FLINT: A SIMULTANEOUS EMULATOR FOR DISCONTINUOUS AND SMOOTH STRUCTURES WITH PINNS

Anonymous authors
Paper under double-blind review

ABSTRACT

Physics-informed neural networks (PINNs) have made substantial progress on hyperbolic conservation laws containing sharp discontinuities. However, accurately capturing discontinuities while preserving transported fine-scale structure remains a challenge. We show that this failure worsens as the spatial frequency of initial data increases, exposing a core limitation of global space–time PINNs. We introduce FLINT (**FL**ux balance for **IN**itial-data **T**ransport), which combines several recent advances in the PINN literature with our novel contributions: discontinuity-selective residual weighting and variable-height control volumes anchored directly to initial conditions. The PDE residual is down-weighted at points where sharp gradients suggest either a shock or contact discontinuity, while the divergence theorem transforms local conservation into global conservation laws over control volumes, which propagate initial data into later times. We evaluate FLINT on 1D benchmark problems for compressible Euler equations, including the newly proposed Sod-k diagnostic (a shock tube with tunable initial density oscillations), the Shu–Osher problem, and a 1D cylindrical blast wave. Across these benchmarks, FLINT captures wave amplitude and phase while maintaining shock and contact locations and global conservation, outperforming baselines in every metric. On Shu–Osher, FLINT reduces wave-region L_2 error by $4-6\times$ and wave amplitude error by $7-31\times$ relative to the stronger baseline, while accurately handling geometric source terms and coordinate singularities in the blast wave. By addressing these 1D transport challenges, FLINT establishes a necessary foundation for scaling global space–time PINNs to complex multi-dimensional compressible flows.

1 INTRODUCTION

Discontinuous problems are ubiquitous in nature, including compressible fluid dynamics, gas dynamics, shallow-water flows, and related transport phenomena (Lax, 1973; LeVeque, 2002; Dafermos, 2016; Toro, 2009). Many fluid and gas systems are well-approximated by partial differential equations (PDEs) that admit discontinuous solutions (Lax, 1973; Dafermos, 2016). One archetypal system is comprised of the compressible Euler's equations of fluid dynamics, which represent the local conservation of mass, momentum, and energy Toro (2009). Here, discontinuities appear in two forms: (1) *contact discontinuities*, where the pressure is constant but the density and/or (transverse) velocity jumps in space, and (2) *shock waves*, where the pressure jumps discontinuously (Toro, 2009; LeVeque, 2002). Along with rarefaction waves, these phenomena represent an Eigenbasis for the solution to the linearized Euler equations. They are colloquially called *waves* or *signals* (Toro, 2009).

Both shocks and contacts represent *unresolved physics*. That is, the continuum prescription provided by the Euler equations breaks down at these points and an atomistic or kinetic picture is required to describe the underlying phenomena (Degond, 2004). Despite this incompleteness, discontinuous solutions are ubiquitous. They must be treated and simulations admitting discontinuities are performed routinely in domain science and engineering. A common approach to handling these discontinuities is to replace the *differential* equation with an *integral* or *distributional* form (Lax, 1973; Dafermos, 2016; Vol'pert, 1967; Le Floch, 1988). This serves two purposes. First, while differentiation of a discontinuous function is ill-defined, integration is perfectly well-behaved (Vol'pert, 1967; Le Floch, 1988). Second, integration of a local conservation law yields global conservation, which may be enforced up to machine precision by construction. Enforcing conservation is particularly important

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

as, when a physical phenomena is under-resolved, one must rely on laws of Nature that are known to hold to guarantee a trustworthy result (Lax, 1957; Vol'pert, 1967; Le Floch, 1988).

Physics Informed Neural Networks (PINNs) represent the solution to a PDE system as a neural network (Raissi et al., 2019; Lu et al., 2021). The PDE residual may then be evaluated by automatic differentiation (Baydin et al., 2018; Bradbury et al., 2018; Paszke et al., 2019). Differentiation of sharp features without care is ill defined, which manifests as noise and oscillations around the feature as first described by Gibbs (1898) and Michelson (1898). Attacking this problem is well-trodden ground in the PINN literature (Patel et al., 2022; Oubarka et al., 2026; Wang & Yang, 2026; De Ryck et al., 2024; Gao et al., 2026; Liu et al., 2024; Zhao et al., 2026; Ghoreishi & Naderan, 2026). However, these approaches either fail to adequately reduce noise around sharp features or miss fine-scale structure surrounding discontinuities, as evidenced by the test problems we show in Section 4, which require capturing both discontinuities and high-frequency smooth regimes. This work is concerned with addressing these shortcomings, and to do so, we borrow the same techniques used by traditional solvers in domain science. Our contributions are summarized next.

Initial-data anchored conservation. We introduce a loss term that integrates over the surface of variable-height control volumes whose lower temporal face is evaluated from the the prescribed initial state, directly injecting known initial data to later predictions through conservation.

Discontinuity-selective residual weighting. We implement a shock detector and WENO-type smoothness indicator, which we use to reduce the strong-form residual around shocks and discontinuities.

A controlled transport diagnostic. We introduce Sod-k, which isolates the effect high-frequency initial data, and displays progressive loss of accuracy in existing PINN approaches.

Comprehensive evaluation. Across Sod-k, Shu–Osher, and a cylindrical blast wave, FLINT preserves oscillatory structure while maintaining discontinuities and conservation, and extends to the Euler equations with geometric source terms.

2 Related PINN Works that Tackle Discontinuities

PINNs for discontinuous hyperbolic problems have largely followed two complementary directions. One replaces the point-wise PDE residual with an integral conservation constraint that remains well defined across shocks (Patel et al., 2022; Oubarka et al., 2026; Wang & Yang, 2026; De Ryck et al., 2024; Gao et al., 2026). The other retains the strong-form equations but reduces their influence in regions where discontinuities make the residual difficult to optimize (Liu et al., 2024; Zhao et al., 2026; Ghoreishi & Naderan, 2026). Our method combines both ideas, but differs from prior work in two important ways. First, our control volumes are anchored to the known initial conditions and span a wide range of temporal horizons. Second, our residual weight identifies a shock using both a large solution gradient and local compression, rather than either signal alone.

Integral Conservation Constraints: Control volume discretization PINN (cvPINN) (Patel et al., 2022) replaces point-wise collocation with flux balance over a fixed partition of space–time control volumes in a manner similar to finite-volume discretization. Weak and Entropy PINNs (WE-PINNs) (Oubarka et al., 2026) removes the fixed partition and instead samples control volumes of different spatial and temporal sizes throughout the domain during training. Coupled Integral PINN (CI-PINN) (Wang & Yang, 2026) takes a different approach, avoiding explicit control-volume quadrature by introducing an auxiliary network that learns an integral representation of the conserved quantities. **Our control volumes are more targeted.** Rather than sampling general local or multi-scale regions in space–time every control volume begins at time $t = 0$, where the solution is known exactly, and extends to time $h \sim \mathcal{U}(t_B, T)$, where h is drawn from a uniform distribution, $\mathcal{U}$, with bounds t_B (where t_B is a small non-zero value) to final time T. Each conservation constraint directly couples the known initial state to the predicted state at a later time h. Boxes with different heights impose this constraint over many durations, from short-time evolution to volumes spanning the formation and propagation of shocks. Thus, the anchored conservation loss introduces a direct non-local connection between the initial data and later predictions rather than relying only on point-wise gradients.

Discontinuity-Selective Residual Weighting: A second line of work modifies the strong-form residual near discontinuities. Physics-Informed Neural Networks with Equation Weight (PINNs-WE) (Liu et al., 2024) down-weights the governing equations in compressive regions using the velocity divergence. Spatio-Temporal Uncertainty-Modulated PINN (UM-PINN) (Zhao et al., 2026) attenuates the residual according to the magnitude of the predicted solution gradient. Weighted Loss Hard Constraint PINN (WHC-PINN) (Ghoreishi & Naderan, 2026) utilizes the same weighting mechanism as Liu et al. (2024). These methods reduce the influence of regions that can dominate PINN optimization near shocks, but each uses a single local indicator. **Our weight instead checks both a large gradient and compression**. A contact discontinuity or entropy wave can contain a large gradient without local compression, while a smooth compressive wave can have negative velocity divergence without forming a shock. Using either quantity alone can therefore suppress the PDE residual away from the shock itself. By combining steepness and compression, our weighting reduces the strong-form residual at steep compressive fronts while retaining PDE supervision on steep non-shock structures.

3 METHODOLOGY

We solve the *one-dimensional* compressible Euler equations in conservative form,

$$\partial_t U + \frac{1}{J}\partial_x J F(U) = S(U), \quad U = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}, \quad F = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E+p) \end{pmatrix}, \quad E = \frac{p}{\gamma - 1} + \tfrac{1}{2}\rho u^2, \quad (1)$$

where U denotes the *conservative* variables, which are density ρ, momentum ρu, and total energy E, corresponding to the *primitive* variables $Q = (\rho, u, p)^\top$, which are density, velocity, and pressure respectively. F is the flux. Here we assume an ideal gas equation of state, allowing us to relate pressure to internal energy via the adiabatic index γ, which is fixed to 1.4. Equation (1) is solved on the space-time box $(x, t) \in [x_0, x_1] \times [0, T]$, where t is the time from zero to final time T and x is a spatial variable with left and right spatial boundaries, x_0 and x_1, respectively. The prescribed initial conditions are $Q_0(x) = Q(0, x) = (\rho, u, p)(0, x)$, with boundary conditions specified as $Q_L(t) = Q(t, x_0) = (\rho, u, p)(t, x_0)$ and $Q_R(t) = Q(t, x_1) = (\rho, u, p)(t, x_1)$.

The *metric factor* J and *source term* S introduce optional curvilinear coordinates. For the Cartesian coordinates used in the Sod-k and Shu-Osher shock tubes, $J = 1$ and $S = 0$. For the blast wave, the spatial coordinate x is re-interpreted as a radial coordinate, and $J = x$. The source term S then becomes $S(U) = (0, p/x, 0)^\top$ to incorporate the relevant Christoffel symbols (Rezzolla & Zanotti, 2013). Note this *modifies* Equation (1) compared to prior work and introduces a *coordinate singularity* at $x = 0$, which is challenging to treat. To solve equation (1) in a way that captures the complex structure emergent from these equations, we develop a novel global space-time PINN framework, as summarized in Figure 1, the following components of which, will be elaborated below.

Our method uses a neural network

$$\mathcal{G}_\theta : (x, t) \mapsto Q_\theta(x, t) = \big(\rho_\theta(x, t), u_\theta(x, t), p_\theta(x, t)\big)^\top,$$

which maps each space–time coordinate to the primitive variables, given hyperparameters and weights θ. The training objective combines a weighted Euler residual, initial and boundary conditions, conservation over interior space–time control volumes, and initial-data-anchored control volumes that connect the prescribed initial conditions to later times through flux balance (4).

Neural network. We use a fully connected coordinate network with residual connections to map (x, t) to $(\rho_\theta, u_\theta, p_\theta)$. Residual connections improve training stability and gradient propagation in deep networks (He et al., 2016a;b) and have also been used to address optimization difficulties in PINNs (Wang et al., 2021). Physically, density and pressure are greater than zero. To ensure positive density and pressure, we set $\rho = \rho_{\min} + \mathrm{softplus}(\rho_\theta)$, $u = u_\theta$, and $p = p_{\min} + \mathrm{softplus}(p_\theta)$ with $\rho_{\min} = p_{\min} = 1e^{-5}$. The offsets impose a strict positive lower bound and prevents numerical underflow. The parameters that define the neural networks used can be found in Appendix F.

