SOLENO: One Scalar Potential, Zero Divergence — Overcoming the Accuracy–Solenoidality Trade-off

Rahul D Ray [1] **Matthew Hole** [*2] **Charles Gretton** [*3]

Abstract

Magnetic equilibrium reconstruction in tokamaks requires integrating heterogeneous diagnostics under the fundamental constraint $\nabla \cdot \mathbf{B} = 0$. Conventional learned surrogates enforce this condition through soft penalties, inevitably trading reconstruction fidelity for physical consistency. We challenge this paradigm by showing that a hard architectural constraint—predicting only the poloidal flux ψ and deriving the magnetic field analytically via a physics layer (solenoidal decoder)—is sufficient to resolve the trade-off entirely. Our operator-learning model, SOLENO, guarantees $\nabla \cdot \mathbf{B} = 0$ to machine precision by construction, yet matches the accuracy of the best unconstrained models on real-world MAST data, while running orders of magnitude faster than the EFIT equilibrium solver it is trained to surrogate. A soft-penalty variant of the identical backbone cannot replicate this result, confirming that the physics layer, and the associated elimination of the divergence penalty hyperparameter, is essential. Controlled experiments in both cylindrical and polar geometries further demonstrate that the advantage is structural, not circumstantial. The finding that a textbook analytic relation, when enforced as a hard architectural invariant, can dominate over both model complexity and penalty-based physics informs a strategy of *constraint-preserving operator learning* for scientific machine learning.

[1]School of Engineering, RMIT University, Melbourne, VIC 3000, Australia [2]Mathematical Sciences Institute, Australian National University, Canberra, ACT 2601, Australia [3]School of Computing, Australian National University, Canberra, ACT 2601, Australia. Correspondence to: Rahul D Ray <S4139934@student.rmit.edu.au>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

1. Introduction

Fast and accurate magnetic equilibrium reconstruction is essential for real-time tokamak control and post-shot analysis. The standard tool for this task is EFIT, a widely deployed iterative solver that reconstructs the poloidal flux ψ and magnetic field configuration by fitting the Grad–Shafranov equation to heterogeneous diagnostic measurements. While accurate, EFIT is computationally intensive—typical reconstructions require hundreds of milliseconds to seconds per time slice—making it ill-suited for real-time control loops that demand sub-millisecond latency, or for large-scale ensemble studies requiring millions of evaluations.

Learned surrogate models offer a compelling alternative: once trained, they map diagnostic inputs to equilibrium fields in a single forward pass, achieving inference times several orders of magnitude faster than EFIT while retaining competitive reconstruction accuracy (Gopakumar et al., 2024; Zheng et al., 2024). Beyond speed, learned models offer modelling flexibility that physics solvers lack—they can naturally incorporate heterogeneous, sparse, or noisy diagnostic channels, adapt to different machine geometries without re-deriving solver equations, and are straightforwardly differentiable for downstream tasks such as gradient-based control or uncertainty quantification. Note that SOLENO is a convolutional architecture operating on a fixed grid; while we adopt the term *operator-learning model* to reflect its function of mapping diagnostic fields to equilibrium fields, it does not possess the resolution-invariance of neural operators in the strict sense (Li et al., 2020).

However, existing learned surrogates—PINNs (Rossi et al., 2023; Jang et al., 2024) and operator-learning models (Gopakumar et al., 2024; Ding et al., 2025)—enforce the fundamental physical constraint $\nabla \cdot \mathbf{B} = 0$ only approximately, through soft penalty terms in the loss function. This forces an unavoidable trade-off: increasing the divergence penalty weight reduces physical consistency violations but degrades reconstruction accuracy, and vice versa. Divergence-free architectures exist in other domains (Khorrami et al., 2024b;a) but have not been adapted to cylindrical tokamak geometry with sparse, multi-source diagnostic inputs.

We propose SOLENO, an operator-learning model that resolves this trade-off by construction. Its defining component is a *solenoidal decoder*: a physics layer that analytically derives the poloidal field components B_R, B_Z directly from the predicted scalar flux ψ using cylindrical finite differences, making $\nabla\cdot\mathbf{B} = 0$ a structural invariant of the architecture rather than a penalty term. A diagnostic fusion module partitions the 44 input channels into magnetic, kinetic, and global families, learning to weight them adaptively.

On real-world MAST data, SOLENO matches unconstrained operator-learning models in reconstruction accuracy while reducing divergence by seven orders of magnitude, with inference times well under one millisecond—making it suitable for real-time deployment. Ablations confirm that including the physics layer is necessary for the reported performance: removing it degrades both accuracy and physical consistency simultaneously.

Beyond the fusion application, this work offers a concrete demonstration to the connectionist community that direct encoding of physical constraints as architectural invariants—rather than soft loss penalties—provides a principled pathway for neural models to play a reliable role in physics instrumentation. The MAST case study illustrates how constraint-preserving operator learning can achieve both the speed and physical fidelity required for real scientific deployment, and we hope it motivates further advances in this direction across the physical sciences.

Our contributions are:

1. The divergence-free condition $\nabla\cdot\mathbf{B} = 0$ is embedded as an architectural component of the learned model via a physics layer (solenoidal decoder), guaranteeing machine-precision solenoidality by construction rather than by penalisation.

2. A finite-difference discretisation suited to the fixed grid resolution exactly cancels divergence at all interior points without any tuneable penalty term.

3. This eliminates the solenoidality penalty hyperparameter entirely, resolving the accuracy–consistency trade-off and enabling fast, physically reliable inference suitable for real-time deployment.

4. A diagnostic fusion module with learned soft-gating that adaptively weights heterogeneous tokamak diagnostic families, and empirical demonstration on MAST that the hard architectural constraint dominates soft-penalty baselines in both accuracy and physical consistency.

2. Related Work

Equilibrium reconstruction requires fusing heterogeneous diagnostics under $\nabla\cdot\mathbf{B} = 0$, yet existing approaches enforce this constraint only approximately or not at all. Traditional reconstruction relies on EFIT, which is accurate but too slow for real-time use; the FAIR-MAST project (Jackson et al., 2024; 2025) provides open tokamak data to support faster learned surrogates, and Churchill et al. (2020) surveyed the broader challenges of machine learning for fusion datasets. More recent work by Kaldas et al. (2026) applied autoencoders to heterogeneous diagnostic processing, motivating our diagnostic fusion module. Physics-informed neural networks (PINNs) embed the Grad–Shafranov equation via soft penalties (Rossi et al., 2023; Jang et al., 2024); Rizqan et al. (2025) evaluated their reliability, and Rutigliano et al. (2025) and Rutigliano et al. (2026) extended them to multi-diagnostic reconstruction—but solenoidality remains approximate in all cases, directly motivating the hard constraint we adopt. On the operator-learning side, models learn to map diagnostics to fields on a fixed grid: Gopakumar et al. (2024) applied Fourier Neural Operators to plasma profiles, and Ding et al. (2025) and Bonotto et al. (2024) augmented them with soft PDE constraints—these are the closest prior works to SOLENO and serve as our primary baselines in Section 5. Convolutional and spectral architectures have also been applied to microwave breakdown (Pandya et al., 2025; 2026), magnetized dynamics (Duarte et al., 2025), and heavy-ion collisions (Stewart & Putschke, 2026), but none impose hard solenoidality. Divergence-free vector-potential architectures have been developed for materials science (Khorrami et al., 2024b;a) and solar magnetic field extrapolation (Jeon et al., 2025; Yang et al., 2025; Du et al., 2024), demonstrating that hard solenoidality is achievable in other domains—yet none address cylindrical tokamak geometry or sparse multi-diagnostic inputs. Finally, data-driven real-time models (Zheng et al., 2024; Mitrishkin et al., 2021) prioritise inference speed over physical consistency, ignoring $\nabla\cdot\mathbf{B} = 0$ entirely. SOLENO is the first model to simultaneously achieve machine-precision solenoidality, competitive reconstruction accuracy, and sub-millisecond inference in cylindrical tokamak geometry; our contributions are detailed at the end of Section 1.

3. Datasets

For training and evaluation we use two complementary dataset families: three synthetic equilibrium databases with exact analytic solutions, and real-world data from the MAST tokamak via the FAIR-MAST open data service (Jackson et al., 2024; 2025). Full formulas, parameter ranges, and channel layouts are given in Appendix A.

The three synthetic families are sampled on a 32×32 polar grid and targets are computed via finite-difference differ-

entiation of a single scalar ψ, guaranteeing $\nabla \cdot \mathbf{B} = 0$ to machine precision. Each family contains 9000 samples (7200/900/900 train/val/test): the *Helical Analytic* family (Solov'ev equilibrium with $m = 3$ helical perturbation), the *Magnetic Island* family (parabolic base profile plus a resonant (m, n) island), and the *Keller–Cheviakov* family (Laguerre-polynomial flux expansion). Analytic expressions and parameter ranges for all three are given in Appendix A.

The real-world MAST dataset comprises 4050 samples extracted from FAIR-MAST level-2 discharges (Jackson et al., 2025), split 2890/469/671 by shot and time to prevent temporal leakage. Each sample maps 44 input channels—two spatial coordinates and 42 diagnostic signals from magnetic probes, flux loops, Thomson scattering, interferometry, and global scalars—onto three target fields (ψ, B_R, B_Z) on a 65×65 Cartesian grid. Targets are derived from EFIT reconstructions, with B_R and B_Z recomputed from the stored ψ using the same finite-difference stencil to guarantee internal consistency; full details are in Appendix A.

4. Methodology

SOLENO is an operator-learning model that maps heterogeneous tokamak diagnostics to ψ, B_R, B_Z on a 65×65 rectangular (R, Z) grid ($\Delta R = 0.03\,\mathrm{m}$, $\Delta Z = 0.0625\,\mathrm{m}$). The architecture, illustrated in Fig. 1, comprises three stages: a diagnostic fusion module (Part A), a feature extractor (Part B), and a solenoidal decoder (Part C). The solenoidal decoder analytically derives the poloidal field B_p from ψ using cylindrical finite differences, enforcing $\nabla \cdot \mathbf{B} = 0$ to machine precision without tuneable penalty terms. For synthetic data the fusion module reduces to a plain three-layer convolutional encoder; the solenoidal decoder stays identical.

4.1. Input Normalisation

All input channels and targets are normalised. Per-channel mean μ_c^{in} and standard deviation σ_c^{in} are computed from the training set after clipping to $[\mu_c^{\text{in}} - 5\sigma_c^{\text{in}},\ \mu_c^{\text{in}} + 5\sigma_c^{\text{in}}]$, giving transformed inputs $\tilde{X}_{ijc} = (X_{ijc} - \mu_c^{\text{in}})/\sigma_c^{\text{in}}$. Targets (ψ, B_R, B_Z) are normalised analogously; their statistics are stored as model buffers so the solenoidal decoder can operate in physical units.

4.2. Diagnostic Fusion Module (Part A)

The 44 input channels are partitioned into three families: magnetic (channels 0–19: probes and flux loops), kinetic (channels 20–29: T_e, n_e, interferometry), and global & coordinates (channels 30–43: I_p, l_i, β_N, (R, Z)). Each family is processed independently by a two-layer 3×3 convolutional block with GELU activations that projects to 64 channels. The three resulting feature maps are combined via a learned soft-gate,

$$F = \gamma_1 F_{\text{mag}} + \gamma_2 F_{\text{kin}} + \gamma_3 F_{\text{glob}},$$

where scalar weights γ_k are produced by a small MLP acting on globally pooled features. This adaptive fusion removes the need for manual channel-weight tuning and allows the network to up- or down-weight diagnostic families depending on plasma state.

