PHASE: MULTI-REGIME MODELING OF INCOMPRESSIBLE MAGNETOHYDRODYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Magnetohydrodynamics (MHD) is central to plasma modeling in astrophysics, space science, fusion, and engineering, but resolving multiscale MHD dynamics is computationally expensive. Machine-learning surrogates enable fast inference by learning reusable solution operators, yet existing models require separate training for each physical regime, limiting generalization across varying parameter settings. We introduce **PHASE**, a *PHysics-Adaptive Scalable operator with residual Error correction*, designed to model incompressible MHD across varying physical parameters with a single model. PHASE combines transfer learning, regime-aware adaptation, physics-centered learning, and residual refinement to improve both physical fidelity and generalization across MHD regimes. Together, these improvements achieve state-of-the-art prediction accuracy on two-dimensional MHD turbulence by reducing relative L_2 errors on physical fields by more than an order of magnitude compared to prior MHD neural-operator baselines. Moreover, PHASE generalizes successfully to unseen parameter values without retraining, demonstrating the cross-regime adaptability expected from operator learning. We evaluate PHASE beyond point-wise prediction errors using derived physical fields, spectral analysis, and distribution statistics, consistently observing improved physical fidelity. We further show that our framework can accurately simulate MHD instabilities by testing it on the Kelvin–Helmholtz instability, demonstrating the robustness of our method.

1 INTRODUCTION

Magnetohydrodynamics (MHD) describes the coupled evolution of electrically conducting fluids and magnetic fields, forming the basis for studying plasmas in astrophysics, space science, fusion, and engineering (Kulsrud, 2005). Predicting MHD dynamics is especially challenging in turbulent and instability-driven regimes, where nonlinear interactions between fluids and magnetic fields generate multiscale and intermittent structures. Traditionally, these dynamics are simulated by numerically solving the governing partial differential equations (PDEs) using finite-element methods (Pencil Code Collaboration et al., 2021; Stone et al., 2020; Fryxell et al., 2000). For many MHD problems, accurately resolving the relevant spatial and temporal scales is computationally expensive, and exploring different initial conditions and physical parameters requires many independent simulations.

Machine-learning-based surrogates instead learn a reusable solution map from initial states and physical parameters to resulting trajectories, amortizing simulation cost over many problem instances and enabling rapid inference. Neural operators (NO) provide a natural framework for this task by learning mappings between function spaces rather than solutions to individual problem instances (Li et al., 2020). Physics-informed neural operators (PINO) further incorporate the governing equations during training, improving physical consistency and reducing reliance on simulation data (Li et al., 2022). Nevertheless, accurately learning strongly turbulent and magnetically coupled dynamics remains challenging.

Current machine-learning surrogates for incompressible MHD primarily employ physics-informed Fourier neural operators (FNO) (Rosofsky & Huerta, 2023). These models perform well in relatively smooth-laminar regimes but lose accuracy as the kinetic Reynolds number, Re, increases and sharper turbulent structures develop. Recent advances augment FNO with conditional diffusion to

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

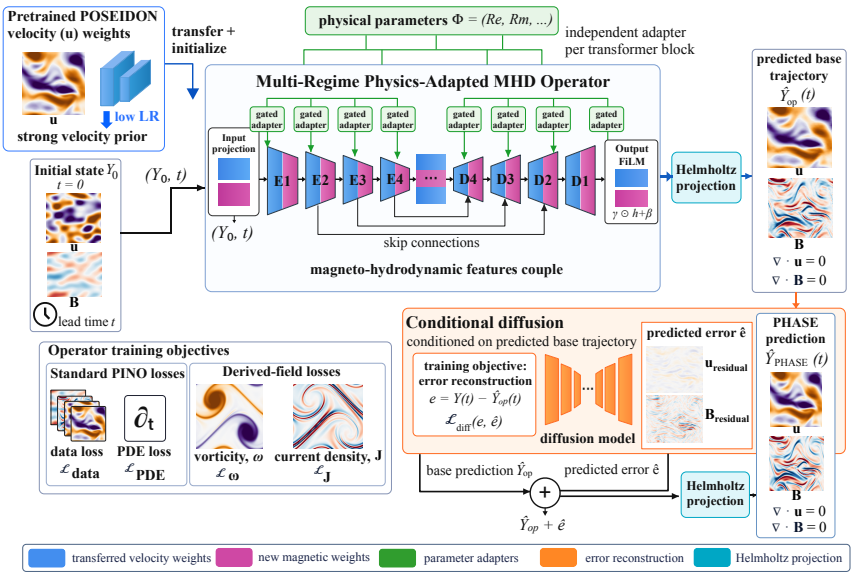

Figure 1: **Overview of PHASE.** PHASE initializes from pretrained fluid representations and augments them with MHD-specific magnetic field parameters. Regime-dependent adapters condition the operator on the physical parameters (Re, Rm), while Helmholtz projection structurally enforces divergence-free constraints, and adding derived-field objectives improves physical fidelity. Finally, a conditional diffusion model learns the operator's residual error and refines unresolved structures.

predict turbulent dynamics (Kacmaz et al., 2025). Despite these improvements, these models are trained separately for each Re, limiting their ability to generalize across physical regimes or predict solutions at previously unseen parameter values. Evaluation is typically limited to velocity and magnetic fields in decaying MHD turbulence, which may not reveal errors in small-scale vortices and magnetic current sheets. These features are more directly characterized by their spatial derivatives, through the vorticity and current-density fields. Furthermore, it remains uncertain whether existing surrogates can model qualitatively distinct, instability-driven MHD dynamics.

Motivated by these limitations, we introduce **PHASE**, a *PHysics-Adaptive Scalable operator with residual Error correction*. PHASE is formulated as a parameter-conditioned MHD operator whose internal representations adapt to changing physical regimes, rather than requiring a separate model for each parameter setting. We initialize the model using pretrained fluid representations from the scalable Operator Transformer (scOT) model in POSEIDON (Herde et al., 2024), and extend the operator to learn coupled velocity–magnetic field dynamics. Physics-based structural constraints and objectives improve the consistency of the predicted fields and their small-scale structures, while a conditional diffusion model refines the residual errors left by the deterministic operator. The complete framework is presented in § 4 & illustrated in Fig. 1.

PHASE is evaluated across both seen and unseen Reynolds numbers, as well as on freely decaying turbulence and instability-driven dynamics. In addition to point-wise field errors, the evaluation assesses whether the learned trajectories preserve physically meaningful derivative fields, multi-scale spectra, distributional statistics, and divergence constraints, as described in § 5.1. The results demonstrate that PHASE substantially outperforms SOTA FNO-based MHD surrogates, generalizes across physical regimes, and captures dynamics beyond the turbulence scenarios considered during model development § 5.2. The main contributions of this work are summarized as follows:

- **Transfer learning**: We show that pretrained fluid-dynamics representations can be effectively transferred to MHD by adapting POSEIDON's velocity prior and extending the scOT backbone to learn magnetic field evolution and velocity–magnetic field coupling. This cross-physics transfer yields **state-of-the-art performance on two-dimensional (2-D) incompressible MHD turbulence** while improving training efficiency (§ 4.1, 5.2.1).

- **Multi-regime modeling**: We introduce parameter-conditioned residual adapters that *adapt* the operator backbone to changing physical regimes, rather than relearning a separate model from scratch for each setting. This enables a **single physics-adapted model** to operate across laminar and strongly turbulent MHD regimes, improving prediction accuracy and **generalizing** correctly to unseen parameter values (§ 4.1, 5.2.1, & 5.2.2).
- **Physics-centered learning**: We combine explicit loss supervision of derived fields such as vorticity & current-density with hard divergence-free constraints through Helmholtz projection. Together, these improve the prediction of **small-scale structure, spectral fidelity, & field distributions**, and enable accurate prediction beyond decaying turbulence to **instability-driven MHD dynamics** (§ 4.2, 5.2.3).

2 Related Work

PDE foundation models & transfer learning. Recent work has explored foundation models for PDEs through multi-physics pretraining and transfer to downstream solution operators (McCabe et al., 2023; Hao et al., 2024; Alkin et al., 2024; Herde et al., 2024; Morel et al., 2025). In particular, POSEIDON pretrains the multiscale scOT architecture on compressible and incompressible Navier–Stokes trajectories, achieving improved accuracy and sample efficiency across diverse PDE solutions (Herde et al., 2024). However, the applicability of such foundation models to MHD remains unexplored.

ML Surrogates for MHD. Rosofsky & Huerta (2023) introduced a PINO surrogate for 2-D incompressible MHD. Their tensorized Fourier neural operator (tFNO) combines training on MHD simulation data with losses derived from the governing equations and physical constraints, achieving correct predictions for laminar flows with $Re \leq 250$. However, its performance degrades in turbulent regimes, where it struggles to learn small-scale velocity and magnetic field structures. Building on this work, Kacmaz et al. (2025) combined the physics-informed tFNO with a conditional diffusion model, henceforth referred to as DINO. Although diffusion improves the recovery of turbulent structures, prediction errors still increase with Re, implying that strongly turbulent regimes remain challenging. Moreover, both frameworks train separate models for each Reynolds number and focus exclusively on decaying turbulence, leaving generalization across physical parameter regimes and applicability to other MHD problems largely unexplored.

3 Background: Incompressible Magnetohydrodynamics

We consider the incompressible MHD equations, which describe the coupled evolution of a conducting fluid velocity field, $\mathbf{u}$, and magnetic field, $\mathbf{B}$:

$$\partial_t \mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} = -\nabla\left(p + \frac{|\mathbf{B}|^2}{2}\right) + (\mathbf{B}\cdot\nabla)\mathbf{B} + \nu\nabla^2\mathbf{u}, \tag{1}$$

$$\partial_t \mathbf{B} + (\mathbf{u}\cdot\nabla)\mathbf{B} = (\mathbf{B}\cdot\nabla)\mathbf{u} + \eta\nabla^2\mathbf{B}, \tag{2}$$

$$\nabla\cdot\mathbf{u} = 0, \qquad \nabla\cdot\mathbf{B} = 0. \tag{3}$$

Here p is the fluid pressure, ν is the kinematic viscosity, and η is the magnetic diffusivity. The first equation governs fluid motion, including feedback from the magnetic field, while the second governs how the magnetic field is advected and diffused by the flow. The divergence constraints enforce incompressible fluid motion and ensure no magnetic monopoles exist. The kinetic and magnetic Reynolds numbers,

$$Re = \frac{UL}{\nu}, \qquad Rm = \frac{UL}{\eta}, \tag{4}$$

control the balance between advection and dissipation in the velocity and magnetic fields, respectively. Here, U and L denote characteristic velocity and length scales, respectively. Larger Re and Rm correspond to weaker viscous and resistive dissipation, allowing finer spatial structures and stronger turbulent dynamics to develop. Smaller Re and Rm correspond to stronger dissipation regimes, which suppresses small-scale structures and produces laminar velocity fields and smooth magnetic fields. Their ratio defines the magnetic Prandtl number, $Pm = Rm/Re$. For our scope, we set $Re = Rm$, and hence $Pm = 1$ in all our experiments. Two derived fields are particularly important for characterizing small-scale MHD structure:

$$\omega = \nabla \times \mathbf{u}, \qquad \mathbf{J} = \nabla \times \mathbf{B}. \tag{5}$$

The vorticity, ω, measures local rotation of the fluid, while the current-density, $\mathbf{J}$, measures spatial variation of the magnetic field. As both depend on spatial derivatives, they amplify small-scale prediction errors and therefore provide a stricter test of physical fidelity than the primary velocity and magnetic fields alone.