Compression-weighted PDE residual. We enforce the Euler equations in strong form in smooth regions while reducing their influence near compressive shocks. Let

$$\mathcal{X}_{\mathrm{pde}} = \{(x_i, t_i)\}_{i=1}^{N_{\mathrm{pde}}}, \qquad (x_i, t_i) \in (x_0, x_1) \times (0, T] \quad \forall i,$$

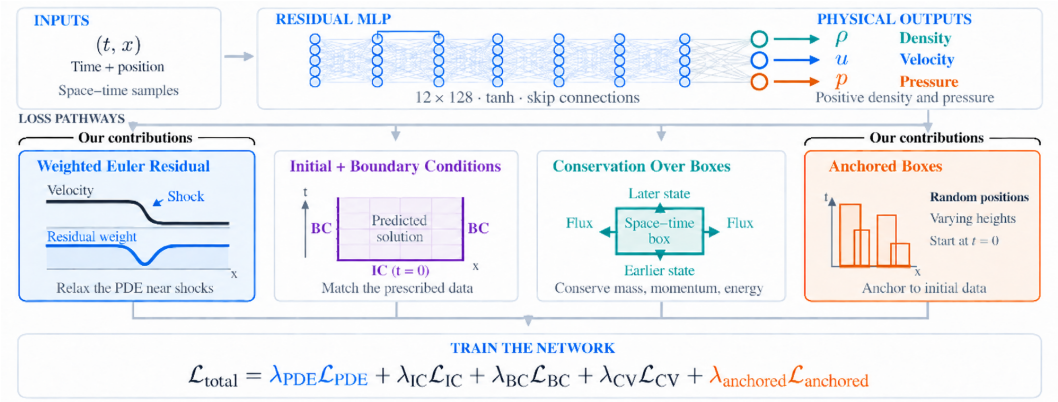


Figure 1: **Overview of FLINT.** A multilayer perceptron maps (t, x) to density, velocity, and pressure. Training combines a compression-weighted residual, initial and boundary condition losses, integral conservation over interior control volumes, and variable-height control volumes anchored to the prescribed initial condition.

denote the set of N_{pde} interior collocation points. At each point, we evaluate the residual $r_i = \partial_t U_i + 1/J\partial_x JF(U_i) - S(U_i)$. We expand on Liu et al. (2024); Zhao et al. (2026); Ghoreishi & Naderan (2026) and define

$$\mathcal{L}_{\text{PDE}} = \frac{1}{N_{\text{pde}}} \sum_{i=1}^{N_{\text{pde}}} \|\lambda_i r_i\|_2^2, \qquad \lambda_i = \frac{1}{1 + \varepsilon D_i g_i},$$
$$D_i = L \left| \frac{\partial_x \rho_i}{\sqrt{1+\rho_i^2}} + \frac{\partial_x u_i}{\sqrt{1+u_i^2}} + \frac{\partial_x p_i}{\sqrt{1+p_i^2}} \right|, \quad g_i = \frac{c_i^2}{c_i^2 + K^2}, \quad c_i = \max(-\partial_x u_i, 0), \tag{2}$$

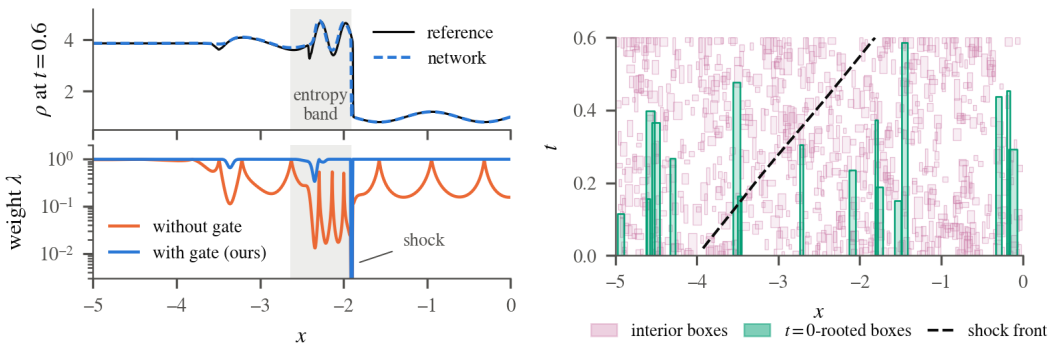
where $L = x_1 - x_0$. The detector D_i measures steep gradients in the primitive variables,while c_i measures compression. The gate g_i distinguishes steep compressive regions from steep but non-compressive features. Shocks are both steep and compressive, while contact discontinuities, smooth waves, and rarefaction fans need not be. Thus, λ_i decreases only at shocks, leaving the Euler residual strongly enforced elsewhere. The sensor strength ε controls the overall strength of the weighting, while the gate threshold K sets the compression scale of the gate, with $g_i = 1/2$ when $c_i = K$. ε and K are tabulated in Appendix F. Here and below, random samples are taken from the uniform distribution over the relevant interval. Figure 2a illustrates the effect of this gated residual weighting on the Shu–Osher problem, which is discussed further in Section 4.

Initial and boundary conditions. Following the standard PINN formulation (Raissi et al., 2019), we enforce the prescribed initial and boundary conditions using normalized squared-error losses. For the initial condition, we randomly sample N_{IC} points on $t = 0$ and penalize the difference between $Q_\theta(0, x)$ and $Q_0(x)$. For the boundary conditions, we randomly sample N_{BC} points on $x = x_0$ and $x = x_1$ and similarly enforce $Q_L(t)$ and $Q_R(t)$. We denote the resulting losses by $\mathcal{L}_{\text{IC}}$ and $\mathcal{L}_{\text{BC}}$.

Interior control volumes. Point-wise residuals enforce the equations (1) locally. Following control-volume formulations for conservation-law PINNs (Patel et al., 2022; Oubarka et al., 2026), we also enforce integral conservation over finite space–time regions. For one control volume $B = [x_L, x_R] \times [t_B, t_T]$, where t_B and t_T are the lower and upper temporal boundaries and x_L and x_R are the left and right spatial boundaries, respectively, integrating (1) gives

$$C(B) = \int_{x_L}^{x_R} J(x)\big[U(t_T, x) - U(t_B, x)\big]\, dx + \int_{t_B}^{t_T} \big[JF(U)\big]_{x_L}^{x_R} dt - \int_{t_B}^{t_T} \int_{x_L}^{x_R} J(x) S(U)\, dx\, dt = 0, \tag{3}$$

where $\big[JF(U)\big]_{x_L}^{x_R} = J(x_R)F(U(t, x_R)) - J(x_L)F(U(t, x_L))$. For the Cartesian coordinates used in Sod-k and Shu–Osher, $J = 1$ and $S = 0$, reducing (3) to the standard space–time flux balance.



(a) Density and weight λ at $t = 0.6$.

(b) Control volumes in space-time.

Figure 2: A trained Shu–Osher network ($x \in [-5, 0]$, $t \leq 0.6$). (a) With the gate, $\lambda \approx 1$ except at the shock; without it, λ also drops to $\sim$0.01 in the oscillator post-shock region (shaded) and $\sim$0.15 ahead of the shock, where the equations hold. (b) Interior boxes (pink); $t = 0$-rooted boxes (green) placed at random x. The network's own shock front (dashed) is shown for reference.

Unlike the point-wise strong form, this integral balance remains well defined across discontinuities because it depends on integrated state changes and boundary fluxes rather than point-wise derivatives (Toro, 2009). Let $\mathcal{B}_{\text{int}} = \{B_j\}_{j=1}^{N_{\text{CV}}}$, with $B_j = [x_{L,j}, x_{R,j}] \times [t_{B,j}, t_{T,j}] \subset (x_0, x_1) \times (0, T]$, denote a sampled interior control volumes, and N_{CV} be the total number of these interior control volumes. We sample their locations randomly and their heights and widths log-uniformly, and approximate each face integral using second-order (trapezoidal) quadrature with 33 points. We minimize $\mathcal{L}_{\text{CV}} = N_{\text{CV}}^{-1} \sum_{j=1}^{N_{\text{CV}}} \|C(B_j)/|B_j|\|_2^2$, where $|B_j|$ is the space–time area.

Initial-data-anchored control volumes. *Our defining contribution is a second set of control volumes that are anchored to the prescribed initial condition at $t = 0$.* Prior control-volume PINNs enforce conservation over local space–time regions (Patel et al., 2022; Oubarka et al., 2026). We additionally sample

$$\mathcal{B}_{\text{anchor}} = \{B_j^{\text{anchor}}\}_{j=1}^{N_{\text{anchor}}}, \quad B_j^{\text{anchor}} = [x_{L,j}, x_{L,j} + w_j] \times [0, h_j],$$

where N_{anchor} is the total number of anchored control volumes, $x_{L,j}$ is the left spatial edge, w_j is the spatial width, and h_j is the temporal extent of the jth anchored volume. We sample $x_{L,j} \sim \mathcal{U}(x_0, x_1 - w_j)$, $w_j \sim \mathcal{U}(w_{\min}, w_{\max})$, and $h_j \sim \mathcal{U}(h_{\min}, h_{\max})$, where $\mathcal{U}$ is the uniform distribution. The sampling bounds $w_{\min}$, $w_{\max}$, $h_{\min}$, and $h_{\max}$ are specified in Appendix F. The lower face of each volume is evaluated from the prescribed initial state Q_0. Each volume therefore directly connects known data at $t = 0$ to the predicted solution at a later time through the conservation balance in (3). To our knowledge, this dedicated use of randomly placed, variable-height initial-data-anchored control volumes has not been explored in prior control-volume PINNs. We define $\mathcal{L}_{\text{anchor}}$ analogously to $\mathcal{L}_{\text{CV}}$ over $\mathcal{B}_{\text{anchor}}$. Figure 2b shows an example set of both rooted boxes and interior boxes for the same Shu-Osher problem as in Figure 2a. The bottom face, $t = 0$, is integrated exactly in closed form. The other three faces use second-order (trapezoidal) quadrature on 385 nodes. We demonstrate the impact of the proposed anchored CV with initial data with a study in Section C.1.

Optimization. PINNs are commonly trained with L-BFGS (Raissi et al., 2019), first-order Adam followed by L-BFGS refinement (Ghoreishi & Naderan, 2026; Lu et al., 2021; Liu et al., 2024; Zhao et al., 2026), and Adam (Patel et al., 2022). PINN objectives can exhibit poorly aligned gradients across competing loss terms, which can make first-order optimization difficult (Wang et al., 2025). We therefore optimize FLINT with SOAP (Vyas et al., 2025), which combines Adam-style second-moment updates with Shampoo-based matrix preconditioning, which has been shown to improve gradient alignment and training stability in PINNs (Wang et al., 2025). The final training objective is

$$\mathcal{L} = \lambda_{\text{PDE}}\mathcal{L}_{\text{PDE}} + \lambda_{\text{IC}}\mathcal{L}_{\text{IC}} + \lambda_{\text{BC}}\mathcal{L}_{\text{BC}} + \lambda_{\text{CV}}\mathcal{L}_{\text{CV}} + \lambda_{\text{anchor}}\mathcal{L}_{\text{anchor}}. \tag{4}$$

The loss weights and optimizer hyperparameters are specified in Appendix F.

3.1 THE INTUITION BEHIND FLINT

PINNs as applied to hyperbolic PDEs are *spacetime* methods. Rather than stepping an instantaneous solution in time, they construct a function that they demand satisfies the PDE for all space and time. This transforms the time integration problem tackled by traditional solvers into an optimization one. A similar transformation may be made in traditional numerical methods by integrating a conservation law implicitly (Beam & Warming, 1976), with the simplest implicit integrator being the Backward Euler (1768) method (Curtiss & Hirschfelder, 1952). While implicit integrators are unconditionally stable (i.e., they always produce an answer), they can "step over" fast dynamics (Heath, 2018). This phenomenon may explain why traditional PINNs capture low-frequency structure but struggle for high frequencies. By enforcing conservation on spacetime boxes that touch the initial conditions, FLINT forces the model to incorporate this structure. Zhao et al. (2026) describe their shock gate as analogous to flux limiters in traditional solvers. Limiters replace a high-accuracy solution, which is subject to slow convergence via the Gibbs (1898) phenomenon, with a lower-order solution that suppresses spurious oscillations. Our residual gate, while similar in spirit, is more similar to an ENO (Harten et al., 1987) or WENO (Liu et al., 1994) scheme, where the method attempts to discard the set of measure zero in the domain where gradients are ill-defined. In this spirit, the discontinuity detector D_i from Equation (2) can be interpreted as a first-order WENO-type smoothness indicator.

4 EXPERIMENTAL RESULTS

We evaluate FLINT on three challenging problems, two of which require modeling fine-scale wave structure through nonlinear hyperbolic dynamics. We first consider Sod-k ($k = 50$) and the Shu–Osher problem Shu & Osher (1989); Zhao et al. (2026). Introduced here as a novel PINN benchmark, Sod-k extends the classical Sod shock tube by adding a tunable, high-frequency density perturbation. Shu–Osher, which has received only limited attention in the PINN literature, tests a moving Mach-3 shock interacting with an entropy wave. Finally, the cylindrical blast wave further evaluates FLINT in the presence of complex geometric source terms and coordinate singularities. Together, these benchmarks evaluate whether a solver can preserve smooth structure, handle curvilinear geometries, and resolve sharp discontinuities while adhering to conservation laws. Reference solutions are generated using fifth-order WENO5 reconstruction (Jiang & Shu, 1996) with global Lax–Friedrichs flux splitting and third-order SSP-RK time integration (Shu & Osher, 1989), while the cylindical blast wave reference is generated using MUSCL–HLLC (minmod-limited MUSCL reconstruction (van Leer, 1979; Sweby, 1984) with the HLLC approximate Riemann solver (Toro et al., 1994; Toro, 2009) and SSP-RK3 time stepping (Gottlieb et al., 2001).

We restrict quantitative comparisons to prior methods with publicly available author implementations that directly correspond to the two core components of FLINT. Under this criterion, we compare against PINNs-WE (Liu et al., 2024) and WE-PINN (Oubarka et al., 2026). In this section, PINNs-WE and WE-PINN use their default optimizers and hyperparameters. An evaluation of these baselines trained with SOAP is provided in Section C.2. Closely related methods, including cvPINN (Patel et al., 2022), UM-PINN (Zhao et al., 2026), and CI-PINN (Wang & Yang, 2026), are cited as relevant prior work but are not benchmarked because no public author implementation is available. The metrics used to measure solution quality are discussed in E.