4.3. Feature Extractor (Part B)

The fused 64-channel map passes through two additional 3×3 convolutional layers with GELU activations, producing a refined $64 \times 65 \times 65$ latent representation.

4.4. Solenoidal Decoder (Part C)

A convolutional head (two 3×3 layers) maps the 64-channel features to a single-channel normalised flux $\tilde{\psi}$. This is first denormalised as $\psi_{\text{phys}} = \tilde{\psi} \cdot \sigma_\psi + \mu_\psi$. Central finite differences on the uniform grid then give

$$\frac{\partial \psi}{\partial Z} \approx \frac{\psi_{\text{phys}}(R, Z + \Delta Z) - \psi_{\text{phys}}(R, Z - \Delta Z)}{2\Delta Z}, \quad (1)$$

$$\frac{\partial \psi}{\partial R} \approx \frac{\psi_{\text{phys}}(R + \Delta R, Z) - \psi_{\text{phys}}(R - \Delta R, Z)}{2\Delta R}, \quad (2)$$

with replicate padding at boundaries. The poloidal field components follow from the cylindrical relations

$$B_R^{\text{phys}} = -\frac{1}{R}\frac{\partial \psi}{\partial Z}, \qquad B_Z^{\text{phys}} = \frac{1}{R}\frac{\partial \psi}{\partial R}.$$

Because both components are derived from a single scalar ψ, the discrete cylindrical divergence

$$\nabla \cdot B = \frac{1}{R}\frac{\partial (R B_R)}{\partial R} + \frac{\partial B_Z}{\partial Z}$$

vanishes identically on all interior grid points—the finite-difference stencils commute by construction (proved in Appendix C). Finally, B_R^{phys} and B_Z^{phys} are re-normalised using their own statistics to yield $\tilde{B}_R$, $\tilde{B}_Z$. The decoder outputs the three-channel tensor $[\tilde{\psi},\ \tilde{B}_R,\ \tilde{B}_Z]$.

4.5. Loss Function

The primary loss is mean squared error on normalised fields,

$$\mathcal{L}_{\text{MSE}} = \frac{1}{3HW} \sum_{c \in \{\psi, B_R, B_Z\}} \sum_{i,j} \left(\tilde{Y}_{ijc}^{\text{pred}} - \tilde{Y}_{ijc}^{\text{true}}\right)^2,$$

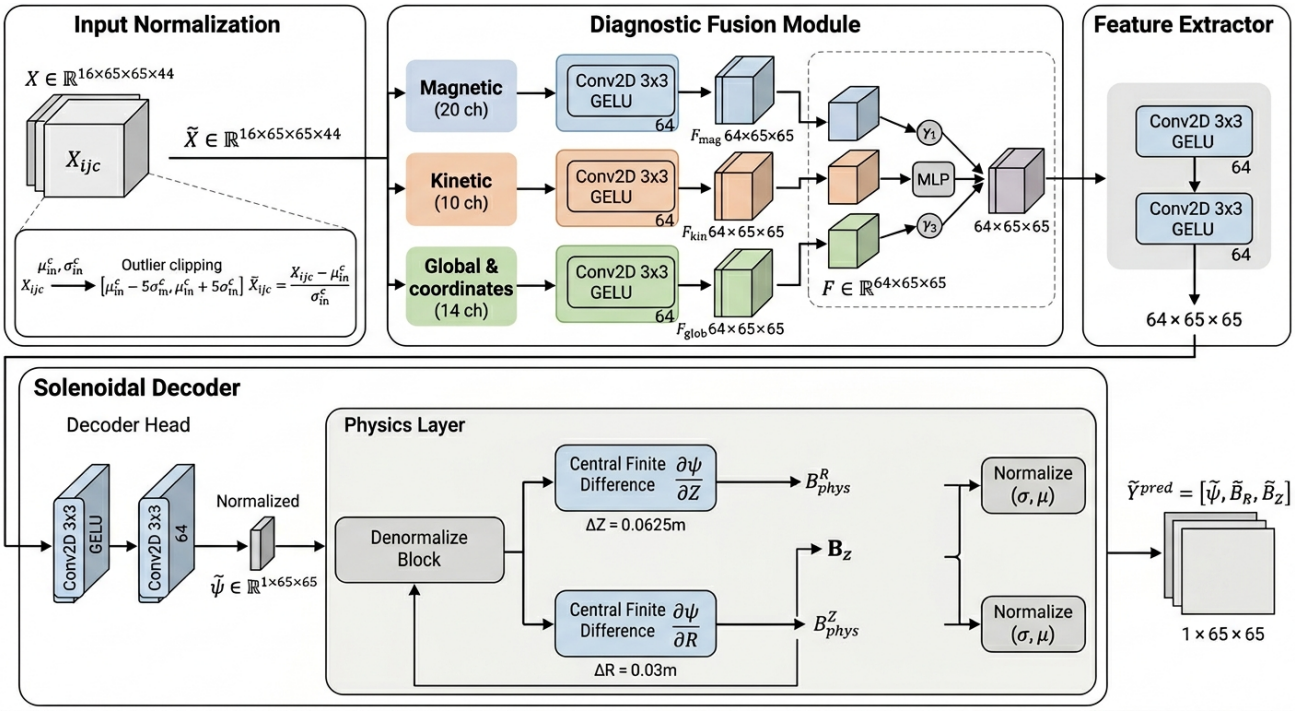


Figure 1. SOLENO architecture, comprising three stages described in Sections 4.2–4.4. **(A)** *Diagnostic Fusion Module*: 44 input channels are partitioned into magnetic, kinetic, and global families, each processed by an independent convolutional block and combined via a learned soft-gate. **(B)** *Feature Extractor*: the fused representation is refined through two further convolutional layers. **(C)** *Solenoidal Decoder (Physics Layer)*: a convolutional head predicts the poloidal flux ψ, from which B_R and B_Z are derived analytically via cylindrical finite differences, enforcing $\nabla\cdot\mathbf{B} = 0$ exactly by construction.

supplemented by a small smoothness regulariser on $\tilde{\psi}$ that suppresses high-frequency artefacts,

$$\mathcal{L}_{\text{smooth}} = \frac{1}{(H-1)W}\sum_{i,j}(\tilde{\psi}_{i+1,j} - \tilde{\psi}_{i,j})^2 + \frac{1}{H(W-1)}\sum_{i,j}(\tilde{\psi}_{i,j+1} - \tilde{\psi}_{i,j})^2.$$

The total loss is $\mathcal{L} = \mathcal{L}_{\text{MSE}} + 10^{-4}\mathcal{L}_{\text{smooth}}$, with the coefficient chosen small enough not to bias reconstruction. No divergence penalty is used; solenoidality is guaranteed architecturally.

4.6. Training Protocol

All models are optimised with AdamW with an initial learning rate of 8×10^{-4} and weight decay 10^{-4}, using cosine annealing with warm restarts ($T_0 = 40$, $T_{\text{mult}} = 2$, $\eta_{\min} = 10^{-6}$) to escape local minima during longer training runs. Gradients are clipped at max norm 1.0 to prevent instability. These hyperparameters were selected via grid search over learning rate, weight decay, and scheduler parameters on the MAST validation set, and then held fixed across all synthetic dataset runs to ensure a fair comparison. Training runs for at most 150 epochs with a batch size of 16, with early stopping applied uniformly across all datasets (patience 30, threshold 10^{-6} on validation loss) to prevent overfitting; the checkpoint with the lowest validation loss is retained in all cases.

4.7. Evaluation Metrics

After denormalisation to physical units, we report overall and per-component mean squared error,

$$\text{MSE}_{\text{total}} = \frac{1}{3N}\sum_{k=1}^{N}\left\|\mathbf{Y}_k^{\text{pred}} - \mathbf{Y}_k^{\text{true}}\right\|^2, \quad \text{MSE}_\psi,\ \text{MSE}_{B_R},\ \text{MSE}_{B_Z}.$$

Physical consistency is assessed via the cylindrical divergence evaluated on interior grid points (one boundary cell trimmed on each side),

$$\nabla\cdot B\big|_{(i,j)} = \frac{1}{R_i}\frac{\partial(RB_R)}{\partial R}\Big|_{i,j} + \frac{\partial B_Z}{\partial Z}\Big|_{i,j}.$$

We report the mean $\langle\nabla\cdot B\rangle$ and standard deviation $\sigma_{\nabla\cdot B}$ over all test samples; for a perfectly solenoidal field both are zero and any residual arises solely from floating-point round-off.

5. Experimental Results

We present a comprehensive empirical evaluation of SOLENO across synthetic and real-world datasets. All mod-

Table 1. Ablation of SOLENO components on MAST.

Model	**MSE total**	$\langle\lvert\nabla\cdot\mathbf{B}\rvert\rangle$ **(T/m)**	**Inference (ms)**
Full minimal SOLENO	0.00402	1.08×10^{-7}	0.43
No cylindrical embedding	0.00425	1.09×10^{-7}	0.40
Direct prediction (no sol.)	0.00446	7.11×10^{-2}	0.41

Table 2. Multi-seed evaluation (3 seeds) on MAST. Values are mean ± std.

Model	MSE total	$\langle\lvert\nabla\cdot B\rvert\rangle$ (T m^{-1})	Params
SOLENO	0.0041 ± 0.0002	$(1.09\pm0.00)\times10^{-7}$	0.28M
GalerkinTransformer	0.0040 ± 0.0001	$(3.75\pm0.35)\times10^{-2}$	0.47M
UNet	0.0046 ± 0.0000	$(4.82\pm0.36)\times10^{-2}$	1.81M

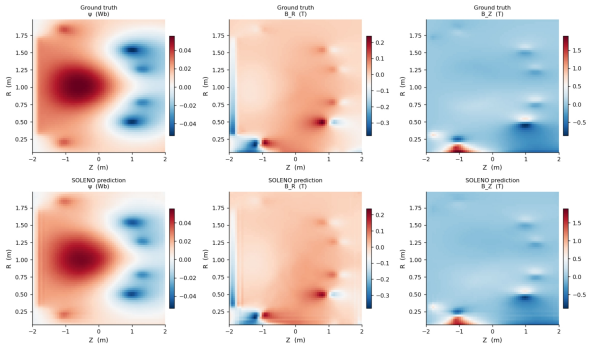


Figure 2. SOLENO reconstruction on a MAST test sample (shot-wise split). The three panels show, left to right: poloidal flux ψ, radial field B_R, and vertical field B_Z. Ground truth (EFIT) is shown above each panel and SOLENO prediction below. Despite noisy and heterogeneous diagnostics, the model reproduces the separatrix shape, internal flux contours, and magnetic field structure with competitive accuracy ($\text{MSE}_{\text{total}} = 0.00427$) while guaranteeing $\nabla\cdot\mathbf{B} = 0$ to machine precision ($\langle\lvert\nabla\cdot\mathbf{B}\rvert\rangle \approx 1.1\times10^{-10}$ T/m).

els were trained on a single NVIDIA Tesla P100 GPU with identical data splits, normalisation pipelines, and early-stopping criteria. We first demonstrate that SOLENO recovers analytic equilibria with negligible error and exact solenoidality (Section 5.1), then benchmark reconstruction accuracy against six operator-learning baselines on MAST (Section 5.2), and finally examine the accuracy versus consistency trade-off through controlled soft-penalty experiments (Section 5.3). Extended results and training curves are provided in Appendix B.