4 PHASE: PHYSICS-ADAPTED MHD OPERATOR

PHASE models the evolution of the 2-D incompressible MHD state,

$$Y(t) = (\mathbf{u}(t), \mathbf{B}(t))$$

as a parametrized solution operator $\mathcal{G}_\theta$ with initial state Y_0, lead time t, & external physical parameters $\mathbf{\Phi}$ as the input:

$$\widehat{Y}(t) = \mathcal{G}_\theta\left(\mathbf{Y}_0, t, \mathbf{\Phi}\right), \tag{6}$$

In the 2-D incompressible MHD context discussed in § 3, $\mathbf{u} = (u_x, u_y)$, $\mathbf{B} = (B_x, B_y)$, and $\mathbf{\Phi} = (Re, Rm)$. While we focus on 2-D problems in this work, we note that the PHASE framework can be readily extended to 3-D MHD. The framework consists of a multi-regime physics-adapted operator transformer that predicts the primary MHD evolution, followed by a conditional diffusion model that corrects the remaining prediction error. PHASE additionally incorporates parameter conditioning, hard physical constraints and losses on derived fields, so that a single model can accurately represent multiple MHD regimes and problems. An overview is shown in Fig. 1. Next, we discuss each component of our framework.

4.1 TRANSFER LEARNING AND MULTI-REGIME CONDITIONING

We leverage transfer learning (TL) from POSEIDON, whose scalable Operator Transformer (scOT) is pretrained on fluid-dynamics data (Herde et al., 2024). PHASE initializes its velocity-related weights from the corresponding pretrained POSEIDON weights, providing a prior for fluid evolution, while newly initialized magnetic field weights extend the operator to jointly predict $(\mathbf{u}, \mathbf{B})$. During MHD fine-tuning, we use a smaller learning rate for the transferred velocity weights and a larger learning rate for the magnetic field weights. This allows the pretrained fluid representations to adapt gradually to the MHD domain while prioritizing the learning of magnetic field evolution and velocity–magnetic field coupling. Further implementation and training details are provided in App. A.6.1.

TL alone does not address variation across physical regimes. We therefore condition the operator on $\mathbf{\Phi} = (Re, Rm)$, allowing a single model to adapt its internal representations as the governing physical parameters change. We introduce lightweight parameter-dependent residual adapters throughout the scOT hierarchy. For a hidden representation h_ℓ, the conditioned update takes the form

$$h_\ell \leftarrow h_\ell + g_\ell(\mathbf{\Phi}) \odot E_\ell(h_\ell), \tag{7}$$

where E_ℓ is a bottleneck residual adapter and g_ℓ is a learned gate determined by the physical parameters. A final feature-wise modulation adjusts the physical output channels:

$$\widehat{Y} = \widehat{Y} + \alpha_{\text{out}}[\gamma(\mathbf{\Phi}) \odot \widehat{Y} + \beta(\mathbf{\Phi})], \tag{8}$$

where $\gamma(\Phi)$ and $\beta(\Phi)$ are learned shallow MLPs. Together, the deep residual adapters and output modulation allow a single operator to adapt its internal and output representations across physical regimes. Eq. 7 is inspired by the parameter-efficient adapters in ML literature, which introduce lightweight residual updates around a shared pretrained backbone (Houlsby et al., 2019). We further condition these residual updates on the physical parameters through learned gates. Eq. 8 follows the FiLM principle of feature-wise affine conditioning (Perez et al., 2018), which allows external parameters to rescale and shift representation. See App. A.6.2 for conditioning details. This design preserves the pretrained operator while learning parameter-dependent corrections across laminar and turbulent regimes. Full adapter and optimization details are given in App. A.6. Moreover, we would like to highlight that naive external parameter conditioning by introducing additional input channels in the operator is less effective. The isolated effect of TL, our multi-regime conditioning & other ablations on our results are depicted in Fig. 2.

4.2 PHYSICS-INFORMED IMPROVEMENTS

A 2-D magnetic field can be represented using an out-of-plane magnetic vector potential A as, $\mathbf{B} = \nabla \times (A\hat{\mathbf{z}})$. This representation guarantees $\nabla \cdot \mathbf{B} = 0$ by construction and is used by MHD numerical solvers and prior ML surrogates (Rosofsky & Huerta, 2023). PHASE directly predicts the physical vector fields ($\mathbf{u}$, $\mathbf{B}$). This avoids reconstructing the magnetic field from a learned potential, for which derived quantities such as the current-density require higher-order differentiation. This representation particularly improves predictions of the magnetic field and the derived current-density field. In incompressible MHD, both velocity and magnetic fields must satisfy $\nabla \cdot \mathbf{u} = 0$ and $\nabla \cdot \mathbf{B} = 0$. Following recent work on divergence-free projection (Li et al., 2026), PHASE applies a Helmholtz projection to both predicted vector fields. For either $\mathbf{q} \in \mathbf{u}, \mathbf{B}$, each Fourier mode is projected onto its divergence-free component:

$$\widehat{\mathbf{q}}_{\perp}(\mathbf{k}) = \left(\mathbf{I} - \frac{\mathbf{k}\mathbf{k}^{\top}}{|\mathbf{k}|_2^2}\right) \widehat{\mathbf{q}}(\mathbf{k}). \tag{9}$$

The projection enforces the divergence constraints up to high numerical precision and is applied to both the operator prediction and the diffusion-corrected output. We further train the deterministic operator with a physics-informed objective

$$\mathcal{L}_{\text{PHASE}} = \lambda_{\text{data}}\mathcal{L}_{\text{data}} + \mathcal{L}_{\text{ic}} + \lambda_{\text{PDE}}\mathcal{L}_{\text{PDE}} + \lambda_{\omega}\mathcal{L}_{\omega} + \lambda_{J}\mathcal{L}_{J}, \tag{10}$$

where $\mathcal{L}_{\text{PDE}}$ penalizes residuals of the governing MHD equations, while $\mathcal{L}_{\text{data}}$ penalizes errors in the primary fields. $\mathcal{L}_{\text{ic}}$ penalizes the component-wise primary field prediction error at the initial time, $t = 0$. Both are standard components of physics-informed surrogate training. The newly introduced losses, $\mathcal{L}_{\omega}$ and $\mathcal{L}_{J}$, supervise vorticity and current-density learning, respectively. These terms explicitly target derivative-sensitive, small-scale vorticity and current-density structures that may not be adequately captured by primary field reconstruction errors alone. We choose λ_{data} and λ_{PDE} based on working configurations in prior work, while tuning the new hyperparameters λ_{ω} and λ_{J} (see App. A.4 and A.5 for further details). The contributions of direct magnetic field prediction, Helmholtz projection, and the derived-field-based physics loss are shown in Fig. 2.

4.3 RESIDUAL DIFFUSION CORRECTION

The neural operator captures the dominant deterministic evolution, but its remaining errors become increasingly structured in strongly turbulent regimes, where fine-scale vortices and magnetic field structures are difficult to reproduce. Inspired by prior MHD surrogates, we use a diffusion model to recover small-scale features unresolved by the neural operator prediction (Kacmaz et al., 2025; Lippe et al., 2023). However, unlike Kacmaz et al. (2025), whose conditional diffusion model reconstructs the full MHD trajectory, we train the diffusion model to predict only the residual between the neural-operator prediction and the DNS trajectory, focusing on error calibration. We model the residual between the operator prediction and the ground-truth trajectory as e, and train a conditional diffusion model to predict $\hat{e}$ while conditioning on the operator trajectory, as shown in Fig. 1 (Ho et al., 2020; Song et al., 2020; Kacmaz et al., 2025). The residual and final PHASE prediction are

$$e = Y - \widehat{Y}_{\text{op}}, \qquad \widehat{Y}_{\text{PHASE}} = \Pi_{\text{div}}\left(\widehat{Y}_{\text{op}} + \hat{e}\right), \tag{11}$$

where Π_{div} denotes the Helmholtz projection (Eq. 9) applied separately to the velocity and magnetic fields. By learning only the correction, the diffusion stage focuses its capacity on structures unresolved by the deterministic operator rather than relearning the full MHD evolution. Details of the diffusion objective is provided in App. A.4.

5 EXPERIMENTS & RESULTS

5.1 EXPERIMENTAL SETUP

Training Data and physical regimes. We generate ground-truth trajectories by solving the 2-D incompressible MHD equations with periodic boundary conditions using the Dedalus spectral solver (Burns et al., 2020). The simulations evolve the velocity components (u_x, u_y) and magnetic vector

potential (A), from which the divergence-free magnetic field is recovered as $\mathbf{B} = \nabla \times A$. PHASE is trained directly on the four primary field channels (u_x, u_y, B_x, B_y). All simulations use $Re = Rm$, corresponding to magnetic Prandtl number $Pm = 1$. Numerical solver settings, initial conditions, and data-generation details are provided in App. A.1. We generate data for freely decaying MHD turbulence and Kelvin-Helmholtz instability over

$$Re_{\text{train}} \in \mathcal{S}_{Re} = \{80, 200, 400, 650, 1000, 1500, 2050, 2750, 3600, 4500\},$$

spanning viscous to strong turbulence regimes. These values are selected to cover the corresponding range of dissipation scales; the selection procedure is described in App. A.2. Each trajectory begins from independently sampled divergence-free velocity and magnetic fields and evolves without external forcing. Using these settings, we test whether an ML surrogate can reproduce multiscale MHD dynamics across a broad range of physical regimes.

Baselines and training setup. We compare PHASE against the physics-informed FNO surrogate of Rosofsky & Huerta (2023) and the DINO model of Kacmaz et al. (2025), which augments a physics-informed tFNO with a conditional diffusion model. The latter DINO model represents the strongest prior baseline in our 2-D incompressible MHD setting. We perform an ablation study at $Re = 1000$ to isolate the contributions of transfer learning, multi-regime conditioning, four-channel prediction, physics constraints, new derived-field-based physics loss and diffusion refinement. This is summarized in Fig. 2, which reports the aggregate relative L_2 errors for the primary fields (u_x, u_y, B_x, B_y) and derived fields (ω, J) and the percentage improvement over the previously evaluated model (see App. A.3). Fig. 2 shows that the gains arise from the complete PHASE framework: the scOT backbone alone underperforms DINO, while transfer learning and gated multi-regime adapters provide successive improvements over training from scratch and naive multi-regime conditioning. Four-channel prediction with Helmholtz projection and physics losses substantially improves the derived fields, and diffusion refinement produces the final PHASE model with the lowest primary- and derived-field errors. We evaluate both a single-regime (SR) PHASE trained on data for specific Re and a multi-regime (MR) model trained jointly across the full Reynolds-number range $\mathcal{S}_{Re}$. The MR model is warm-started from the SR $Re = 1000$ checkpoint before parameter-conditioned fine-tuning. Full architecture, optimization, and learning-rate details are provided in App. A.6.

