

Supplementary Information: Field-Residual Learning Does Not Reliably Improve a Strong Fixed-Budget Riemann Starter

Ramesh Kolluru
Senior CFD Consultant and Independent Researcher, India
rameshkolluru43@gmail.com
ORCID: [0000-0001-5447-8550](https://orcid.org/0000-0001-5447-8550)

24 September 2026

This supplement collects the confirmatory seed-level hard-aware outcomes, audit tables, additional holdout fields, Monte Carlo detail, gradient checks, and environment notes that support the main manuscript. Cross-references use the same labels via the `xr` package.

1 Confirmatory hard-aware seed-level outcomes

Table 1 lists the ten confirmatory trajectories. Both rules select from the same stored run. Δ_s is the paired hard-anchor holdout median versus deterministic $Q_0 = 0.034$. A practical gain is $\Delta_s < -0.01$. The two overall-median columns are the nine-anchor holdout medians of the checkpoints selected by each rule; no seed beats Q_0 on that metric under either rule.

Table 1 Seed-wise confirmatory checkpoint selection around Q_0 . Both rules are applied to the same full-epoch trajectory. Δ_s is the hard-anchor paired difference versus Q_0 . Overall median columns are the nine-anchor holdout medians of the selected maps.

Seed	Overall epoch	Overall Δ_s	Overall med.	Comp. epoch	Comp. Δ_s	Comp. overall
0	16	+0.000	0.001	23	-0.020	0.002
1	5	+0.014	0.003	9	+0.033	0.024
2	24	+0.014	0.014	18	+0.022	0.014
3	13	+0.062	0.006	28	-0.022	0.001
4	30	+0.040	0.025	19	+0.022	0.003
5	27	+0.034	0.002	6	+0.015	0.030
6	17	+0.042	0.023	26	+0.010	0.020
7	16	+0.009	0.006	20	+0.010	0.002
8	28	+0.034	0.024	8	+0.023	0.001
9	8	+0.035	0.019	35	+0.005	0.001
Mean	—	+0.028	—	—	+0.010	—
Median	—	+0.034	—	—	+0.013	—
$\Delta_s < -0.01$	—	0/10	—	—	2/10	—

2 Star walk-off and primitive errors

Table 2 Absolute star correction $|\log p_* - \log p_{Q_0}|$ and $|u_* - u_{Q_0}|$ for the seed-0 residual (WS-2) and the field-only ablation.

Case	$ \Delta \log p_* $	$ \Delta u_* $	$ \Delta \log p_* _{\text{field}}$	$ \Delta u_* _{\text{field}}$
Sod	0.008	0.001	0.008	0.002
Toro 123	0.008	0.098	0.005	0.006
Left blast	0.013	0.001	0.013	0.003
Right blast	0.013	0.016	0.011	0.011
Collision	0.008	0.015	0.006	0.021
Stat. shock	0.005	0.001	0.004	0.001
Stat. contact	0.000	0.000	0.000	0.000
Slow contact	0.000	0.000	0.000	0.000
Slow shock	0.009	0.001	0.007	0.000
Mean / max	0.007/0.013	0.015/0.098	0.006/0.013	0.005/0.021
Hard mean / max	0.011/0.013	0.011/0.016	0.010/0.013	0.012/0.021

Table 3 Primitive Rel- L^2 for Q_0 versus the seed-0 residual, and WS-2 front errors. A dash means the diagnostic is not defined. No positivity failure occurs on either structured map.

Case	ρ_{Q_0}	ρ_θ	u_{Q_0}	u_θ	p_{Q_0}	p_θ	x_c	x_s
Sod	0.000	0.003	0.000	0.033	0.000	0.005	0.002	0.003
Toro 123	0.000	0.000	0.000	0.013	0.000	0.000	0.035	—
Left blast	0.000	0.122	0.000	0.061	0.000	0.023	0.008	0.052
Right blast	0.061	0.109	0.031	0.044	0.011	0.016	0.006	0.038
Collision	0.046	0.102	0.022	0.049	0.045	0.099	0.001	0.007
Stat. shock	0.000	0.002	0.000	0.001	0.000	0.002	0.004	0.004
Stat. contact	0.000	0.000	0.000	0.000	0.000	0.000	0.004	—
Slow contact	0.000	0.000	0.000	0.000	0.000	0.000	0.006	—
Slow shock	0.000	0.031	0.000	0.033	0.000	0.039	0.003	0.005

2.1 Why field training moves away from Q_0

The claim that a log-field objective walks the star off Q_0 is an observed training outcome. The representative per-anchor primitive errors are reported in the main manuscript, and the star corrections are listed in Table 2. It is not a claim that the first gradient of L_{field} at the frozen Q_0 star is anti-aligned with a Newton teacher. Figure 1 and Table 4 evaluate $\partial L / \partial \log p_*$ and $\partial L / \partial u_*$ at that frozen star, for the full log-field loss, plateau-only, density-only, velocity-only, and pressure-only components, and for direct star supervision.

On every non-contact anchor the two-vector cosine between L_{field} and L_{star} is at least 0.93 (contacts are exactly zero: the gate and the skip vanish). The first SGD step at Q_0 is therefore not an anti-teacher step. Alignment of that step with the Newton direction $(\log p^{\text{ex}} - \log p_{Q_0}, u^{\text{ex}} - u_{Q_0})$ is nevertheless poor on the blasts for *both* losses (left blast ≈ 0.19 ; right blast ≈ 0.40). The acoustic-scaled teacher is a weak Newton step there as well.

The component split is informative on the right blast: the full-field u_* gradient and the plateau-only u_* gradient have opposite signs, while $\partial / \partial \log p_*$ remains large and same-signed, so the two-vector cosine stays near one. Pressure dominates $\partial L_{\text{field}} / \partial \log p_*$ on that family. Walk-off during training is therefore a trajectory- and mixed-batch effect, not a local sign flip of L_{field} at initialization. The confirmatory full-epoch checkpoint-selection experiment in the main paper shows that sparse validation cadence was an artifact of the original logging schedule; hard-aware

selection reduces degradation but does not remove the lack of repeatable improvement over Q_0 . Figure 2 of the main manuscript shows the selection half of that story on the original seed-0 run. The present diagnosis establishes the failure mode empirically and excludes several simple explanations—insufficient residual expressivity, contact leakage, PDE-critic dependence, first-step anti-alignment, and front-cell metric artefacts—but it does not yet identify a unique optimisation mechanism. That gap is consistent with the known mismatch between strong-form collocation residuals and entropy solutions of hyperbolic conservation laws [1–4].

Table 4 Cosine similarity at the frozen Q_0 star between the field-loss and star-loss 2-vectors ($\partial/\partial \log p_*$, $\partial/\partial u_*$), and between each first step and the Newton direction. Contacts are identically zero.

Anchor	field vs star	field vs Newton	star vs Newton
Sod	0.93	0.96	1.00
Toro 123	1.00	1.00	1.00
Left blast	1.00	0.19	0.15
Right blast	1.00	0.40	0.42
Collision	1.00	0.73	0.73
Stat. shock	1.00	1.00	1.00
Slow shock	1.00	1.00	0.99

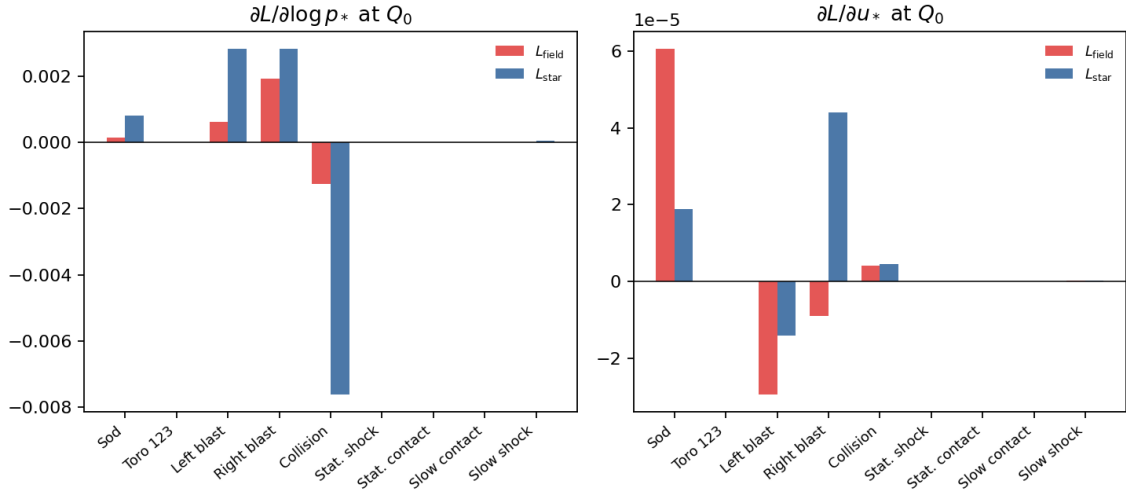


Fig. 1 Initial gradients of the log-field loss and the Newton-star teacher at the frozen Q_0 star. The first field step is not anti-teacher; blast families are poorly aligned with the Newton direction for both losses.