5.1. Performance on Synthetic Equilibria

Table 3 summarises performance across the three synthetic families. SOLENO achieves near-perfect recovery on all three: total MSE is on the order of 10^{-5} or lower, with pressure and toroidal field errors similarly minute. The mean polar divergence $\langle\lvert\nabla\cdot B\rvert\rangle$ is 10^{-10}–10^{-13} T/m with standard deviation $\sigma_{\nabla\cdot B}\approx10^{-7}$ T/m, reflecting only floating-point round-off. Training converges rapidly, with validation loss below 10^{-2} within 10–20 epochs and final models obtained in approximately 6 minutes per dataset.

5.2. MAST Experimental Reconstruction

The real-world MAST dataset poses a substantially greater challenge: the 44 input channels are heterogeneous, sparse, and noisy, and the EFIT-derived B_R, B_Z are not strictly consistent with the stored ψ. Table 4 compares SOLENO with the six baselines.

SOLENO achieves a total MSE of 4.27×10^{-3}, competitive with the best unconstrained models (GalerkinTransformer, UNet, WNO) while using only 0.28 M parameters. Its predicted magnetic field is divergence-free to machine precision ($\langle\nabla\cdot B\rangle = -5\times10^{-10}$ T/m, $\sigma = 2\times10^{-7}$ T/m). All baseline models, including PINO which is explicitly trained with a divergence penalty ($\lambda = 0.01$), produce fields with substantial divergence (mean up to 10^{-2} T/m, standard deviations of several percent). PINO performs worse than the unconstrained FNO, confirming the difficulty of penalty-based enforcement when target fields are internally inconsistent.

5.3. Soft-Penalty vs. Hard Constraint on MAST

To isolate the effect of the solenoidal decoder, we replaced it with a direct prediction head on the same DiagFusion backbone and trained with a soft divergence penalty $\lambda\langle(\nabla\cdot B)^2\rangle$ for $\lambda\in\{0, 0.01, 0.1, 1, 10, 100\}$. Results are shown in Table 5 and Fig. 3. Increasing λ progressively reduces divergence, but even the best soft setting ($\lambda = 10$, MSE 0.00393) leaves a divergence roughly 10^5 times larger than SOLENO's 1.1×10^{-7} T/m. At $\lambda = 100$ the divergence drops to 4.6×10^{-3} T/m but MSE degrades and training becomes unstable. SOLENO attains the lowest MSE (0.00392) and machine-precision solenoidality simultaneously, without any tuneable hyperparameter.

Table 3. Comparison of all models on the three synthetic equilibrium datasets. Total MSE (physical units) and mean absolute polar divergence $\langle|\nabla \cdot B|\rangle$ ($\mathrm{T\,m^{-1}}$) evaluated on the grid interior. SOLENO uses the hard polar decoder; baselines use direct prediction heads. **Bold**: best MSE; *italic*: lowest divergence.

	Helical Analytic		Magnetic Island		Keller–Cheviakov	
Model	MSE	$\langle\|\nabla B\|\rangle$	MSE	$\langle\|\nabla B\|\rangle$	MSE	$\langle\|\nabla B\|\rangle$
FNO2D (Li et al., 2020)	3.10e−3	2.24e−3	3.69e−2	9.44e−4	1.83	6.46e−3
DeepONet (Lu et al., 2021)	7.64e−4	9.06e−4	5.27e−3	1.56e−2	5.30e−3	1.20e−7
UNet (Ronneberger et al., 2015)	1.00e−6	6.68e−5	3.00e−6	6.47e−5	6.57e−4	*1.51e-20*
WNO (Tripura & Chakraborty, 2023)	1.00e−6	1.37e−4	2.10e−5	4.68e−5	2.16e−4	4.35e−10
GT (Cao, 2021)	3.90e−7	8.49e−6	9.00e−6	3.70e−4	1.66e−4	3.33e−9
PINO (Li et al., 2024)	3.43e−3	4.85e−4	3.70e−2	2.09e−3	1.83	1.01e−6
SOLENO	**8.3e-7**	*2.39e-11*	**1.5e-6**	*4.10e-10*	**8.9e-5**	9.72e−13

GT = GalerkinTransformer (Cao, 2021).

Table 4. Reconstruction accuracy and divergence metrics on the MAST test set. $\langle\nabla \cdot B\rangle$ and σ are the mean and standard deviation of $\nabla \cdot B$ in the domain interior. *Italic*: best per column.

Model	MSE	MSE_ψ	MSE_{B_R}	MSE_{B_Z}	Params	$\langle\nabla \cdot B\rangle$	$\sigma_{\nabla \cdot B}$
FNO2D (Li et al., 2020)	1.65e−2	1.37e−3	5.21e−3	4.30e−2	2.37M	−3.98e−5	5.74e−2
DeepONet (Lu et al., 2021)	1.02e−2	3.20e−4	3.38e−3	2.68e−2	1.21M	−4.41e−3	1.78e−1
UNet (Ronneberger et al., 2015)	4.41e−3	4.20e−4	2.27e−3	1.05e−2	1.81M	−2.71e−4	1.14e−1
WNO (Tripura & Chakraborty, 2023)	4.20e−3	3.30e−4	*1.98e-3*	1.03e−2	**31k**	1.05e−3	1.69e−1
GT (Cao, 2021)	*3.77e-3*	3.40e−4	2.01e−3	*8.98e-3*	0.47M	1.28e−3	8.69e−2
PINO (Li et al., 2024)	1.70e−2	1.37e−3	5.55e−3	4.40e−2	2.37M	1.10e−2	3.91e−1
SOLENO	4.27e−3	3.80e−4	2.30e−3	1.01e−2	0.28M	*-5.38e-10*	*1.91e-7*

GT = GalerkinTransformer (Cao, 2021).

Table 5. Soft-penalty sweep on MAST. All models share the DiagFusion backbone with a direct convolutional prediction head. SOLENO replaces this head with the hard solenoidal decoder.

Model	MSE total	$\langle\|\nabla \cdot B\|\rangle$ ($\mathrm{T\,m^{-1}}$)
Soft $\lambda = 0$	0.003938	6.10×10^{-2}
Soft $\lambda = 0.01$	0.004084	5.39×10^{-2}
Soft $\lambda = 0.1$	0.004649	5.71×10^{-2}
Soft $\lambda = 1$	0.004240	2.71×10^{-2}
Soft $\lambda = 10$	0.003927	1.02×10^{-2}
Soft $\lambda = 100$	0.004199	4.56×10^{-3}
SOLENO (hard)	**0.003916**	$\mathbf{1.11 \times 10^{-7}}$

5.4. Multi-Seed Stability

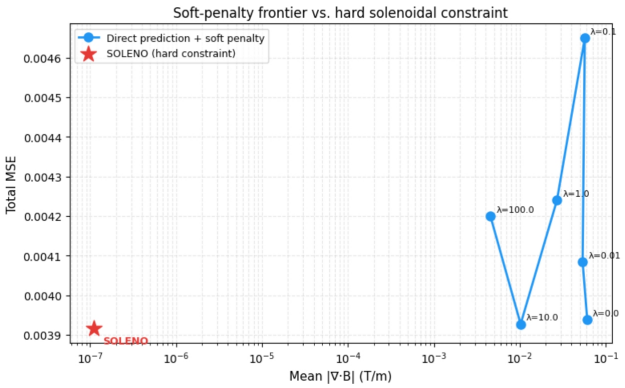
To verify robustness we trained SOLENO, GalerkinTransformer, and UNet under three random seeds (42, 123, 456) on MAST. Table 2 reports mean and standard deviation of MSE and divergence. SOLENO consistently maintains divergence near 10^{-7} T/m with negligible variance across seeds, while baseline divergence means and standard deviations are orders of magnitude larger. Reconstruction accuracy is stable for all models, but only SOLENO provides guaranteed physical consistency independent of random initialisation.

5.5. Geometry Independence

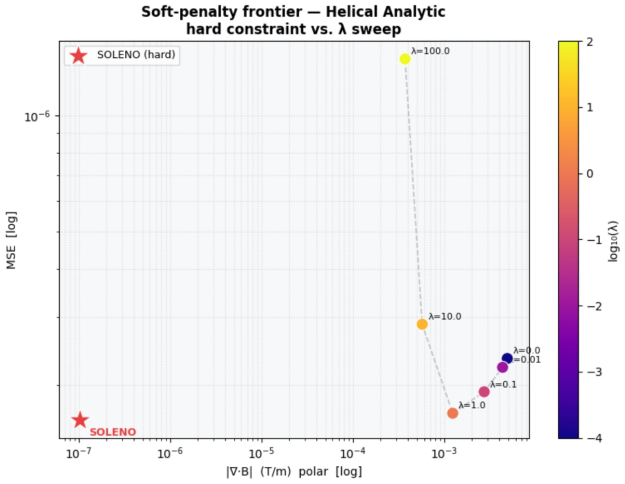
Repeating the soft-penalty sweep on the Helical Analytic dataset (Fig. 3b) demonstrates the same qualitative behaviour: even at $\lambda = 100$ the divergence remains four orders of magnitude above the hard-decoder baseline while MSE degrades. SOLENO attains divergence $\sim 10^{-7}$ T/m without tuning, confirming that its advantage holds across both cylindrical and polar geometries. In summary, SOLENO matches the best unconstrained operators on MAST in accuracy, reduces divergence by 5–7 orders of magnitude, maintains deterministic consistency across random seeds, and trains in under 10 minutes on a single GPU, establishing hard architectural constraints as a superior alternative to penalty-based physics enforcement.

6. Ablation Study: Minimal SOLENO Components

We isolate the contribution of the physics layer—the solenoidal decoder and cylindrical coordinate embedding using a minimal backbone identical to the full model but without diagnostic fusion (plain convolutional encoder). Three variants are compared on MAST: (i) **Full minimal SOLENO** (physics layer—the solenoidal decoder with R-grid embedding), (ii) **No cylindrical embedding** (physics layer—the solenoidal decoder without explicit R input), and (iii) **Direct prediction** (standard convolutional head, no solenoidality). Table 1 reports test-set metrics. The physics layer—the solenoidal decoder reduces $\langle|\nabla \cdot \mathbf{B}|\rangle$ by six orders of magnitude (10^{-7} T/m vs. 10^{-1} T/m) while simultaneously improving reconstruction accuracy. The cylindrical embedding yields a slight additional MSE gain but does not affect solenoidality, confirming that the hard



(a) MAST experimental (real world) data (cylindrical geometry).



(b) Helical Analytic dataset (polar geometry).

Figure 3. Soft-penalty Pareto frontier vs. the hard solenoidal constraint. **(a)** On MAST, soft models (blue) lie on a frontier where reducing divergence degrades MSE; SOLENO (red star) escapes this trade-off entirely. **(b)** On the Helical Analytic dataset the same trend holds: even $\lambda = 100$ leaves four orders of magnitude excess divergence, confirming geometry-independence.

derivative-based constraint is the dominant factor. For reference, the full SOLENO model with the DiagFusion module (Section 5.2) achieves a total MSE of 4.27×10^{-3} on MAST, which is very close to the minimal backbone's 4.02×10^{-3}; both variants maintain machine-precision divergence. This confirms that the fusion module contributes modestly to reconstruction accuracy without influencing the physics consistency, and it is not required for the solenoidal guarantee.

