940 422, doi: <https://doi.org/10.1016/j.jcp.2004.08.021>
- 941 Dai, W., & Woodward, P. R. 1998, THE
942 ASTROPHYSICAL JOURNAL, 494,
943 317. <https://iopscience.iop.org/article/10.1086/305176/meta>
- 944 Davis, S. W., & Tchekhovskoy, A. 2020,
945 Annual Review of Astronomy and
946 Astrophysics, 58, 407, doi: [10.1146/annurev-astro-081817-051905](https://doi.org/10.1146/annurev-astro-081817-051905)
- 947 Dedner, A., Kemm, F., Kröner, D., et al.
948 2002, Journal of Computational
949 Physics, 175, 645,
950 doi: [10.1006/jcph.2001.6961](https://doi.org/10.1006/jcph.2001.6961)
- 951 Del Zanna, L., Velli, M., & Londrillo, P.
952 2001, Astronomy & Astrophysics, 367,
953 705, doi: [10.1051/0004-6361:20000455](https://doi.org/10.1051/0004-6361:20000455)
- 954 Dobbs, C. L., & Price, D. J. 2007,
955 Monthly Notices of the Royal
956 Astronomical Society, 383, 497,
957 doi: [10.1111/j.1365-2966.2007.12591.x](https://doi.org/10.1111/j.1365-2966.2007.12591.x)
- 958 Einfeldt, B., Munz, C. D., Roe, P. L., &
959 Wgreen, B. 1991, JOURNAL OF
960 COMPUTATIONAL PHYSICS, 92,
961 213. <https://ui.adsabs.harvard.edu/abs/1991JCoPh..92..273E/abstract>

- Evans, C. R., & Hawley, J. F. 1988, *Astrophysical Journal*, 332, 659, doi: [10.1086/166684](https://doi.org/10.1086/166684)
- Falle, S. A. E. G. 1991, *Monthly Notices of the Royal Astronomical Society*, 250, 581, doi: [10.1093/mnras/250.3.581](https://doi.org/10.1093/mnras/250.3.581)
- Felker, K. G., & Stone, J. 2018, *Journal of Computational Physics*, 375, 1365, doi: [10.1016/j.jcp.2018.08.025](https://doi.org/10.1016/j.jcp.2018.08.025)
- Galishnikova, A. K., Kunz, M. W., & Schekochihin, A. A. 2022, Tearing instability and current-sheet disruption in the turbulent dynamo, arXiv. <http://arxiv.org/abs/2201.07757>
- Gardiner, T. A., & Stone, J. M. 2005, *Journal of Computational Physics*, 205, 509, doi: [10.1016/j.jcp.2004.11.016](https://doi.org/10.1016/j.jcp.2004.11.016)
- . 2008, *Journal of Computational Physics*, 227, 4123, doi: [10.1016/j.jcp.2007.12.017](https://doi.org/10.1016/j.jcp.2007.12.017)
- Gent, F. A., Mac Low, M.-M., Käpylä, M. J., & Singh, N. K. 2021, *The Astrophysical Journal Letters*, 910, L15, doi: [10.3847/2041-8213/abed59](https://doi.org/10.3847/2041-8213/abed59)
- Girichidis, P., Pfrommer, C., Hanasz, M., & Naab, T. 2019, *Monthly Notices of the Royal Astronomical Society*, stz2961, doi: [10.1093/mnras/stz2961](https://doi.org/10.1093/mnras/stz2961)
- Girichidis, P., Pfrommer, C., Pakmor, R., & Springel, V. 2022, *Monthly Notices of the Royal Astronomical Society*, 510, 3917, doi: [10.1093/mnras/stab3462](https://doi.org/10.1093/mnras/stab3462)
- Girichidis, P., Naab, T., Walch, S., et al. 2016, *The Astrophysical Journal*, 816, L19, doi: [10.3847/2041-8205/816/2/L19](https://doi.org/10.3847/2041-8205/816/2/L19)
- Godunov, S. K., & Bohachevsky, I. 1959, *Matematičeskij sbornik*, 47(89), 271. <https://hal.science/hal-01620642>
- Goldberg, D. 1991, *ACM computing surveys (CSUR)*, 23, 5
- Grand, R. J. J., Gómez, F. A., Marinacci, F., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, stx071, doi: [10.1093/mnras/stx071](https://doi.org/10.1093/mnras/stx071)