2.2 Broader Riemann draws

Table 5 evaluates the same maps on the locked protocol-v2 validation, in-family, and detuned-family members. These are a small locked diagnostic extension ($n = 18$ each), not a large-distribution claim. Hard entries use the six members from families $\{3, 4, 5\}$; validation has three hard members. On both draws, Q_0 is close to the Newton-star oracle under the reported medians (in-family hard 0.123 versus 0.121; detuned-family hard 0.129 versus 0.129). A case-resampled bootstrap ($B = 10,000$ draws; Table 6) puts the paired median difference median($\text{Rel-}L^2(Q_0) - \text{Rel-}L^2(\text{oracle})$) at 0 on in-family $[0, 2.5 \times 10^{-6}]$ and 6×10^{-7} on detuned-family $[0, 3.9 \times 10^{-6}]$. The hard-only paired intervals include zero (in-family $[-0.002, 0.015]$; detuned-family $[-0.008, 0.001]$). Absolute-median intervals are wide, as expected for $n = 18$. Field residuals on Q_0 do not improve

those hard medians. A residual on PVRS slightly lowers the in-family hard median (0.119 versus 0.123) and raises the detuned-family hard median (0.146 versus 0.129). The Newton-star-only network is close to the oracle on the in-family set and stays within 0.002 on the detuned-family set. Figure 2 shows the paired errors, separated by soft and hard regimes.

Table 5 Locked protocol-v2 splits other than the nine named anchors. Entries are overall / hard median $\text{Rel-}L^2$. Hard members are families $\{3, 4, 5\}$: 3 on val, 6 on each 18-member draw. Detuned-family is a $3\times$ widened, family-conditioned perturbation of the same nine anchors.

Method	Val (9/3)	In-family (18/6)	Detuned (18/6)
Q_0	0.042/0.089	0.034/0.123	0.026/0.129
Exact Newton star	0.030/0.082	0.033/0.121	0.024/0.129
Residual on Q_0	0.019/0.076	0.033/0.127	0.024/0.134
Residual on PVRS	0.021/0.081	0.042/0.119	0.056/0.146
Star-sup. only	0.030/0.082	0.033/0.121	0.028/0.131

Table 6 Bootstrap 95% intervals for Q_0 minus the Newton-star oracle on the locked 18-member sets (10^4 case resamples). The interval on the paired overall median includes values at the float noise floor; the hard-only intervals include zero.

Paired median difference	In-family	Detuned-family
All members ($n = 18$)	0 $[0, 2.5 \times 10^{-6}]$	6×10^{-7} $[0, 3.9 \times 10^{-6}]$
Hard members ($n = 6$)	0.002 $[-0.002, 0.015]$	2×10^{-6} $[-0.008, 0.001]$

The residual on Q_0 has a lower validation median than Q_0 because that checkpoint was selected on validation. The in-family and detuned-family members, which were not used for selection, do not reproduce a hard-family gain. Those ~ 0.03 open-split medians for Q_0 and the Newton oracle are largely the same Toro-versus- $\mathcal{S}$ assignment convention as the collision floor (Section 6.1 of the main manuscript): against $\mathcal{S}(\text{Newton})$ labels the oracle is 0 and Q_0 falls to $\sim 10^{-5}$. The nine-anchor hard-case error 0.034 is not that artefact; it is the right blast, identical on both conventions.

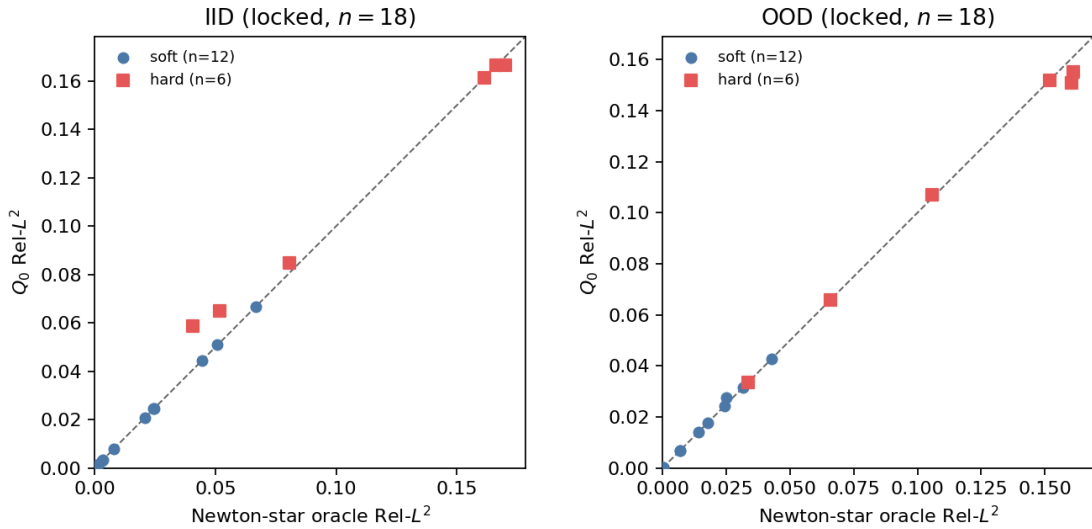


Fig. 2 Paired $\text{Rel-}L^2$ of Q_0 against the Newton-star oracle on the locked in-family and detuned-family members. Soft families (circles) and hard families (squares). The dashed line is $y = x$.

2.3 Locked Monte Carlo evaluation

The 18-member in-family and detuned-family sets remain a small locked diagnostic extension. Table 7 and Figures 3–4 evaluate the same *seed-0* learned maps on a separately locked Monte Carlo draw: 100 in-family and 100 $3\times$ detuned-family members per family ($n = 900 + 900$), seed 20260819, independent of protocol-v2. Five-seed evidence for those residuals is reported on the nine anchors, not on this 1800-member draw. The learned-map Monte Carlo scores are therefore seed-0 diagnostic extensions, not multi-seed estimates. The draw is family-balanced. It is still a wider local perturbation of the same nine ideal-gas Riemann regimes, not a qualitatively new PDE class. It is not the near-vacuum family of Section 6.6 of the main manuscript. No draw was rejected. Audit thresholds $p < 10^{-3}$ or $\rho < 10^{-2}$ were never hit; the lowest detuned-family blast pressure is $\approx 3.6 \times 10^{-3}$, and the lowest Toro-123 detuned-family IC pressure is ≈ 0.22 .

On the 900 in-family members, Q_0 has mean 0.048, median 0.026, $p_{90} = 0.133$, $p_{95} = 0.157$, and worst 0.207 (hard median 0.110). The exact-star oracle is mean 0.047, median 0.024, $p_{95} = 0.155$, worst 0.207 (hard 0.109). The paired median Q_0 minus exact is 0 overall and on the hard subset. That aggregate closeness does not erase case-specific errors on the nine-anchor set (right blast, collision). The field residual on Q_0 is slightly worse (median 0.029, hard 0.112; paired median $Q_0 - \text{WS-2} = -0.001$). The PVRS residual is worse still (median 0.033, hard 0.113). Star-only supervision matches the oracle.

On the 900 detuned-family members the same pattern holds: Q_0 0.026/0.161/0.205 (hard 0.117) versus exact 0.025/0.161/0.210 (hard 0.116). The field residual remains slightly worse. The PVRS residual degrades more clearly (median 0.047, hard 0.130). Star-only supervision again matches the oracle. These percentiles confirm the nine-anchor and 18-member reading on a larger locked sample from the *same* family boxes. They do not establish a general Riemann-distribution result.

Table 7 Locked Monte Carlo, 100 members per family per split ($n = 900$ each). Learned residuals are the seed-0 checkpoints, not five-seed means. Entries are overall median / p_{95} / worst, then hard median. Paired rows are $\text{Rel-}L^2(Q_0) - \text{Rel-}L^2(\cdot)$; a negative paired median means Q_0 is better.

