
Duhamel Neural Operator for Learning Phase-Field Dynamics

Matthew Ju^{1,2} Yingheng Tang² Kang'an Wang^{3,4} Haozhi Sha⁵

¹College of Computing, Data Science, and Society, University of California, Berkeley, CA, USA

²Applied Mathematics and Computational Research Division,
Lawrence Berkeley National Laboratory, Berkeley, CA, USA

³Department of Materials Science and Engineering, University of California, Berkeley, CA, USA

⁴Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

⁵Department of Physics and Astronomy and California NanoSystems Institute,
University of California, Los Angeles, CA, USA

{matthew_ju, kangan.wang}@berkeley.edu, ytang4@lbl.gov, shz23@g.ucla.edu

Abstract

Phase-field simulators are used to predict how polarization fields evolve, a task that has been challenging due to their high computational cost. Machine learning surrogates have quicker inference times, but tend to struggle in approximating the amplification of incubation noise. We introduce the Duhamel Neural Operator (Duhamel-NO), a neural operator architecture consisting of a Fourier state-space model and a next-step U-Shaped Neural Operator, whose update is built on Duhamel’s principle applied to the Time-Dependent Ginzburg–Landau (TDGL) theory. Duhamel-NO reproduces the reference solver’s (FERROX) trajectories up to three orders of magnitude faster while exhibiting the most physically consistent predictions and significantly lower error than the best of seven tuned baselines, achieving a near-perfect correlation zero-shot generalization on unseen materials, initializations, and free energy landscapes. Additionally, we construct an open benchmark for ferroelectric phase-field dynamics. Data & Code: github.com/matthew-ju/Duhamel-Neural-Operator

1 Introduction

Ferroelectrics store information in the direction of their spontaneous polarization, which an electric field can switch. Hafnia-based films stay ferroelectric at thicknesses compatible with silicon processing, so they are being adopted for memory and logic. The standard model of their domain dynamics is the phase-field method, in which the out-of-plane polarization $P_z(\mathbf{r}, t)$ relaxes by time-dependent Ginzburg–Landau (TDGL) gradient flow, coupled to the electrostatics of the device [Chen, 2002]. Graphics processing unit (GPU) solvers such as FERROX [Kumar et al., 2023] are accurate, but one 3D trajectory is expensive, and a study over n values of each Landau coefficient needs n^3 of them.

A trajectory starts from small random noise, 0.5–2% of the saturated polarization P_S of the training materials. During incubation, a band of Fourier modes of this noise grows exponentially while the other modes decay. Domains then nucleate and coarsen. Which noise modes win during incubation decides the final pattern, so a surrogate that errs in the first frames cannot recover it. This is where neural surrogates are weakest, and for truncated neural operators the weakness is structural. A Fourier neural operator (FNO) keeps a fixed set of low Fourier modes, so above that cut-off its linear response is one constant. This caps its pattern correlation at frame 1 at 0.853 on our materials; a trained FNO reaches 0.826.

We derive an architecture that removes this cap. The TDGL right-hand side splits exactly into a linear and a nonlinear part. Duhamel’s principle then writes the next frame as a linear propagator, which is exactly a diagonal Fourier state-space model (SSM) computed in closed form, plus an integral of the nonlinear force, which a neural operator learns.

Our contributions.

- The Duhamel Neural Operator (Duhamel-NO, Section 3), which has no truncation cap (Section B.7).
- Without retraining and from unseen noise, Duhamel-NO reaches correlation 0.982 on held-out materials and 0.978 outside the training range, with 23–26% lower error than the best of seven tuned baselines, at $1814\times$ the throughput of the solver.
- To our knowledge, the first public benchmark for ferroelectric phase-field dynamics (Section G).

Related work. Our update is the exponential-Euler member of exponential time differencing [Cox and Matthews, 2002, Hochbruck and Ostermann, 2010]. Other neural operators also integrate a Fourier symbol exactly [Zhang et al., 2025, Li et al., 2026], but their symbol is learned and only damps. Ours must grow on a band of modes, is measured from simulations, and is interpolated across materials (Section A). Earlier ferroelectric surrogates [Alhada-Lahbabi et al., 2023, 2024] were trained and tested at fixed coefficients and on initial noise seen during training, a leakage that inflates accuracy [Kapoor and Narayanan, 2023].