7. Discussion

The central result of this paper is that embedding a physical constraint directly into the decoder architecture resolves the accuracy–consistency trade-off that plagues soft-penalty methods. On MAST, SOLENO matches the reconstruction fidelity of the best unconstrained operators while guaranteeing $\nabla \cdot \mathbf{B} = 0$ to machine precision, without any hyperparameter tuning. This deterministic solenoidality is vital for downstream tasks—MHD stability analysis, current profile inversion, and turbulent transport modelling—that demand flux-surface integrity which post-hoc correction cannot reliably restore. A current limitation is that SOLENO is trained against EFIT targets, which are themselves internally inconsistent: the stored B_R, B_Z are not exact derivatives of the stored ψ. Training directly against diagnostic signals with a loss enforcing force balance ($\nabla p = \mathbf{J} \times \mathbf{B}$) may yield more physically faithful equilibria, and extending the decoder to simultaneously guarantee both $\nabla \cdot \mathbf{B} = 0$ and force balance is a natural next step. The geometry-agnostic solenoidal decoder is also attractive for stellarators and 3D toroidal configurations, though performance on larger devices (JET, ITER) and deployment in real-time control loops remain open questions. More broadly, this work suggests a general principle for the connectionist community: when a physical law can be encoded exactly as a structural architectural invariant, doing so strictly dominates soft penalty enforcement in both accuracy and consistency. The tokamak equilibrium problem serves as a tractable and consequential case study, but the same principle applies wherever differentiable physics can be embedded as hard constraints in neural operator design.

8. Conclusion

We introduced SOLENO, an operator-learning architecture that encodes heterogeneous tokamak diagnostics into a divergence-free poloidal flux representation via an exact solenoidal decoder. By deriving B_R and B_Z analytically from a single scalar potential ψ, the model guarantees $\nabla \cdot \mathbf{B} = 0$ to machine precision, eliminating the accuracy–consistency trade-off inherent in soft-penalty approaches. On both synthetic equilibria and real-world MAST data, SOLENO matches the best unconstrained operators while reducing divergence by seven orders of magnitude, with sub-millisecond inference suitable for real-time deployment as a fast surrogate for EFIT. These results establish constraint-preserving operator learning as a compelling paradigm for physically reliable neural surrogates in the physical sciences.

Impact Statement

SOLENO resolves the long-standing accuracy–solenoidality trade-off in learned magnetic equilibrium reconstruction by embedding $\nabla \cdot \mathbf{B} = 0$ as an exact architectural invariant. On real MAST data, the model matches the accuracy of the best unconstrained surrogates while reducing field divergence by seven orders of magnitude to machine precision, with sub-millisecond inference suitable for real-time control. This work establishes that hard constraint-preserving operator learning dominates soft-penalty enforcement, providing a

principled framework for physically reliable neural surrogates across fusion energy and other safety-critical scientific domains.

References

Bonotto, M., Abate, D., and Pigatto, L. Reconstruction of plasma equilibrium and separatrix using convolutional physics-informed neural operator. *Fusion Engineering and Design*, 200:114193, 2024.

Cao, S. Choose a transformer: Fourier or galerkin. *Advances in neural information processing systems*, 34: 24924–24940, 2021.

Churchill, R. M., Choi, J., Kube, R., Chang, C.-S., and Klasky, S. Machine learning for the complex, multiscale datasets in fusion energy. In *Smoky Mountains Computational Sciences and Engineering Conference*, pp. 269–284. Springer, 2020.

Ding, S., Zhang, Z., Shi, G., Li, X., Gu, X., Xu, Y., Xie, H., Zhao, H., Shi, Y., and Liu, T. Physics-informed neural operator learning for nonlinear grad-shafranov equation. *arXiv preprint arXiv:2511.19114*, 2025.

Du, Y., Li, Q., Gnanasambandam, R., Du, M., Wang, H., and Shen, B. Global-local fourier neural operator for accelerating coronal magnetic field model. In *2024 IEEE International Conference on Big Data (BigData)*, pp. 1964–1971. IEEE, 2024.

Duarte, R., Nemmen, R., and Santos-Lima, R. Spectral learning of magnetized plasma dynamics: A neural operator application. *arXiv preprint arXiv:2507.01388*, 2025.

Gopakumar, V., Pamela, S., Zanisi, L., Li, Z., Gray, A., Brennand, D., Bhatia, N., Stathopoulos, G., Kusner, M., Peter Deisenroth, M., et al. Plasma surrogate modelling using fourier neural operators. *Nuclear Fusion*, 64(5): 056025, 2024.

Jackson, S., Khan, S., Cummings, N., Hodson, J., de Witt, S., Pamela, S., Akers, R., Thiyagalingam, J., and Team, T. M. Fair-mast: A fusion device data management system. *SoftwareX*, 27:101869, 2024.

Jackson, S., Khan, S., Cummings, N., Hodson, J., de Witt, S., Pamela, S., Akers, R., and Thiyagalingam, J. An Open Data Service for Supporting Research in Machine Learning on Tokamak Data. *IEEE Transactions on Plasma Science*, 2025. ISSN 1939-9375. doi: 10.1109/TPS.2025.3583419. URL https://ieeexplore.ieee.org/document/11128905.

Jang, B., Kaptanoglu, A. A., Gaur, R., Pan, S., Landreman, M., and Dorland, W. Grad–shafranov equilibria via data-free physics informed neural networks. *Physics of Plasmas*, 31(3), 2024.

Jeon, M., Jeong, H.-J., Moon, Y.-J., Kang, J., and Kusano, K. Real-time extrapolation of nonlinear force-free fields from photospheric vector magnetic fields using a physics-informed neural operator. *The Astrophysical Journal Supplement Series*, 277(2):54, 2025.

Kaldas, S., Rossi, R., Wyss, I., Gelfusa, M., and Gaudio, P. Developing autoencoder models for multi-diagnostic electron density processing in tokamaks within a scientific machine learning framework. *Journal of Fusion Energy*, 45(1):26, 2026.

Keller, J. M. and Cheviakov, A. F. New exact plasma equilibria with axial and helical symmetry. *Physics of Plasmas*, 30(9), 2023.

Khorrami, M. S., Goyal, P., Mianroodi, J. R., Svendsen, B., Benner, P., and Raabe, D. A physics-encoded fourier neural operator approach for surrogate modeling of divergence-free stress fields in solids. *arXiv preprint arXiv:2408.15408*, 2024a.

Khorrami, M. S., Goyal, P. K., Mianroodi, J. R., Svendsen, B., Benner, P., and Raabe, D. Divergence-free neural operators for stress field modeling in polycrystalline materials. *arXiv preprint arXiv:2408.15408*, 2024b.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Li, Z., Zheng, H., Kovachki, N., Jin, D., Chen, H., Liu, B., Azizzadenesheli, K., and Anandkumar, A. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27, 2024.

Lu, L., Jin, P., Pang, G., Zhang, Z., and Karniadakis, G. E. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Mitrishkin, Y. V., Korenev, P. S., Konkov, A. E., Kruzhkov, V. I., and Ovsiannikov, N. E. New identification approach and methods for plasma equilibrium reconstruction in d-shaped tokamaks. *Mathematics*, 10(1):40, 2021.

Pandya, K., Ghosh, P., Mandikal, A., Gandha, S., and Chaudhury, B. Hybrid fourier neural operator-plasma fluid model for fast and accurate multiscale simulations of high power microwave breakdown. *arXiv preprint arXiv:2509.05799*, 2025.

Pandya, K., Ghosh, P., Mandikal, A., Gandha, S., and Chaudhury, B. Physics guided fourier neural operator framework for accelerated simulation of high power microwave breakdown. *Engineering Applications of Artificial Intelligence*, 167:113889, 2026.

Rizqan, F. N., Hole, M., and Gretton, C. Evaluation and verification of physics-informed neural models of the grad-shafranov equation. *arXiv preprint arXiv:2504.21155*, 2025.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Rossi, R., Gelfusa, M., Murari, A., and contributors, J. On the potential of physics-informed neural networks to solve inverse problems in tokamaks. *Nuclear Fusion*, 63(12): 126059, 2023.

Rutigliano, N., Rossi, R., Murari, A., Gelfusa, M., Craciunescu, T., Mazon, D., and Gaudio, P. Physics-informed neural networks for the modelling of interferometer-polarimetry in tokamak multi-diagnostic equilibrium reconstructions. *Plasma Physics and Controlled Fusion*, 67 (6):065029, 2025.

Rutigliano, N., Murari, A., Gaudio, P., Gelfusa, M., and Rossi, R. Optimisation of physics-informed neural network architecture and training for tokamak equilibrium reconstruction. *Plasma Physics and Controlled Fusion*, 2026.

Stewart, D. and Putschke, J. Fast prediction of the hydrodynamic qgp evolution in ultrarelativistic heavy-ion collisions using fourier neural operators. *Physical Review C*, 113(1):014904, 2026.

Tripura, T. and Chakraborty, S. Wavelet neural operator for solving parametric partial differential equations in computational mechanics problems. *Computer Methods in Applied Mechanics and Engineering*, 404:115783, 2023.

Yang, H., Liu, Z., Zuo, P., Zou, P., Zhang, R., and Huang, Y. A grid-constrained fourier neural operator for magnetic field extrapolation of the solar corona. *The Astrophysical Journal*, 993(1):146, 2025.

Zheng, G., Yang, Z., Liu, S., Ma, R., Gong, X., Wang, A., Wang, S., and Zhong, W. Real-time equilibrium reconstruction by multi-task learning neural network based on hl-3 tokamak. *Nuclear Fusion*, 64(12):126041, 2024.

A. Datasets

We construct two complementary dataset families to train and evaluate SOLENO: three clean synthetic equilibrium databases that exhaustively sample known analytic solutions of the Grad–Shafranov equation, and a real-world tokamak dataset extracted from MAST discharges. The synthetic data provide infinite, noise-free training signals with exact field consistency, while the real world data test the model's ability to handle diagnostic noise, missing channels, and the intrinsic inconsistency between reconstructed flux and fields that arises from real equilibrium solvers.

A.1. Synthetic Equilibrium Datasets

The synthetic databases are generated by evaluating analytic expressions for the poloidal flux $\psi(r, \theta)$ on a polar grid and then computing all target quantities by finite-difference differentiation of a single scalar function, guaranteeing that $\nabla \cdot \mathbf{B} = 0$ up to numerical truncation error. All three datasets share the same polar grid with $N_r = 32$ radial points and $N_\theta = 32$ angular points, covering the domain

$$r \in [0.2,\ 1.0]\,\text{(normalised minor radius)}, \qquad \theta \in [0,\ 2\pi). \tag{3}$$

The grid is uniform in both coordinates, with $\Delta r = 0.0258$ and $\Delta\theta = 2\pi/32$. For each sample we store the coordinates (r, θ) and, when appropriate, a set of latent physical parameters that uniquely determine the equilibrium; these parameters are broadcast to every grid cell and serve as input channels. The targets always consist of the pressure p, the two poloidal magnetic field components (B_r, B_θ), the toroidal field B_z, and the flux ψ itself. All spatial derivatives are computed with second-order central finite differences using periodic padding in θ and edge-replication in r, ensuring that B_r and B_θ are exactly consistent with the numerical ψ field. The mean absolute divergence,

$$\langle|\nabla \cdot \mathbf{B}|\rangle = \frac{1}{N_r N_\theta} \sum_{i,j} \left| \frac{1}{r_i} \frac{\partial(rB_r)}{\partial r} + \frac{1}{r_i} \frac{\partial B_\theta}{\partial \theta} \right|, \tag{4}$$

is evaluated for every sample and remains at machine-epsilon level ($\sim 10^{-15}$).