4.1 SOD SHOCK TUBE WITH SINUSOIDAL DENSITY PERTURBATIONS

We first consider a controlled modification of the Sod shock tube that introduces a sinusoidal density perturbation into the right state of the initial conditions. We solve the 1D Euler equations on $(x, t) \in [0, 1] \times [0, 0.2]$ with

$$(\rho, u, p)(x, 0) = \begin{cases} (1, 0, 1), & x < 0.5 \\ (0.125(1 + 0.2\sin(kx)), 0, 0.1), & x \geq 0.5. \end{cases} \tag{5}$$

We impose Dirichlet boundary conditions by fixing the primitive variables at both spatial boundaries to their initial values: $(\rho, u, p)(0, t) = (1, 0, 1), \quad (\rho, u, p)(1, t) = \big(0.125(1 + 0.2\sin(k)), 0, 0.1\big)$, where $k = 50$ for this experiment. Appendix A.1 details experiments for additional values of k.

FLINT correctly predicts both the causal structure of the shock and the smooth structure of the sine wave. Figure 3 shows that FLINT closely follows the WENO5 (Jiang & Shu, 1996; Shu

Table 1: **Solution and wave-capturing quality on Sod-K ($k = 50$).** All metrics are defined in Appendix E. ↓ means lower is better and bold font represents **best** per column. Each entry is the mean ± the standard deviation over 3 seeds.

	Space–time L_2 ↓			L_2 at T ↓			Wave L_2 at T ↓			Wave amplitude↓			$\mathcal{R}_{\text{cons}}^{\max}$ (10^{-3})↓		
Method	ρ	u	p	ρ	u	p	ρ	u	p	ρ	u	p	ρ	ρu	E
PINNs-WE	0.051 ±0.033	0.164 ±0.042	0.035 ±0.010	0.058 ±0.037	0.149 ±0.045	0.038 ±0.011	0.212 ±0.110	0.081 ±0.030	0.079 ±0.030	0.59 ±0.13	4.84 ±4.50	3.17 ±2.48	3.78 ±3.09	2.99 ±2.13	10.85 ±4.31
WE-PINN	0.028 ±0.001	0.139 ±0.008	0.022 ±0.001	0.034 ±0.001	0.165 ±0.010	0.034 ±0.002	0.154 ±0.006	0.061 ±0.021	0.051 ±0.013	0.41 ±0.16	2.70 ±3.44	1.78 ±2.46	6.89 ±1.08	1.48 ±0.86	**1.80** ±0.74
FLINT (Ours)	**0.016** ±0.005	**0.102** ±0.021	**0.018** ±0.006	**0.016** ±0.005	**0.082** ±0.059	**0.018** ±0.010	**0.035** ±0.013	**0.008** ±0.004	**0.010** ±0.003	**0.03** ±0.01	**0.09** ±0.05	**0.12** ±0.05	**1.22** ±0.71	**1.23** ±1.03	3.21 ±0.42
WENO5	—												*0.05*	*0.09*	*0.20*

& Osher, 1989) reference across density, velocity, and pressure. In the enlarged region, FLINT reproduces the amplitude and phase of the wave while maintaining the contact discontinuity and shock locations. In contrast, competing methods attenuate the wave, distort its phase, or produce larger errors around the discontinuities. Importantly, this improvement does not come at the expense of global conserved quantities. Error magnitudes and the maximum global conservation residual $R_{\text{cons}}^{\max}$ are tabulated for the Sod-k problem in Table 1, where FLINT achieves the lowest mean value among the evaluated methods.

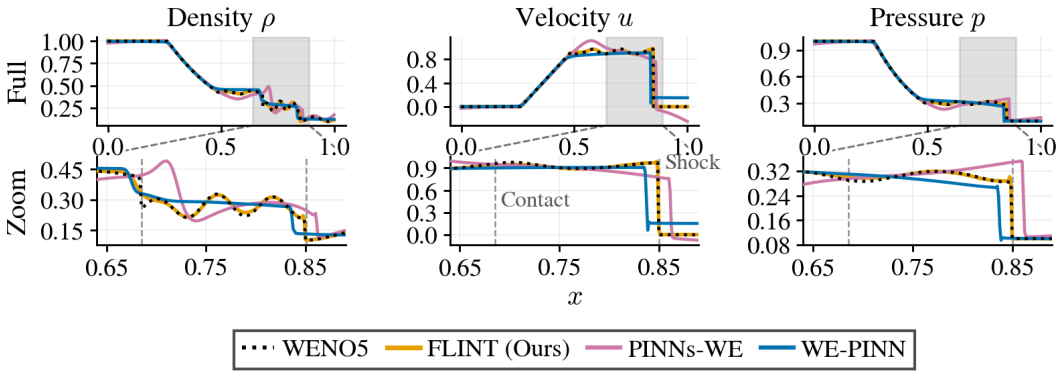


Figure 3: **Sod shock tube with a sinusoidal density wave at $k = 50$.** Predicted density ρ, velocity u, and pressure p at $t = 0.2$ are compared with the WENO5 reference solution for FLINT, PINNs-WE, and WE-PINN. The top row shows the solution over the full spatial domain, while the bottom row enlarges the region containing the oscillatory wave, contact discontinuity, and shock.

4.2 Shu-Osher

We next consider the Shu-Osher (Shu & Osher, 1989) problem, a well-known and challenging test problem in the hydrodynamics solver literature that has limited attention in PINNs literature. It models the interaction between a Mach-3 shock initially located at $x_s = -4$ and an entropy wave (described by a variation in density but fixed pressure) on the domain $(x, t) \in [-5, 0] \times [0, 0.6]$:

$$(\rho, u, p)(x, 0) = \begin{cases} (3.857143, 2.629369, 10.33333), & x < -4, \\ (1 + 0.2\sin(5x), 0, 1), & x \geq -4. \end{cases} \tag{6}$$

FLINT captures the post-shock wave structure that emerges from the interaction of the shock with the entropy wave. Figure 4 shows the strongest separation between FLINT and the competing PINN methods. FLINT closely follows the WENO5 (Jiang & Shu, 1996; Shu & Osher, 1989) oscillations in density, velocity, and pressure, capturing their amplitude and phase behind the shock while maintaining its location. Competing methods substantially smooth or distort this structure. Figure 5 demonstrates a consistently small error in global conserved quantities. Error magnitudes

Table 2: **Solution and wave-capturing quality on Shu–Osher.** All metrics are defined in Appendix E. ↓ means lower is better and bold font represents **best** per column. Each entry is the mean ± the standard deviation over 3 seeds.

Method	Space–time L_2 ↓			L_2 at T ↓			Wave L_2 at T ↓			Wave amplitude↓			$\mathcal{R}_{\text{cons}}^{\max}$ (10^{-2})↓		
	ρ	u	p	ρ	u	p	ρ	u	p	ρ	u	p	ρ	ρu	E
PINNs-WE	0.099 ±0.012	0.126 ±0.037	0.113 ±0.024	0.125 ±0.023	0.159 ±0.055	0.142 ±0.036	0.085 ±0.019	0.033 ±0.007	0.064 ±0.015	0.88 ±0.19	0.85 ±0.24	0.89 ±0.17	12.82 ±4.32	56.35 ±3.87	172.20 ±24.20
WE-PINN	0.085 ±0.002	0.093 ±0.003	0.089 ±0.002	0.079 ±0.008	0.067 ±0.035	0.071 ±0.019	0.074 ±0.001	0.029 ±0.001	0.055 ±0.001	0.99 ±0.02	0.99 ±0.02	0.98 ±0.03	9.81 ±8.60	5.58 ±2.91	15.19 ±2.23
FLINT (Ours)	**0.064** ±0.002	**0.086** ±0.002	**0.077** ±0.002	**0.027** ±0.007	**0.025** ±0.027	**0.024** ±0.016	**0.019** ±0.002	**0.005** ±0.001	**0.011** ±0.003	**0.12** ±0.03	**0.03** ±0.01	**0.03** ±0.01	**0.97** ±0.19	**1.21** ±0.29	**2.35** ±1.45
WENO5						—							*<0.01*	*0.57*	*1.32*

and the maximum global conservation residual $R_{\text{cons}}^{\max}$ are tabulated in Table 9, where FLINT achieves the lowest mean value for every reported metric.

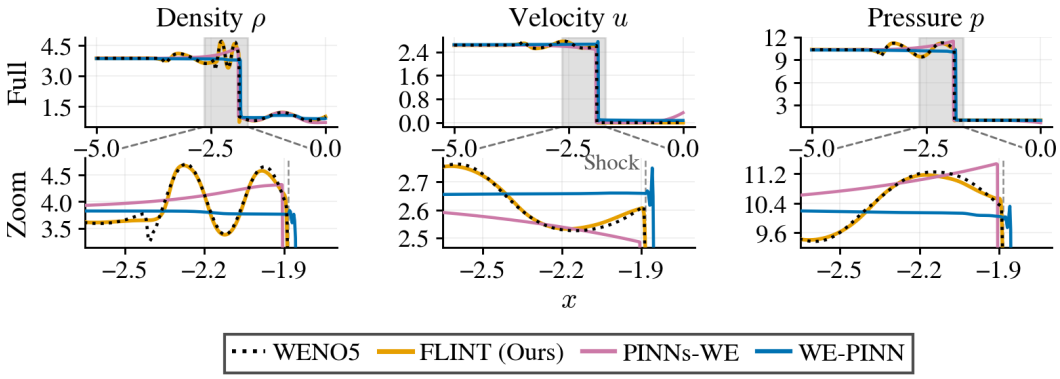


Figure 4: **Shu–Osher solution at** $t = 0.6$. Predicted density ρ, velocity u, and pressure p are compared with the WENO5 reference solution for FLINT, PINNs-WE, and WE-PINN. The top row shows the solution over the full spatial domain, while the bottom row enlarges the region behind the shock to display the oscillatory wave behavior.

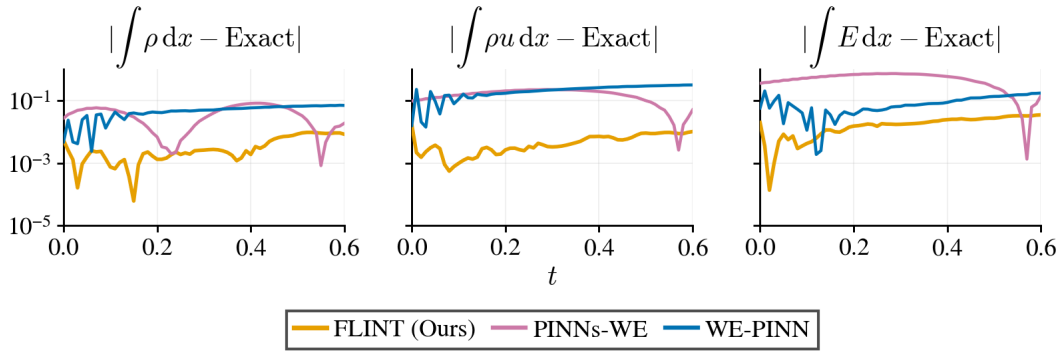


Figure 5: **Global integral error on the Shu–Osher problem.** Absolute errors between predictions and reference spatial integrals of mass, momentum, and total energy are shown over time on a logarithmic scale for FLINT, PINNs-WE, and WE-PINN.

4.3 Cylindrical Blast Wave

Finally, we consider the blast wave in cylindrical coordinates, where (1) is the general form of Euler's equations that includes geometric source terms for the curvilinear coordinates. This common setting

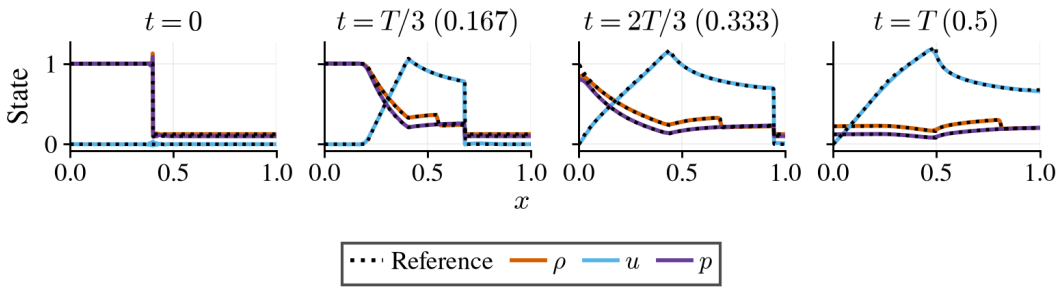


Figure 6: **2D Cylindrical blast wave in 1D radial form.** Solution plots of density ρ, radial velocity u, and pressure p are plotted against the radial coordinate x.

may be considered a simple "multi-physics" problem where the geometric terms represent additional physics. We note that while the blast wave is a two-dimensional structure, we solve it in one spatial dimension by leveraging symmetry, and to our knowledge, it has not been attempted in the PINN literature. The initial conditions comprise an over-pressured region in $x < 0.4$:

$$(\rho, u, p)(x, 0) = \begin{cases} (1, 0, 1), & x < 0.4, \\ (0.125, 0, 0.1), & x \geq 0.4 \end{cases}. \tag{7}$$

We solve on the spacetime domain $(x, t) \in (0, 1] \times [0, 0.5]$. Note the open interval in x. $x = 0$ is a coordinate singularity, where equation (1) diverges when the geometric factors are included. Commensurately, the boundary condition at $x = 0$ becomes a *reflecting* mixed Dirichlet-von Neumann boundary condition such that $\partial_x \rho(x = 0, t) = \partial_x E(x = 0, t) = 0$ and $u(x = 0, t) = 0$. Figure 6 shows primitive variables at four different times as a function of x. The shock, rarefaction, and contact are all well-captured, and each solution shows the expected curvature due to geometric factors. **As the curvilinear form of the Euler equations required for this problem are not implemented in our baselines, we are unable to show them for comparison.** Quantitative errors over five random seeds are reported in Appendix A.4.