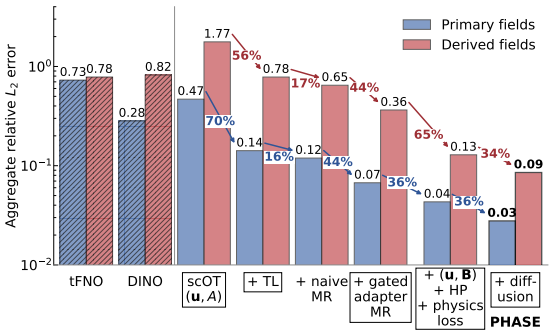

Figure 2: Aggregate relative L_2 errors for the primary and derived fields across model ablations evaluated at $Re = 1000$, with percentage improvements between successive models. Hatched bars denote the previous tFNO and DINO baselines. TL, MR, and HP denote transfer learning, multi-regime training, and Helmholtz projection, respectively.

Evaluation metrics. We evaluate both pointwise accuracy and physical fidelity. Relative L_2 errors are reported for the primary velocity and magnetic fields and for the derived fields, vorticity and current-density, which provide a more sensitive measure of small-scale velocity and magnetic field structures. We also measure divergence error for both $\mathbf{u}$ and $\mathbf{B}$ (App. A.3), multiscale spectral agreement, and distributional statistics including standard deviation and kurtosis (App. A.7). These complementary diagnostics test whether the learned MHD trajectories reproduce not only field values but also their spectral scaling, fluctuation amplitudes, and intermittent structures.

5.2 Results

5.2.1 State-of-the-Art Turbulence Prediction

First, we demonstrate that PHASE improves upon the previous state-of-the-art in predicting MHD dynamics in the turbulent regime. Fig. 3 (a) shows the aggregate relative L_2 errors for the primary and derived fields in the turbulent regime with $Re = 1000$, comparing the DINO model with SR and

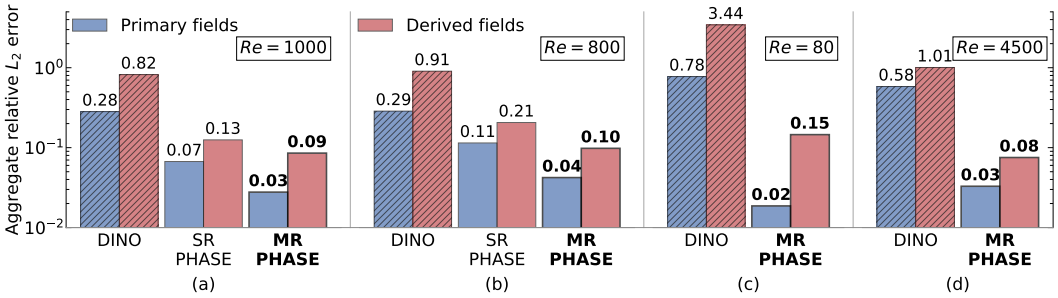

Figure 3: Aggregate relative L_2 errors for the primary and derived fields. Panels (a)-(d) show results at $Re = 1000$, 800, 80, and 4500, respectively; the single-regime model is included for $Re = 1000$ and its unseen $Re = 800$ generalization case.

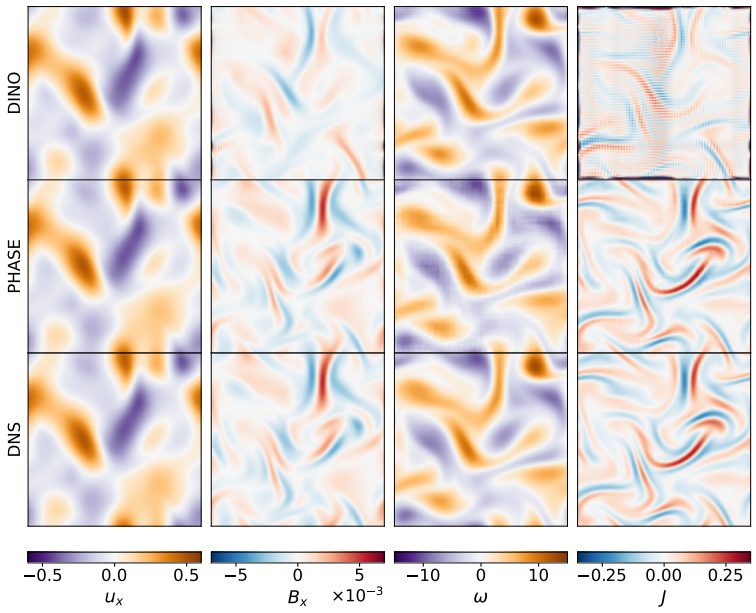

Figure 4: The x-components of the velocity and magnetic fields, vorticity, and current-density for $Re = 1000$ at the final evolution time, $t = 1.0$. The first row shows the DINO prediction, the second row shows the MR PHASE prediction, and the third row shows the DNS solution from the MHD simulation. We find that PHASE accurately captures structures in the primary and derived fields .

MR PHASE. MR PHASE performs best across all metrics, achieving roughly an order-of-magnitude reduction in errors across all fields compared to the previous state-of-the-art DINO model.

Fig. 4 compares the DINO and MR PHASE predictions with the DNS ground truth at the final simulation time, $t = 1$. PHASE shows excellent agreement with the DNS for both primary and derived fields, whereas DINO exhibits pronounced inaccuracies, particularly in the current-density field. DINO predicts the magnetic vector potential, from which the magnetic field and current-density are obtained through first- and second-order spatial derivatives, respectively, amplifying errors. PHASE instead predicts both magnetic field components directly and includes dedicated physics losses for the derived fields (§ 4.2), capturing the magnetic field and current-density structures accurately. Although DINO ensures a divergence-free magnetic field through the vector-potential representation, it enforces velocity incompressibility through a soft loss, which does not guarantee divergence-free predictions. For the solutions shown, DINO yields $\nabla \cdot u \approx 0.15$. In contrast, Helmholtz projection (§ 4.2) yields $\nabla \cdot u \approx 9 \times 10^{-6}$ and $\nabla \cdot B \approx 6 \times 10^{-8}$ for PHASE, satisfying both constraints to high numerical accuracy.

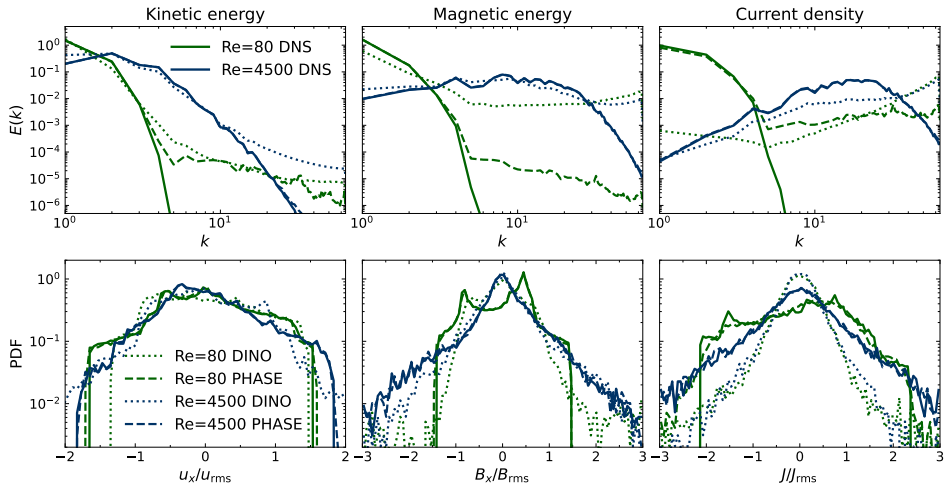

Figure 5: Spectra and PDFs for the DNS, DINO, and MR PHASE models at $Re = 80$ and 4500. The first row shows the normalized kinetic energy, magnetic energy, and current-density spectra, each scaled by its total spectral power. The second row shows the rms-normalized PDFs of u_x, B_x, and J. Compared to the DINO model, PHASE shows good predictions for all spectra and PDFs in the laminar and strongly turbulent regimes.

EXTREMELY VISCOUS AND TURBULENT REGIMES

Next, we evaluate PHASE's performance in the laminar and strongly turbulent regimes. In Fig. 3, we show the DINO and MR PHASE performance in the viscous regime ($Re = Rm = 80$) and the strongly turbulent regime ($Re = Rm = 4500$). Our model achieves excellent performance in both extreme parameter regimes. To assess whether PHASE captures the distribution of kinetic and magnetic energy across spatial scales, we examine the kinetic energy, magnetic energy, and current-density spectra for $Re = 80$ and 4500 in Fig. 5, alongside the DINO predictions. We also examine the probability density functions (PDFs) of the primary and derived fields to evaluate if PHASE captures their statistical distributions and magnetic intermittency. The PDFs of u_x, B_x, and J show excellent agreement with the DNS at both $Re = 80$ and $Re = 4500$. PHASE also reproduces the kinetic energy, magnetic energy, and current-density spectra well, particularly in the strongly turbulent $Re = 4500$ regime, where the DINO model fails. At high wavenumbers (small length-scales), however, PHASE over-predicts the spectral power across all three spectra and this limitation is most pronounced at $Re = 80$. Further discussion on the spectra and PDFs are provided in App. A.7.2.

5.2.2 GENERALIZATION ACROSS REYNOLDS NUMBER

The generalization of the $Re = 1000$ trained DINO and SR PHASE models, together with MR PHASE, to an unseen Reynolds number, $Re = 800$, is evaluated in Fig. 3 (b). While DINO fails to capture the dynamics at an unseen Re, both PHASE models perform substantially better. MR PHASE achieves the lowest errors across all primary and derived fields, demonstrating the strongest generalization to unseen Reynolds numbers (see App. A.7.1 and A.7.2). This shows that unlike traditional MHD solvers, which require a separate simulation for each parameter value, PHASE can predict correct dynamics at unseen physical parameters correctly without retraining or running new numerical simulations. This generalization capability is a key advantage of our ML surrogate.

5.2.3 INSTABILITY DYNAMICS

To test whether PHASE can model dynamics beyond freely decaying turbulence, we train and evaluate PHASE on the MHD Kelvin–Helmholtz (KH) instability. KH provides a more stringent test because the weak transverse perturbation velocity must be resolved in the presence of a much stronger shear flow, unlike in isotropic turbulence, where the two velocity components have comparable amplitudes. The model must capture the amplification of this perturbation through the linear KH growth

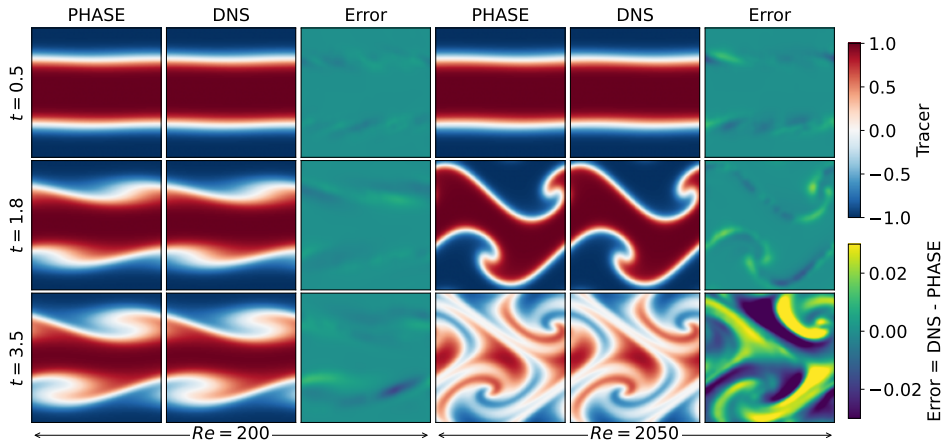

Figure 6: Tracer evolution for the KH instability test in the viscous ($Re = 200$) and turbulent ($Re = 2050$) regime. Columns compare PHASE predictions, DNS, and errors (DNS − PHASE), while rows show the initial phase ($t = 0.5$), linear growth ($t = 1.8$), and the nonlinear mixing stage ($t = 3.5$) of the KH instability. PHASE can accurately capture spatial and temporal KH instability dynamics across viscous and turbulent regimes.

stage, nonlinear vortex roll-up, saturation, and the subsequent decay of the instability. We evaluate PHASE in both viscous and turbulent KH regimes. To examine these different stages, we track a passive tracer that is advected and diffused by the flow; further details are provided in App. A.1. The tracer provides a direct visualization of shear-layer deformation ($t = 0.5$), vortex formation ($t = 1.8$), and mixing ($t = 3.5$). Fig. 6 compares the PHASE predictions, DNS solutions, and corresponding errors for the viscous ($Re = 200$) and turbulent ($Re = 2050$) regimes. PHASE closely reproduces the tracer evolution in both regimes, capturing the entire time evolution of the KH instability (see App. A.8 for further details).

