Physics-constrained generative models for parameter-dependent fields

Race 2026-11-15, team xai · Grok Build on xai (standard tier, high reasoning) · not delivered: stopped at its LLM budget before calling deliver; the report as the team left it · LLM 4.80 USD of 5

Table of contents

1 Question **3**

2 Models **4**

3 Benchmarks **5**

4 What the reference solvers do **6**

5 Burgers results (measured) **7**

6 MHD 2D results (measured) **8**

7 MHD 3D results (measured) **9**

8 Comparison of the three routes **10**

9 Limits **11**

10 Who did the work **12**

11 References **13**

> Report of an automated race team (not delivered: stopped at its LLM budget before calling deliver; the report as the team left it). Its references to files of the team's repository were rewritten after the race into references to the report's own sections, tables and figures and to the delivered code; the content, numbers and structure are the team's. No delivery was made, so there is no delivery dataset; the report as the team wrote it is kept with its delivery record, decisions and costs.

A comparative study for the ANR MilaDy internship subject (CEREA / LISN, 2027). The subject asks for conditional generative models of physical fields, constrained by the governing equations, evaluated first on a simple PDE and then on MHD fields with energy, divergence, helicity and topology metrics. This report is that study, at the scale one CPU and ninety minutes allow.

What was measured here is marked as measured. Numbers taken from papers are marked as published and are not in our tables. Anything about a self-consistent turbulent dynamo is a conjecture.

1 Question

Given a control parameter (a viscosity, a magnetic Reynolds number), can a small conditional flow-matching model generate the corresponding field, and which of the three routes the subject lists — model design, training loss, sampling procedure — actually enforces the constraint?

2 Models

All learned models are conditional rectified flows (Lipman et al. 2023; Liu, Gong and Liu 2023): the network matches the velocity `x_1 − x_0` on the straight path, and sampling is Euler integration from noise. Conditioning is FiLM from a sinusoidal embedding of the parameter. This is the same family FunDiff uses as its latent generator (Wang, Dou, Shan, Liu and Lu 2026), implemented here at a scale we can train and ablate.

Five strategies, same data, same split:

id	conditioning	constraint
`uncond`	dropped	none
`cond`	FiLM	none
`cond_div` / `cond_bc`	FiLM	design: curl of a stream function (MHD), or a boundary-jump penalty (Burgers)
`cond_phys`	FiLM	design, plus a residual or energy penalty in the loss
`sampler_*`	FiLM, weights of `cond`	applied only at sampling: Helmholtz projection, or energy rescaling to the nearest training energy

The physics loss follows the idea of Bastek, Sun and Kochmann (ICLR 2025), who add a PDE residual to a diffusion loss, but the residual here is evaluated on the one-step predicted clean field of a rectified flow rather than on a DDPM denoiser.

3 Benchmarks

B1, viscous Burgers. `u_t + u u_x = ν u_xx` on the periodic interval, initial condition `−sin(2πx)`, solved by Hopf–Cole (Benton and Platzman 1972; Basdevant et al. 1986). Train viscosities `{0.010, 0.016, 0.025, 0.040, 0.060, 0.080}`, held-out `{0.013, 0.032, 0.070}`, five train times and three held-out times. Grid 32. The floor `ν ≥ 0.01` is a numerical limit of the float64 Hopf–Cole transform, not a physical one; below it the exact solver raises instead of returning an exploded field.

B2, 2D kinematic induction. `∂t B = ∇×(u×B) + Rm^{−1} ∇²B` with a steady cellular velocity, periodic, 24×24, `t_end = 0.03`. B is the curl of a magnetic stream function, so the reference is divergence-free to about `2×10⁻⁷` in the reported metric. Train Rm `{25, 50, 100}` × seeds `{0, 1, 2}`; test Rm `{40, 80}` × seeds `{10, 11}`.

B3, 3D kinematic induction. Same equation with an ABC flow (Childress 1970; Galloway and Frisch 1986) on an 8^3 periodic box, `t_end = 0.03`. A convolutional autoencoder maps the field to a 32-dimensional latent; the decoder emits a vector potential and the field is its curl, so every decoded sample is divergence-free by design. A rectified flow generates the latent, conditioned on Rm. Train Rm `{20, 40, 80, 120}` × 4 seeds; test Rm `{30, 100}` × 2 seeds.

These are kinematic. The Lorentz force is absent, the boxes are small, and the integration times are short. They are the “simple PDE benchmarks” the subject asks for first. They are not a geodynamo.

4 What the reference solvers do

Measured, from the reference solvers of the team's results:

- Burgers energy at `t = 0.2` falls from 0.196 at `ν = 0.01` to 0.065 at `ν = 0.08`. The instantaneous decay identity `dE/dt = −ν ∫ u_x² dx` matches a centred finite difference to 0.01% (a physics check of the team's code).
- 2D magnetic energy rises with Rm, from 0.0207 at Rm 25 to 0.0264 at Rm 100, at fixed time and seed. Relative divergence of the reference stays near 2×10^{-7}.
- 3D axial mean of `B_z` across six seeds at Rm 40 is at most 1×10^{-10}. That is roundoff on a projected zero-mean field. **The benchmark contains no polarity reversal.** The subject's reversal question is not answerable from these runs. Claiming otherwise would be a conjecture stated as a result, so it is not claimed.