5 Conclusion, Limitations, and Outlook

To summarize, we have presented FLINT, a novel PINN framework that introduces initial-data anchored control volumes and a shock-selective residual weighting. We combine this technology with prior innovations and show it outperforms relevant baselines on several novel and under-explored problems that stress a network's ability to capture both sharp features and smooth fine-scale structure.

The computational expense of FLINT (and all PINNs) is still large compared to a traditional (e.g., finite volumes) solver. The control volume method also grows more expensive if more spatial structure needs to be captured, as the number of sample points must be large enough to sample that structure. For example, training the Shu-Osher problem required $\sim 10^5$ samples. However, we note as PINN architectures randomly sample the space, batch accuracy likely scales as shot noise, $1/\sqrt{N}$ for N samples. This is independent of dimension, in contrast to mesh-based methods, where the cost scales exponentially with dimension. In this framing a PINN may be interpreted as a Monte Carlo method (Caflisch, 1998), and FLINT may begin to outperform grid-based solvers for high-dimensional PDEs where a grid may be prohibitively expensive. Some traditional implicit solvers enforce conservation by the topology of their differential operators, rather than in a loss term. (See Press et al. (2007), and Toro (2009), and the many references therein.) These methods can achieve global conservation that is accurate up to machine precision. **The PINN-based approach, including ours, is less amenable to this treatment**, as the differential operator is not a stencil.

Our method does not contain **any explicit treatment for entropy**, which would be naturally incorporated via numerical dissipation in, e.g., a finite volume solver. Incorporating, e.g., the Rankine (1870)-Hugoniot (1887) conditions (Hugoniot, 1889). However, Section A.2, shows one example of FLINT choosing the entropy-admissible solution. Finally, while many hyperparameters generalize across problems, as is state-of-the-art in the PINN literature, others **hyperparameters** were hand-tuned per-problem. (We refer the reader to Table 10 in Appendix F.) Formulations that admit hyperparameters that generalize more broadly remain a goal.

AI USE STATEMENT

In this work, we used generative AI tools to assist with the verification of FLINT via unit tests, as well as to assist with plotting. Baseline codes were not always turn-key and AI tools were used to debug, verify, and configure them. We also used generative AI tools to assist with literature search and to identify potentially relevant prior work for the authors to review. All AI-assisted code and diagnostic tools were reviewed and tested by the authors, and all literature recommendations were independently checked against the original sources. The authors made all final decisions regarding the methodology, experiments, analysis, and conclusions and take responsibility for the full content of this work.

REPRODUCIBILITY STATEMENT

We have taken several steps to support reproducibility. The implementation has been streamlined into a unified codebase with shared components across the benchmark problems and baseline evaluations. The main text specifies the governing equations, benchmark definitions, and training objective, while Appendix F reports the model, sampling, control-volume, loss, and optimization hyperparameters used for each experiment. Appendix E defines the evaluation metrics, and the remaining appendices provide additional benchmark results and ablations. We also provide a notebook that documents the workflow and implementation details for one of the spotlight results. During development, the authors reviewed the code and plotting scripts where applicable to check that the implemented experiments were consistent with the intended methodology.

REFERENCES

Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: A survey. *Journal of Machine Learning Research*, 18(153):1–43, 2018.

Richard M Beam and R.F Warming. An implicit finite-difference algorithm for hyperbolic systems in conservation-law form. *Journal of Computational Physics*, 22(1):87–110, 1976. ISSN 0021-9991. doi: https://doi.org/10.1016/0021-9991(76)90110-8. URL https://www.sciencedirect.com/science/article/pii/0021999176901108.

John P. Boyd. *Chebyshev and Fourier Spectral Methods*. Dover Publications, 2nd, revised edition, 2001. ISBN 9780486411835.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Yash Katariya, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: Composable transformations of Python+NumPy programs, 2018. URL https://github.com/jax-ml/jax.

Russel E. Caflisch. Monte carlo and quasi-monte carlo methods. *Acta Numerica*, 7:1–49, 1998. doi: 10.1017/S0962492900002804.

Charles F. Curtiss and Joseph O. Hirschfelder. Integration of stiff equations. *Proceedings of the National Academy of Sciences*, 38(3):235–243, 1952. doi: 10.1073/pnas.38.3.235.

Constantine M. Dafermos. *Hyperbolic Conservation Laws in Continuum Physics*, volume 325 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 4th edition, 2016. ISBN 9783662494493. doi: 10.1007/978-3-662-49451-6.

Arka Daw, Jie Bu, Sifan Wang, Paris Perdikaris, and Anuj Karpatne. Mitigating propagation failures in physics-informed neural networks using retain-resample-release (R3) sampling. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 7264–7302. PMLR, 2023. URL https://proceedings.mlr.press/v202/daw23a.html.

Tim De Ryck, Siddhartha Mishra, and Roberto Molinaro. wpinns: Weak physics informed neural networks for approximating entropy solutions of hyperbolic conservation laws. *SIAM Journal on Numerical Analysis*, 62(2):811–841, 2024. doi: 10.1137/22M1522504.

Pierre Degond. *Macroscopic limits of the Boltzmann equation: a review*, pp. 3–57. Birkhäuser Boston, Boston, MA, 2004. ISBN 978-0-8176-8200-2. doi: 10.1007/978-0-8176-8200-2_1. URL https://doi.org/10.1007/978-0-8176-8200-2_1.

Leonhard Euler. *Institutionum Calculi Integralis, Volumen Primum*. Academia Imperialis Scientiarum, Petropoli, 1768. Problem 85, Sections 650–653, pp. 424–425.

Gerald B. Folland. *Introduction to Partial Differential Equations*. Princeton University Press, 2nd edition, 1995. ISBN 9780691043616. URL https://press.princeton.edu/books/hardcover/9780691043616/introduction-to-partial-differential-equations.

Qi Gao, Kuang Huang, and Xuan Di. Efficient weak-entropy pinn for solving hyperbolic conservation laws, 2026.

Mahshid Sadat Ghoreishi and Hamid Naderan. Physics-informed neural network with weighted loss and hard constraints for hyperbolic conservation laws. *Scientific Reports*, 16:4201, 2026. doi: 10.1038/s41598-025-34263-1.

J. Willard Gibbs. Fourier's series. *Nature*, 59:200, 1898. doi: 10.1038/059200b0. URL https://doi.org/10.1038/059200b0.

S. K. Godunov. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Matematicheskii Sbornik*, 47(89):271–306, 1959.

Sigal Gottlieb, Chi-Wang Shu, and Eitan Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Review*, 43(1):89–112, 2001. doi: 10.1137/S003614450036757X.

Ami Harten, Björn Engquist, Stanley Osher, and Sukumar R. Chakravarthy. Uniformly high order accurate essentially non-oscillatory schemes, iii. *Journal of Computational Physics*, 71(2):231–303, 1987. doi: 10.1016/0021-9991(87)90031-3.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016a.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pp. 630–645, 2016b. doi: 10.1007/978-3-319-46493-0_38.

Michael T. Heath. *Scientific Computing: An Introductory Survey*. Society for Industrial and Applied Mathematics, Philadelphia, revised second edition, 2018. ISBN 978-1-61197-557-4. doi: 10.1137/1.9781611975581.

Pierre Henri Hugoniot. Mémoire sur la propagation du mouvement dans les corps et plus spécialement dans les gaz parfaits (première partie). *Journal de l'École Polytechnique*, (57):3–97, 1887.

Pierre Henri Hugoniot. Mémoire sur la propagation du mouvement dans les corps et plus spécialement dans les gaz parfaits (deuxième partie). *Journal de l'École Polytechnique*, (58):1–125, 1889.

Guang-Shan Jiang and Chi-Wang Shu. Efficient implementation of weighted eno schemes. *Journal of Computational Physics*, 126(1):202–228, 1996. doi: 10.1006/jcph.1996.0130.

Haina Jiang, Liam Wang, Peng-Chen Chen, Min Seop Kwak, Seungryong Kim, Brian Bell, and Jeong Joon Park. Error-conditioned neural solvers, 2026.

Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, Robert M. Kirby, and Michael W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems*, volume 34, pp. 26548–26560, 2021.

P. D. Lax. Hyperbolic systems of conservation laws ii. *Communications on Pure and Applied Mathematics*, 10(4):537–566, 1957. doi: https://doi.org/10.1002/cpa.3160100406. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160100406.

Peter D. Lax. *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, volume 11 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, 1973. ISBN 9780898711776. doi: 10.1137/1.9781611970562.

Philippe Le Floch. Entropy weak solutions to nonlinear hyperbolic systems under nonconservative form. *Communications in Partial Differential Equations*, 13(6):669–727, 1988. doi: 10.1080/03605308808820557. URL https://doi.org/10.1080/03605308808820557.

Randall J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2002. ISBN 9780521009249. doi: 10.1017/CBO9780511791253.

Li Liu, Shengping Liu, Hui Xie, Fansheng Xiong, Tengchao Yu, Mengjuan Xiao, Lufeng Liu, and Heng Yong. Discontinuity computing using physics-informed neural networks. *Journal of Scientific Computing*, 98(1):22, 2024. doi: 10.1007/s10915-023-02412-1.

Xu-Dong Liu, Stanley Osher, and Tony Chan. Weighted essentially non-oscillatory schemes. *Journal of Computational Physics*, 115(1):200–212, 1994. doi: 10.1006/jcph.1994.1187.

Lu Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis. DeepXDE: A deep learning library for solving differential equations. *SIAM Review*, 63(1):208–228, 2021. doi: 10.1137/19M1274067.

Albert A. Michelson. Fourier's series. *Nature*, 59:200, 1898. doi: 10.1038/059200a0. URL https://doi.org/10.1038/059200a0.

Ismail Oubarka, Imad Kissami, Mohamed Boubekeur, Fayssal Benkhaldoun, Aziz Madrane, and Zakaria Saadi. Weak and entropy physics-informed neural networks for conservation laws, 2026.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, pp. 8024–8035. Curran Associates, Inc., 2019.

Ravi G. Patel, Indu Manickam, Nathaniel A. Trask, Mitchell A. Wood, Myoungkyu Lee, Ignacio Tomas, and Eric C. Cyr. Thermodynamically consistent physics-informed neural networks for hyperbolic systems. *Journal of Computational Physics*, 449:110754, 2022. doi: 10.1016/j.jcp.2021.110754.

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3 edition, 2007. ISBN 978-0-521-88068-8.

Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

William John Macquorn Rankine. On the thermodynamic theory of waves of finite longitudinal disturbance. *Philosophical Transactions of the Royal Society of London*, 160:277–288, 1870. doi: 10.1098/rstl.1870.0015.

Luciano Rezzolla and Olindo Zanotti. *Relativistic Hydrodynamics*. Oxford University Press, 2013. ISBN 978-0-19-852890-6. doi: 10.1093/acprof:oso/9780198528906.001.0001.

Chi-Wang Shu and Stanley Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes, ii. *Journal of Computational Physics*, 83(1):32–78, 1989. doi: 10.1016/0021-9991(89)90222-2.

P. K. Sweby. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM Journal on Numerical Analysis*, 21(5):995–1011, 1984. doi: 10.1137/0721062.

E. F. Toro, M. Spruce, and W. Speares. Restoration of the contact surface in the HLL-Riemann solver. *Shock Waves*, 4(1):25–34, 1994. doi: 10.1007/BF01414629.

Eleuterio F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*. Springer, Berlin, Heidelberg, 3 edition, 2009. ISBN 978-3-540-49834-6. doi: 10.1007/b79761.

Bram van Leer. Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. *Journal of Computational Physics*, 32(1):101–136, 1979. doi: 10.1016/0021-9991(79)90145-1.

A I Vol'pert. The spaces and quasilinear equations. *Mathematics of the USSR-Sbornik*, 2(2):225, feb 1967. doi: 10.1070/SM1967v002n02ABEH002340. URL https://doi.org/10.1070/SM1967v002n02ABEH002340.

Nikhil Vyas, Depen Morwani, Rosie Zhao, Itai Shapira, David Brandfonbrener, Lucas Janson, and Sham Kakade. SOAP: Improving and stabilizing shampoo using adam for language modeling. In *International Conference on Learning Representations*, 2025.

Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021. doi: 10.1137/20M1318043.

Sifan Wang, Ananyae Kumar Bhartari, Bowen Li, and Paris Perdikaris. Gradient alignment in physics-informed neural networks: A second-order optimization perspective. In *Advances in Neural Information Processing Systems*, volume 38, pp. 186682–186732, 2025. doi: 10.52202/085713-5615.

Sifan Wang, Shawn Koohy, Yiping Lu, and Paris Perdikaris. When PINNs go wrong: Pseudo-time stepping against spurious solutions. *arXiv preprint arXiv:2604.23528*, 2026. doi: 10.48550/arXiv.2604.23528. URL https://arxiv.org/abs/2604.23528.