6 Conclusion and Future Work

We introduced PHASE, a physics-adapted multi-regime neural operator for incompressible MHD. PHASE transfers pretrained velocity representations from the fluid-dynamics foundation model POSEIDON and extends the scOT backbone to learn magnetic field evolution. Together with residual diffusion refinement and physics-structured improvements, this model achieves state-of-the-art accuracy on MHD turbulence predictions, reducing relative L_2 errors by more than an order of magnitude compared with previous MHD neural-operator surrogates from Kacmaz et al. (2025).

Parameter-conditioned residual adapters enable a single model to adapt across laminar and strongly turbulent MHD regimes and generalize to unseen parameter values without retraining. Multi-regime modeling not only improves cross-regime generalization but also outperforms the corresponding single-regime model at a fixed Reynolds number. Multi-regime training is therefore beneficial even when the model is applied only to a single physical regime. Explicit losses on vorticity and current-density further improve the recovery of small-scale structures, spectral scaling, and field distributions. PHASE also captures the evolution of the Kelvin–Helmholtz instability, demonstrating applicability beyond decaying turbulence to instability-driven dynamics. These results establish PHASE as a step toward a unified MHD model across physical regimes.

PHASE is currently developed and evaluated only for 2-D incompressible MHD. This study does not address compressibility, supersonic flows, shock formation, or three-dimensional dynamics. We also restrict our experiments to $Re = Rm$. Therefore, the model's performance when viscous and resistive scales vary independently remains unexplored. As found in Fig. 5, PHASE over-predicts spectral power at high wavenumbers. Future work will investigate targeted spectral losses and reinforcement-learning-based refinement to improve these predictions. We will also extend PHASE to compressible MHD, including supersonic Mach number regimes, varying magnetic Prandtl numbers, and broader classes of plasma instabilities.

AI USE STATEMENT

We have not used generative AI tools for generating synthetic datasets; Designing research methodology or experiments; Implementing methods or interpreting results; Cleaning and reformatting datasets. The rest of the required disclosures like developing theoretical models or conceptual frameworks formulating mathematical claims or assisting with proofs; Proposing or refining hypotheses, are not applicable to this work. Additionally, we used generative AI tools for drafting parts of the research paper; Editing text to improve readability; Editing software code and scientific figures. We have reviewed all AI-assisted work, formatting edits to the main paper were proof-read multiple times along with verifying the core logic of the code post edits. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide details pertaining to data generation, loss & training details in App. A.1, App. A.4 & App. A.6 respectively. Readers interested in the code implementation can refer to the anonymous link to the PHASE repository [link]. Furthermore, one can find the anonymized trained model weights for MR PHASE [here]. We consider the resources provided sufficient for the review of our work. Additionally, the complete HuggingFace space with all baseline model weights & configurations will be disclosed after review.

REFERENCES

Benedikt Alkin, Andreas Fürst, Simon Schmid, Lukas Gruber, Markus Holzleitner, and Johannes Brandstetter. Universal Physics Transformers: A Framework For Efficiently Scaling Neural Operators. *arXiv e-prints*, art. arXiv:2402.12365, February 2024. doi: 10.48550/arXiv.2402.12365.

Keaton J. Burns, Geoffrey M. Vasil, Jeffrey S. Oishi, Daniel Lecoanet, and Benjamin P. Brown. Dedalus: A flexible framework for numerical simulations with spectral methods. *Physical Review Research*, 2(2):023068, April 2020. doi: 10.1103/PhysRevResearch.2.023068.

B. Fryxell, K. Olson, P. Ricker, F. X. Timmes, M. Zingale, D. Q. Lamb, P. MacNeice, R. Rosner, J. W. Truran, and H. Tufo. FLASH: An Adaptive Mesh Hydrodynamics Code for Modeling Astrophysical Thermonuclear Flashes. *The Astrophysical Journal Supplement Series*, 131(1): 273–334, Nov 2000. doi: 10.1086/317361.

Zhongkai Hao, Chang Su, Songming Liu, Julius Berner, Chengyang Ying, Hang Su, Anima Anandkumar, Jian Song, and Jun Zhu. DPOT: Auto-Regressive Denoising Operator Transformer for Large-Scale PDE Pre-Training. *arXiv e-prints*, art. arXiv:2403.03542, March 2024. doi: 10.48550/arXiv.2403.03542.

Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient Foundation Models for PDEs. *arXiv e-prints*, art. arXiv:2405.19101, May 2024. doi: 10.48550/arXiv.2405.19101.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. *arXiv e-prints*, art. arXiv:2006.11239, June 2020. doi: 10.48550/arXiv.2006.11239.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for NLP. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2790–2799. PMLR, 09–15 Jun 2019. URL `https://proceedings.mlr.press/v97/houlsby19a.html`.

Semih Kacmaz, E. A. Huerta, and Roland Haas. Resolving turbulent magnetohydrodynamics: a hybrid operator-diffusion framework. *Machine Learning: Science and Technology*, 6(3):035057, September 2025. doi: 10.1088/2632-2153/ae054c.

Russell M. Kulsrud. *Plasma physics for astrophysics*. 2005.

Xigui Li, Hongwei Zhang, Ruoxi Jiang, Deshu Chen, Chensen Lin, Limei Han, Yuan Qi, Xin Guo, and Yuan Cheng. Project and Generate: Divergence-Free Neural Operators for Incompressible Flows. *arXiv e-prints*, art. arXiv:2603.24500, March 2026. doi: 10.48550/arXiv.2603.24500.

Zhijie Li, Wenhui Peng, Zelong Yuan, and Jianchun Wang. Fourier neural operator approach to large eddy simulation of three-dimensional turbulence. *Theoretical and Applied Mechanics Letters*, 12 (6):100389, November 2022. doi: 10.1016/j.taml.2022.100389.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier Neural Operator for Parametric Partial Differential Equations. *arXiv e-prints*, art. arXiv:2010.08895, October 2020. doi: 10.48550/arXiv.2010.08895.

Phillip Lippe, Bas Veeling, Paris Perdikaris, Richard Turner, and Johannes Brandstetter. PDE-refiner: Achieving accurate long rollouts with neural PDE solvers. *Advances in Neural Information Processing Systems*, 36:67398–67433, 2023.

Michael McCabe, Bruno Régaldo-Saint Blancard, Liam Holden Parker, Ruben Ohana, Miles Cranmer, Alberto Bietti, Michael Eickenberg, Siavash Golkar, Geraud Krawezik, Francois Lanusse, Mariel Pettee, Tiberiu Tesileanu, Kyunghyun Cho, and Shirley Ho. Multiple Physics Pretraining for Physical Surrogate Models. *arXiv e-prints*, art. arXiv:2310.02994, October 2023. doi: 10.48550/arXiv.2310.02994.

Rudy Morel, Jiequn Han, and Edouard Oyallon. DISCO: learning to DISCover an evolution Operator for multi-physics-agnostic prediction. In *International Conference on Machine Learning (ICML)*, 2025. URL `https://proceedings.mlr.press/v267/morel25a.html`.

Pencil Code Collaboration, Axel Brandenburg, Anders Johansen, Philippe Bourdin, Wolfgang Dobler, Wladimir Lyra, Matthias Rheinhardt, Sven Bingert, Nils Haugen, Antony Mee, Frederick Gent, Natalia Babkovskaia, Chao-Chin Yang, Tobias Heinemann, Boris Dintrans, Dhrubaditya Mitra, Simon Candelaresi, Jörn Warnecke, Petri Käpylä, Andreas Schreiber, Piyali Chatterjee, Maarit Käpylä, Xiang-Yu Li, Jonas Krüger, Jørgen Aarnes, Graeme Sarson, Jeffrey Oishi, Jennifer Schober, Raphaël Plasson, Christer Sandin, Ewa Karchniwy, Luiz Rodrigues, Alexander Hubbard, Gustavo Guerrero, Andrew Snodin, Illa Losada, Johannes Pekkilä, and Chengeng Qian. The Pencil Code, a modular MPI code for partial differential equations and particles: multipurpose and multiuser-maintained. *The Journal of Open Source Software*, 6(58):2807, February 2021. doi: 10.21105/joss.02807.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron C. Courville. Film: Visual reasoning with a general conditioning layer. In *AAAI*, 2018.

Shawn G. Rosofsky and E. A. Huerta. Magnetohydrodynamics with physics informed neural operators. *Machine Learning: Science and Technology*, 4(3):035002, September 2023. doi: 10.1088/2632-2153/ace30a.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv e-prints*, art. arXiv:2011.13456, November 2020. doi: 10.48550/arXiv.2011.13456.

James M. Stone, Kengo Tomida, Christopher J. White, and Kyle G. Felker. The Athena++ Adaptive Mesh Refinement Framework: Design and Magnetohydrodynamic Solvers. *The Astrophysical Journal Supplement Series*, 249(1):4, July 2020. doi: 10.3847/1538-4365/ab929b.

A Appendix

A.1 MHD Data generation

We solve the 2-D incompressible MHD equations on the periodic unit square, $(x, y) \in [0, 1]^2$, using Dedalus (Burns et al., 2020). The spatial discretization uses $N_x = N_y = 128$. At each time step, the solver evolves the velocity components u_x and u_y and the out-of-plane magnetic vector potential A. The in-plane magnetic field is reconstructed as $\mathbf{B} = \nabla \times A$, which satisfies $\nabla \cdot \mathbf{B} = 0$ by construction. Although the simulations evolve (u_x, u_y, A), the PHASE model uses the four physical field components (u_x, u_y, B_x, B_y). For visualization, we additionally evolve a passive tracer, s, according to

$$\partial_t s + \mathbf{u} \cdot \nabla s = \nu \nabla^2 s, \tag{12}$$

which we particularly use to study instabilities. The tracer is not included among the input or target channels of PHASE.

For decaying turbulence, the initial conditions are sampled from smooth Gaussian random fields, following the prescription in Rosofsky & Huerta (2023). The fields are generated in Fourier space and satisfy periodic boundary conditions. We sample a stream function ψ and magnetic potential A and construct

$$\mathbf{u} = \nabla \times (\psi \hat{\mathbf{z}}) = (\partial_y \psi, -\partial_x \psi)\,, \qquad \mathbf{B} = \nabla \times (A \hat{\mathbf{z}}) = (\partial_y A, -\partial_x A)\,. \tag{13}$$

This construction ensures that both initial fields satisfy their divergence-free constraints. For each physical-parameter configuration, we generate $1{,}000$ trajectories with independently sampled initial conditions. The code uses the fourth-order Runge–Kutta time stepping with $\Delta t = 10^{-3}$. For all our simulations we use $Re = Rm$, i.e. $Pm = 1$

The decaying-turbulence trajectories are evolved without external forcing, so the flow evolves freely under nonlinear advection, Lorentz coupling, viscosity, and magnetic diffusion. Each trajectory is integrated until $t = 1$, and solution snapshots are written every ten solver steps, giving an output timestep of $\Delta t_{\text{out}} = 10^{-2}$.