- Grete, P., Glines, F. W., & O'Shea, B. W. 2021, *IEEE Transactions on Parallel and Distributed Systems*, 32, 85, doi: [10.1109/TPDS.2020.3010016](https://doi.org/10.1109/TPDS.2020.3010016)
- Grete, P., Dolence, J. C., Miller, J. M., et al. 2023, *The International Journal of High Performance Computing Applications*, 37, 465, doi: [10.1177/10943420221143775](https://doi.org/10.1177/10943420221143775)
- Han, J. 2017, *Annual Review of Astronomy and Astrophysics*, 55, 111, doi: [10.1146/annurev-astro-091916-055221](https://doi.org/10.1146/annurev-astro-091916-055221)
- Hanasz, M., Lesch, H., Naab, T., et al. 2013, *The Astrophysical Journal*, 777, L38, doi: [10.1088/2041-8205/777/2/L38](https://doi.org/10.1088/2041-8205/777/2/L38)
- Hanasz, M., Wóltński, D., & Kowalik, K. 2009, *The Astrophysical Journal*, 706, L155, doi: [10.1088/0004-637X/706/1/L155](https://doi.org/10.1088/0004-637X/706/1/L155)
- Holmen, J. K., Grete, P., & Vergara, V. G. M. 2023, in *CUG Conference Proceedings*, Helsinki, Finland. <https://www.ornl.gov/publication/early-experiences-olcf-frontier-system-athenapk-and-parthos>
- Hopkins, P. F., Chan, T. K., Garrison-Kimmel, S., et al. 2020, *Monthly Notices of the Royal Astronomical Society*, 492, 3465, doi: [10.1093/mnras/stz3321](https://doi.org/10.1093/mnras/stz3321)
- Hunter, J. D. 2007, *Computing in science & engineering*, 9, 90
- Kaspi, V. M., & Beloborodov, A. M. 2017, *Annual Review of Astronomy and Astrophysics*, 55, 261, doi: [10.1146/annurev-astro-081915-023329](https://doi.org/10.1146/annurev-astro-081915-023329)
- Kim, C.-G., & Ostriker, E. C. 2015, *The Astrophysical Journal*, 815, 67, doi: [10.1088/0004-637X/815/1/67](https://doi.org/10.1088/0004-637X/815/1/67)
- LeVeque, R. J. 2002, *Finite volume methods for hyperbolic problems*, Vol. 31 (Cambridge university press)
- Liska, M. T. P., Chatterjee, K., Issa, D., et al. 2022, *The Astrophysical Journal Supplement Series*, 263, 26, doi: [10.3847/1538-4365/ac9966](https://doi.org/10.3847/1538-4365/ac9966)
- Martin-Alvarez, S., Devriendt, J., Slyz, A., & Teyssier, R. 2018, *Monthly Notices of the Royal Astronomical Society*, 479, 3343, doi: [10.1093/mnras/sty1623](https://doi.org/10.1093/mnras/sty1623)

- Martin-Alvarez, S., Slyz, A., Devriendt, J., & Gómez-Guijarro, C. 2020, Monthly Notices of the Royal Astronomical Society, 495, 4475, doi: [10.1093/mnras/staa1438](https://doi.org/10.1093/mnras/staa1438)
- Mignone, A., & Tzeferacos, P. 2010, Journal of Computational Physics, 229, 2117, doi: [10.1016/j.jcp.2009.11.026](https://doi.org/10.1016/j.jcp.2009.11.026)
- Miniati, F., & Martin, D. F. 2011, Astrophysical Journal, Supplement Series, 195, doi: [10.1088/0067-0049/195/1/5](https://doi.org/10.1088/0067-0049/195/1/5)
- Miyoshi, T., & Kusano, K. 2005, Journal of Computational Physics, 208, 315, doi: [10.1016/j.jcp.2005.02.017](https://doi.org/10.1016/j.jcp.2005.02.017)
- Muller, J.-M., Brisebarre, N., De Dinechin, F., et al. 2018, Handbook of floating-point arithmetic (Springer)
- Naab, T., & Ostriker, J. P. 2017, Annual Review of Astronomy and Astrophysics, 55, 59, doi: [10.1146/annurev-astro-081913-040019](https://doi.org/10.1146/annurev-astro-081913-040019)