Method	In-family	$3\times$ detuned-family
Q_0	0.026/0.157/0.207 (0.110)	0.026/0.161/0.205 (0.117)
Exact Newton star	0.024/0.155/0.207 (0.109)	0.025/0.161/0.210 (0.116)
Residual on Q_0	0.029/0.155/0.213 (0.112)	0.029/0.162/0.215 (0.118)
Residual on PVRS	0.033/0.153/0.214 (0.113)	0.047/0.169/0.227 (0.130)
Star-sup. only	0.025/0.155/0.207 (0.109)	0.026/0.161/0.210 (0.117)
Paired Q_0 –exact (median)	0	0
Paired Q_0 –field residual (median)	−0.001	−0.001
Paired Q_0 –PVRS res. (median)	−0.001	−0.004

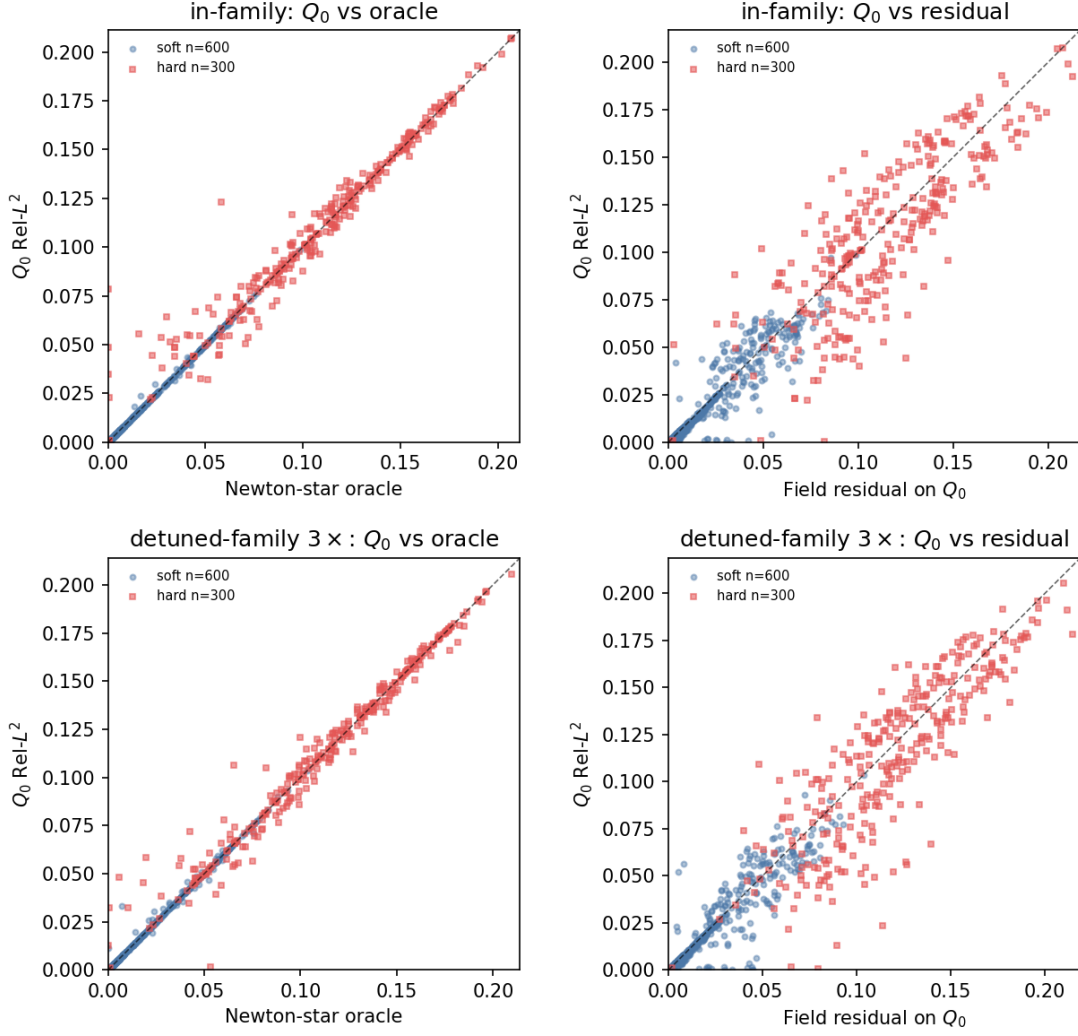


Fig. 3 Paired Monte Carlo Rel- L^2 of Q_0 against the Newton-star oracle and the seed-0 field residual on Q_0 (900 in-family + 900 $3 \times$ detuned-family). Soft families are circles; hard families are squares. The dashed line is $y = x$.

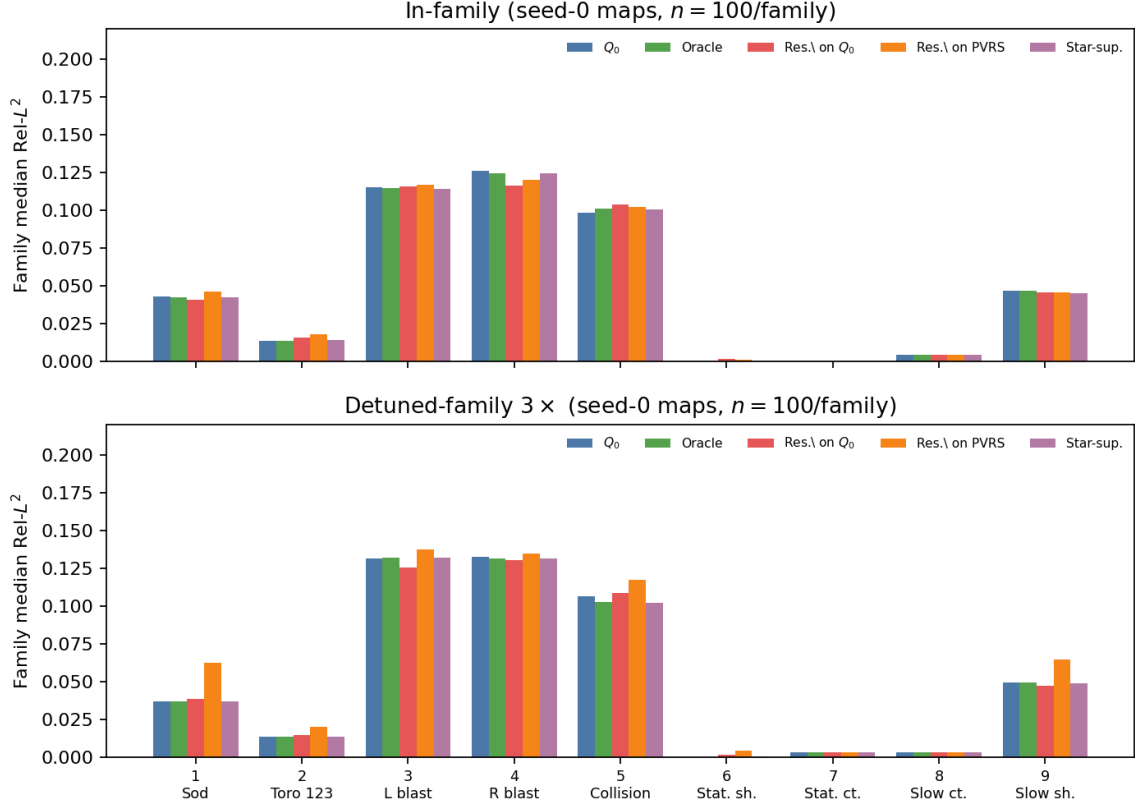


Fig. 4 Family-stratified Monte Carlo medians for the seed-0 maps. Detuned-family is the $3\times$ widened box around the same nine anchors, not a new regime class.

2.4 Matched Q_0 fields

Figure 5 overlays exact, Q_0 , and the seed-0 field residual at $t = T$. Where Q_0 is already on the grid, the residual is either invisible or a small star bias (Sod, Toro 123 velocity shelf, blast density peaks). Contacts coincide for all three maps.