Tao Wang, Bo Zhao, Sicun Gao, and Rose Yu. Understanding the difficulty of solving Cauchy problems with PINNs. In *Proceedings of the 6th Annual Learning for Dynamics and Control Conference*, volume 242 of *Proceedings of Machine Learning Research*, pp. 453–465. PMLR, 2024.

Yeping Wang and Shihao Yang. Coupled integral pinn for discontinuity, 2026.

Chenhui Xu, Dancheng Liu, Amir Nassereldine, and Jinjun Xiong. FP64 is all you need: Rethinking failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-4785.

Darui Zhao, Ze Tao, and Fujun Liu. Spatio-temporal uncertainty-modulated physics-informed neural networks for solving hyperbolic conservation laws with strong shocks, 2026.

A Additional Experiments on Hyperbolic PDEs Across Various Configurations

A.1 Frequency-controlled Sod-k benchmark

We use Sod-k to test how the methods respond as the oscillatory structure in the initial condition becomes increasingly fine. We vary only the frequency parameter in Equation (5) and consider $k \in \{5, 20, 100\}$. Within each method, all remaining problem and training settings are held fixed across the three frequencies. FLINT is trained with SOAP, while PINNs-WE and WE-PINN use their respective default optimization settings.

These experiments provides a controlled increase in the amount of fine-scale structure that must be captured while leaving the underlying shock-tube problem unchanged. We evaluate density, velocity, and pressure over the full space–time domain, the final-time T, and within the region between the contact and shock, where the wave structures of the problem are located.

Observed Solution Quality at Final Time T Figure 7 compares the final-time density, velocity, and pressure with the WENO5 reference solution. At $k = 5$, all three PINN methods recover the large-scale solution, including the contact and shock locations. Differences are already visible in the enlarged wave region, where FLINT more closely follows the reference across the three primitive variables.

At $k = 20$, the distinction becomes more clear. FLINT continues to closely match the oscillatory structure in density, velocity, and pressure, while PINNs-WE and WE-PINN attenuate the wave structure. Observing the full domain, their solutions remain plausible and retain the discontinuities, indicating that the growing changes are concentrated in the fine-scale structure rather than in the large-scale.

The differences are strongly observed at $k = 100$. FLINT closely matches the repeated oscillations between the contact and shock with much better phase and amplitude agreement in comparison WENO5. PINNs-WE and WE-PINN still capture the coarse shock-tube structure but produce flatter solution curves throughout the same wave region. Increasing k therefore separates the ability to model the large-scale discontinuities from the ability to represent the finer structure carried between them.

Performance across frequency Table 3 quantifies the behavior observed in Figure 7. FLINT obtains the lowest value in every reported metric and primitive variables across all three frequencies. The difference becomes more clear in the wave-region and amplitude metrics, which isolate the ability to model the oscillatory structure from the surrounding shock-tube solution.

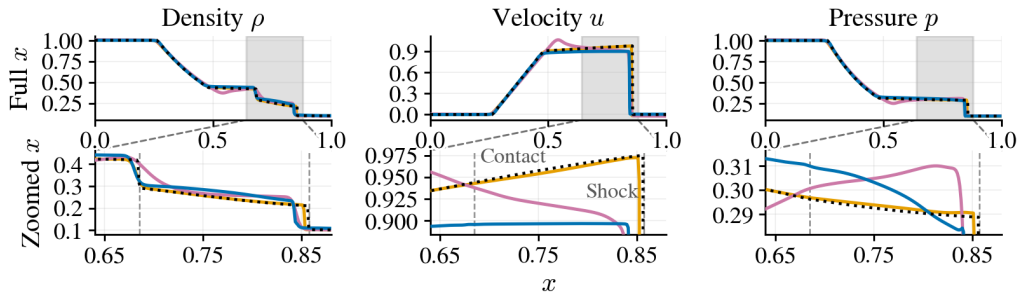
A.2 Inviscid Burgers Equation

We evaluate FLINT on inviscid Burgers' equation, a standard benchmark for shock-capturing PINNs (Liu et al., 2024; Zhao et al., 2026; Toro, 2009). This provides a simpler test of shock modeling and conservation before graduating to the Euler experiments. We solve

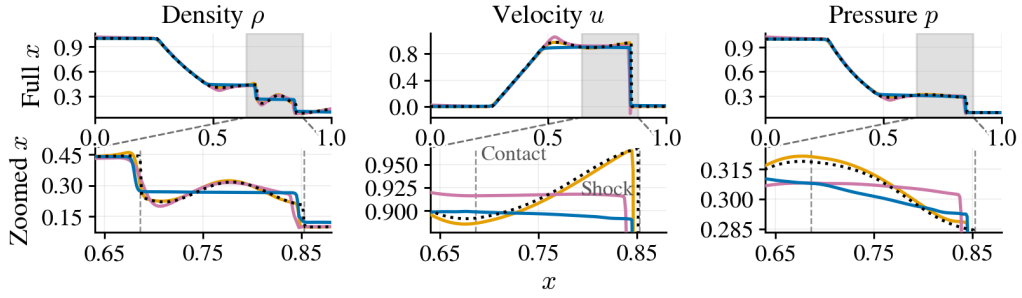
$$\partial_t u + \partial_x \left(\tfrac{1}{2} u^2\right) = 0, \tag{8}$$

on $x \in [0, 2]$ and $t \in [0, 1]$, with $u(x, 0) = -\sin(\pi(x - 1))$ and $u(0, t) = u(2, t) = 0$. The smooth initial condition develops a stationary entropy shock at $x = 1$ after the time $t = 1/\pi$. We generate the reference solution using a first-order Godunov method (Godunov, 1959).

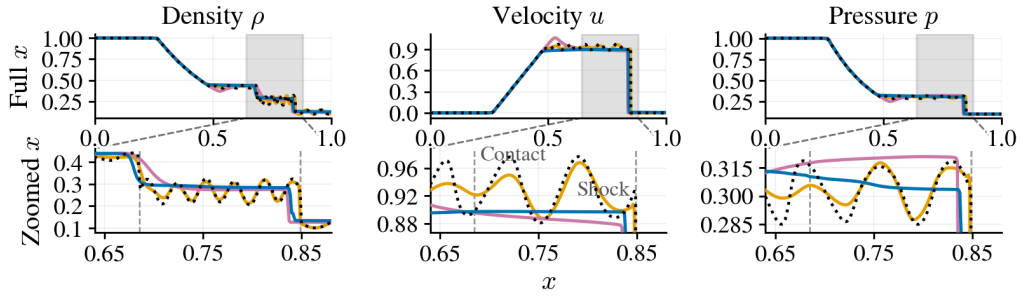
FLINT most accurately represents the shock, while all three methods agree closely in the smooth regions. Figure 8 shows that FLINT closely matches the reference shock at $t = 1$, whereas PINNs-WE slightly misplaces the shock front and WE-PINN produces a wider and more smooth transition. The corresponding space–time L_2 errors are 0.016, 0.038, and 0.083 for FLINT, PINN-WE, and WE-PINN, respectively, with the largest differences concentrated near the discontinuity.



(a) Final-time solution for $k = 5$.



(b) Final-time solution for $k = 20$.



(c) Final-time solution for $k = 100$.

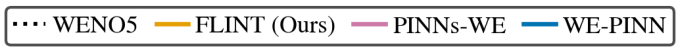


Figure 7: **Solution quality at final time $T = 0.2$ as the initial-condition frequency increases.** Density ρ, velocity u, and pressure p are shown for $k \in \{5, 20, 100\}$. In each subfigure, the upper row shows the full spatial domain and the lower row enlarges the shaded interval. Dashed vertical lines in the enlarged views mark the reference contact and shock locations. The dotted black curve is the WENO5 reference solution. The models compared are FLINT, PINNs-WE, and WE-PINN.

FLINT also preserves the global conservation more accurately. For this problem, zero boundary flux and a zero initial integral imply the exact conservation law $\int_0^2 u(t,x)\,\mathrm{d}x = 0$ for all time. As shown in Figure 9, the maximum absolute conservation error is 5.5×10^{-4} for FLINT, compared with 1.8×10^{-3} for PINN-WE and 1.4×10^{-2} for WE-PINN. Thus, on this standard benchmark, FLINT both represents the shock more accurately and better preserves the conserved quantity.

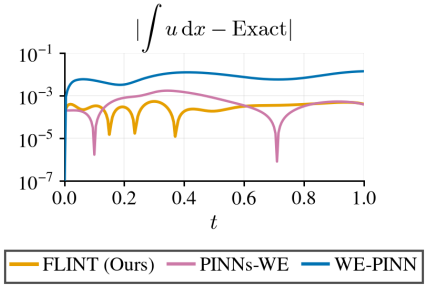


Figure 9: **Global conservation on Burgers.** Absolute error in $\int_0^2 u(t,x)\,\mathrm{d}x$ over time.

Table 3: **Solution and wave-capturing performance across the initial-condition frequency k.** All metrics are defined in Appendix E. ↓ lower is better and the bold font represents the **best** per column within each k. Each entry is the mean ± the standard deviation over 3 seeds.

	Space–time L_2 ↓			L_2 at T ↓			Wave-region L_2 at T ↓			Wave amplitude↓		
Method	ρ	u	p	ρ	u	p	ρ	u	p	ρ	u	p
$k = 5$												
PINNs-WE	0.031 ±0.004	0.177 ±0.050	0.034 ±0.004	0.039 ±0.004	0.172 ±0.087	0.045 ±0.016	0.117 ±0.028	0.128 ±0.120	0.099 ±0.071	0.840 ±0.445	11.979 ±19.291	11.333 ±19.152
WE-PINN	0.016 ±0.001	0.144 ±0.007	**0.020** ±0.001	0.028 ±0.001	0.188 ±0.008	0.038 ±0.002	0.107 ±0.005	0.070 ±0.003	0.032 ±0.001	0.063 ±0.053	0.949 ±0.023	3.126 ±0.199
FLINT (Ours)	**0.015** ±0.003	**0.128** ±0.044	0.020 ±0.004	**0.016** ±0.005	**0.117** ±0.045	**0.024** ±0.008	**0.010** ±0.009	**0.005** ±0.005	**0.007** ±0.008	**0.017** ±0.011	**0.197** ±0.282	**0.088** ±0.041
$k = 20$												
PINNs-WE	0.042 ±0.021	0.171 ±0.062	0.037 ±0.012	0.042 ±0.019	0.173 ±0.036	0.040 ±0.011	0.136 ±0.074	0.067 ±0.050	0.048 ±0.022	0.469 ±0.295	5.031 ±5.239	2.618 ±2.113
WE-PINN	0.027 ±0.000	0.110 ±0.003	0.019 ±0.000	0.034 ±0.001	**0.101** ±0.015	0.024 ±0.002	0.130 ±0.001	0.036 ±0.005	0.028 ±0.003	0.933 ±0.013	0.923 ±0.027	0.634 ±0.076
FLINT (Ours)	**0.016** ±0.001	**0.106** ±0.011	**0.019** ±0.003	**0.016** ±0.001	0.107 ±0.016	**0.022** ±0.002	**0.020** ±0.007	**0.004** ±0.001	**0.005** ±0.002	**0.040** ±0.020	**0.085** ±0.015	**0.032** ±0.033
$k = 100$												
PINNs-WE	0.050 ±0.024	0.175 ±0.035	0.038 ±0.010	0.060 ±0.021	0.168 ±0.023	0.046 ±0.009	0.222 ±0.049	0.087 ±0.024	0.089 ±0.014	0.517 ±0.304	5.905 ±2.718	4.168 ±1.489
WE-PINN	0.028 ±0.001	0.140 ±0.012	0.022 ±0.002	0.044 ±0.001	0.154 ±0.011	0.037 ±0.003	0.171 ±0.006	0.056 ±0.007	0.048 ±0.012	0.550 ±0.412	1.470 ±0.906	1.506 ±1.323
FLINT (Ours)	**0.017** ±0.003	**0.113** ±0.025	**0.018** ±0.001	**0.023** ±0.005	**0.101** ±0.014	**0.024** ±0.003	**0.075** ±0.038	**0.014** ±0.003	**0.013** ±0.001	**0.144** ±0.023	**0.148** ±0.063	**0.109** ±0.021

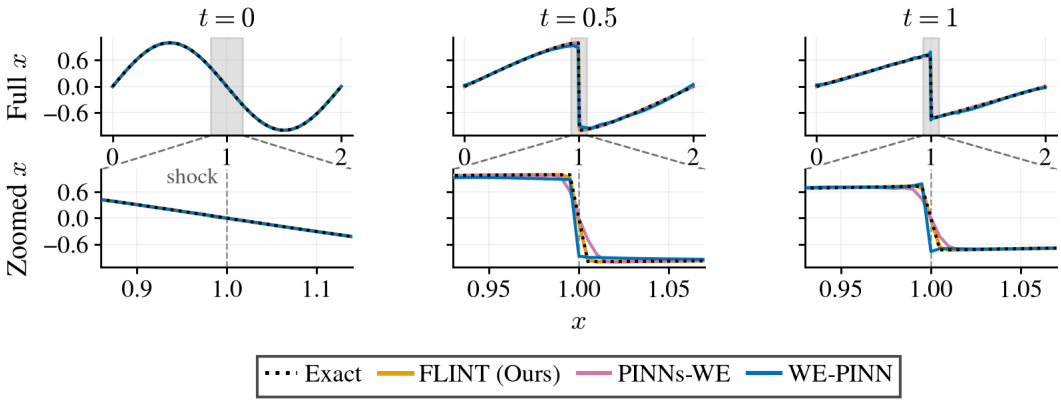


Figure 8: **Shock formation on the inviscid Burgers benchmark.** Solutions at $t = 0$, 0.5, and 1. The top row shows the full domain, while the bottom row magnifies the region around the stationary shock at $x = 1$.