For the Kelvin–Helmholtz (KH) instability, we use the same periodic BCs, but replace the random velocity initialization by a two-layer shear profile. The streamwise velocity is initialized as

$$u_x(y) = U_0 \left[\tanh\left(\frac{y - y_1}{\delta} \right) - \tanh\left(\frac{y - y_2}{\delta} \right) - 1 \right], \tag{14}$$

with shear layers centered at $y_1 = 1/4$ and $y_2 = 3/4$. The transverse velocity is seeded with a small perturbation to trigger the instability. In our runs we use $U_0 = 1$, perturbation amplitude $\epsilon = 10^{-2}$, and shear-layer width $\delta = 0.05$. The magnetic field is initialized as a divergence-free random field through the same magnetic potential construction used for the turbulence data. The KH trajectories are evolved until $t = 5$, to capture the growth and saturation of the instability.

A.2 Multi-regime Reynolds number selection

As the plasma is described by 2-D incompressible MHD, and the magnetic field is extremely weak compared to the velocity field, the small-scale cutoff is estimated using the 2-D enstrophy-dissipation scale,

$$\chi = \nu \left\langle |\nabla \omega|^2 \right\rangle_{x,y,t}, \tag{15}$$

where the average is taken over space and time. The corresponding viscous scale is

$$\ell_\chi = \left(\frac{\nu^3}{\chi} \right)^{1/6}. \tag{16}$$

For our unit periodic domain, $k_{\text{phys}} = 2\pi k_{\text{shell}}$; therefore, the cutoff is

$$k_\chi = \frac{1}{2\pi} \left(\frac{\chi}{\nu^3} \right)^{1/6}. \tag{17}$$

We select $Re \in \{80, 200, 400, 650, 1000, 1500, 2050, 2750, 3600, 4500\}$ to approximately uniformly sample the estimated enstrophy-dissipation cutoff wavenumber k_χ, spanning viscous to strongly turbulent regimes. We restrict our study to $Pm = Rm/Re = 1$, and therefore set $Rm = Re$.

Table 1: Field-wise and aggregate relative L_2 errors for the model ablations. The tFNO and DINO results are the baselines from previous studies. We define $u_{\text{avg}} = \frac{1}{2}[L_2(u_x) + L_2(u_y)]$ and $B_{\text{avg}} = \frac{1}{2}[L_2(B_x) + L_2(B_y)]$. The primary field aggregate is $\mathbf{P} = \frac{1}{2}[u_{\text{avg}} + B_{\text{avg}}]$, and the derived aggregate is $\mathbf{D} = \frac{1}{2}[L_2(\omega) + L_2(J)]$. The final two columns report the root-mean-square divergence errors.

Ser. No.	Model	Avg. rel. L_2				Aggregate rel. L_2		RMS div. errors	
		u_{avg}	B_{avg}	$\boldsymbol{\omega}$	$\mathbf{J}$	$\mathbf{P}$	$\mathbf{D}$	$\nabla\cdot\mathbf{u}$	$\nabla\cdot\mathbf{B}$
1	tFNO	0.523	0.935	0.571	0.995	0.729	0.783	2.3×10^{-1}	3.2×10^{-8}
2	DINO	0.035	0.530	0.081	1.561	0.283	0.821	1.5×10^{-1}	6.3×10^{-8}
3	scOT $(\mathbf{u}, A)$	0.261	0.676	0.530	3.009	0.468	1.770	1.6×10^{0}	7.8×10^{-8}
4	3 + TL	0.033	0.250	0.135	1.430	0.142	0.782	4.0×10^{-1}	7.6×10^{-8}
5	4 + naive MR	0.028	0.210	0.126	1.168	0.119	0.647	3.8×10^{-1}	7.6×10^{-8}
6	4 + gated adapter MR 6 + $(\mathbf{u}, \mathbf{B})$	0.020	0.115	0.098	0.631	0.067	0.364	3.2×10^{-1}	7.5×10^{-8}
7	+ Helmholtz projection + physics loss	0.025	0.062	0.114	0.143	0.043	0.129	9.1×10^{-6}	6.1×10^{-8}
8	**PHASE**: 7 + diffusion	**0.012**	**0.044**	**0.063**	**0.108**	**0.028**	**0.085**	9.1×10^{-6}	6.0×10^{-8}

A.3 Ablation table

We compare PHASE with the tFNO and DINO baselines from previous MHD surrogate studies (Rosofsky & Huerta, 2023; Kacmaz et al., 2025). Unlike the training procedure in Rosofsky & Huerta (2023), we train tFNO from scratch, without a warm start, to enable a direct comparison with the scOT baseline trained without transfer learning or warm start. Detailed ablation results are reported in Tab. 1. The ablation study isolates the effects of velocity-weight transfer from POSEIDON; naive and gated-adapter multi-regime conditioning; the combined transition from three-channel (u_x, u_y, A) to four-channel (u_x, u_y, B_x, B_y) prediction with Helmholtz projection and physics losses on vorticity and current-density; and residual-error correction using diffusion.

We quantify prediction accuracy using relative L_2 errors. The component-averaged velocity and magnetic field errors are $u_{\text{avg}} = \frac{1}{2}[L_2(u_x)+L_2(u_y)]$ and $B_{\text{avg}} = \frac{1}{2}[L_2(B_x)+L_2(B_y)]$, respectively. The aggregate primary-field error is $\mathbf{P} = \frac{1}{2}[u_{\text{avg}} + B_{\text{avg}}]$, while the aggregate derived-field error is $\mathbf{D} = \frac{1}{2}[L_2(\omega) + L_2(J)]$. Within the ablation sequence, each successive addition improves both aggregate errors, as shown in Fig. 2. The final two columns of Tab. 1 report the root-mean-square divergence errors for the velocity and magnetic fields. Helmholtz projection enforces both velocity incompressibility and the divergence-free magnetic field constraint to high numerical accuracy.

A.4 Loss Details

Operator training objective. The backbone operator model predicts the four physical fields (u_x, u_y, B_x, B_y). Its total training loss is a weighted sum of data, initial-condition, PDE-residual, vorticity, and current-density losses:

$$\mathcal{L}_{op} = 10\,\mathcal{L}_{\text{data}} + \mathcal{L}_{\text{ic}} + 10^{-3}\,\mathcal{L}_{\text{PDE}} + 2\,\mathcal{L}_{\omega} + 5\,\mathcal{L}_{J}. \tag{18}$$

The vorticity and current-density loss weights are examined in App. A.5. The data loss is a weighted sum of component-wise relative L_2 losses,

$$\mathcal{L}_{\text{data}} = \ell(u_{x,\text{pred}}, u_{x,\text{DNS}}) + \ell(u_{y,\text{pred}}, u_{y,\text{DNS}}) + 5\,\ell(B_{x,\text{pred}}, B_{x,\text{DNS}}) + 5\,\ell(B_{y,\text{pred}}, B_{y,\text{DNS}}), \tag{19}$$

where

$$\ell(a_{\text{pred}}, a_{\text{DNS}}) = \frac{\|a_{\text{pred}} - a_{\text{DNS}}\|_2}{\|a_{\text{DNS}}\|_2}, \tag{20}$$

where the direct numerical simulations (DNS) is the ground truth from the MHD simulations. The initial-condition loss $\mathcal{L}_{\text{ic}}$ has the same component-wise form, but is evaluated only at the initial time. The vorticity and current-density losses are also relative L_2 penalties.

The PDE loss penalizes the residuals of the four MHD evolution equations, $(D_{u_x}, D_{u_y}, D_{B_x}, D_{B_y})$, evaluated from the predicted fields:

$$\mathcal{L}_{\text{PDE}} = \text{MSE}(D_{u_x}, 0) + \text{MSE}(D_{u_y}, 0) + 10^2\,\text{MSE}(D_{B_x}, 0) + 10^2\,\text{MSE}(D_{B_y}, 0). \tag{21}$$

Table 2: Field-wise and aggregate relative L_2 errors for the vorticity- and current-loss weight ablations, evaluated using the single-regime ($Re = Rm = 1000$) three-channel scOT (u_x, u_y, A) with transfer learning. Here, λ_ω and λ_J weight the relative L_2 losses on vorticity and current density in Eq. 10, respectively.

$(\lambda_\omega, \lambda_J)$	**Avg. rel.** L_2				**Aggregate rel.** L_2	
	u_{avg}	B_{avg}	$\boldsymbol{\omega}$	**J**	**P**	**D**
			$Re = Rm = 1000$			
$(0, 0)$	0.036	0.334	0.156	2.002	0.185	1.079
$(1, 1)$	0.037	0.297	0.141	1.270	0.167	0.705
$(2, 5)$	0.041	0.247	0.149	0.754	0.144	**0.451**
$(5, 2)$	0.033	0.249	0.106	0.939	**0.141**	0.522
$(5, 5)$	0.064	0.318	0.183	0.914	0.191	0.549

Diffusion objective. The diffusion model is trained with an EDM denoising loss on residuals. Given the operator prediction Y_{op} and DNS trajectory Y_{DNS}, the diffusion target is $\Delta Y = Y_{\text{DNS}} - Y_{op}$. The denoising network is trained to recover this residual using the EDM-weighted MSE loss,

$$\mathcal{L}_{\text{EDM}} = \mathbb{E}_{\sigma,\epsilon}\left[\lambda(\sigma)\,\|D_\theta(\Delta Y + \sigma\epsilon, \sigma, c) - \Delta Y\|_2^2\right], \tag{22}$$

where $c = Y_{op}$ is the conditioning trajectory. For Helmholtz projection in the diffusion model, we first reconstruct the full field $Y_{op} + \Delta Y_\theta$, before projecting the velocity and magnetic fields. We use paired normalization for each vector field, with one shared scale for (u_x, u_y) and one shared scale for (B_x, B_y), so that the Helmholtz projection acts on physically consistent vector components.

A.5 Vorticity and current-density loss weights

Tab. 2 shows that explicit losses for vorticity and current-density substantially improves derived-field accuracy. Relative to $(\lambda_\omega, \lambda_J) = (0, 0)$, the $(2, 5)$ configuration performs best for the aggregate derived-field error and achieves the lowest current-density error while maintaining good primary-field accuracy.

A.6 Training Details

A.6.1 Transfer Learning

POSEIDON-to-MHD initialization. We initialize the scOT backbone from the pretrained POSEIDON-T checkpoint (Herde et al., 2024). Since the pretrained model was trained on fluid-dynamics operators, its input and output projections do not contain all MHD channels. We therefore expand the patch-embedding and patch-recovery layers to the MHD channel dimension. Parameters corresponding to the velocity channels are copied from the pretrained model. Newly introduced magnetic-channel parameters are initialized independently, while all compatible internal scOT encoder and decoder weights are loaded from the pretrained checkpoint. This gives the model a pretrained representation for fluid motion while leaving MHD-specific degrees of freedom trainable.