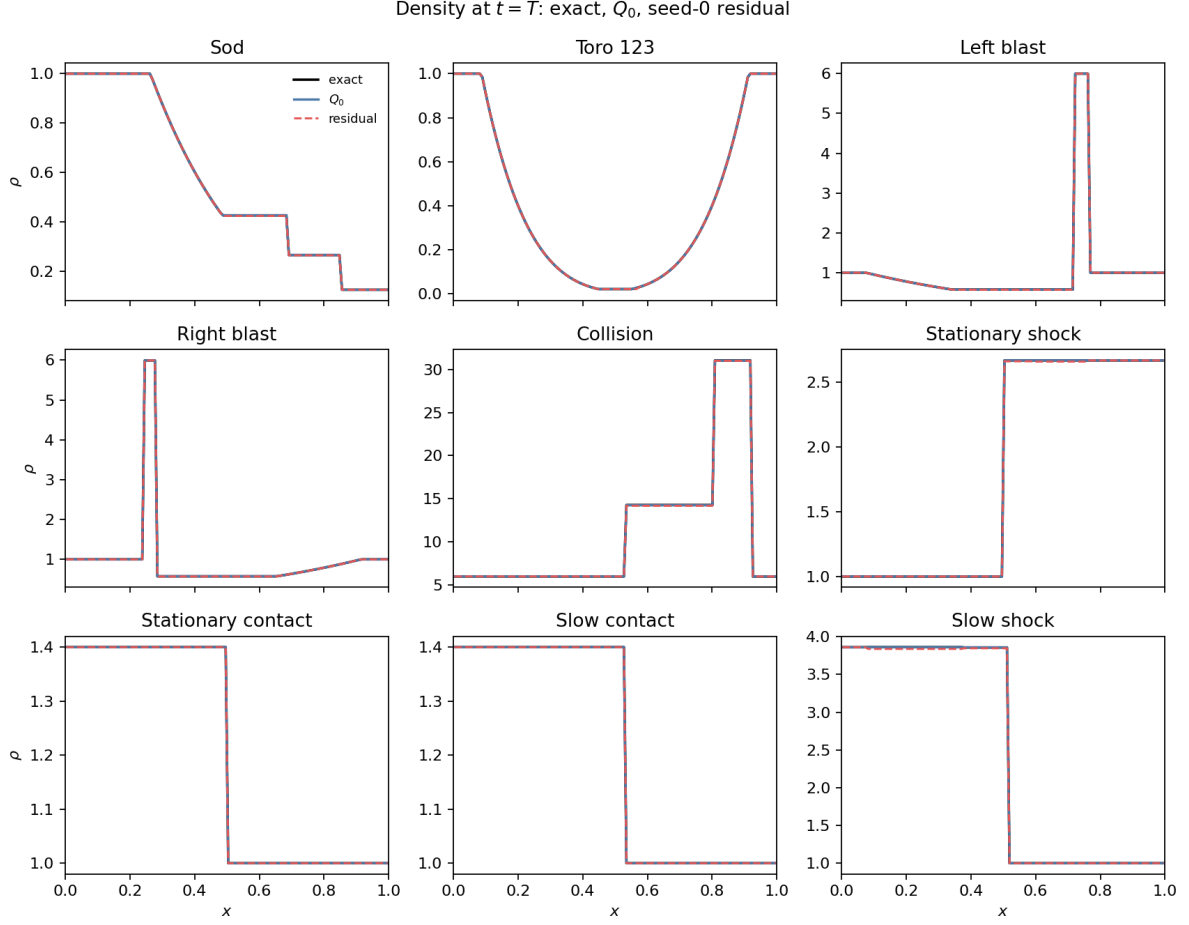


Fig. 5 Density at $t = T$ on the nine classic anchors: exact (black), Q_0 (blue), seed-0 field residual (red dashed).

A Remaining classic-anchor fields

Figures 6–12 complete the nine-anchor IC-only comparison for the seed-0 field residual (red dashed) against the exact Toro label (black). Q_0 numbers are reported in Table 8 of the main manuscript and Table 3 of this supplement.

Toro 123 (soft). Double rarefaction, near-vacuum star. Q_0 uses the two-rarefaction branch and has $u_* = 0$ and $\text{Rel-}L^2 \sim 0$. The field residual is 0.004 with a trained u_* leak (0.098 versus 0).

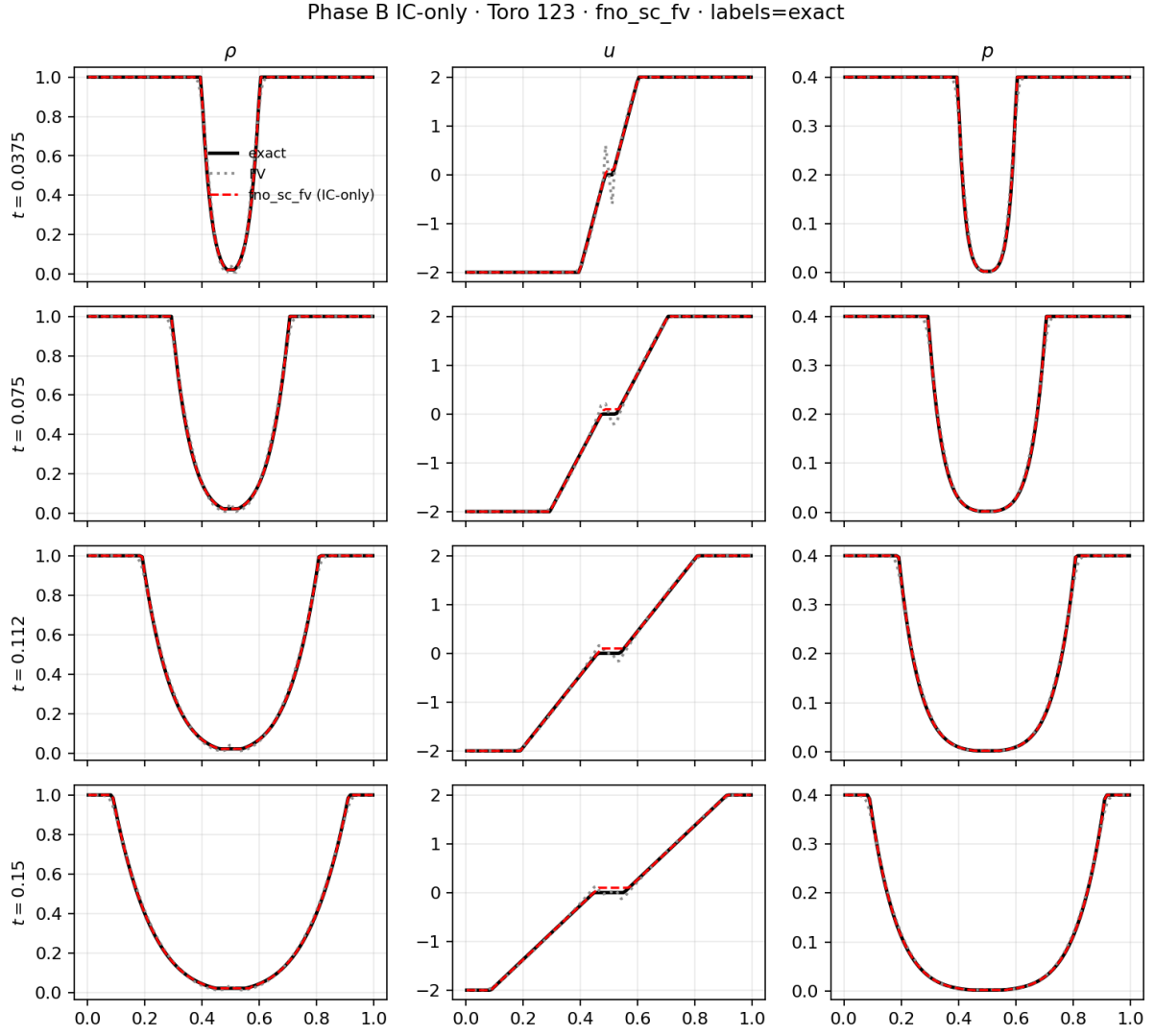


Fig. 6 Toro 123 (family 2).

Left blast (hard). Q_0 is 0.0004 with $p_* \approx 462$. Trained Rel- L^2 0.068 with $p_* = 455$ versus 461.