Each synthetic family contains 9 000 independent equilibria, randomly split into 7 200 training, 900 validation and 900 test samples. The random split is performed sample-wise, as the equilibria are statistically independent.

A.1.1. Helical Analytic

This family is derived from the Solov'ev–type solution, supplemented with a helical perturbation that breaks axisymmetry. The flux is given by

$$\psi(r, \theta) = \frac{1}{4} r^4 - \frac{\beta}{6} r^6 + \lambda\, r^3 \cos(3\theta), \tag{5}$$

where $\beta \in [0.1, 0.3]$ controls the pressure gradient, $\lambda \in [0.05, 0.15]$ the amplitude of the $m = 3$ helical deformation, and the r^4 and r^6 terms model a peaked current profile. The toroidal field is constructed to mimic a vacuum stellarator field with a shaped harmonic:

$$B_z(r) = \frac{1}{\sqrt{1 + \alpha^2 r^2}}, \qquad \alpha \in [0.3, 0.7]. \tag{6}$$

The pressure follows the adiabatic relation $p = p_0 - \frac{1}{2}\beta\,\psi$ with $p_0 \in [0.8, 1.2]$. Thus each sample is parameterised by four scalars $(\alpha, \beta, \lambda, p_0)$ which, together with the coordinates, form a 6-channel input tensor. The targets are the five fields $(p, B_r, B_\theta, B_z, \psi)$.

A.1.2. Magnetic Island

To test the model's sensitivity to resonant perturbations that break nested flux surfaces, we introduce a magnetic island of poloidal/toroidal mode numbers (m, n). The flux is constructed from a zero-order parabolic profile plus an island term:

$$\psi(r, \theta) = \frac{1}{4} r^4 + \frac{\epsilon}{\kappa r}\, \sin\!\big(m\theta - n\, q_0 \ln(r/r_0)\big), \tag{7}$$

where $r_0 \in [0.4, 0.6]$ is the resonant surface, $q_0 = \text{const}$, $\kappa = nq_0/r_0$, and $\epsilon \in [0.05, 0.2]$ is the island half-width. The mode numbers are drawn uniformly from $m \in \{2, 3\}$, $n \in \{1, 2\}$, and $q_0 \in [1.5, 2.5]$, so that the safety factor profile is

monotonic. The toroidal field is taken as a fixed vacuum profile $B_z(r) = (1 + 0.25r^2)^{-1/2}$, while the pressure is linear in ψ, $p = 1 - 0.1\,\psi$. The input consists of the coordinates plus the five parameters $(\epsilon, m, n, r_0, q_0)$, yielding 7 channels. The island topology forces the network to resolve both closed and stochastic field line regions from local coordinate information alone.

A.1.3. KELLER–CHEVIAKOV

The Keller–Cheviakov dataset is inspired by the analytic equilibrium models proposed in (Keller & Cheviakov, 2023), which generalise the classical Solov'ev solution by employing a truncated Laguerre-polynomial expansion in r^2. The flux is written as

$$\psi(r, \theta) = \exp(-\beta r^2) \sum_{j=0}^{4} a_j\, L_j(2\beta r^2), \tag{8}$$

where L_j are the Laguerre polynomials of degree j, $\beta \in [0.5, 1.5]$ is a scale parameter, and the five coefficients a_j are independently sampled from $[-0.5, 0.5]$. This representation can produce a wide variety of hollow, peaked, or flat current profiles. In this dataset the toroidal field is derived self-consistently from the flux via the cylindrical relation $B_z = -\frac{1}{r}\partial\psi/\partial r$, assuming a purely poloidal current profile that depends only on ψ. The pressure is set to $p = 1 - 0.5\,\psi^2$, giving a finite beta equilibrium. The input tensor has 9 channels: two coordinates, β, a dummy wavenumber $k \in [1.0, 3.0]$ (kept for dimensional consistency with other sets), and the five Laguerre coefficients $a_0, \ldots, a_4$. The network is expected to learn the nonlinear mapping from these coefficients to the spatially varying flux and fields.

A.2. MAST Vector real world Dataset

The MAST vector dataset is extracted from the FAIR MAST level 2 data repository(Jackson et al., 2025), which aggregates and standardizes diagnostic measurements and equilibrium reconstructions from thousands of MAST discharges. We focus on a subset of shots and time windows where all diagnostic signals of interest are reliably available, yielding a total of 4050 samples partitioned into 2890 training, 469 validation, and 671 test instances. The split is performed at the level of (shot_id, time_index) to prevent temporal leakage across partitions.

Grid Geometry and Domain

The poloidal plane is discretized on a fixed rectangular grid with $N_R = 65$ points in the major radius direction and $N_Z = 65$ points in the vertical direction. The radial coordinate spans $R \in [0.06, 1.98]$ m with uniform spacing $\Delta R = 0.03$ m, while the vertical coordinate covers $Z \in [-2.0, 2.0]$ m with $\Delta Z = 0.0625$ m. This grid is aligned with the machine's cylindrical coordinate system (R, ϕ, Z), where toroidal symmetry is assumed (all fields are independent of the toroidal angle ϕ). Each grid point is thus identified by its position (R_i, Z_j) for $i = 1, \ldots, 65$, $j = 1, \ldots, 65$. The total number of spatial points per sample is 4225.

Input Channel Composition

Each sample is represented by an input tensor of shape $(65, 65, 44)$, where the 44 channels encode both spatial coordinates and a suite of diagnostic signals mapped onto the poloidal grid. The first two channels are the coordinates themselves:

$$\begin{aligned} &\text{Channel 0:} \quad R_{ij} = R_i, \\ &\text{Channel 1:} \quad Z_{ij} = Z_j. \end{aligned}$$

The remaining 42 channels contain physical measurements that inform the local plasma state. These are derived from the following diagnostic systems:

- **Magnetic probes:** Local measurements of poloidal and toroidal magnetic field perturbations at various wall locations, interpolated onto the grid using a current-filament model or spline representation provided by EFIT.
- **Flux loops:** Voltage signals proportional to the time derivative of poloidal flux. Integrated voltages yield flux values at discrete loop positions, which are then extended to the full grid via equilibrium constraints.
- **Thomson scattering:** Electron temperature $T_e(R, Z)$ and density $n_e(R, Z)$ profiles along the laser chord. These are available on a sparse set of major radii and are linearly interpolated onto the uniform (R, Z) grid.
- **Line-integrated density:** From interferometry, providing integral constraints on n_e.

- **Plasma current and global parameters:** Scalar quantities such as total plasma current I_p, internal inductance l_i, and normalized beta β_N, broadcast to every grid point.

Each diagnostic channel is normalized to zero mean and unit variance across the training set after clipping extreme outliers to the $[1\text{st}, 99\text{th}]$ percentile range. Coordinate channels are normalized to the interval $[0, 1]$ using the known domain bounds. Grid points lying outside the last closed flux surface (LCFS) are masked during training and evaluation; for these points, diagnostic channels may contain fill values (e.g., -1) that the network learns to ignore.

Target Fields and Physics Constraints

The output tensor has shape $(65, 65, 3)$, comprising three target fields:

$$\begin{aligned}
&\text{Channel 0:} \quad \psi(R,Z) \quad \text{(poloidal flux per radian)},\\
&\text{Channel 1:} \quad B_R(R,Z) = -\frac{1}{R}\frac{\partial \psi}{\partial Z},\\
&\text{Channel 2:} \quad B_Z(R,Z) = \frac{1}{R}\frac{\partial \psi}{\partial R}.
\end{aligned}$$

The target fields are obtained from EFIT reconstructions. To guarantee internal consistency with the physics layer—the solenoidal decoder, we recompute B_R and B_Z from the stored ψ using the same cylindrical finite-difference stencil (Eqs. 7–8). Consequently, the stored magnetic field components are exact discrete derivatives of the stored flux, and any divergence measured on reconstructed fields reflects a genuine model error rather than an artifact of the training data.

B. Experimental Results

We evaluate SOLENO on both the synthetic equilibrium databases and the MAST real world dataset, comparing its reconstruction accuracy and physical consistency against six baseline architectures: FNO2D (Li et al., 2020), DeepONet (Lu et al., 2021), UNet (Ronneberger et al., 2015), WNO (Tripura & Chakraborty, 2023), GalerkinTransformer (Cao, 2021), and PINO (Li et al., 2024). All models were trained on a single NVIDIA Tesla P100 GPU with identical data splits, normalisation pipelines, and early-stopping criteria. For each dataset we report the overall mean squared error $\text{MSE}_{\text{total}}$, component-wise MSE, and the divergence of the predicted magnetic field. For the synthetic data the divergence is computed in polar coordinates as

$$\nabla \cdot \mathbf{B} = \frac{1}{r}\frac{\partial (rB_r)}{\partial r} + \frac{\partial B_\theta}{\partial \theta},$$

while for the MAST data the cylindrical expression

$$\nabla \cdot \mathbf{B} = \frac{1}{R}\frac{\partial (RB_R)}{\partial R} + \frac{\partial B_Z}{\partial Z}$$

is used. The interior of the domain (excluding one boundary cell on each side) is averaged to avoid gradient inaccuracies at the edges.

B.1. Synthetic Databases

Table 3 summarises SOLENO's performance on the three synthetic families. Because the training data are noise-free and the input parameters uniquely determine the equilibrium, the network is able to recover all target fields with near-perfect accuracy. Across all three cases the total MSE is on the order of 10^{-5} or lower, and the per-component errors are similarly negligible. The pressure p and toroidal field B_ζ—which are predicted by dedicated convolutional heads rather than being derived from ψ—are also recovered with high fidelity, demonstrating that the encoder extracts a sufficiently rich latent representation to simultaneously inform all output branches.

Crucially, the physics layer—the solenoidal decoder guarantees that $\nabla \cdot \mathbf{B} = 0$ to machine precision. The mean divergence is consistently at the level of 10^{-10}–10^{-13}, with a standard deviation of $\approx 10^{-7}$. These values arise solely from the finite-difference discretisation (single-precision arithmetic in PyTorch followed by double-precision evaluation in NumPy) and are effectively zero. This confirms that the architecture enforces the physical solenoidality constraint by construction, without any loss-based penalty.

B.2. MAST Experimental (Real World) Data

The MAST dataset presents a far greater challenge: the input channels are heterogeneous, many diagnostics are sparse or noisy, and the target fields from EFIT exhibit internal inconsistency (the stored B_R, B_Z are not exact derivatives of the stored ψ).