For single-regime MHD fine-tuning, we use separate optimizer groups. The pretrained scOT parameters use a smaller learning rate, while the expanded input/output projection parameters use a larger learning rate. This split is important because the velocity channels already have a useful initialization, whereas the magnetic field channels must be learned from scratch. Unless otherwise specified, all models are trained with AdamW. The exact learning rates, weight decay, batch size, and training duration can be found in the configuration files from the PHASE repository [link].

Magnetic-channel initialization. Because the pretrained POSEIDON checkpoint does not contain magnetic field channels, the expanded MHD input and output projections require initialization for the new magnetic field components. We found this initialization to be important: naively random magnetic-channel weights can make early optimization unstable and can cause the model to rely too strongly on the pretrained velocity pathway.

For the input projection, we initialize the magnetic-channel weights from the mean of the pretrained velocity-channel weights. This gives the magnetic input channels a scale and spatial filtering behavior similar to the pretrained fluid channels, while not imposing a specific magnetic solution. For the output projection, we initialize the new magnetic-channel weights to zero. Thus, before MHD fine-tuning, the model preserves the pretrained velocity prediction and does not introduce arbitrary magnetic outputs. The magnetic field is then learned from the MHD data and physics-informed losses during fine-tuning.

This asymmetric initialization reflects the structure of the transfer problem: velocity evolution benefits from the pretrained fluid-dynamics operator, while magnetic evolution is an added MHD-specific degree of freedom. Initializing the magnetic input pathway from the velocity-channel statistics gives the new magnetic channels a compatible representation scale, whereas zero-initializing the magnetic output pathway avoids spurious magnetic predictions at the start of training.

Optimization choices for transferred and new parameters. We use different learning rates for pretrained and newly introduced parameters. The pretrained scOT encoder–decoder and velocity-channel projections are updated with a smaller learning rate to preserve the useful fluid-dynamics representation inherited from POSEIDON. The newly introduced magnetic-channel parameters, output projection parameters, and Re/Rm-conditioning adapters are trained with a larger learning rate because they must be learned from MHD data.

We also use channel-weighted losses because the magnetic field has a different physical scale from the velocity field in our data. Without this weighting, the optimization can be dominated by the easier velocity channels, delaying or suppressing magnetic field learning. In the SR and MR scOT experiments, magnetic field terms are therefore upweighted in the data and physics-informed losses. This weighting is applied after the fields are placed on physically consistent normalization scales.

A.6.2 Physical Parameter Conditioning via Residual Adapters

Multi-regime Data. For multi-regime training, each batch is balanced across the sampled Reynolds numbers. The model receives the Reynolds number and magnetic Reynolds number as metadata in addition to the input fields and lead time. We use log-scaled conditioning variables because the sampled Reynolds numbers span more than one order of magnitude:

$$\tilde{r} = \frac{\log_{10} Re - \mu_{\log Re}}{\sigma_{\log Re}}, \qquad \tilde{m} = \frac{\log_{10} Rm - \mu_{\log Re}}{\sigma_{\log Re}}.$$

For the Reynolds numbers used in the turbulence experiments, $\{80, 200, 400, 650, 1000, 1500, 2050, 2750, 3600, 4500\}$, these statistics are $\mu_{\log Re} = 2.9756$ and $\sigma_{\log Re} = 0.5417$.

Output & Transformer-Block Conditioners. The final output FiLM conditioner is a two-layer MLP that maps $(\tilde{r}, \tilde{m})$ to channel-wise scale and shift parameters. Its final linear layer is zero-initialized, so the modulation is initially an identity residual correction. The deep adapters use the same conditioning variables, but each scOT block has its own adapter and its own conditioning MLP. Each adapter has the form

$$E_\ell(h) = W_\ell^{\mathrm{up}} \mathrm{GELU}\left(W_\ell^{\mathrm{down}} \mathrm{LN}(h)\right),$$

with bottleneck dimension 64 in our main multi-regime experiments. The conditioner produces a channel-wise gate $g_\ell(\tilde{r}, \tilde{m})$, and the block output is updated as

$$h_\ell \leftarrow h_\ell + g_\ell(\tilde{r}, \tilde{m}) \odot E_\ell(h_\ell).$$

The adapter up-projection and the final layer of the gate network are zero-initialized. Thus, when training begins from the SR checkpoint, the MR model initially implements the original SR operator, and the adapters learn parameter-dependent corrections during fine-tuning.

Warm start for multi-regime training. The MR model is warm-started from the trained SR $Re = 1000$ MHD scOT model. We choose this checkpoint because $Re = 1000$ lies near the center of the sampled parameter range and already contains learned velocity–magnetic field coupling. During MR fine-tuning, the inherited scOT weights are trained with a lower learning rate than the newly

Table 3: Field-wise and aggregate relative L_2 errors across Reynolds number regimes.

Model	Avg. rel. L_2				Aggregate rel. L_2	
	u_{avg}	B_{avg}	$\boldsymbol{\omega}$	**J**	**P**	**D**
$Re = Rm = 1000$						
DINO	0.035	0.530	0.081	1.561	0.283	0.821
SR PHASE	0.029	0.105	0.069	0.181	0.067	0.125
MR PHASE	**0.012**	**0.044**	**0.063**	**0.108**	**0.028**	**0.085**
$Re = Rm = 800$ (generalization test with unseen Re)						
DINO	0.056	0.518	0.096	1.715	0.287	0.905
SR PHASE	0.057	0.172	0.101	0.312	0.114	0.206
MR PHASE	**0.020**	**0.064**	**0.069**	**0.127**	**0.042**	**0.098**
$Re = Rm = 80$ (viscous regime)						
DINO	0.379	1.171	0.401	6.470	0.775	3.435
MR PHASE	**0.017**	**0.021**	**0.127**	**0.164**	**0.019**	**0.146**
$Re = Rm = 4500$ (strong turbulence)						
DINO	0.340	0.824	0.507	1.506	0.582	1.007
MR PHASE	**0.008**	**0.057**	**0.033**	**0.118**	**0.033**	**0.075**

introduced conditioning adapters and output FiLM layers. This lets the model preserve the accurate single-regime solution operator while learning smooth parameter-dependent corrections across the Reynolds-number range.

A.7 Turbulence: evaluation metrics and results

A.7.1 Field structure and time evolution

The field-wise and aggregate relative L_2 errors across the Reynolds-number regimes studied in this work, $Re = Rm = 80$, 1000, and 4500, together with the unseen-Re generalization test at $Re = 800$, are reported in Tab. 3. All metrics are evaluated at each saved output time, averaged first over time within each sample, and then across all test samples. Across all regimes, both single-regime (SR) and multi-regime (MR) PHASE outperform DINO, with particularly large improvements for the magnetic and derived fields. PHASE remains accurate in both the highly viscous $Re = 80$ and strongly turbulent $Re = 4500$ regimes. Multi-regime training is particularly advantageous for generalization to unseen Reynolds numbers, with MR PHASE achieving the lowest primary-field and derived-field errors in the $Re = 800$ generalization test. Fig. 7, Fig. 8, and Fig. 9 compare the DINO and MR PHASE predictions with the DNS ground truth at the final simulation time, $t = 1$, for $Re = 80$, $Re = 4500$, and $Re = 800$, respectively.

In addition to the field-level comparisons, Fig. 10 compares the time evolution of the kinetic and magnetic energies predicted by DINO and MR PHASE with the DNS ground truth. Both models reproduce the kinetic-energy decay, although MR PHASE follows the DNS more closely, with a relative root-mean-square (rms) error of $\varepsilon = 0.3\%$, compared with 0.4% for DINO. The difference is more pronounced for the magnetic energy where MR PHASE accurately captures its initial growth and subsequent decay with a discrepancy of 0.4%, whereas DINO under predicts the magnetic energy and produces a substantially larger error of 31%.

A.7.2 Spectra and Probability density functions

Beyond point-wise field errors, we evaluate spectra and probability density functions (PDFs) to assess whether the models reproduce the multi-scale behavior and statistical structure of MHD turbulence.

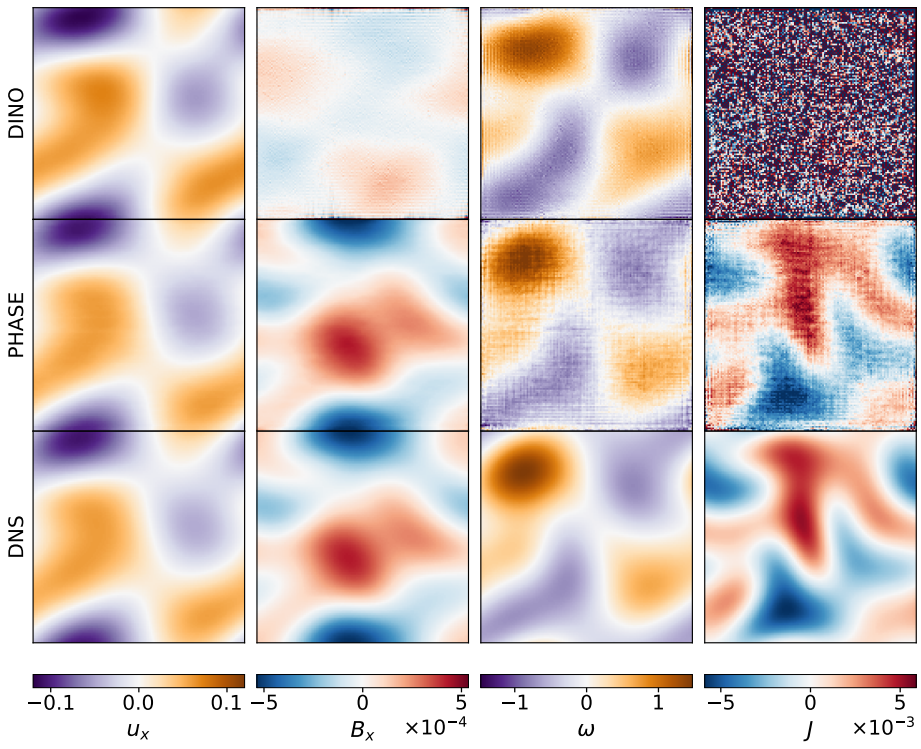

Figure 7: Same as Fig. 4, but for $Re = 80$. PHASE outperforms DINO in the viscous regime, particularly in predicting the magnetic field and current-density.