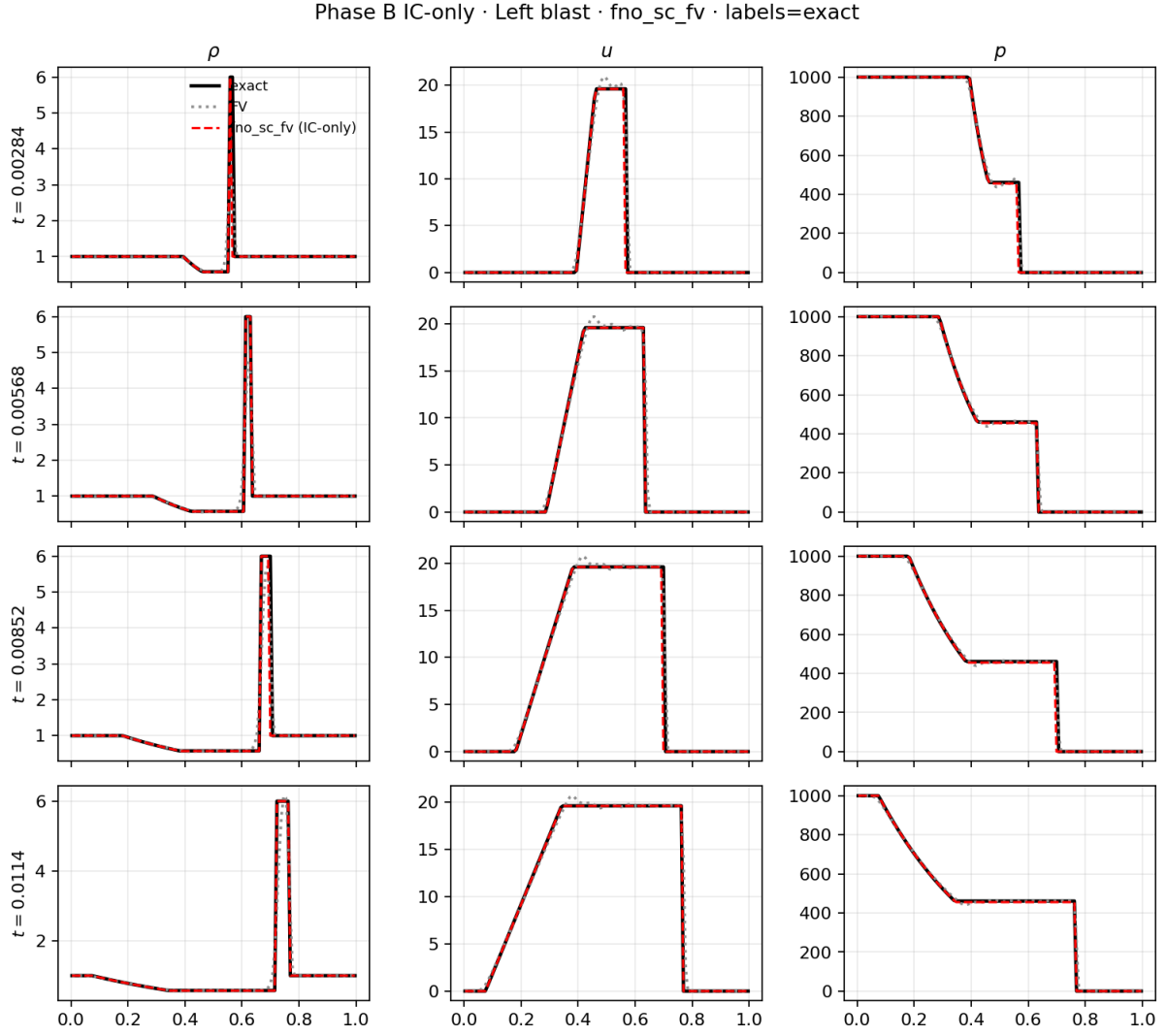


Fig. 7 Left blast (family 3).

Right blast (hard). One of the two families where Q_0 remains visibly imperfect (0.034; exact-star oracle 0.000). Trained Rel- L^2 0.056.

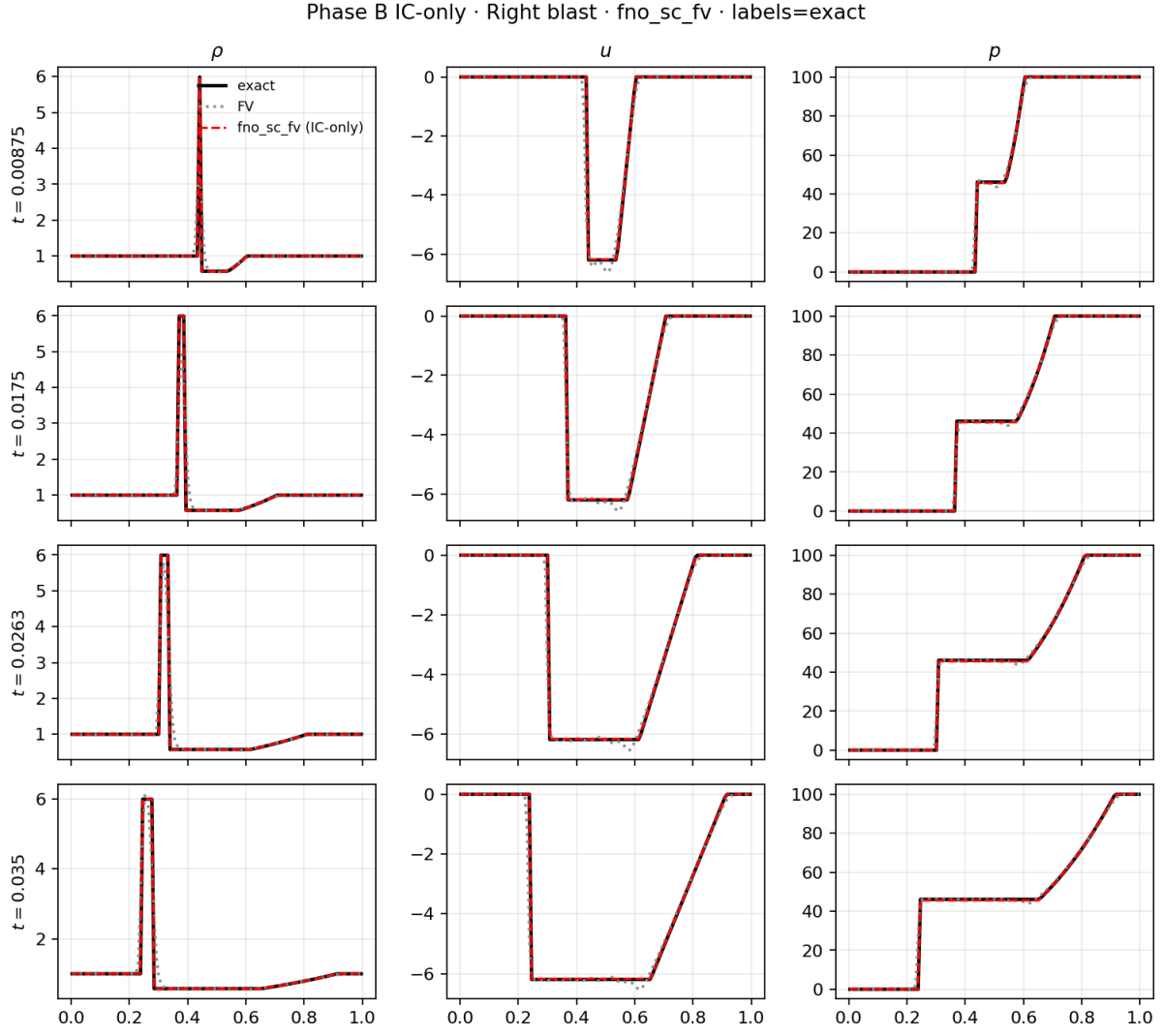


Fig. 8 Right blast (family 4).

Stationary shock (soft). Q_0 is exact. Trained Rel- L^2 0.002.

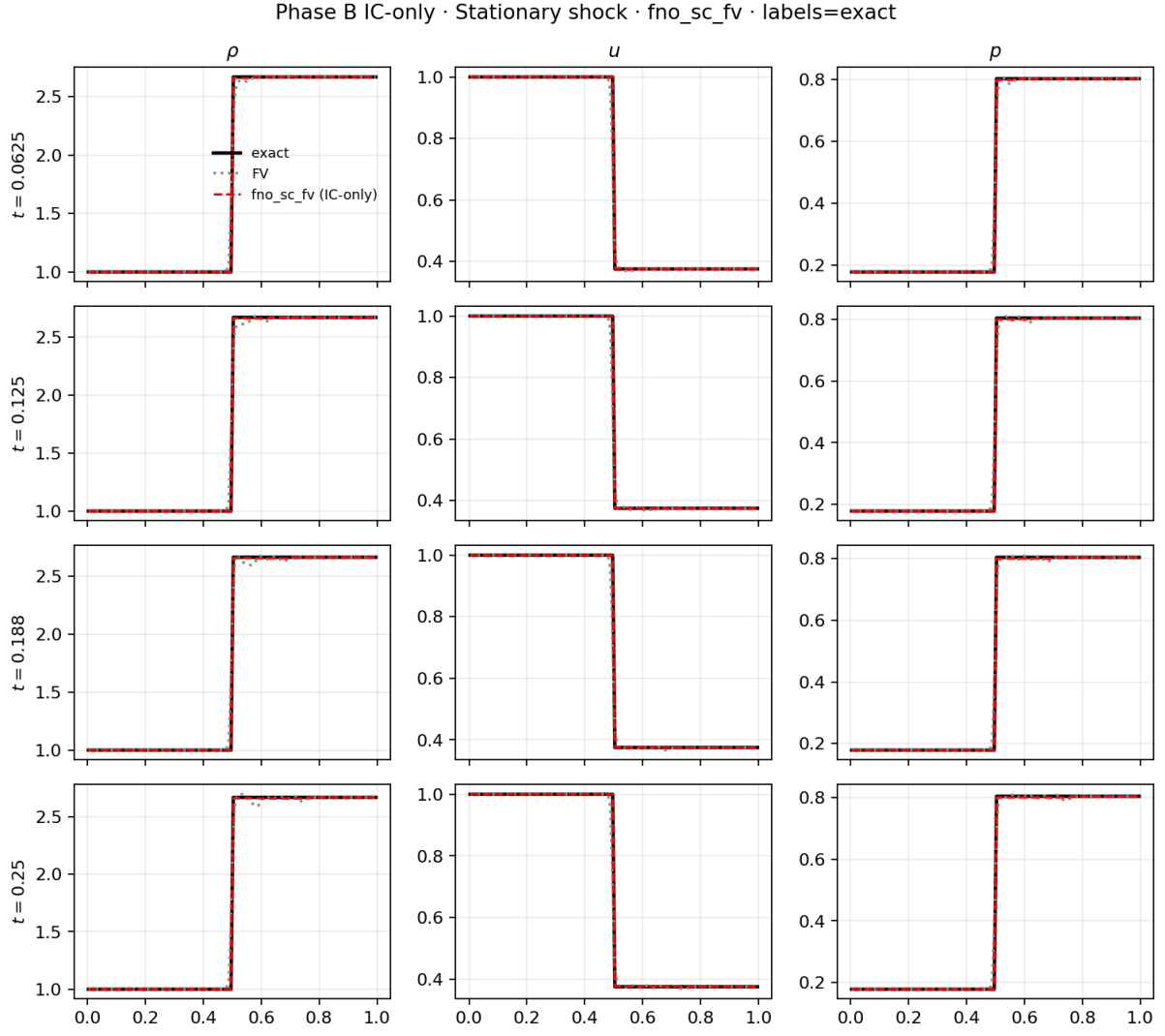


Fig. 9 Stationary shock (family 6).