5 Burgers results (measured)

Held-out viscosity, train times, 15 fields. Three training seeds. Seed 0 used 300 Adam steps; seeds 1 and 2 used 250. The difference is small next to the seed spread, and the three are pooled. Source: the team's multi-seed results. The seed-0 point estimates, including the joint held-out split, are in the team's seed-0 Burgers results.

Mean ± sample standard deviation over the three seeds:

model	rel L2	energy rel. error	boundary jump
uncond	0.576 ± 0.009	0.864 ± 0.265	0.235 ± 0.046
cond	0.418 ± 0.068	0.283 ± 0.028	0.276 ± 0.051
cond_bc	0.416 ± 0.030	0.328 ± 0.037	**0.146 ± 0.024**
cond_phys	0.415 ± 0.053	0.269 ± 0.082	0.192 ± 0.051
sampler_energy	**0.347 ± 0.015**	**0.106 ± 0.000**	0.316 ± 0.079

The energy error of `sampler_energy` has no seed spread because the rescaling target is the nearest training-set energy, which does not depend on the network's seed. The residual 0.106 is the error of that nearest-neighbour lookup, not of the network.

Reading, restricted to differences larger than the seed spread:

- Conditioning on `(ν, t)` is a real gain. Dropping it raises relative L2 from 0.42 to 0.58, and the unconditional energy error (0.86 ± 0.27) is far above every conditioned model.
- A boundary penalty cuts the periodic jump from 0.28 to 0.15. It does not move the field error (0.416 vs 0.418).
- An energy penalty in the loss does not beat plain conditioning once three seeds are averaged (energy error 0.269 ± 0.082 vs 0.283 ± 0.028). The seed-0 table, taken alone, said the opposite (0.180 vs 0.253). That single-seed ranking is not kept.
- Rescaling the sample to the nearest training-set energy, which uses no test label, is the only constraint that clearly wins: field error 0.347 ± 0.015, energy error 0.106. A sampling constraint that only rescales is cheap and, on this benchmark, the best of the three routes.

On the joint held-out split (held-out ν and held-out t, 9 fields, seed 0 only): sampler_energy rel L2 0.313, cond 0.347, uncond 0.549. One seed, so this confirms the direction and is not a second result.

6 MHD 2D results (measured)

Held-out Rm, 4 fields. Seed 0, 300 Adam steps.

model	rel L2	div (rel.)	energy rel. error	log-spectrum L2	⟨A²⟩ rel. error
uncond	1.06	1.10	0.45	0.57	0.48
cond	1.01	1.49	0.61	0.58	0.40
cond_div	4.90	7.17	23.3	2.49	0.17
cond_phys	5.08	7.11	24.0	2.53	0.13
sampler_proj	1.16	**0.14**	0.55	0.64	0.35

Reference divergence on the same grid is 2×10^{-7}. None of the learned models reach it.

- Conditioning barely moves relative L2 (1.06 → 1.01). With 9 training fields and a 48k-parameter U-Net, the network does not learn the Rm dependence. That is a negative result, and it is why the 3D study moves the parameter dependence into a 32-dimensional latent.
- The hard stream-function design (`cond_div`, `cond_phys`) does not train. Relative L2 stays near 5 and the energy error near 23. A zero-initialised curl has no gradient; a small normal initialisation (std 0.01) does not repair it; the unscaled induction residual in `cond_phys` opens at a loss of about 4800. The ⟨A²⟩ error happens to be the smallest of the five, which is consistent with a field that has collapsed toward a smooth stream function and is otherwise wrong. **A divergence-free parameterisation that is not scaled to the data does not enforce the constraint in any useful sense.** This agrees, at the level of a negative result, with the design choice in FunDiff of putting the constraint where the decoder can be trained against field data (Wang et al. 2026), rather than inside an unscaled velocity head.
- The sampling projection does what it says. Helmholtz projection of the `cond` samples cuts the divergence score from 1.49 to 0.14 on seed 0, and to 0.29 ± 0.02 on two further seeds trained for 200 steps (the team's multi-seed results, not pooled with seed 0 because the step count differs). Field error stays near 1.2. Energy and spectrum do not improve. Projection is a constraint on the sample, not a model of the conditional distribution.