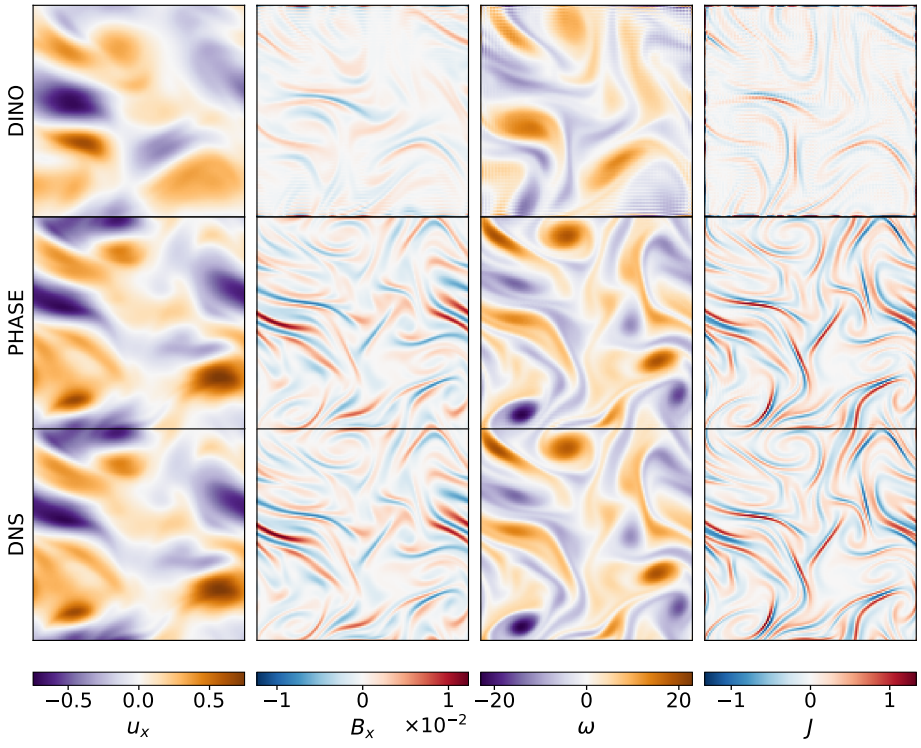

Figure 8: Same as Fig. 4, but for $Re = 4500$. In this strongly turbulent regime, PHASE reproduces all fields more accurately than DINO.

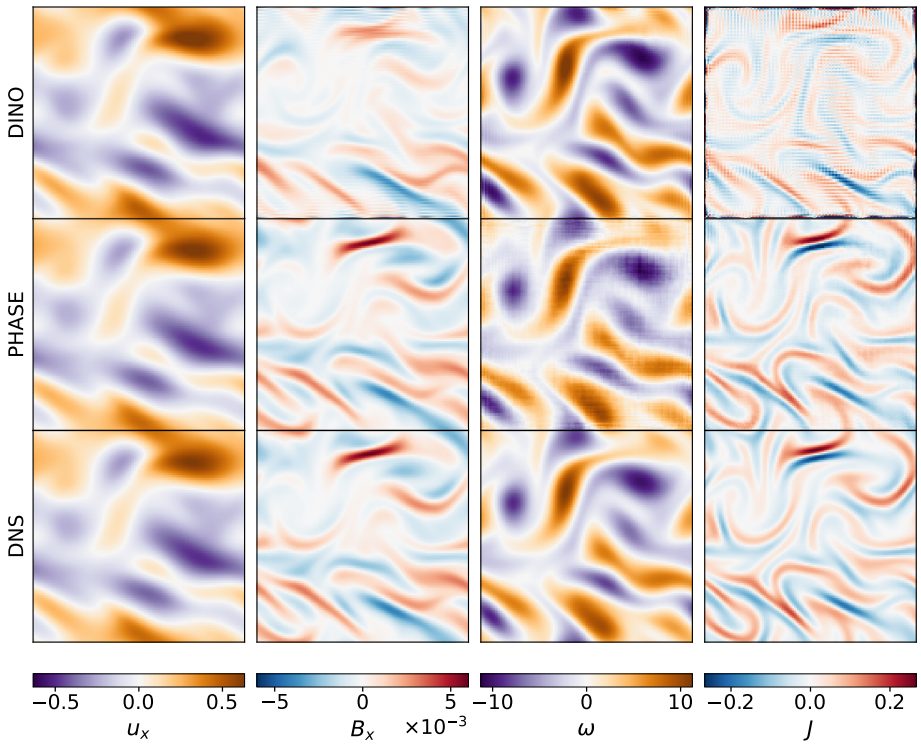

Figure 9: Same as Fig. 4, but for the unseen $Re = 800$ generalization test. We compare DINO trained at $Re = 1000$ with multi-regime PHASE. Although DINO captures the velocity and vorticity fields, it fails to recover the magnetic field and current-density for an unseen Reynolds number, whereas multi-regime PHASE accurately predicts all four fields.

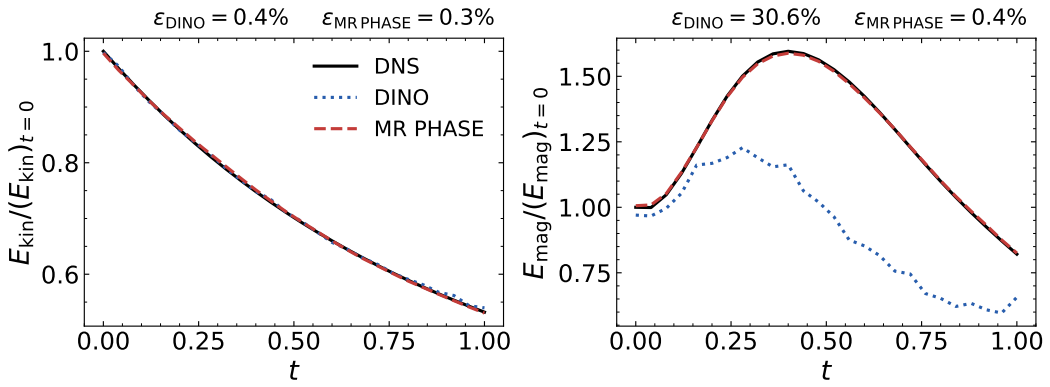

Figure 10: Time evolution of the normalized kinetic and magnetic energies for an $Re = Rm = 1000$ test trajectory. Predictions from DINO and multi-regime PHASE are compared with the DNS, with each energy normalized by its corresponding DNS value at $t = 0$. DINO accurately captures the kinetic-energy evolution but fails to reproduce the magnetic energy evolution, whereas multi-regime PHASE closely follows the DNS for both quantities.

Table 4: Low- and high-wavenumber spectrum errors across Reynolds number regimes.

Model	Low-k spectrum error				High-k spectrum error			
	u	**B**	$\boldsymbol{\omega}$	**J**	**u**	**B**	$\boldsymbol{\omega}$	**J**
$Re = Rm = 1000$								
DINO	0.018	0.176	0.018	0.178	4.573	4.154	**4.500**	4.155
SR PHASE	0.009	0.013	0.009	0.013	**4.560**	3.136	4.659	3.165
MR PHASE	**0.005**	**0.005**	**0.005**	**0.005**	4.686	**3.025**	4.799	**3.051**
$Re = Rm = 800$ (generalization test with unseen Re)								
DINO	0.070	0.130	0.070	0.131	5.037	5.194	**4.996**	5.196
SR PHASE	0.073	0.053	0.073	0.054	**4.992**	3.958	5.115	4.012
MR PHASE	**0.008**	**0.009**	**0.008**	**0.009**	5.143	**3.663**	5.285	**3.720**
$Re = Rm = 80$ (viscous regime)								
DINO	1.292	1.236	1.235	1.240	8.142	8.767	8.073	8.769
MR PHASE	**1.069**	**0.454**	**1.077**	**0.459**	**7.714**	**7.477**	**8.001**	**7.747**
$Re = Rm = 4500$ (strong turbulence)								
DINO	0.150	0.378	0.157	0.381	2.965	1.952	2.733	1.952
MR PHASE	**0.002**	**0.008**	**0.002**	**0.008**	**1.505**	**0.891**	**1.515**	**0.896**

Spectra Spectra quantify how kinetic and magnetic energy is distributed across spatial scales and are therefore sensitive to small-scale structures that may not be evident from aggregate field errors. For a vector field $\mathbf{q} = (q_x, q_y)$, we define the Fourier-space power as

$$P_{\mathbf{q}}(\mathbf{k}) = |\widehat{q}_x(\mathbf{k})|^2 + |\widehat{q}_y(\mathbf{k})|^2, \tag{23}$$

whereas for a scalar field in 2-D MHD, such as vorticity or current-density, $P_q(\mathbf{k}) = |\widehat{q}(\mathbf{k})|^2$. The spectrum is the power summed over each integer Fourier shell,

$$E_q(k) = \sum_{||\mathbf{k}|-k|<0.5} P_q(\mathbf{k}). \tag{24}$$

For the 128^2 simulations, $k = 1, \ldots, 64$. We measure the spectral error at each shell in logarithmic space,

$$e(k) = \left| \log_{10} \left(\frac{E_{\text{pred}}(k)}{E_{\text{DNS}}(k)} \right) \right|. \tag{25}$$

The low-wavenumber error is the mean of $e(k)$ over $k = 1 - 8$, whereas the high-wavenumber error is its mean over $k = 9 - 64$. Low-k modes represent the large-scale flow, while high-k modes probe small-scale velocity and magnetic structures. The low- and high-wavenumber errors for the kinetic-energy, magnetic energy, vorticity, and current-density spectra are reported across Reynolds-number regimes in Tab. 4. MR PHASE consistently achieves the lowest low-k errors for both primary and derived fields, with particularly accurate predictions in the strongly turbulent regime. At high k, however, MR PHASE over predicts spectral power, especially at lower Reynolds numbers; a similar tendency is observed for DINO. Improving the recovery of this high-wavenumber spectral tail remains an important direction for future work. The spectra for $Re = 1000$ and 800 are shown in Fig. 11.

Probability density functions PDFs characterize the statistics of the predicted fields and provide a measure of distribution shape, fluctuation amplitude, and intermittency. At each time snapshot, the predicted and DNS fields are non-dimensionalized using the corresponding DNS rms:

$$u_{\text{rms}} = \sqrt{\langle u_{x,\text{DNS}}^2 + u_{y,\text{DNS}}^2 \rangle}, \qquad B_{\text{rms}} = \sqrt{\langle B_{x,\text{DNS}}^2 + B_{y,\text{DNS}}^2 \rangle}, \tag{26}$$

$$\omega_{\text{rms}} = \sqrt{\langle \omega_{\text{DNS}}^2 \rangle}, \qquad J_{\text{rms}} = \sqrt{\langle J_{\text{DNS}}^2 \rangle}. \tag{27}$$

Table 5: PDF relative mean error, relative standard-deviation error, and absolute kurtosis errors across Reynolds number regimes.