Stationary contact (soft). The gate forces $g = 0$, so $\text{Rel-}L^2 = 0$.

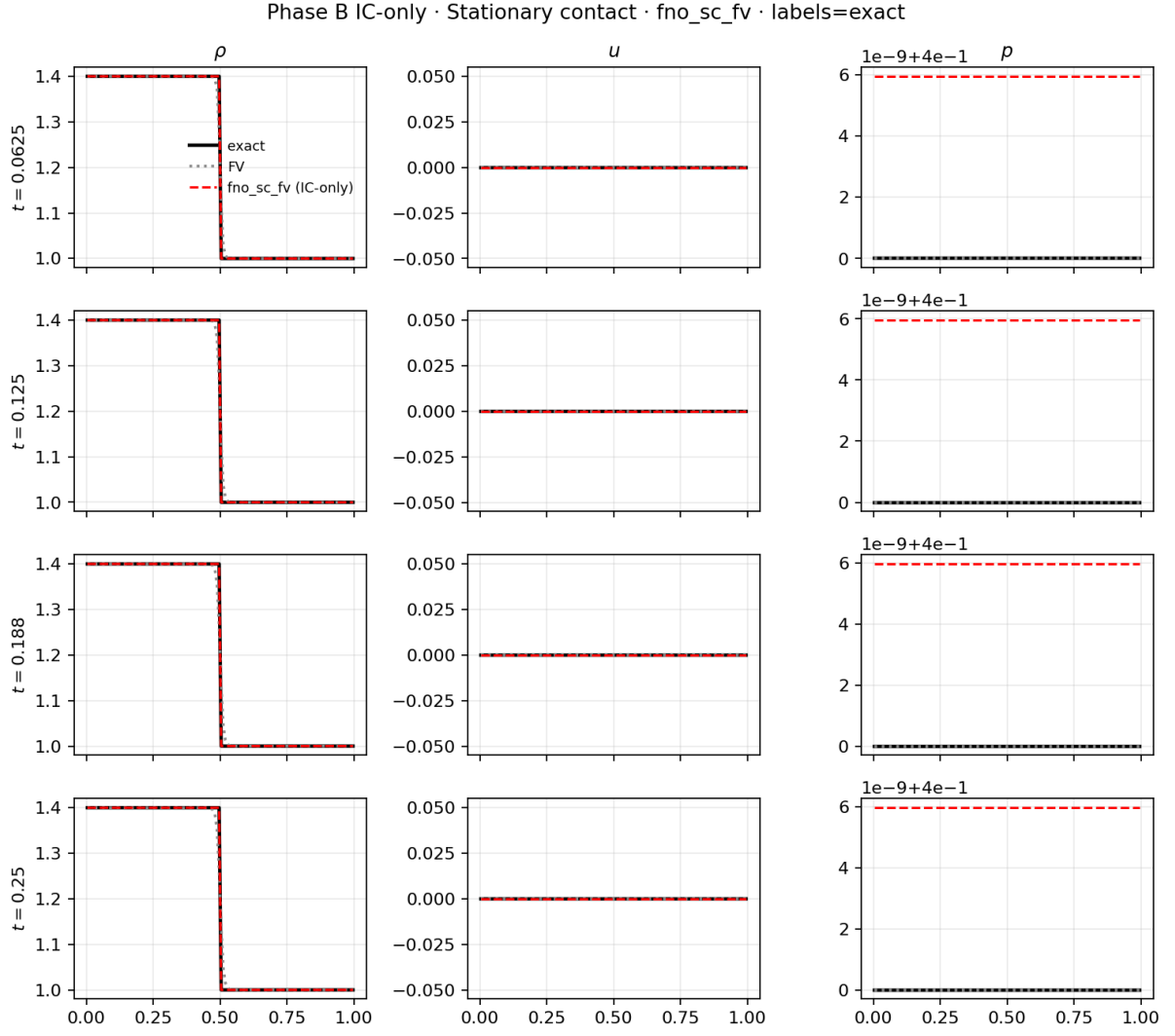


Fig. 10 Stationary contact (family 7).

Slow contact (soft). The gate vanishes, so $\text{Rel-}L^2 = 0$.

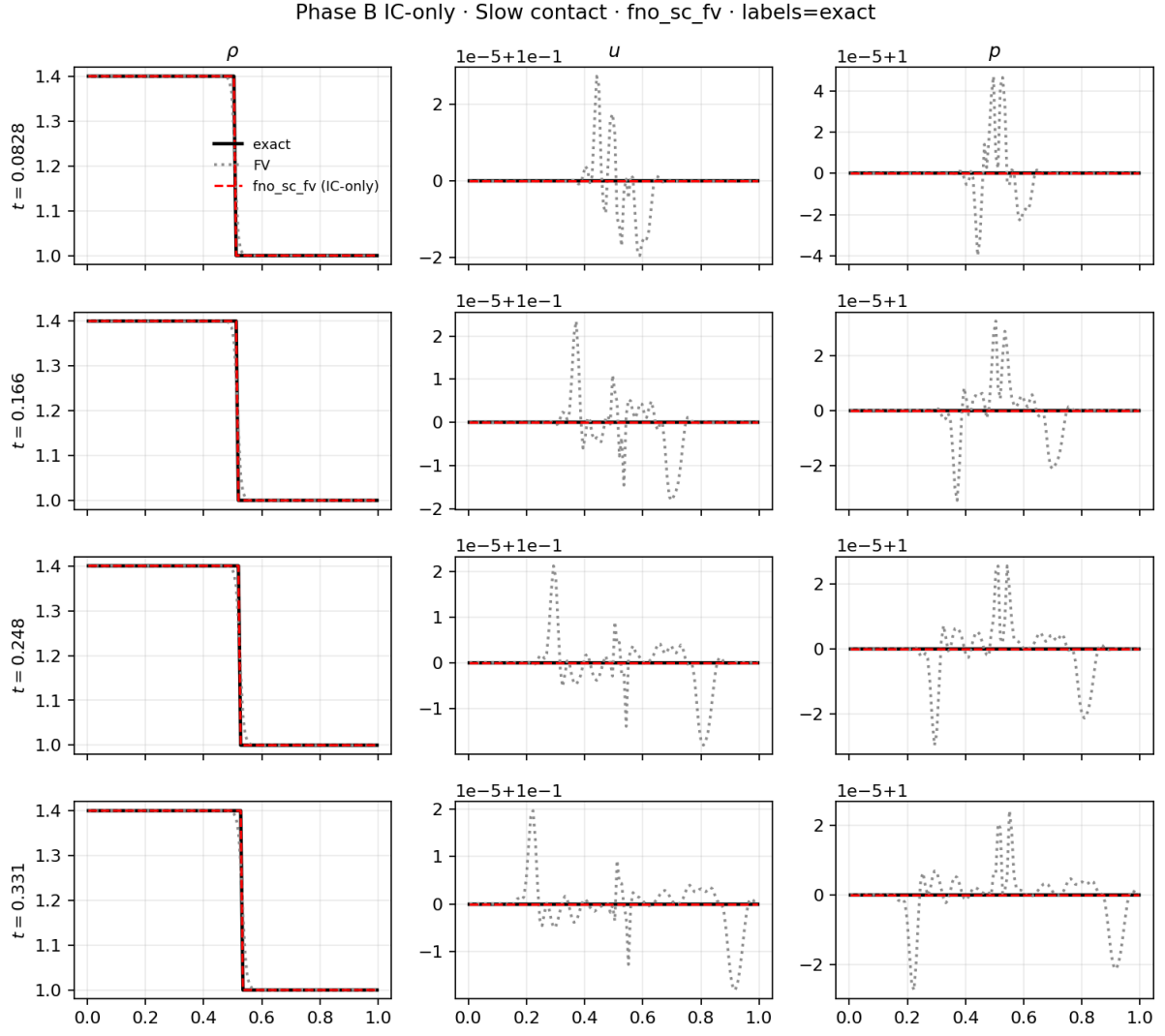


Fig. 11 Slow contact (family 8).

Slow shock (soft). Q_0 is exact. Trained Rel- L^2 0.034.