- Nakariakov, V. M., & Kolotkov, D. Y. 2020, Annual Review of Astronomy and Astrophysics, 58, 441, doi: [10.1146/annurev-astro-032320-042940](https://doi.org/10.1146/annurev-astro-032320-042940)
- Ntormousi, E., Tassis, K., Del Sordo, F., Fragkoudi, F., & Pakmor, R. 2020, Astronomy & Astrophysics, 641, A165, doi: [10.1051/0004-6361/202037835](https://doi.org/10.1051/0004-6361/202037835)
- Núñez-Castiñeyra, A., Grenier, I. A., Bournaud, F., et al. 2022, Cosmic-ray diffusion and the multi-phase interstellar medium in a dwarf galaxy. I. Large-scale properties and γ -ray luminosities, arXiv. <http://arxiv.org/abs/2205.08163>
- Orszag, S. A., & Tang, C. M. 1979, Journal of Fluid Mechanics, 90, 129, doi: [10.1017/S002211207900210X](https://doi.org/10.1017/S002211207900210X)
- Pakmor, R., Pfrommer, C., Simpson, C. M., & Springel, V. 2016, The Astrophysical Journal, 824, L30, doi: [10.3847/2041-8205/824/2/L30](https://doi.org/10.3847/2041-8205/824/2/L30)
- Pakmor, R., & Springel, V. 2013, Monthly Notices of the Royal Astronomical Society, 432, 176, doi: [10.1093/mnras/stt428](https://doi.org/10.1093/mnras/stt428)
- Pakmor, R., Gómez, F. A., Grand, R. J. J., et al. 2017, Monthly Notices of the Royal Astronomical Society, 469, 3185, doi: [10.1093/mnras/stx1074](https://doi.org/10.1093/mnras/stx1074)
- Pariev, V. I., & Colgate, S. A. 2007, The Astrophysical Journal, 658, 114, doi: [10.1086/510734](https://doi.org/10.1086/510734)
- Pariev, V. I., Colgate, S. A., & Finn, J. M. 2007, The Astrophysical Journal, 658, 129, doi: [10.1086/510735](https://doi.org/10.1086/510735)
- Pfrommer, C., Pakmor, R., Simpson, C. M., & Springel, V. 2017, The Astrophysical Journal, 847, L13, doi: [10.3847/2041-8213/aa8bb1](https://doi.org/10.3847/2041-8213/aa8bb1)
- Philippov, A., & Kramer, M. 2022, Annual Review of Astronomy and Astrophysics, 60, 495, doi: [10.1146/annurev-astro-052920-112338](https://doi.org/10.1146/annurev-astro-052920-112338)
- Powell, K. G. 1997, An Approximate Riemann Solver for Magnetohydrodynamics (Berlin, Heidelberg: Springer Berlin Heidelberg), 570–583, doi: [10.1007/978-3-642-60543-7_23](https://doi.org/10.1007/978-3-642-60543-7_23)
- Ryu, D., Jones, T., & Frank, A. 1995, Astrophysical Journal, 452, 785
- Ryu, D., & Jones, T. W. 1995, The Astrophysical Journal, 422, 228
- Schive, H.-Y., ZuHone, J. A., Goldbaum, N. J., et al. 2018, Monthly Notices of the Royal Astronomical Society, 481, 4815, doi: [10.1093/mnras/sty2586](https://doi.org/10.1093/mnras/sty2586)
- Schneider, E. E., & Robertson, B. E. 2015, Astrophysical Journal, Supplement Series, 217, 24, doi: [10.1088/0067-0049/217/2/24](https://doi.org/10.1088/0067-0049/217/2/24)
- Shankar, S., Mösta, P., Brandt, S. R., et al. 2022, GRaM-X: A new GPU-accelerated dynamical spacetime GRMHD code for Exascale computing with the Einstein Toolkit, arXiv. <http://arxiv.org/abs/2210.17509>
- Shin, M.-S., Stone, J. M., & Snyder, G. F. 2008, APJ, 680, 336, doi: [10.1086/587775](https://doi.org/10.1086/587775)
- Simpson, C. M., Pakmor, R., Marinacci, F., et al. 2016, The Astrophysical Journal, 827, L29, doi: [10.3847/2041-8205/827/2/L29](https://doi.org/10.3847/2041-8205/827/2/L29)