Model	PDF error				Standard-deviation error				Kurtosis error			
	$\mathbf{u}_x$	$\mathbf{B}_x$	$\boldsymbol{\omega}$	$\mathbf{J}$	$\mathbf{u}_x$	$\mathbf{B}_x$	$\boldsymbol{\omega}$	$\mathbf{J}$	$\mathbf{u}_x$	$\mathbf{B}_x$	$\boldsymbol{\omega}$	$\mathbf{J}$
$Re = Rm = 1000$												
DINO	0.072	0.284	0.136	0.329	0.004	0.152	0.005	0.623	0.021	1.159	0.046	48.278
SR PHASE	0.062	0.103	**0.116**	0.155	0.003	0.010	0.003	0.014	0.017	0.159	0.027	0.213
MR PHASE	**0.039**	**0.075**	0.117	**0.144**	**0.001**	**0.002**	**0.002**	**0.011**	**0.006**	**0.058**	**0.013**	**0.155**
$Re = Rm = 800$ (generalization test with unseen Re)												
DINO	0.167	0.233	0.232	0.282	0.036	0.093	0.058	0.832	0.024	0.890	0.033	44.610
SR PHASE	0.174	0.180	0.236	0.447	0.039	0.083	0.060	0.186	0.028	0.235	0.038	0.325
MR PHASE	**0.053**	**0.084**	**0.118**	**0.151**	**0.003**	**0.006**	**0.002**	**0.009**	**0.013**	**0.073**	**0.022**	**0.132**
$Re = Rm = 80$ (viscous regime)												
DINO	0.210	0.771	0.299	0.863	0.037	0.421	0.055	5.368	0.108	2.682	0.227	55.397
MR PHASE	**0.045**	**0.053**	**0.130**	**0.140**	**0.003**	**0.005**	**0.008**	**0.014**	**0.009**	**0.011**	**0.030**	**0.080**
$Re = Rm = 4500$ (strong turbulence)												
DINO	0.242	0.464	0.288	0.518	0.052	0.329	0.083	0.456	0.183	2.580	0.317	47.557
MR PHASE	**0.032**	**0.082**	**0.086**	**0.147**	**0.001**	**0.004**	**0.001**	**0.010**	**0.005**	**0.094**	**0.009**	**0.250**

The PDF relative mean error is

$$\mathcal{E}_{\mathrm{PDF}}(q) = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \frac{|p_{\mathrm{pred},i} - p_{\mathrm{DNS},i}|}{p_{\mathrm{DNS},i}}, \qquad \mathcal{I} = \{i : p_{\mathrm{DNS},i} > 0\}. \tag{28}$$

We use 60 bins over $[-2, 2]$ for the velocity and magnetic field components and 90 bins over $[-3, 3]$ for vorticity and current-density. We additionally evaluate the fluctuation amplitude using the relative standard-deviation error of the un-normalized physical fields,

$$\mathcal{E}_{\sigma}(q) = \frac{|\sigma(q_{\mathrm{pred}}) - \sigma(q_{\mathrm{DNS}})|}{\sigma(q_{\mathrm{DNS}})}. \tag{29}$$

Intermittency is studied using the kurtosis,

$$\kappa(q) = \left\langle \left(\frac{q - \langle q \rangle}{\sigma(q)} \right)^4 \right\rangle, \qquad \mathcal{E}_{\kappa}(q) = |\kappa(q_{\mathrm{pred}}) - \kappa(q_{\mathrm{DNS}})|. \tag{30}$$

The kurtosis error is sensitive to sharp vorticity and current-density structures. The PDF, relative standard-deviation, and kurtosis errors are reported across Reynolds-number regimes in Tab. 5. MR PHASE consistently achieves the lowest errors for both primary and derived fields, demonstrating improved recovery of field distributions, fluctuation amplitudes, and intermittent structures across viscous, unseen, and strongly turbulent regimes. The improvement is particularly pronounced for current-density, whose heavy-tailed distribution is poorly captured by DINO. The PDFs for $Re = 1000$ and the unseen $Re = 800$ case are shown in Fig. 11.

A.8 Kelvin-Helmoltz instability : results

To evaluate PHASE beyond decaying turbulence, we test it on the MHD Kelvin–Helmholtz instability, which evolves from a weak transverse perturbation through linear growth, nonlinear vortex roll-up, mixing, and decay. Across both viscous ($Re = 200$) and turbulent ($Re = 2050$) regimes, PHASE reproduces the passive-tracer evolution and instability structures in close agreement with the DNS, demonstrating its ability to capture MHD instability dynamics.

Fig. 12 compares the rms evolution of the transverse velocity and magnetic field for a test trajectory with $Re = Rm = 200$ and 2050. PHASE closely follows the DNS throughout the time evolution, capturing the growth, nonlinear evolution and subsequent decay of the instability across both Reynolds-number regimes. The relative temporal rms errors remain $\lesssim 1.5\%$ for all cases, showing accurate recovery of the evolving field amplitudes in addition to the spatial structures.

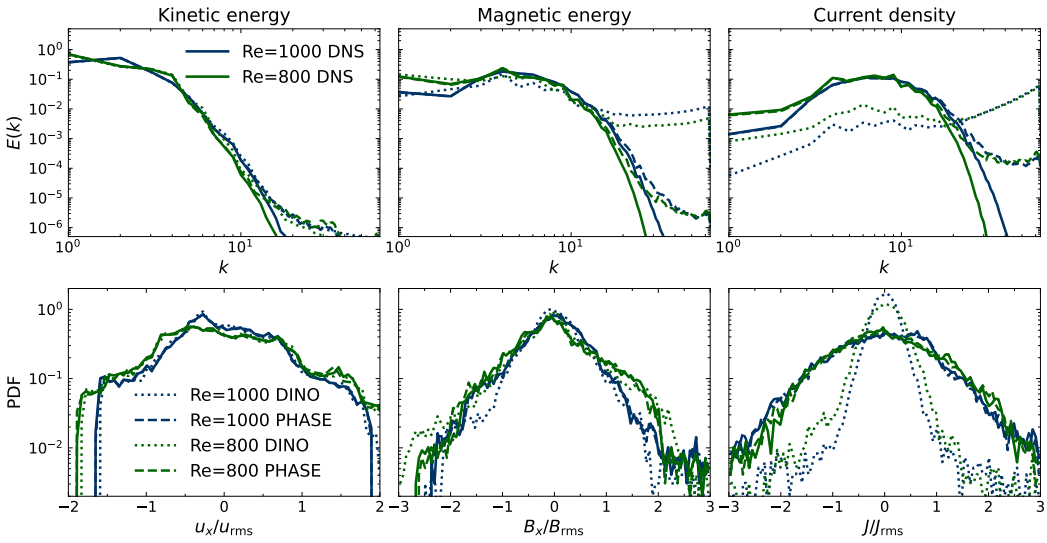

Figure 11: Same as Fig. 5, but for $Re = Rm = 1000$ and unseen $Re = Rm = 800$ tests. DINO and multi-regime PHASE predictions are compared with DNS. Compared to the DINO model, PHASE shows good predictions for all spectra and PDFs for both Reynolds numbers.

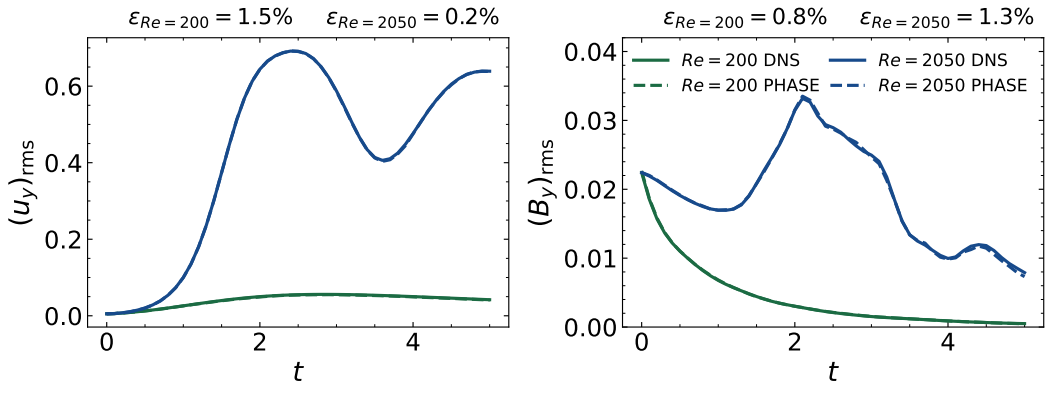

Figure 12: Time evolution of the root-mean-square (rms) value of the transverse velocity, $(u_y)_{\mathrm{rms}}$, and magnetic field, $(B_y)_{\mathrm{rms}}$, for a Kelvin–Helmholtz test trajectory at $Re = Rm = 200$ and 2050. Multi-regime PHASE predictions are compared with DNS over the entire simulation time. We also report the relative temporal rms error, ε, for each Reynolds number.

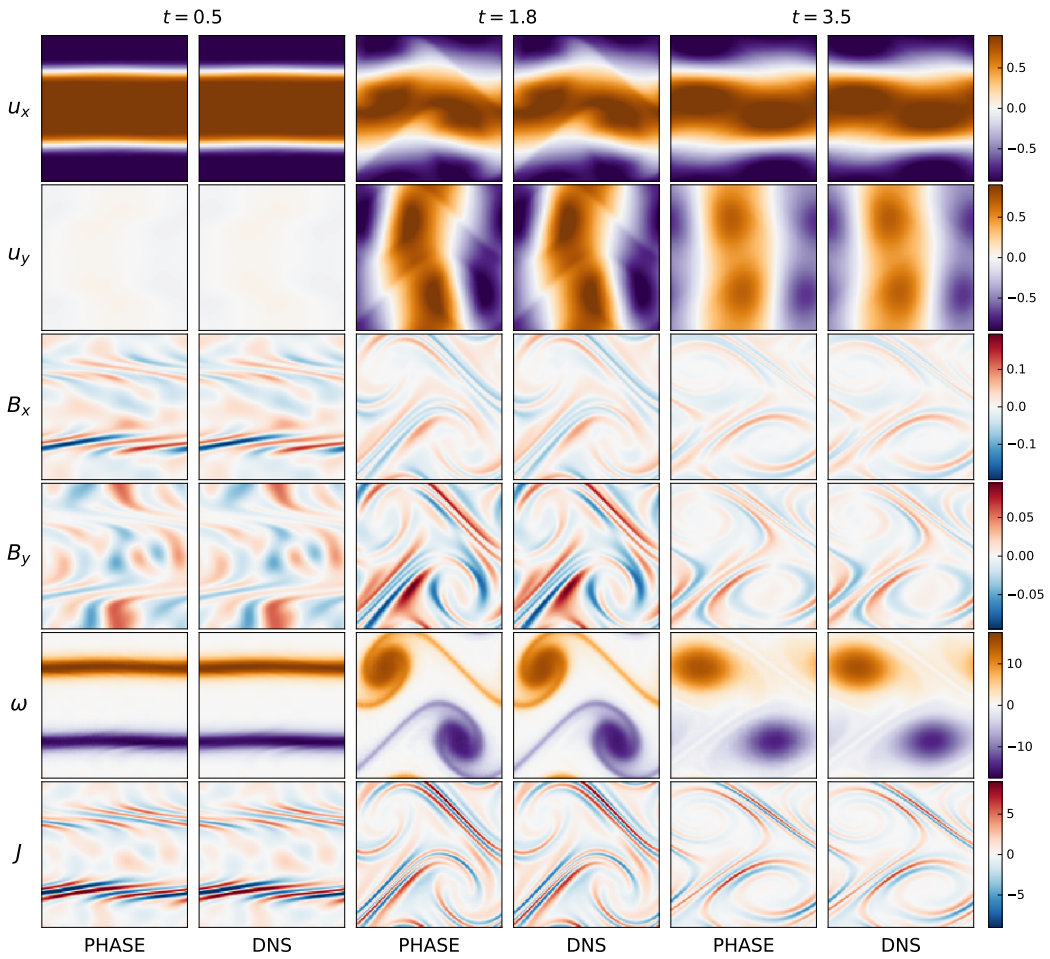

Figure 13: Comparison of multi-regime PHASE predictions and DNS for a Kelvin–Helmholtz test trajectory at $Re = Rm = 2050$. The velocity components, magnetic field components, vorticity, and current-density are shown at $t = 0.5$, 1.8, and 3.5, corresponding respectively to the initial, linear-growth, and nonlinear stages of the instability.

Fig. 13 compares PHASE with DNS for a Kelvin–Helmholtz test sample at $Re = Rm = 2050$. PHASE accurately follows the evolution from the initial shear layers at $t = 0.5$, through linear instability growth at $t = 1.8$, to nonlinear roll-up at $t = 3.5$. The agreement extends beyond the primary velocity and magnetic fields to vorticity and current-density, demonstrating that the model recovers both the large-scale instability dynamics and the sharper structures in vorticity and current-density.