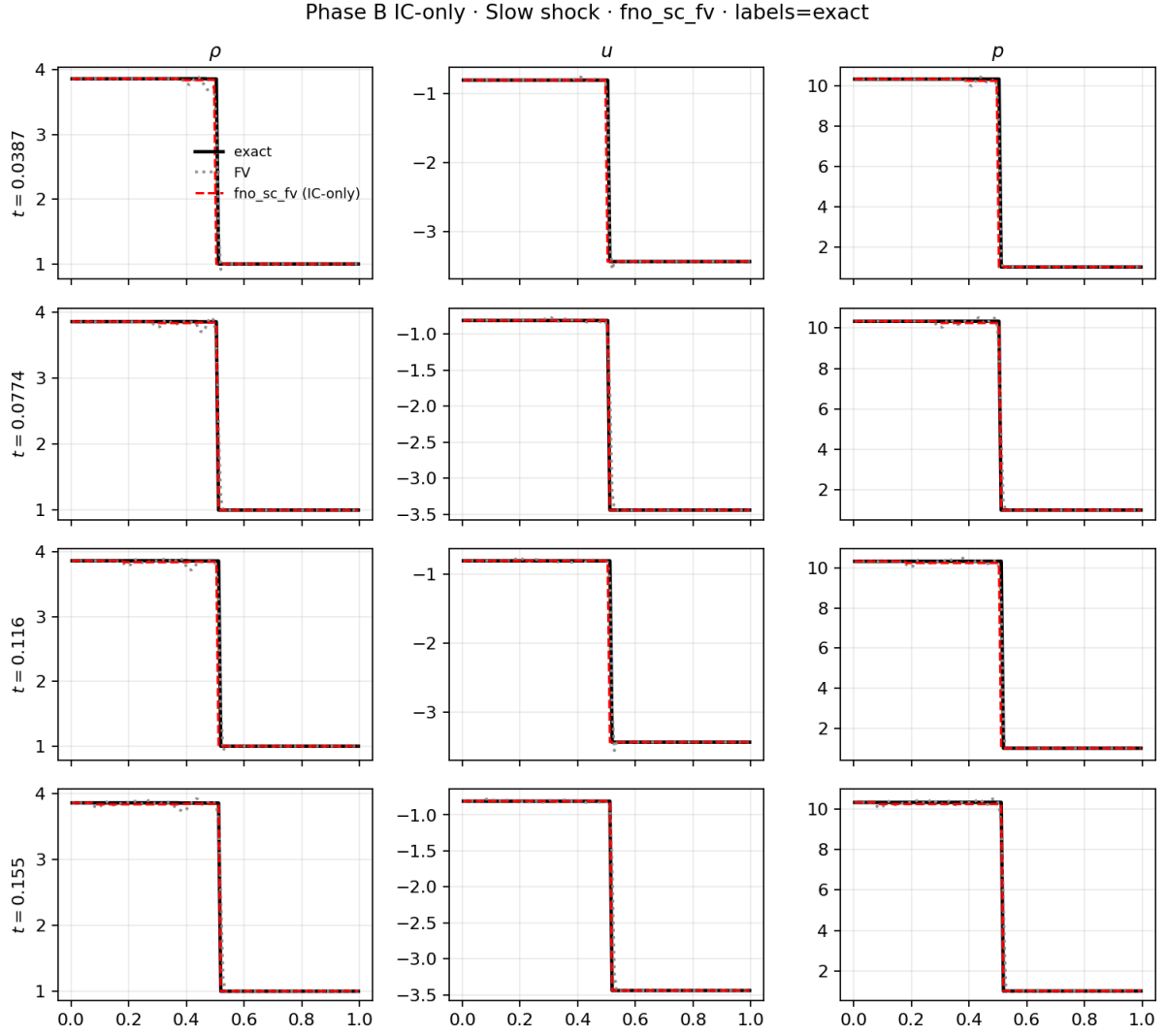


Fig. 12 Slow shock (family 9).

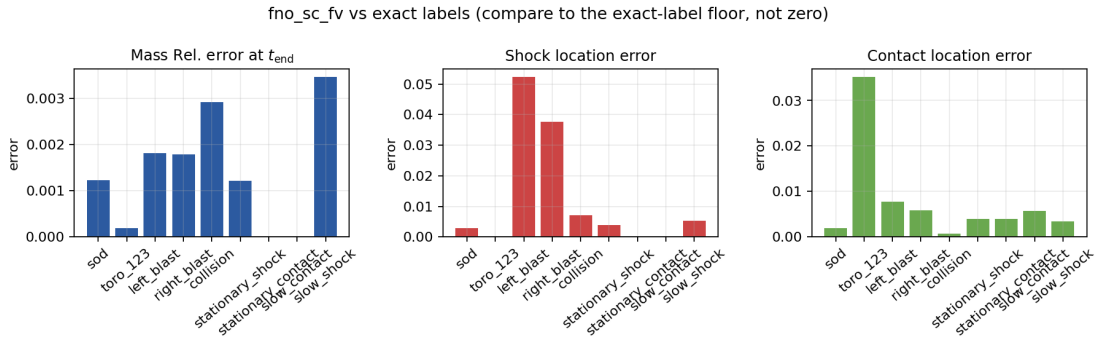


Fig. 13 Physics diagnostics on the classic set. Median shock / contact x -err is 0.006/0.004; median RH relative residual is 0.0015.



Fig. 14 Holdout median Rel- L^2 for the trained WS-2 run.

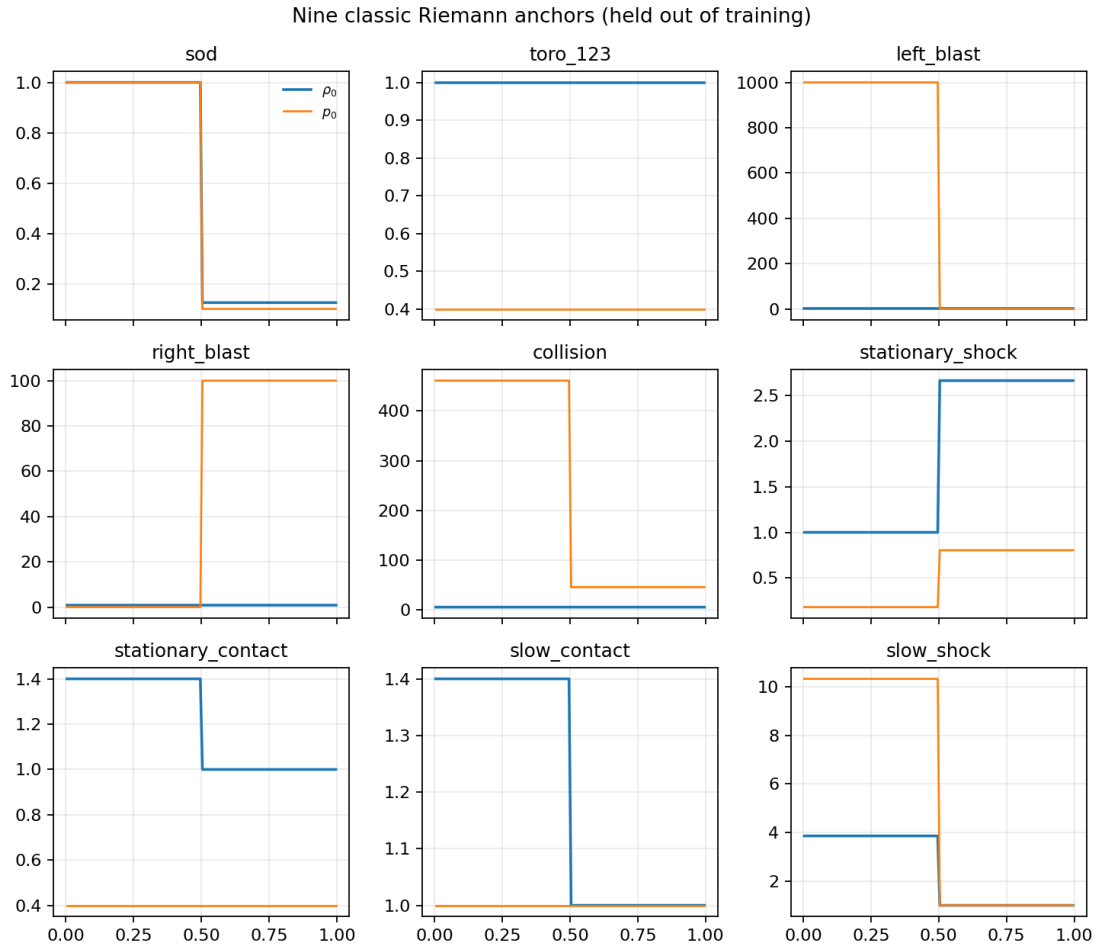


Fig. 15 The nine classic Riemann anchors (held out of training).