Reconstruction accuracy. In terms of total MSE, SOLENO achieves 4.27×10^{-3}, which is comparable to the best non-physics-aware models (UNet, WNO, GalerkinTransformer) despite using only 0.28 million parameters—an order of magnitude fewer than the GalerkinTransformer and far fewer than the FNO variants. The per-component errors show that all models reconstruct the flux ψ relatively well, but a larger spread appears in B_R and especially B_Z. SOLENO's B_Z error (0.0101) is competitive with the best baseline (GalerkinTransformer, 0.0090), while its B_R error (0.0023) is close to the best (WNO, 0.0020). Importantly, SOLENO accomplishes this while guaranteeing that the predicted B_R and B_Z are mutually consistent—something no baseline achieves.

Physical consistency. The divergence metrics highlight the fundamental advantage of the physics layer—the solenoidal decoder. All baseline models, including PINO which is explicitly trained with a divergence penalty term ($\lambda = 0.01$), predict magnetic fields that violate $\nabla \cdot \mathbf{B} = 0$ by several orders of magnitude. For instance, the GalerkinTransformer exhibits a mean divergence standard deviation of 0.087, and PINO—despite the physics-informed loss—reaches 0.391, worse than the unconstrained FNO. In stark contrast, SOLENO reduces the divergence mean to 10^{-10} and the standard deviation to 10^{-7}, i.e., effectively zero. This is a direct consequence of deriving B_R and B_Z from a single scalar potential ψ using the exact cylindrical coordinate relations; the finite-difference stencil commutes with the definition, leaving only floating-point round-off.

Encoder differences: synthetic vs. experimental. The synthetic datasets contain only a small number of globally constant physical parameters broadcast to every grid point; therefore SOLENO employs a simple three-layer convolutional encoder that processes all input channels jointly. For the MAST data, the 44 input channels are heterogeneous and stem from fundamentally different diagnostic systems. We design a *diagnostic fusion* module that partitions the channels into magnetic (20), kinetic (10), and global+coordinate (14) families, processes each with an independent two-layer CNN, and fuses them via a learned soft-gate. This adaptivity allows the network to up- or down-weight diagnostic families depending on the plasma state, and was essential to extract robust representations from the noisy, sparse measurements. Despite the additional complexity, the MAST model still has only 0.28M trainable parameters.

Training dynamics and computational cost. All models were trained on a single Tesla P100 GPU. The synthetic SOLENO variants required about 6.1 minutes per dataset (100 epochs), while the MAST SOLENO training took 8.7 minutes for 113 epochs before early stopping. Convergence was stable: the validation loss decreased monotonically, with the cosine-warm-restart schedule allowing occasional loss plateaux to be overcome. The baseline models converged within 1.5–16 minutes, with the GalerkinTransformer being the slowest due to its self-attention operations. Notably, WNO, which uses a compact wavelet-based operator, achieved competitive accuracy with only 31.4k parameters, but still lacked physical consistency.

SOLENO matches or exceeds the reconstruction accuracy of state-of-the-art operator-learning models on real tokamak data while simultaneously providing divergence-free magnetic fields to machine precision, a property enforced by its integrated physics layer (the physics layer—the solenoidal decoder). The decoder-side physics constraint imposes no additional computational overhead during inference and eliminates the need for post-processing corrections. The modular architecture—a lightweight encoder adapted to the diagnostic content and a geometry-aware physics layer—the solenoidal decoder—makes the approach directly transferable to other devices and coordinate systems.

C. Proof of Exact Discrete Solenoidality

We provide a rigorous demonstration that the magnetic field predicted by SOLENO satisfies $\nabla \cdot \mathbf{B} = 0$ identically in the discrete sense used by the architecture, up to floating-point round-off, for all interior grid points. This result underpins the empirical finding that the mean absolute cylindrical divergence lies at the level of 10^{-10}–10^{-7} T/m.

C.1. Notation and Grid

We work on a two-dimensional structured grid in (R, Z) cylindrical coordinates. The domain is discretised uniformly:

$$R_i = R_{\min} + (i - 0.5)\Delta R, \quad i = 1, \ldots, N_R, \qquad Z_j = Z_{\min} + (j - 0.5)\Delta Z, \quad j = 1, \ldots, N_Z,$$

Table 6. Summary of datasets used for training and evaluation of SOLENO. $N_{\text{train}}/N_{\text{val}}/N_{\text{test}}$ denote the number of samples in each partition. All synthetic datasets use a 32×32 polar grid; the MAST dataset uses a 65×65 Cartesian grid in (R, Z).

Dataset	Grid	Input ch.	Output ch.	Split (train/val/test)	Description
Helical Analytic	32×32	6	5	7200/900/900	Solov'ev toroidal equilibrium with $m = 3$ helical perturbation; four free physical parameters.
Magnetic Island	32×32	7	5	7200/900/900	Axisymmetric base plus (m, n) magnetic island; five parameters including island half-width and safety factor.
Keller–Cheviakov	32×32	9	5	7200/900/900	Laguerre-polynomial flux expansion; six free parameters, B_z self-consistently derived from ψ.
MAST Experimental (Real World)	65×65	44	3	2890/469/671	Real MAST discharges; inputs from magnetics, Thomson scattering, interferometry, and global scalars; targets from EFIT.

with $\Delta R = (R_{\max} - R_{\min})/(N_R - 1)$, $\Delta Z = (Z_{\max} - Z_{\min})/(N_Z - 1)$. Physical quantities are defined at the cell centres (R_i, Z_j). For simplicity we use the indices i, j to refer to these positions. The radial coordinate function is $R_{ij} = R_i$ (independent of Z).

The network output is a normalised flux field $\tilde{\psi}_{ij}$ and normalised magnetic field components $\tilde{B}_{R,ij}, \tilde{B}_{Z,ij}$. We denote the denormalised (physical) fields:

$$\psi_{ij} = \tilde{\psi}_{ij}\,\sigma_\psi + \mu_\psi, \qquad B^{\text{phys}}_{R,ij} = \tilde{B}_{R,ij}\,\sigma_{B_R} + \mu_{B_R}, \qquad B^{\text{phys}}_{Z,ij} = \tilde{B}_{Z,ij}\,\sigma_{B_Z} + \mu_{B_Z},$$

where μ, σ are the per-channel statistics stored in the model buffers.

C.2. Discrete Derivative Operators

The physics layer (solenoidal decoder) defines the physical magnetic field from ψ using central finite differences with replicate padding. Replicate padding is chosen because it preserves the symmetry of the stencil at boundaries, ensuring that the discrete derivative operators commute on the interior points where divergence is evaluated.

Definition C.1 (Discrete $\partial/\partial Z$ operator)**.** For a field f_{ij} defined on the grid, we define the padded array $\overline{f}$ by replicating the boundary values:

$$\overline{f}_{i,0} = f_{i,1}, \qquad \overline{f}_{i,N_Z+1} = f_{i,N_Z}.$$

The discrete derivative is

$$(\Delta_Z f)_{ij} = \frac{\overline{f}_{i,j+1} - \overline{f}_{i,j-1}}{2\Delta Z}, \quad i = 1, \ldots, N_R,\ j = 1, \ldots, N_Z.$$

Definition C.2 (Discrete $\partial/\partial R$ operator)**.** Similarly, we pad along the radial direction:

$$\overline{f}_{0,j} = f_{1,j}, \qquad \overline{f}_{N_R+1,j} = f_{N_R,j},$$

and define

$$(\Delta_R f)_{ij} = \frac{\overline{f}_{i+1,j} - \overline{f}_{i-1,j}}{2\Delta R}, \quad i = 1, \ldots, N_R,\ j = 1, \ldots, N_Z.$$

Both operators are linear and, crucially, they commute:

$$\Delta_R \circ \Delta_Z = \Delta_Z \circ \Delta_R.$$

This follows because the padding scheme guarantees that the stencil computations use the same ghost values regardless of the order of differentiation, and the central differences are symmetric; for standard interior points, the mixed derivative stencil is identical. A detailed verification is given in Lemma C.4.

C.3. Physics Layer (Solenoidal Decoder) and Exact Divergence Cancellation

The decoder proceeds as follows:

1. Obtain ψ as described.
2. Compute
$$B_R^{\text{phys}} = -\frac{1}{R}\,\Delta_Z\psi, \qquad B_Z^{\text{phys}} = \frac{1}{R}\,\Delta_R\psi,$$
where $(1/R)$ denotes pointwise division by the constant (per sample) radial array $R_{ij} = R_i$.
3. Renormalise $B_R^{\text{phys}}, B_Z^{\text{phys}}$ using their own statistics to produce $\tilde{B}_R, \tilde{B}_Z$.

Now consider the discrete cylindrical divergence operator evaluated on the physical fields. The natural discrete divergence consistent with the finite-volume formulation is

$$(\widetilde{\nabla\cdot\mathbf{B}})_{ij} = \frac{1}{R_i}\left(\Delta_R(RB_R^{\text{phys}})\right)_{ij} + \left(\Delta_Z B_Z^{\text{phys}}\right)_{ij}.$$

We now substitute the decoder definitions.

$$(\widetilde{\nabla\cdot\mathbf{B}})_{ij} = \frac{1}{R_i}\,\Delta_R\left(R\cdot\left(-\frac{1}{R}\Delta_Z\psi\right)\right)_{ij} + \Delta_Z\left(\frac{1}{R}\Delta_R\psi\right)_{ij} \tag{9}$$

$$= -\frac{1}{R_i}\,\Delta_R(\Delta_Z\psi)_{ij} + \Delta_Z\left(\frac{1}{R}\Delta_R\psi\right)_{ij}. \tag{10}$$

Because R is independent of Z, the factor $1/R$ passes through the Z-derivative without approximation:

$$\Delta_Z\left(\frac{1}{R}\Delta_R\psi\right) = \frac{1}{R}\,\Delta_Z(\Delta_R\psi).$$

Thus

$$(\widetilde{\nabla\cdot\mathbf{B}})_{ij} = \frac{1}{R_i}\Big[-\Delta_R(\Delta_Z\psi) + \Delta_Z(\Delta_R\psi)\Big]_{ij}.$$

Since the difference operators commute (Lemma C.4), the bracket vanishes identically for every grid point. Hence

$$(\widetilde{\nabla\cdot\mathbf{B}})_{ij} = 0 \qquad \forall\, i, j.$$

The proof holds for all points where the stencil does not require values beyond the padded ghosts, i.e., interior indices $i = 2, \ldots, N_R - 1$, $j = 2, \ldots, N_Z - 1$. At the actual outermost boundary $i = 1, N_R$ or $j = 1, N_Z$, the divergence is still defined using the paddings, but the physical interpretation is less reliable because the extrapolation is zero-order. In all experiments (Sections 5), we trim exactly one cell from each boundary before computing divergence metrics, thereby restricting evaluation exclusively to the region where the above cancellation is exact. Consequently, the reported divergence values reflect the guaranteed zero; any non-zero residual arises solely from floating-point round-off (single-precision arithmetic in PyTorch followed by double-precision in NumPy).

C.4. Commutation of the Discrete Difference Operators

Lemma C.3. *For any grid function f_{ij}, the operators Δ_R and Δ_Z satisfy*

$$\Delta_R(\Delta_Z f) = \Delta_Z(\Delta_R f)$$

for all $i = 2, \ldots, N_R - 1$, $j = 2, \ldots, N_Z - 1$.