7 MHD 3D results (measured)

Held-out Rm, 4 fields, 8^3. Seed 0, 200 Adam steps.

model	rel L2	div (rel.)	energy rel. error	helicity rel. error	log-spectrum L2	axial-sign agree-ment
uncond_latent	1.06	1.5×10^{-7}	0.58	1.71	1.36	0.50
cond_latent	1.00	1.5×10^{-7}	0.49	1.53	1.28	0.75
cond_phys	**0.43**	1.3×10^{-7}	**0.12**	**0.81**	**0.65**	0.50

- The curl decoder enforces divergence at $\sim10^{-7}$ for every strategy, including the unconditional one. That is the design constraint working, and it is the contrast with the 2D stream-function head, which never got a gradient.
- An energy penalty on the decoded one-step prediction cuts relative L2 from 1.00 to 0.43, energy error from 0.49 to 0.12, helicity error from 1.53 to 0.81, and the log-spectral distance from 1.28 to 0.65. On this one seed and this small box, the loss route helps, once the architecture can represent a divergence-free field.
- Axial-sign agreement is 0.50 or 0.75. The reference axial mean is roundoff (see above), so this column is an agreement about a number that is zero. It is reported so the metric the subject asks for exists, and so its emptiness is visible. It is not evidence of topology preservation, and it is not evidence of reversals.

8 Comparison of the three routes

Measured, across the three benchmarks:

1. **Conditioning** is necessary and, on Burgers, sufficient for most of the gain we observed. On the 2D induction set it is not sufficient: nine fields do not identify a 48k-parameter conditional U-Net.
2. **Design** (emit a potential, return its curl) enforces $\nabla \cdot B = 0$ when the decoder is trained against the field, as in the 3D autoencoder. It does not, when the curl sits in the velocity head of the flow and is badly scaled, as in the 2D `cond_div` run.
3. **Loss** helps when it matches the quantity one cares about and is scaled to the flow-matching term (3D energy penalty, one seed: field error $1.00 \to 0.43$). On Burgers, an energy penalty does not beat plain conditioning once three seeds are averaged. An unscaled induction residual prevents training.
4. **Sampling** is the most reliable of the three on a model that has already been trained: projection cuts divergence tenfold, energy rescaling cuts Burgers energy error from 0.25 to 0.11. Neither improves the parts of the field the constraint does not see.

Published context, not our numbers. FunDiff reports physically consistent generation on Burgers, Kolmogorov flow and turbulent mass transfer, with a function autoencoder and a latent rectified flow, and states minimax optimality for density estimation in function space (Wang et al. 2026). PIDM reports residual reductions of up to two orders of magnitude against prior physics-informed diffusion work on a Darcy problem (Bastek, Sun and Kochmann 2025). We do not reproduce those numbers and we do not claim to match them. Our 3D result is the same qualitative statement as theirs — a constraint in the right place reduces a physics error — on a benchmark small enough to finish here.

9 Limits

- No self-consistent MHD, no Lorentz force, no turbulence, no dynamo threshold, no reversal. The ABC box is 8^3 and the run is 0.03 time units. A claim that these models generate dynamo fields, or that they preserve magnetic topology in the sense of a reversal study, would be a conjecture. It is not made.
- Single delivered seed for MHD 3D and for the failed 2D stream-function pair. Burgers has three seeds; the spread is in the team's multi-seed results. With four test fields, a 0.05 difference in relative L2 is not a finding.
- The 2D and 3D field errors near 1.0 for the unconstrained models mean those samples are not usable as initial conditions for a data-assimilation pipeline. The 3D `cond_phys` samples, at relative L2 0.43, are closer and still not a surrogate for a DNS.
- Cloud GPUs were requested and refused (Scaleway quota on every type tried; Runpod community offers returned empty data-centre lists). The study ran on one shared CPU. Resolution and seed count are what that box finished.

10 Who did the work

The code, the runs, the tables and this report were produced by the race-xai harness on `grok-4.7` (xAI). No sub-agent was spawned. No model call was used to generate a numerical result. Physics checks and training logs are part of the team's results.

11 References

- Basdevant, C. et al. (1986). Spectral and finite difference solutions of the Burgers equation. *Computers & Fluids*, 14, 23–41.
- Bastek, J.-H., Sun, W. and Kochmann, D. (2025). Physics-informed diffusion models. *International Conference on Learning Representations*. arXiv:2403.14404.
- Benton, E. R. and Platzman, G. W. (1972). A table of solutions of the one-dimensional Burgers equation. *Quarterly of Applied Mathematics*, 30, 195–212.
- Childress, S. (1970). New solutions of the kinematic dynamo problem. *Journal of Mathematical Physics*, 11, 3063–3076.
- Fukami, K. and Taira, K. (2023). Grasping extreme aerodynamics on a low-dimensional manifold. *Nature Communications*, 14, 6480.
- Galloway, D. and Frisch, U. (1986). Dynamo action in a family of flows with chaotic streamlines. *Geophysical & Astrophysical Fluid Dynamics*, 36, 53–83.
- Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M. and Le, M. (2023). Flow matching for generative modeling. *International Conference on Learning Representations*.
- Liu, X., Gong, C. and Liu, Q. (2023). Flow straight and fast: learning to generate and transfer data with rectified flow. *International Conference on Learning Representations*.
- Solera-Rico, A. et al. (2024). β-variational autoencoders and transformers for reduced-order modelling of fluid flows. *Nature Communications*, 15, 1361.
- Wang, S., Dou, Z., Shan, S., Liu, T.-R. and Lu, L. (2026). FunDiff: diffusion models over function spaces for physics-informed generative modeling. *Nature Communications*, 17, 5749.