- Sod, G. A. 1978, *Journal of Computational Physics*, 27, 1, doi: [10.1016/0021-9991\(78\)90023-2](https://doi.org/10.1016/0021-9991(78)90023-2)
- Stone, J. M., & Gardiner, T. 2009, *New Astronomy*, 14, 139, doi: [10.1016/j.newast.2008.06.003](https://doi.org/10.1016/j.newast.2008.06.003)
- Stone, J. M., Gardiner, T. A., Teuben, P., Hawley, J. F., & Simon, J. B. 2008, *The Astrophysical Journal Supplement Series*, 178, 137. <https://ui.adsabs.harvard.edu/abs/2008ApJS..178..137S/abstract>
- Stone, J. M., Tomida, K., White, C. J., & Felker, K. G. 2020, *The Astrophysical Journal Supplement Series*, 249, 4, doi: [10.3847/1538-4365/ab929b](https://doi.org/10.3847/1538-4365/ab929b)
- Toro, E. F. 2009, *Riemann Solvers and Numerical Methods for Fluid Dynamics Third Edition* (Springer), doi: [10.1007/978-3-540-49834-6](https://doi.org/10.1007/978-3-540-49834-6)
- Torrilhon, M. 2005, *SIAM Journal on Scientific Computing*, 26, 1166
- Tóth, G. 1996, in *Magnetodynamic Phenomena in the Solar Atmosphere: Prototypes of Stellar Magnetic Activity*, ed. Y. Uchida, T. Kosugi, & H. S. Hudson (Dordrecht: Springer Netherlands), 471–472, doi: [10.1007/978-94-009-0315-9_101](https://doi.org/10.1007/978-94-009-0315-9_101)
- Trujillo Bueno, J., & del Pino Alemán, T. 2022, *Annual Review of Astronomy and Astrophysics*, 60, 415, doi: [10.1146/annurev-astro-041122-031043](https://doi.org/10.1146/annurev-astro-041122-031043)
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in science & engineering*, 13, 22
- Van Leer, B. 2006, in 16th AIAA Computational Fluid Dynamics Conference (Orlando, Florida: American Institute of Aeronautics and Astronautics), doi: [10.2514/6.2003-3559](https://doi.org/10.2514/6.2003-3559)
- Werhahn, M., Girichidis, P., Pfrommer, C., & Whittingham, J. 2023, *Monthly Notices of the Royal Astronomical Society*, 525, 4437, doi: [10.1093/mnras/stad2105](https://doi.org/10.1093/mnras/stad2105)
- Werhahn, M., Pfrommer, C., Girichidis, P., & Winner, G. 2021, *Monthly Notices of the Royal Astronomical Society*, 505, 3295, doi: [10.1093/mnras/stab1325](https://doi.org/10.1093/mnras/stab1325)
- Wibking, B. D., & Krumholz, M. R. 2021, arXiv:2105.04136 [astro-ph]. <http://arxiv.org/abs/2105.04136>
- Wilson, G., Bryan, J., Cranston, K., et al. 2017, *PLOS Computational Biology*, 13, e1005510, doi: [10.1371/journal.pcbi.1005510](https://doi.org/10.1371/journal.pcbi.1005510)
- Wilson, G., Aruliah, D. A., Brown, C. T., et al. 2014, *PLOS Biology*, 12, 7
- Yang, H.-Y. K., Ruszkowski, M., Ricker, P. M., Zweibel, E., & Lee, D. 2012, *The Astrophysical Journal*, 761, 185, doi: [10.1088/0004-637X/761/2/185](https://doi.org/10.1088/0004-637X/761/2/185)
- Yoshida, K., Matsukiyo, S., Shimokawa, K., Washimi, H., & Hada, T. 2021, *The Astrophysical Journal*, 916, 29, doi: [10.3847/1538-4357/ac02c2](https://doi.org/10.3847/1538-4357/ac02c2)
- Zingale, M., Almgren, A. S., Sazo, M. B., et al. 2020, *Journal of Physics: Conference Series*, 1623, 012021, doi: [10.1088/1742-6596/1623/1/012021](https://doi.org/10.1088/1742-6596/1623/1/012021)