Entropy admissibility without an explicit entropy constraint. We next test whether FLINT selects the entropy-admissible solution when conservation alone does not uniquely determine the physical solution. Importantly, FLINT does not include an explicit entropy condition in its training objective.

We solve the inviscid Burgers equation

$$\partial_t u + \partial_x \left(\frac{1}{2} u^2 \right) = 0, \qquad (x, t) \in [0, 2] \times [0, 0.5], \tag{9}$$

with

$$u(x, 0) = \begin{cases} 0, & x < 1, \\ 1, & x \geq 1, \end{cases} \qquad u(0, t) = 0, \qquad u(2, t) = 1. \tag{10}$$

Because $u_L < u_R$, the conservation law admits both the physical rarefaction and a non-entropic expansion shock as weak solutions (Toro, 2009). The latter propagates at the Rankine–Hugoniot

Table 4: **Relative L_2 error on the Burgers benchmark.** We report error over the full space–time domain and at $t = 0$, 0.5, and 1. Because Burgers has a single state variable, the evaluation columns correspond to times rather than different fields. WE-PINN has zero error at $t = 0$ because the initial condition is enforced exactly by construction. ↓ indicates lower is better bold font represents the **best** per column.

Method	Space–time↓	$t = 0$ (IC)↓	$t = 0.5$ ↓	$t = 1$ ↓
PINNs-WE	0.0376	0.0028	0.0509	0.0512
WE-PINN	0.0831	**0.0001**	0.0804	0.0970
FLINT (Ours)	**0.0156**	0.0008	**0.0114**	**0.0089**

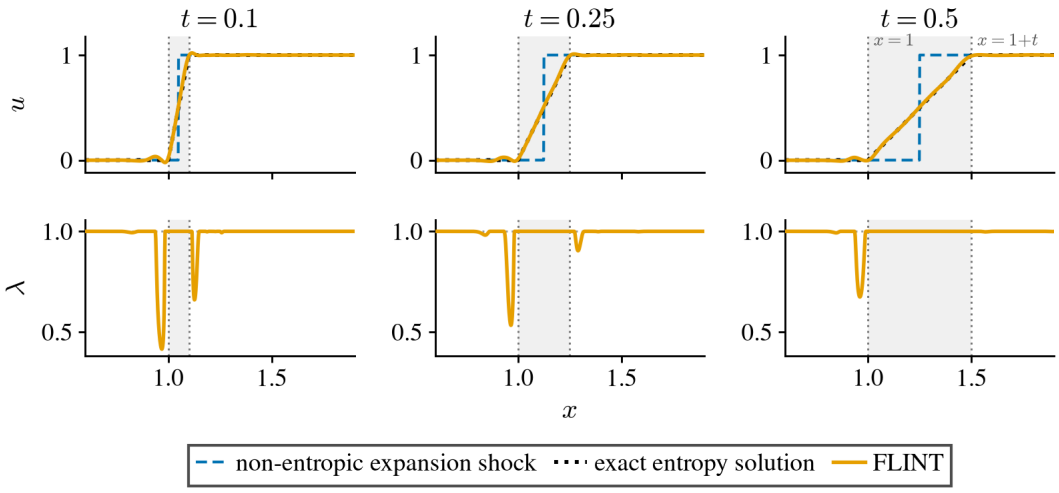


Figure 10: **Entropy selection on the Burgers rarefaction problem.** Top row shows the exact entropy solution, the non-entropic expansion shock, and the FLINT prediction at $t = 0.1$, 0.25, and 0.5. The shaded region marks the exact rarefaction fan $x \in [1, 1 + t]$. Bottom row shows the corresponding FLINT residual weight λ.

speed (Rankine, 1870; Hugoniot, 1887; 1889)

$$s = \frac{f(u_R) - f(u_L)}{u_R - u_L} = \frac{u_L + u_R}{2} = \frac{1}{2}. \tag{11}$$

The entropy-admissible solution is instead

$$u(x,t) = \begin{cases} 0, & x < 1, \\ \dfrac{x-1}{t}, & 1 \le x \le 1+t, \\ 1, & x > 1+t. \end{cases} \tag{12}$$

The reference solution is available in closed form, so no numerical reference solver is required. Figure 10 shows that FLINT selects the entropy-admissible rarefaction rather than the competing expansion shock. This happens despite that there is no explicit entropy constraint during training. The residual weights provide insight into this behavior. Throughout the expanding fan, $u_x > 0$, so the compression detector is inactive and $\lambda = 1$. The strong-form PDE residual therefore remains fully enforced throughout the rarefaction. Together with the compressive-shock experiment above, this shows that FLINT treats compressive and expansive discontinuities differently in a manner consistent with entropy admissibility.

A.3 Sod Shock Tube

We also evaluate FLINT on the 1D Sod shock tube, another standard benchmark for discontinuity-capturing PINNs (Liu et al., 2024; Zhao et al., 2026). The Sod shock tube involves multiple primitive

variables, density ρ, velocity u, and pressure p, which introduces new types of discontinuities (contact discontinuity) and regions where the PDE residuals become large (rarefaction fans) (Liu et al., 2024; Zhao et al., 2026; Toro, 2009). The equations are defined in (1). The initial conditions are given by:

$$(\rho,\ u,\ p)(0,x) = \begin{cases} (1,\ 0,\ 1), & x < 0.5, \\ (0.125,\ 0,\ 0.1), & x \geq 0.5, \end{cases} \tag{13}$$

where $x \in [0,1]$ and $t \in [0, 0.2]$. The experimental setup starts with a thin partition at $x = 0.5$. When the thin partition is removed it sends a rarefaction left, and a contact discontinuity and a shock to the right. We generate the reference with the exact Riemann solution found in Toro (2009).

All three methods produce solutions that closely resemble the reference, but the zoomed in regions in Figure 11 show differences near the discontinuities. FLINT and WE-PINN closely follow the reference solution, while PINNs-WE produces a broader contact transition and places the shock to the left of the shock location in the reference solution.

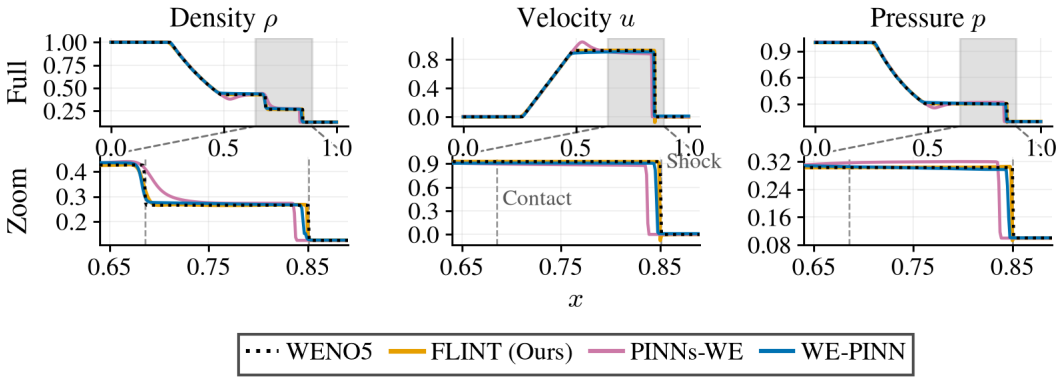


Figure 11: **Sod shock-tube solution plot at $t = 0.2$.** Predicted density ρ, velocity u, and pressure p for FLINT, PINNs-WE, and WE-PINN are compared with the exact Riemann reference solution. The upper row shows the full spatial domain, and the lower row enlarges the shaded interval. Dashed vertical lines in the enlarged views mark the reference contact and shock locations.

The global integral error curves in Figure 12 show a consistently low error for FLINT. FLINT maintains global integral mass, momentum, and energy errors of order 10^{-4} over the observed interval. In comparison, the global integral mass and momentum errors of WE-PINN increase toward 10^{-3} at later times and PINNs-WE has inconsistent behavior throughout the observed interval. Individual baseline errors occasionally fall below those of FLINT, but FLINT maintains consistently small errors across all three conserved quantities rather than only at a few times.

Finally, Table 5 quantifies the observations from the plot. FLINT yields consistently better performance over the baseline methods.

A.4 Cylindrical Blast Wave Quantitative Results

Table 6 reports the relative L_2 error over the full space–time domain and at $T/3$, $2T/3$, and T. Results are averaged over five random seeds. Across all reported variables and evaluation times, the mean relative L_2 error remains below 0.1. The standard deviations quantify variation across the five independent training runs.

B Ablation: PDE Residuals Can Be Insensitive to Wave Error

Representing fine-scale structure requires more than obtaining a small PDE residual. Prior PINN studies report propagation failures during training (Daw et al., 2023) and analyze spurious solutions admitted by residual losses evaluated at finite collocation sets (Wang et al., 2026). To show this issue independently of shocks and the components of FLINT, we observed the linear-advection problem with increasing wave frequency and several loss-weight configurations (Equation (4)).

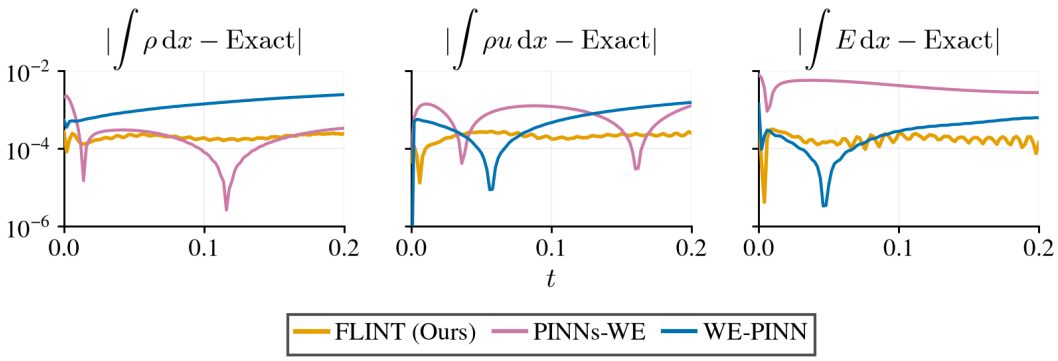


Figure 12: **Global integral error on the Sod shock tube.** Absolute errors in total mass, momentum, and energy are shown over $t \in [0, 0.2]$ for FLINT, PINNs-WE, and WE-PINN. Each error compares the spatial integral of the predicted conserved variable, ρ, ρu, or E, over $x \in [0, 1]$ with its corresponding exact value at that time. The vertical axes use a logarithmic scale.

Table 5: The solution for the Sod shock tube is evaluated over the full space-time domain and at $t = 0.1$ and $t = 0.2$, which are the intermediate and final times, respectively. The original Sod shock tube does not contain any sinusoidal waves in the density initial conditions so we do not report on the ability to capture wave-structure. ↓ means lower is better and bold fond represents the **best** per column. Each entry is the mean ± the standard deviation over 3 seeds

	Space–time $L_2 \downarrow$			L_2 at $t = 0.1 \downarrow$			L_2 at $t = 0.2 \downarrow$		
Method	ρ	u	p	ρ	u	p	ρ	u	p
PINNs-WE	0.042 ±0.021	0.173 ±0.040	0.037 ±0.009	0.041 ±0.018	0.174 ±0.077	0.036 ±0.013	0.048 ±0.019	0.175 ±0.021	0.044 ±0.008
WE-PINN	0.014 ±0.002	0.101 ±0.011	**0.016** ±0.002	0.014 ±0.002	0.107 ±0.015	0.017 ±0.002	0.022 ±0.003	0.112 ±0.014	0.026 ±0.003
FLINT (Ours)	**0.014** ±0.005	**0.089** ±0.009	0.017 ±0.005	**0.013** ±0.003	**0.078** ±0.032	**0.015** ±0.003	**0.013** ±0.002	**0.065** ±0.018	**0.016** ±0.003

Table 6: **Solution quality on the cylindrical blast wave.** The metrics are defined in Appendix E. The performance is measured against a MUSCL-HLLC reference over the radial domain $r \in [0, 1]$. We evaluate the on the full space-time domain, $t = T/3, 2T/3$ and T. ↓ means lower is better. The results are the mean ± the standard deviation over 5 seeds.