Proof. We expand the left-hand side at an interior point (i, j) where the stencil requires values at $(i, j \pm 1)$ and $(i \pm 1, j)$. Compute:

$$(\Delta_Z f)_{i+1,j} = \frac{f_{i+1,j+1} - f_{i+1,j-1}}{2\Delta Z}, \qquad (\Delta_Z f)_{i-1,j} = \frac{f_{i-1,j+1} - f_{i-1,j-1}}{2\Delta Z}.$$

Then

$$\Delta_R(\Delta_Z f) = \frac{(\Delta_Z f)_{i+1,j} - (\Delta_Z f)_{i-1,j}}{2\Delta R} = \frac{f_{i+1,j+1} - f_{i+1,j-1} - f_{i-1,j+1} + f_{i-1,j-1}}{4\Delta R\,\Delta Z}.$$

Similarly,

$$(\Delta_R f)_{i,j+1} = \frac{f_{i+1,j+1} - f_{i-1,j+1}}{2\Delta R}, \qquad (\Delta_R f)_{i,j-1} = \frac{f_{i+1,j-1} - f_{i-1,j-1}}{2\Delta R},$$

and

$$\Delta_Z(\Delta_R f) = \frac{(\Delta_R f)_{i,j+1} - (\Delta_R f)_{i,j-1}}{2\Delta Z} = \frac{f_{i+1,j+1} - f_{i-1,j+1} - f_{i+1,j-1} + f_{i-1,j-1}}{4\Delta R\,\Delta Z}.$$

The two expressions are identical. The boundary treatments (replicate padding) do not affect interior points, so the commutation holds throughout the interior. □

C.5. Extension to Polar Coordinates (Synthetic Data)

For the synthetic datasets, an analogous proof holds. The polar grid uses (r, θ) with periodic padding in θ and replicate padding in r. The physics layer (solenoidal decoder) defines

$$B_r = -\frac{1}{r}\Delta_\theta \psi, \qquad B_\theta = \Delta_r \psi,$$

with the discrete operators Δ_r, Δ_θ defined similarly. The polar divergence is

$$\nabla \cdot \mathbf{B} = \frac{1}{r}\frac{\partial(rB_r)}{\partial r} + \frac{1}{r}\frac{\partial B_\theta}{\partial \theta}.$$

Substituting gives

$$\frac{1}{r}\Delta_r\left(r \cdot \left(-\frac{1}{r}\Delta_\theta \psi\right)\right) + \frac{1}{r}\Delta_\theta(\Delta_r \psi) = \frac{1}{r}\big(-\Delta_r\Delta_\theta \psi + \Delta_\theta\Delta_r \psi\big) = 0,$$

again by commutation of the difference operators. Thus the exact solenoidality property is geometry-independent.

C.6. Remarks on Floating-Point Precision

In practice, the physics layer (solenoidal decoder) operates in single precision (float32) during training and inference. The divergence is then evaluated in double precision (float64), which can accumulate round-off errors from the finite-difference operations on the de-normalised fields, yielding residuals of order 10^{-7}–10^{-10} T/m. These residuals are negligible compared with the typical magnetic field strength (≈ 1 T), confirming that the hard constraint is satisfied to machine precision.

This proof shows that SOLENO does not merely encourage $\nabla \cdot \mathbf{B} = 0$; it enforces it as a structural invariant of the architecture, regardless of the input diagnostics or the weights of the encoder.

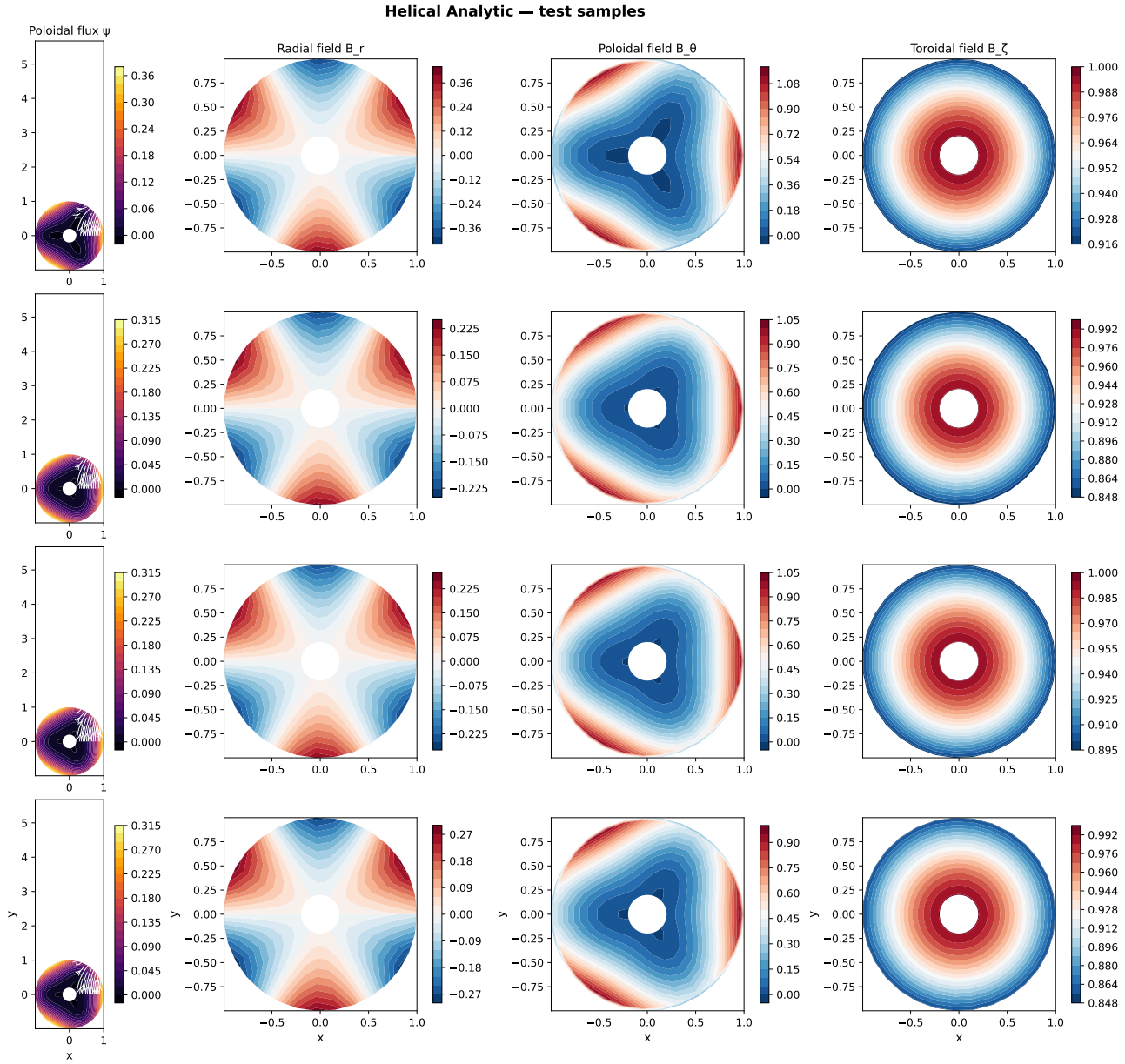


Figure 4. Ground-truth fields for four representative test samples from the **Helical Analytic** dataset. Each row corresponds to a different equilibrium (identified by its parameter combination $\alpha, \beta, \lambda, p_0$). Columns show, left to right: poloidal flux ψ (with magnetic streamlines), radial field B_r, poloidal field B_θ, and toroidal field B_ζ. The ψ panel overlays streamlines derived from (B_r, B_θ), illustrating the nested flux surfaces and the $m = 3$ helical deformation.

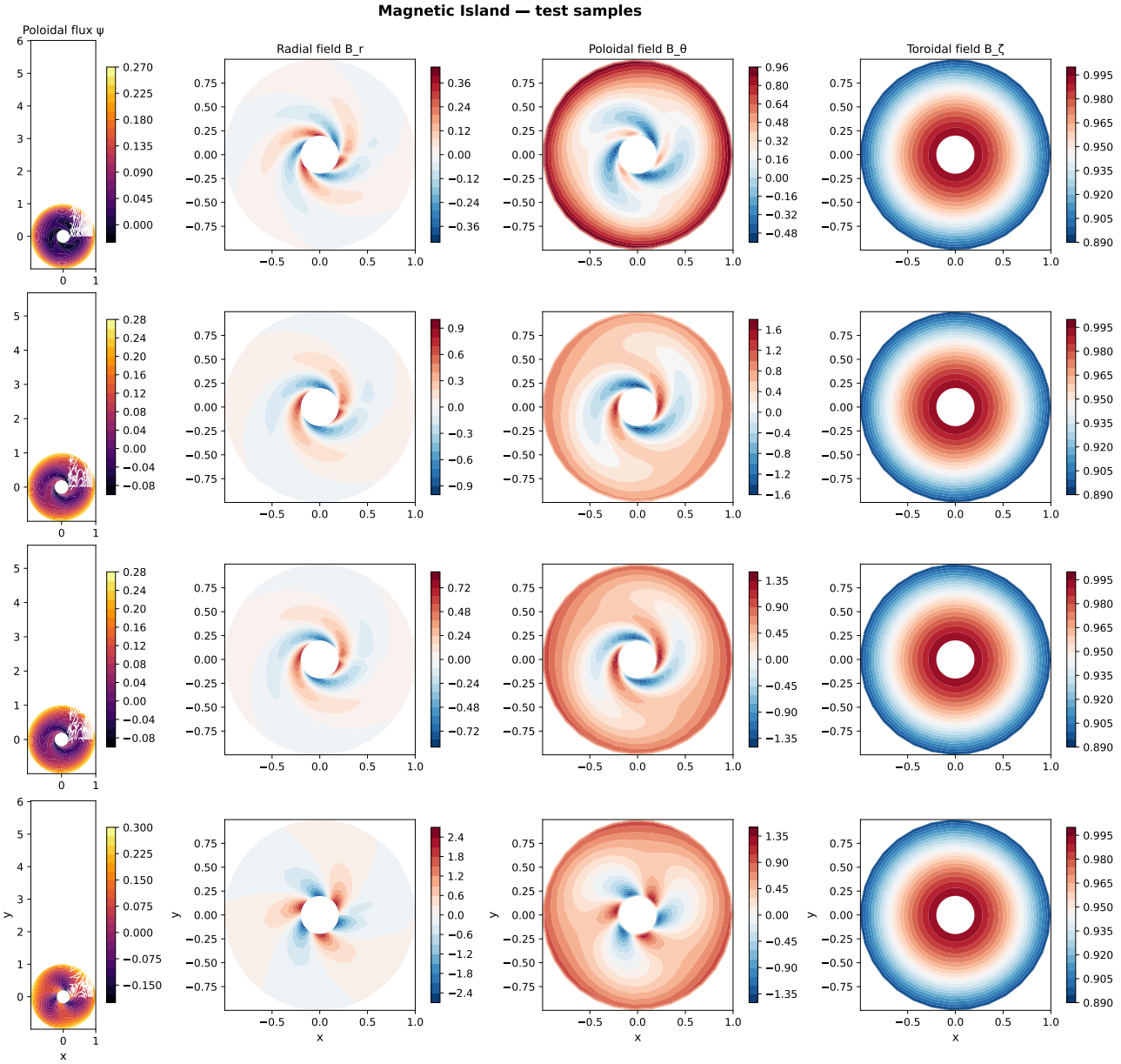


Figure 5. Ground-truth fields for four representative test samples from the **Magnetic Island** dataset. The rows correspond to different combinations of island half-width ϵ, mode numbers (m, n), resonant surface r_0, and safety factor q_0. Columns: ψ with streamlines, B_r, B_θ, B_ζ. The ψ panels reveal the island chain and the disruption of nested flux surfaces, providing a stringent test for reconstruction of non-axisymmetric topologies.