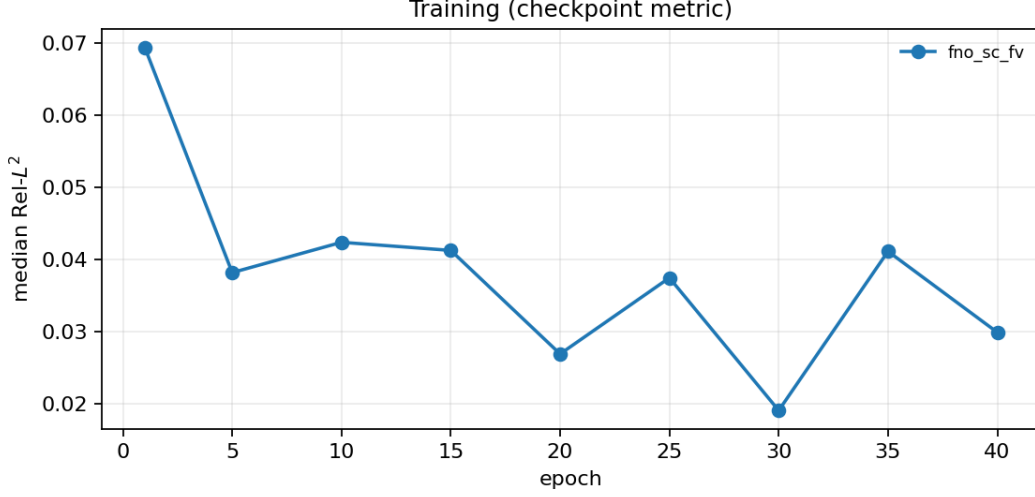


Fig. 16 Original sparse-cadence seed-0 training log, retained for historical diagnosis; it is not the basis of the confirmatory selection result.

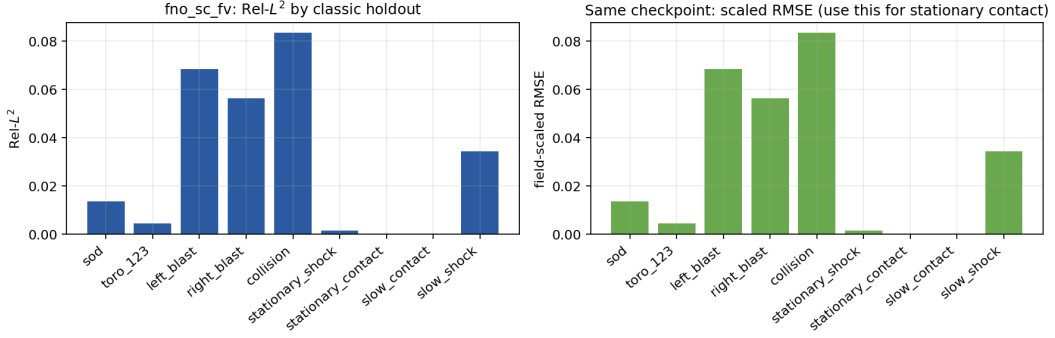


Fig. 17 Classic-anchor Rel- L^2 by case for the trained WS-2 checkpoint.

B Per-seed weak-base and teacher cells

Table 8 reports the seed-level medians that enter Table 13 of the main manuscript. The PVRS mean is not driven by a single favourable run: all five seeds beat frozen PVRS on the hard median. Seed 2 of the one-pass two-shock residual does not beat the frozen classic median 0.035. Star-only collision scores that fall below the 0.027 oracle floor are shock-side assignment artefacts.

Table 8 Per-seed classic / hard / worst Rel- L^2 on the nine anchors (validation-median checkpoint).

Recipe	Seed	Classic	Hard	Worst
Residual on PVRS	0	0.025	0.048	0.059
	1	0.039	0.044	0.127
	2	0.027	0.052	0.097
	3	0.007	0.034	0.059
	4	0.008	0.048	0.059
Residual on TS $N = 1$	0	0.020	0.048	0.059
	1	0.014	0.044	0.069
	2	0.037	0.065	0.068
	3	0.007	0.048	0.050
	4	0.002	0.034	0.052
Star-sup. only	0	1.4×10^{-4}	2.1×10^{-4}	0.022
	1	2.7×10^{-5}	7.4×10^{-5}	0.008
	2	6.9×10^{-5}	1.0×10^{-4}	0.047
	3	1.4×10^{-4}	0.008	0.034
	4	4.9×10^{-5}	8.8×10^{-5}	1.7×10^{-4}

C Front-insensitive metrics and the collision floor

The space-time primitive Rel- L^2 is sensitive to one-cell shock assignment. The main manuscript reports the front-aware robustness summary in Table 10. That table repeats the nine-anchor comparison for Q_0 , the Newton-star oracle, and the seed-0 field residual using (i) the paper primitive Rel- L^2 , (ii) Rel- L^2 of the conserved state $(\rho, \rho u, E)$, (iii) a plateau-only primitive Rel- L^2 on cells whose label $|\partial_x q|$ lies at or below the sample 0.7-quantile, and (iv) the complementary front-band Rel- L^2 .

The remaining notes are the node-level assignment details. On the collision tube the oracle primitive score 0.027 is almost entirely the front band (front 0.147, plateau 2×10^{-8}). That residual is one grid cell at one intermediate time ($t = 0.016$, $x = 0.699$): the right shock sits on a node and the float64 label and the batched sampler $\xi \leq s$ assign opposite sides ($\rho \approx 31$ versus $\rho \approx 6$). The initial and final slices match to 10^{-8} . The Q_0 collision plateau error 0.028 is the remaining Picard star bias. The seed-0 residual is worse on both the plateau (0.076) and the front band (0.223). Right blast is the one family where the residual improves the plateau (0.003 versus Q_0 0.039) while degrading the front band (0.121 versus 2×10^{-4}); the paper primitive score still favours Q_0 .

Newton-star-only collision scores below 0.027 place the right shock on the label’s side of that node. At seed 0 the learned star is already Newton-close ($p_* = 1691.69$ versus 1691.65, $u_* = 8.688$ versus 8.690); the remaining field error is the known floor plus two further nodes ($t = 0.0054$, $x = 0.566$ and $t = 0.027$, $x = 0.832$) from $|\Delta u_*| = 0.0015$. The teacher loss is saturated ($(\Delta u_*/\text{span})^2 \sim 10^{-9}$) and cannot penalise a one-cell assignment.

D Reproducibility details

Timing used Python 3.11.14 and PyTorch 2.3.0 on an Apple M2 (macOS 27.0) under default power settings. Training and the timing results in Table 17 of the main manuscript used the CPU. All principal comparisons use the locked partition, fixed label-generation procedure, and predetermined seeds described in Section 4 of the main manuscript. The two-shock Picard starter is initialised from the PVRS estimate.

The anonymized reproducibility archive available at submission contains the complete implementation, unit tests, data-generation procedure, locked protocol, evaluation scripts, raw outputs, and environment specification. It also records the Monte Carlo randomization seed and the numerical environment used for the reported results.

Tests cover exact-star recovery of Sod, zero-logits matching the detached guess, the acoustic gate blocking a contact velocity leak, log-plateau loss gradients, the star-identity residual at the exact Sod star, the PVRS / two-rarefaction / two-shock branch switches, and a zero star-supervision residual at the exact Sod star.

The Pareto and Q_0 -residual five-seed studies are not learned-improvement results. The PVRS residual improves that weak base; it does not reach Q_0 . The star-supervision-only study uses a Newton teacher at train time.

E Finite-difference checks of $\mathcal{S}$

Table 9 differentiates a primitive field MSE through $\mathcal{S}$ at the frozen Q_0 star ($\varepsilon = 10^{-4}$, centred differences). Agreement is acceptable away from contacts and wave-boundary kinks. Collision $\partial L / \partial u_*$ is a documented kink, not a coding error.

Table 9 Autograd versus centred finite differences of the field MSE at the Q_0 star ($\varepsilon = 10^{-4}$).

Anchor	AD $\partial_{\log p}$	FD $\partial_{\log p}$	AD ∂_u	FD ∂_u
Sod	$+2.30 \times 10^{-5}$	$+2.30 \times 10^{-5}$	$+1.72 \times 10^{-5}$	$+1.72 \times 10^{-5}$
Toro 123	$+1.85 \times 10^{-11}$	$+1.85 \times 10^{-11}$	0	0
Left blast	$+1.28 \times 10^2$	$+1.28 \times 10^2$	-4.24×10^{-3}	-4.21×10^{-3}
Right blast	+2.25	+2.25	-1.63×10^{-3}	-1.79×10^{-3}
Collision	-4.28×10^3	-4.28×10^3	$+1.41 \times 10^{-3}$	0
Stat. shock	$+3.50 \times 10^{-6}$	$+3.51 \times 10^{-6}$	-1.52×10^{-8}	-1.54×10^{-8}
Stat. contact	0	$\sim 10^{-8}$	0	0

References

- [1] 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.
- [2] 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.
- [3] Aidan Chaumet and Jan Giesselmann. Improving weak PINNs for hyperbolic conservation laws: Dual norm computation, boundary conditions and systems. *SMAI Journal of Computational Mathematics*, 10:373–401, 2024. doi: 10.5802/smai-jcm.116.
- [4] Tim De Ryck and Siddhartha Mishra. Numerical analysis of physics-informed neural networks and related models in physics-informed machine learning. *Acta Numerica*, 33:633–713, 2024. doi: 10.1017/S0962492923000089.