	Space–time $L_2 \downarrow$			L_2 at $T/3 \downarrow$			L_2 at $2T/3 \downarrow$			L_2 at $T \downarrow$		
Method	ρ	u	p	ρ	u	p	ρ	u	p	ρ	u	p
FLINT (Ours)	0.021 ±0.003	0.063 ±0.022	0.026 ±0.004	0.017 ±0.004	0.083 ±0.036	0.020 ±0.006	0.050 ±0.006	0.068 ±0.037	0.074 ±0.009	0.023 ±0.005	0.017 ±0.001	0.019 ±0.006

B.1 Linear-Advection Experiment

We solve

$$u_t + cu_x = 0, \qquad (x, t) \in (-6, 6) \times (0, 2), \qquad c = 1, \tag{14}$$

with initial condition

$$u_0(x) = w(x)\,(\sin x + A \sin kx)\,, \qquad A = \frac{1}{2}, \tag{15}$$

and compactly supported window

$$w(x) = \begin{cases} \cos^2\left(\dfrac{\pi x}{2R}\right), & |x| \le R, \\ 0, & |x| > R, \end{cases} \qquad R = 3. \tag{16}$$

The window is C^1, and the exact solution is

$$u^\star(x, t) = u_0(x - ct). \tag{17}$$

The packet remains inside the domain for $0 \leq t \leq 2$ and satisfies the homogeneous boundary conditions $u^\star(\pm 6, t) = 0$.

We train vanilla PINNs (Raissi et al., 2019) without discontinuity residual weighting or control-volume constraints using

$$\mathcal{L} = \lambda_{\text{IC}}\mathcal{L}_{\text{IC}} + \lambda_{\text{PDE}}\mathcal{L}_{\text{PDE}} + \lambda_{\text{BC}}\mathcal{L}_{\text{BC}}. \tag{18}$$

To create a robust experiment, we vary the loss weights for the initial-conditions and PDE residuals. We consider

$$(\lambda_{\text{IC}}, \lambda_{\text{PDE}}) \in \{(0.5, 1), (10, 1), (0.5, 10), (1, 10)\}. \tag{19}$$

For each loss-weight configuration, we vary k in (15) while keeping the remaining experimental settings fixed, except for the number of collocation points. We increase the point count at higher frequencies to reduce the risk of missing the finer waves through insufficient sampling. This helps separate the observed residual–error mismatch from the limitations of a fixed sampling configuration. Finally, at the final time T, we compare L_2 error and the reported PDE residual.

B.2 Discrepancies between PDE Residuals and Error Across Frequencies and Loss Weights

Figure 13 shows that the same behavior persists for all four loss-weight configurations. From $k = 1$ to $k = 512$, the error of the solution at the final time increases, while the observed PDE residual is consistently small. Changing the loss weights affects accuracy, particularly at low and intermediate frequencies, but does not affect the high-frequency discrepancy in these experiments.

Thus, the observed gap is not confined to the original choice of the loss weights and a small reported residual is not sufficient to indicate accurate wave recovery across the tested configurations.

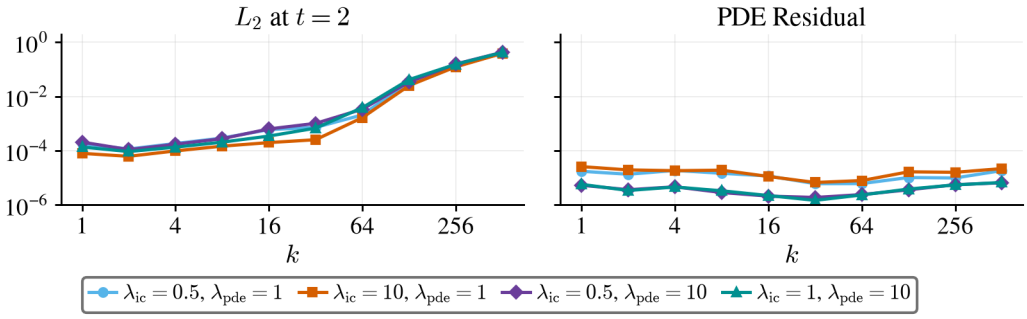


Figure 13: **Advection accuracy across frequencies and loss weights.** The left panel shows L_2 solution error at $t = 2$, and the right panel shows the PDE residual, for frequencies from $k = 1$ to 512. Curves correspond to the four initial-condition and PDE loss-weight configurations in the shared legend. Both panels use logarithmic axes.

B.3 Classical Interpretation and Finite-Sampling Effects

The method of characteristics give the classical solution $u(x, t) = g(x - ct)$ for linear advection, for any (smooth) function g (Folland, 1995). The prescribed initial condition selects the function g. The following proposition shows that the PDE residual can remain zero even when a wave is weakened or removed.

Proposition B.1 (An incorrect wave can still satisfy the PDE)**.** *With w, A, and c defined above, let $\xi = x - ct$ and*

$$u_\alpha(x, t) = w(\xi)\left[\sin\xi + \alpha A \sin(k\xi)\right], \qquad \alpha \in [0, 1]. \tag{20}$$

Every u_α satisfies

$$\partial_t u_\alpha + c\,\partial_x u_\alpha = 0 \tag{21}$$

and the prescribed homogeneous boundary conditions. Nevertheless,

$$u_\alpha - u^\star = (\alpha - 1)Aw(\xi)\sin(k\xi). \tag{22}$$

Thus, $\alpha = 1$ gives the complete waveform, whereas $\alpha = 0$ removes the second oscillatory component entirely without changing the PDE residual.

Proof. Differentiating (20) gives

$$\begin{aligned}\partial_x u_\alpha &= w'(\xi)\,[\sin\xi + \alpha A \sin(k\xi)] \\ &\quad + w(\xi)\,[\cos\xi + \alpha A k \cos(k\xi)]\,, \qquad (23)\\ \partial_t u_\alpha &= -c\,w'(\xi)\,[\sin\xi + \alpha A \sin(k\xi)] \\ &\quad - c\,w(\xi)\,[\cos\xi + \alpha A k \cos(k\xi)]\,. \qquad (24)\end{aligned}$$

These terms cancel in $\partial_t u_\alpha + c\,\partial_x u_\alpha$. All fields vanish at $x = \pm 6$ because $w(x - ct) = 0$ there throughout the simulated time interval. Subtracting $u^\star = u_1$ gives Equation (22). □

The fields in this example satisfy the PDE everywhere, not just at the points used during training. The prescribed initial condition distinguishes these fields, but the PDE residual does not, regardless of its loss weight. The proposition therefore explains why waveform error need not produce a PDE residual.

A separate concern is aliasing. In spectral collocation, distinct Fourier modes can agree at every grid point (Boyd, 2001, Section 11.2). For example, let N be the number of equally spaced grid points on the interval $[0, 2\pi)$, and let x_j denote the location of point j. The points are located at $x_j = 2\pi j/N$, where $j = 0, \ldots, N-1$. For any positive integer frequency m,

$$\sin((m+N)x_j) = \sin(mx_j), \qquad j = 0, \ldots, N-1. \qquad (25)$$

The two waves have the same values at the sampled points but differ between them. This happens because of where the points are placed. In contrast, the fields in the proposition satisfy the PDE everywhere, not just at the sampled points.

Finite collocation points can also hide PDE violations. Under their stated assumptions, Wang et al. (2026) construct incorrect fields that satisfy the initial boundary conditions and have zero residual at a fixed collocation set, but not necessarily between its points. This differs from Proposition B.1, whose fields have zero residual everywhere but differ in their initial conditions.

Our loss-weight experiment demonstrates that residual accuracy and wave accuracy can be strongly separated across the tested frequencies and loss weights. The classical construction explains why the PDE residual alone need not detect the missing wave information, while finite-sampling literature identifies an additional limitation of empirical residual evaluation.

C Ablation Studies

C.1 Why Initial-Data Anchoring Matters

Anchoring ablation design. We test whether the benefit of the anchored control volumes comes from their geometry or from placing the prescribed initial state directly inside the conservation constraint. We keep the remaining training setup fixed and compare three control-volume configurations in Sod-k with $k = 50$. The term face represents any side of the control-volume boxes.

The first is the standard FLINT configuration. Each box begins at $t = 0$, and its lower temporal face is evaluated from the prescribed initial condition $Q_0(x)$. The second keeps exactly the same box geometry, but evaluates the lower face using the network prediction $Q_\theta(0, x)$. The third moves the boxes away from $t = 0$, so that the input to both temporal faces utilizes the network predictions. These comparisons help identify whether FLINT's improved performance comes from using exact initial data on the lower face, placing the boxes at $t = 0$, or both.

Exact versus predicted initial faces. Figure 14 compares the solutions at the final time T. Using the prescribed initial state on the lower face represents the oscillatory structure while maintaining the contact and shock locations. Replacing the same face with $Q_\theta(0, x)$ creates a solution that weakens the wave structure and increases the error near the discontinuities, despite leaving the control-volume locations and geometry unchanged. Lifting the boxes away from $t = 0$ produces

a similar degradation, especially in velocity u and pressure p. The predicted-face comparison is especially informative because the network is trained to satisfy the initial condition through $\mathcal{L}_{\mathrm{IC}}$ in (4). Thus, $Q_\theta(0, x) \approx Q_0(x)$ is encouraged independently of the control-volume loss. Nevertheless, using the network outputs as an input to the $t = 0$ temporal face does not produce the same result as using the prescribed state directly.

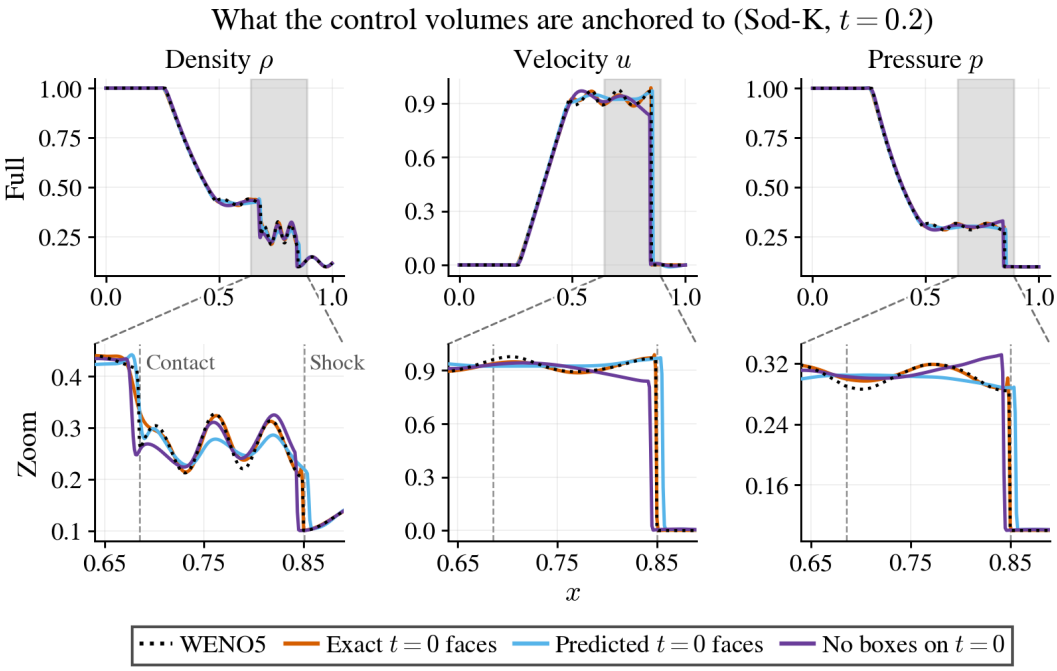


Figure 14: **Initial-data anchoring ablation on Sod-k with $k = 50$.** Density ρ, velocity u, and pressure p at the final time T are shown for three control-volume constructions. The first evaluates the lower face at $t = 0$ from the prescribed initial condition. The second uses the network prediction on the same $t = 0$ face. The third lifts the control volumes away from $t = 0$ so that both temporal faces are evaluated from network predictions. The dotted black curve is the WENO5 reference solution.

Effect of Initial-Data Anchoring On Solution Quality. Table 7 shows that the use of the exact initial data gives the lowest mean error in eight of the nine comparisons, including all space-time and final-time errors. At $t = 0.2$, the density error is 0.016, compared to 0.018 for predicted faces and 0.034 when no boxes are placed on $t = 0$. The exact-data configuration also achieves lower mean final-time errors in velocity and pressure, showing that the improvement extends across all three primitive variables. *Within the tested configurations, these results support using the prescribed initial data directly in the conservation constraints to improve solution accuracy.*

C.2 The Effects of Training Baselines with SOAP

In the main comparisons, each baseline is evaluated using its default optimizer and training settings, while FLINT is trained with SOAP (Vyas et al., 2025). To test whether FLINT's improvement is explained by the optimizer choice alone, we retrain PINNs-WE (Liu et al., 2024) and WE-PINN (Oubarka et al., 2026) with SOAP using the same number of optimization steps as FLINT. The remaining model formulations are unchanged.

Tables 8 and 9 show that matching the optimizer does not remove the performance gap. On both Sod-k and Shu–Osher, FLINT achieves the lowest mean error in every reported metric, including space–time error, final-time error, wave-region error, and wave-amplitude error. The advantage is becomes especially clear in the wave-region and amplitude metrics, which measure the representation of the oscillatory structure.

Table 7: **Performance of the anchored control-volume ablation for Sod-k ($k = 50$).** The bottom face of every box is the analytic integral of the exact initial data or the network's $t = 0$ prediction over that box. In the third case, no box lies on $t = 0$. The metrics are defined in Appendix E. We evaluate on the full space-time domain, $t = 0.1$ and $t = 0.2$, where the last two times are the intermediate and final times, respectively. ↓ means lower is better and bold font represents the **best** per column. Each entry is the mean ± the standard deviation over 3 seeds.