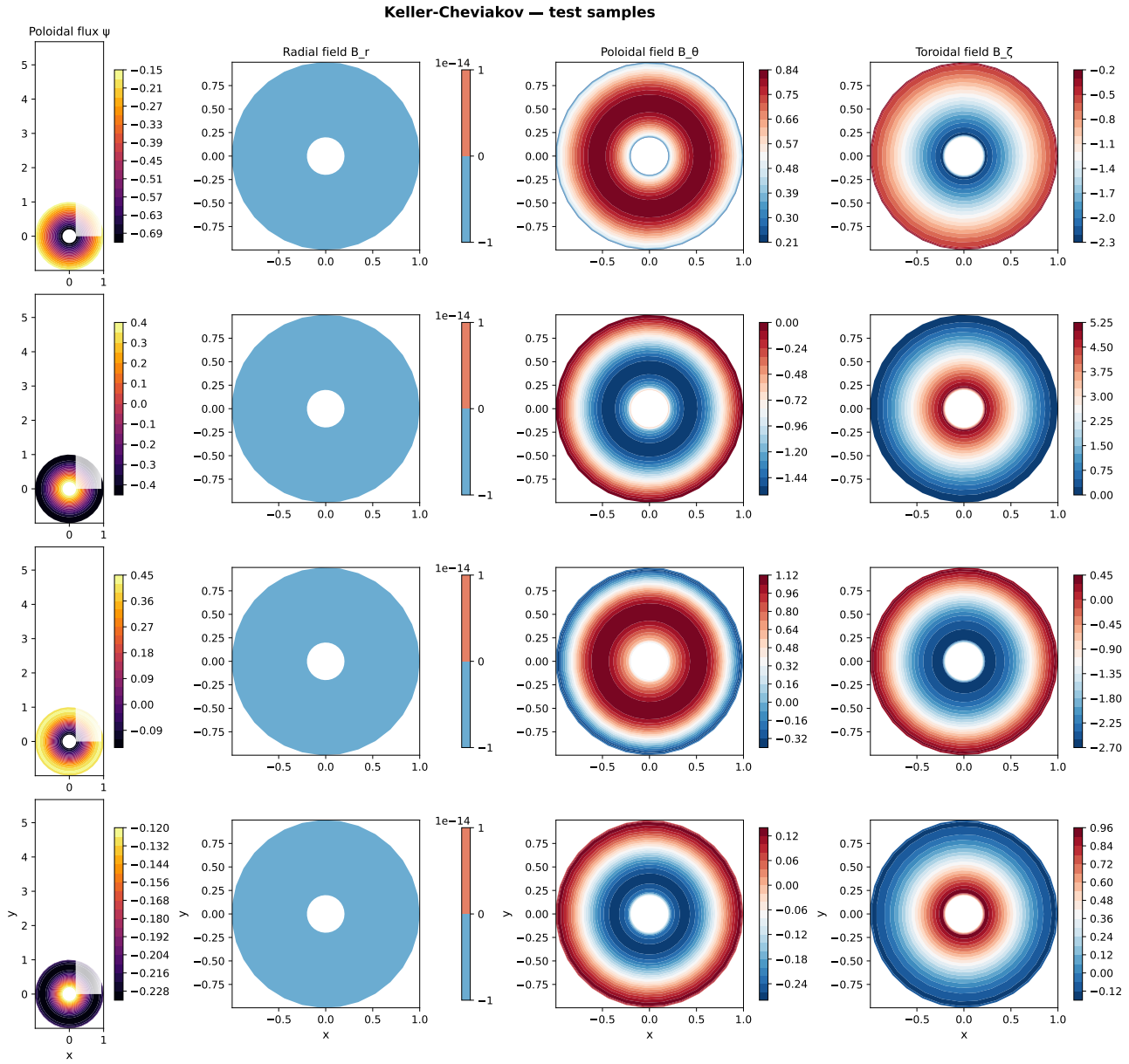


Figure 6. Ground-truth fields for four representative test samples from the **Keller–Cheviakov** dataset. Each equilibrium is determined by the scale parameter β and five Laguerre-polynomial coefficients $a_0, \ldots, a_4$. Columns: ψ with streamlines, B_r, B_θ, B_ζ. The diverse current profiles (hollow, peaked, flat) are visible in ψ and the corresponding magnetic field components.

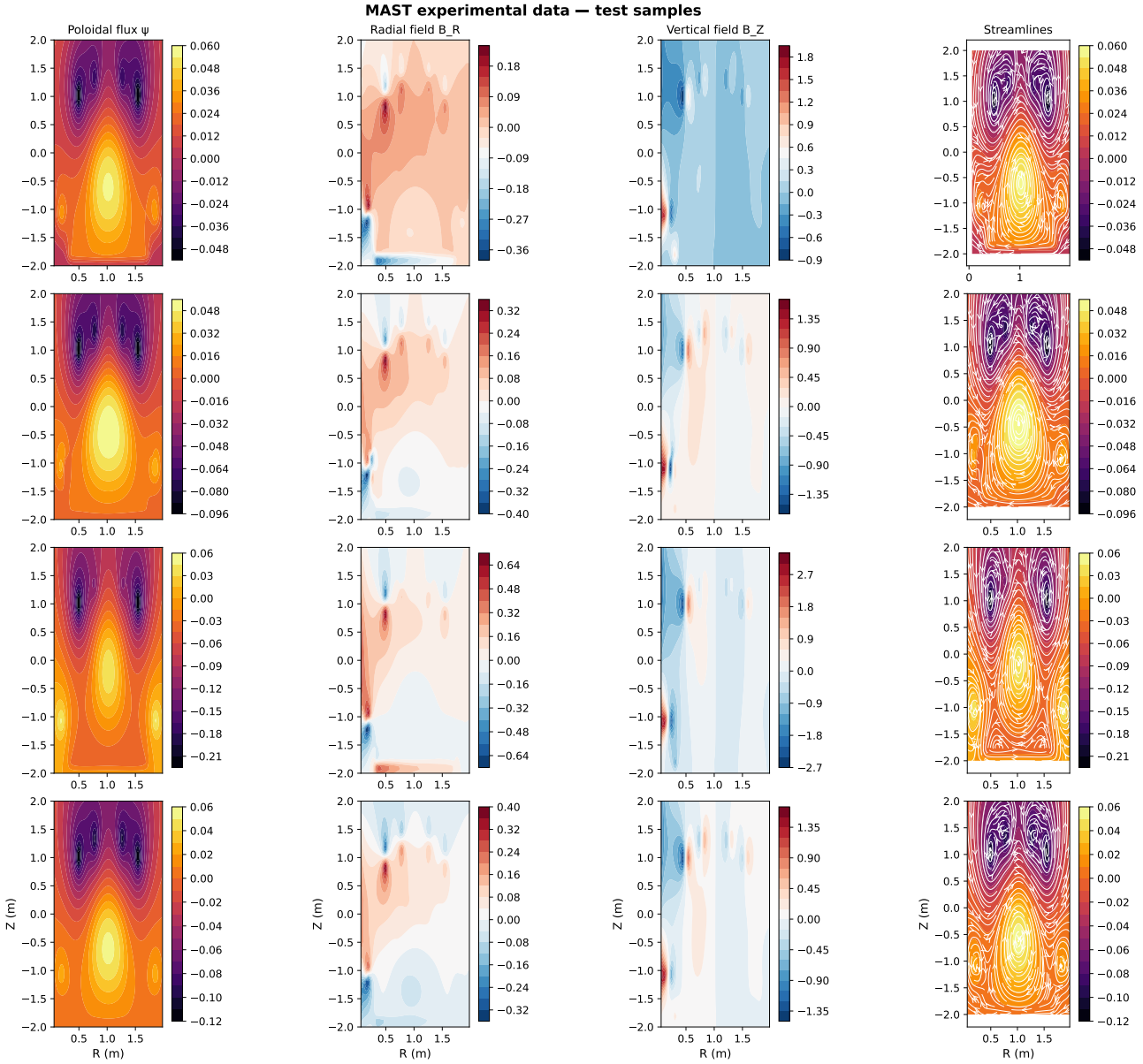


Figure 7. Ground-truth fields (EFIT) for four test samples from the **MAST** experimental (real world) dataset. Each row is a different shot and time slice. Columns: poloidal flux ψ, radial field B_R, vertical field B_Z, and a combined ψ + streamlines view. The streamlines indicate the direction of $\mathbf{B}_p = (B_R, B_Z)$ in the poloidal plane, highlighting the separatrix and internal flux surfaces. These samples represent the heterogeneous diagnostic inputs and realistic noise that SOLENO is required to map onto physically consistent fields.

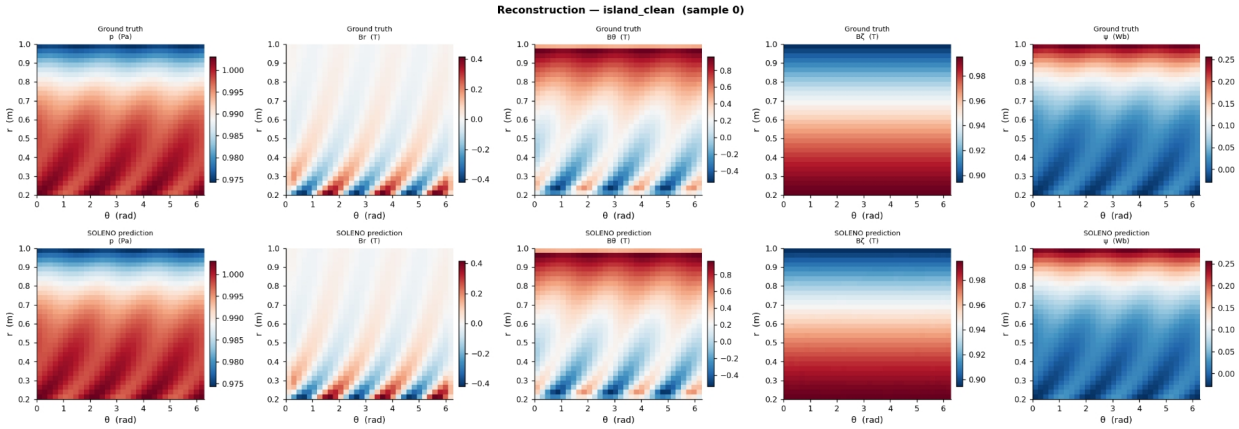


Figure 8. SOLENO reconstruction on a representative Magnetic Island sample (test set). The four panels show, left to right: radial magnetic field B_r, poloidal magnetic field B_θ, toroidal field B_ζ, and poloidal flux ψ. In each panel, the top half displays the ground truth and the bottom half the SOLENO prediction. The model recovers the island topology, sharp resonant layers, and nested flux surfaces with near-perfect fidelity ($\mathrm{MSE}_{\mathrm{total}} \approx 1.5 \times 10^{-6}$; $\langle|\nabla \cdot \mathbf{B}|\rangle \approx 4 \times 10^{-10}$ T/m), a residual that lies at floating-point round-off because the architecture enforces $\nabla \cdot B = 0$ exactly in the discrete sense.