Model	$t = 0$ faces	Space–time L_2 ↓			L_2 at $t = 0.1$ ↓			L_2 at $t = 0.2$ ↓		
		ρ	u	p	ρ	u	p	ρ	u	p
FLINT (Ours)	Exact	**0.016** ±0.005	**0.102** ±0.021	**0.018** ±0.006	**0.018** ±0.005	0.110 ±0.031	**0.019** ±0.006	**0.016** ±0.005	**0.082** ±0.059	**0.018** ±0.010
	Predicted	0.018 ±0.003	0.125 ±0.037	0.019 ±0.005	0.021 ±0.006	0.125 ±0.077	0.021 ±0.010	0.022 ±0.003	0.127 ±0.029	0.026 ±0.005
	None	0.034 ±0.011	0.115 ±0.027	0.024 ±0.004	0.033 ±0.012	**0.102** ±0.066	0.022 ±0.007	0.034 ±0.014	0.098 ±0.050	0.026 ±0.008

These results show that SOAP alone does not explain FLINT's improved performance. Even when all methods are trained with the same optimizer and optimization budget, FLINT retains better performance across both benchmarks.

Table 8: **Solution and wave-capturing quality on Sod-k ($k = 50$).** All metrics are defined in Appendix E. All models are trained with the SOAP optimizer. ↓ means lower is better and bold font represents the **best** per column. Each entry is the mean ± the standard deviation over three seeds.

	Space–time L_2 ↓			L_2 at T ↓			Wave-region L_2 at T ↓			Wave amplitude↓		
Method	ρ	u	p	ρ	u	p	ρ	u	p	ρ	u	p
PINNs-WE	0.078 ±0.039	0.410 ±0.145	0.061 ±0.021	0.130 ±0.083	0.248 ±0.081	0.081 ±0.031	0.472 ±0.209	0.095 ±0.038	0.068 ±0.002	2.098 ±0.753	1.144 ±0.611	0.224 ±0.176
WE-PINN	0.030 ±0.001	0.178 ±0.005	0.026 ±0.001	0.037 ±0.000	0.178 ±0.003	0.040 ±0.000	0.141 ±0.007	0.048 ±0.016	0.096 ±0.006	0.338 ±0.086	2.369 ±1.739	2.387 ±1.880
FLINT (Ours)	**0.016** ±0.005	**0.102** ±0.021	**0.018** ±0.006	**0.016** ±0.005	**0.082** ±0.059	**0.018** ±0.010	**0.035** ±0.013	**0.008** ±0.004	**0.010** ±0.003	**0.033** ±0.014	**0.090** ±0.052	**0.122** ±0.045

Table 9: **Solution and wave-capturing quality on Shu–Osher.** All metrics are defined in Appendix E. All models are trained with the SOAP optimizer. ↓ means lower is better and bold font represents the **best** per column. Each entry is the mean ± the standard deviation over three seeds.

	Space–time L_2 ↓			L_2 at T ↓			Wave-region L_2 at T ↓			Wave amplitude↓		
Method	ρ	u	p	ρ	u	p	ρ	u	p	ρ	u	p
PINNs-WE	0.114 ±0.041	0.189 ±0.128	0.133 ±0.050	0.155 ±0.107	0.232 ±0.214	0.159 ±0.111	0.134 ±0.091	0.073 ±0.060	0.126 ±0.098	0.427 ±0.356	0.788 ±0.522	0.580 ±0.257
WE-PINN	0.078 ±0.001	0.102 ±0.002	0.089 ±0.002	0.095 ±0.002	0.112 ±0.003	0.101 ±0.003	0.077 ±0.001	0.031 ±0.001	0.060 ±0.001	0.953 ±0.021	0.936 ±0.004	0.895 ±0.018
FLINT (Ours)	**0.064** ±0.002	**0.086** ±0.002	**0.077** ±0.002	**0.027** ±0.007	**0.025** ±0.027	**0.024** ±0.016	**0.019** ±0.002	**0.005** ±0.001	**0.011** ±0.003	**0.122** ±0.035	**0.027** ±0.012	**0.035** ±0.012

D Training Dynamics and Loss Contributions

Training and solution trajectories. For Sod-k ($k = 50$) and Shu–Osher, we examine FLINT's training trajectories to assess optimization stability and convergence, and to identify which loss components drive training. A decreasing physics-informed objective does not necessarily imply a more accurate solution, as observed empirically in prior PINN studies (Krishnapriyan et al., 2021; Wang et al., 2024; Xu et al., 2025) and more recently formalized through the residual–reconstruction gap (Jiang et al., 2026).

On both problems, the control-volume terms contribute more to the training objective than the pointwise PDE residual for much of optimization. Figure 15 shows that the total loss decreases by

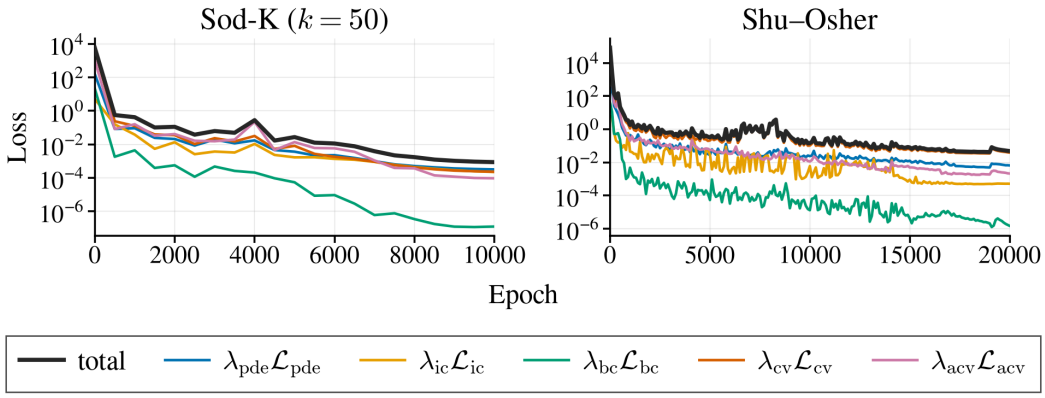


Figure 15: **Training objective decomposition for Sod-k and Shu–Osher.** The total objective and its weighted PDE, initial-condition, boundary, interior control-volume, and anchored control-volume components are shown throughout training. The control-volume terms dominate much of the optimization on both problems, while the boundary contribution remains small. Losses are shown on a logarithmic scale.

more than six orders of magnitude on both Sod-k and Shu–Osher. Over the full trajectory, the two conservation terms account for a median 64% of the total loss on Sod-k and 84% on Shu–Osher, while the PDE residual contributes 37% and 14% of the final loss, respectively. The boundary term remains much smaller because the solution stays close to the prescribed boundary states over the simulated interval. Training is not strictly monotonic and the total loss temporarily increases near epoch 4,000 on Sod-k and epoch 8,300 on Shu–Osher, primarily because of increases in the conservation terms.

E METRICS

These are the metrics definitions used throughout Section 4 and within this appendix.

Relative L_2. For a primitive variable $q \in \{\rho, u, p\}$, we compute the norm

$$L_2[\mathbf{q}_{\text{ref}}, \mathbf{q}_\theta] = \frac{\|\mathbf{q}_\theta - \mathbf{q}_{\text{ref}}\|_2}{\|\mathbf{q}_{\text{ref}}\|_2}, \tag{26}$$

where $\mathbf{q}_\theta = (q_\theta(t_i, x_i))_{i=1}^N$ and $\mathbf{q}_{\text{ref}} = (q_{\text{ref}}(t_i, x_i))_{i=1}^N$ are the model and reference values evaluated at the same N points.

We report (26) over the full space–time domain, at the terminal time T, and over a terminal-time wave region $\mathcal{W}$, denoted Space-Time L_2, L_2 at T, and Wave-region L_2 at T, respectively.

The wave region is defined independently of the model prediction as

$$\mathcal{W} = \begin{cases} [x_c^{\text{ref}}(T) + \delta,\ x_s^{\text{ref}}(T) - \delta], & \text{Sod-}k, \\ [x_s^{\text{ref}}(0) + \delta,\ x_s^{\text{ref}}(T) - \delta], & \text{Shu–Osher}, \end{cases} \qquad \delta = 10\Delta x, \tag{27}$$

where x_c^{ref} and x_s^{ref} denote the reference contact and shock positions, respectively and $\Delta x = (x_1 - x_0)/(N_x - 1)$ is the spatial spacing of the evaluation grid. The margin δ excludes the discontinuity neighborhoods so that the wave-region error measures the fidelity of the transported structure rather than front misalignment, which is evaluated separately. The same region $\mathcal{W}$ is used for the wave-amplitude metric.

Wave amplitude $|1 - A/A_{\text{ref}}|$. For a primitive variable $q \in \{\rho, u, p\}$, define

$$A[\mathbf{q}] = \max_{x \in \mathcal{W}} \mathbf{q}(T, x) - \min_{x \in \mathcal{W}} \mathbf{q}(T, x). \tag{28}$$

We report

$$\text{Amplitude error} = \left| 1 - \frac{A[\mathbf{q}_\theta]}{A[\mathbf{q}_{\text{ref}}]} \right|. \tag{29}$$

This metric directly measures the attenuation or amplification of the transported wave.

Global conservation residual $\mathcal{R}_{\text{cons}}^{\max}$. For a conserved variable $z \in \{\rho, \rho u, E\}$, we compute

$$\mathcal{R}_{\text{cons},z}^{\max} = \max_{t\in[0,T]} \left| \int_{x_0}^{x_1} \left[z_\theta(t,x) - z_\theta(0,x) \right] dx + \int_0^t \left[F_z(\tau, x_1) - F_z(\tau, x_0) \right] d\tau \right|, \tag{30}$$

where the spatial and temporal domains and the physical variables ρ, u, p, and E are defined in Section 2. Here, z_θ is the conserved variable obtained from the network prediction, F_z is its corresponding flux, and τ is the time integration variable. Thus, $\mathcal{R}_{\text{cons},z}^{\max}$ measures the maximum deviation from global conservation over $t \in [0, T]$, independently of initial-condition error because the balance is measured relative to $z_\theta(0, x)$.

F Framework Hyperparameters

This section defines the hyperparameters used in our method and are shown in Table 10.

Table 10: **Hyperparameters of the proposed method.** Every entry represents the configuration of the run we report for that problem.

Hyperparameter	Sod	Sod-k ($k = 50$)	Shu–Osher	Blast Wave
Network architecture				
Depth × width	12×128	12×128	12×128	12×128
Sampling				
Collocation points	10,000	10,000	150,000	17,500
Initial-condition points	4,000	4,000	20,000	5,000
Boundary points (per wall)	200	200	600	280
Control volumes				
Interior control volumes	768	768	1,536	960
Anchored control volumes	128	128	512	128
Interior box width	[0.005, 0.02]	[0.005, 0.02]	[0.03, 0.11]	[0.005, 0.02]
Interior box height	[0.005, 0.022]	[0.005, 0.022]	[0.012, 0.062]	[0.007, 0.035]
Interior face quad nodes	33	33	33	33
Anchored box width	[0.005, 0.02]	[0.005, 0.02]	[0.105, 0.385]	[0.005, 0.02]
Anchored box height	[0.022, 0.2]	[0.022, 0.2]	[0.05, 0.6]	[0.035, 0.5]
Anchored face quad nodes	385	385	385	385
Loss weights				
PDE weight λ_{pde}	1	1	1	1
IC weight λ_{ic}	10	0.5	0.5	0.5
BC weight λ_{bc}	10	10	10	10
Interior CV weight λ_{cv}	1	1	1	1
Anchored CV weight λ_{acv}	1	1	30	1
Residual weighting				
Sensor strength ε	1.5	1.5	1.5	1.5
Gate threshold K	23.66	23.66	3.87	23.66
Optimizer	SOAP	SOAP	SOAP	SOAP
Learning rate	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Epochs	10,000	10,000	20,000	15,000

Table 10 contains two types of entries. Most settings are carried over unchanged from the sod Shock tube configurations and are not re-tuned for each problem, with the exception of Burgers, which is a simpler case. These fixed settings include the network architecture, optimizer, and learning rate, and the weights on the PDE residual, initial-condition, and boundary-condition losses. Three hyperparameters were selected through separate searches. The number of quadrature nodes on the anchored faces was selected by grid search over $\{33, 121, 209, 297, 385\}$, with candidates evaluated by their integration error against a reference. The number of anchored boxes was selected by a

separate grid search from 16 to 1024 at the fixed box size, using the solution performance metric. Finally, the anchored-loss weight was also selected through a grid search for each problem from $\lambda_{\text{anchor}} \in \{0, 0.1, 0.3, 1, 3, 10, 30\}$. The sampling configuration was scaled with the size of the space–time domain rather than tuned independently for each problem. We first established the sampling parameters on the Sod shock tube and reused them unchanged for Sod-k. When the same configuration was applied to Shu-Osher, the larger space–time domain reduced both the density of the collocation points and the relative coverage of the sampled control volumes. We therefore used the Sod configuration as a reference and increased the number of collocation points and control volumes so that the collocation point density and approximate coverage of the control-volumes, relative to the full space–time domain remained comparable. The same scaling principle was applied to the cylindrical blast wave.