2 Problem Setup

The polarization relaxes by gradient flow of a Ginzburg–Landau free energy:

$$\frac{\partial P_z}{\partial t} = -\Gamma \frac{\delta F}{\delta P_z}, \quad F[P_z] = \int \left[\underbrace{\frac{\alpha}{2} P_z^2 + \frac{G_{11}}{2} \|\nabla P_z\|^2}_{\text{quadratic}} + \underbrace{\frac{\beta}{4} P_z^4 + \frac{\gamma}{6} P_z^6}_{\text{higher order}} \right] d\mathbf{r} + U_{\text{elec}}[P_z]. \quad (1)$$

U_{elec} is the electrostatic energy of the device, which is quadratic in P_z . The material enters through three Landau coefficients. The coefficient $\alpha < 0$ sets the depth of the double well, and β and γ set its shape ($\beta < 0$ gives a first-order well). We simulate a 5 nm ferroelectric film on a 4 nm dielectric with FERROX, on a $64 \times 64 \times 10$ grid, for 401 frames from small random noise (Section D).

Datasets. **D1** has 1001 materials, spread evenly by a Latin hypercube [McKay et al., 1979] over $\log |\alpha|$ for the well depth $|\alpha| \in [1.6, 4.0] \times 10^9 \text{ V m/C}$, the saturated polarization $P_S \in [0.10, 0.43] \text{ C/m}^2$, and the well shape $r \in [0, 1.4]$ ($r > 1$ means $\beta < 0$). We train on 745 and hold out 256. **D2** has 300 further materials, partly outside this range, with P_S up to $2.2\times$ the training maximum and first-order wells with r up to 2. Every trajectory starts from its own random noise. This matters because FERROX gives the same initial field for the same seed, and a model trained on shared seeds scores 0.99 on a seen seed but only 0.66–0.82 on an unseen one (Section D).

Metrics. We compare a predicted frame with the true one by their *pattern correlation* (Pearson correlation over voxels) and by the relative error $\|A - B\|_2 / \|B\|_2$. Frame $t=400$ is the settled state. Every prediction is rolled out from the $t=0$ noise alone.

3 Method

Duhamel-NO predicts the next frame from the current one. Its update has two parts that mirror the physics, an exact linear part that grows or damps each Fourier mode, and a learned nonlinear part that stops the growth.

3.1 Splitting the TDGL equation

The quadratic energy in Equation (1) gives forces that are linear in P_z , and the higher-order energy gives nonlinear forces. So the equation splits exactly into

$$\partial_t P_z = \mathcal{L} P_z + \mathcal{N}(P_z), \quad \mathcal{L} P_z = -\Gamma (\alpha P_z - G_{11} \nabla^2 P_z - E_z[P_z]), \quad \mathcal{N}(P_z) = -\Gamma (\beta P_z^3 + \gamma P_z^5), \quad (2)$$

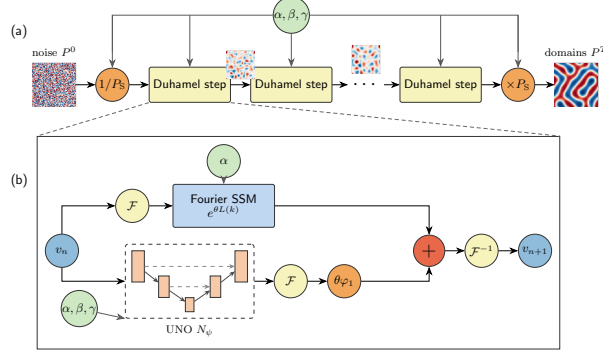


Figure 1: **Duhamel-NO.** (a) The initial noise P^0 is put in material units and advanced by repeated Duhamel steps conditioned on (α, β, γ) . (b) One step (Equation (5)). The Fourier SSM $e^{\theta L(k)}$, set by α , multiplies every mode of v_n by its exact linear factor, and the UNO N_ψ gives the nonlinear correction, which is filtered by $\theta \varphi_1(\theta L)$ and added in Fourier space.

where E_z is the depolarization field, which is linear in P_z . The linear part $\mathcal{L}$ contains everything that matters while the field is small. The nonlinear part $\mathcal{N}$ grows like P_z^3 , so it is negligible during incubation. Once domains form, it is what stops the growth at P_S .

3.2 Duhamel-NO

Duhamel's principle. Without $\mathcal{N}$, the field one frame of length τ later would simply be $e^{\tau \mathcal{L}} P_z(t)$. Duhamel's principle (variation of constants) adds $\mathcal{N}$ back as a source. The force at each instant s inside the frame is carried forward by the linear dynamics for the rest of the frame (Section B.2):

$$P_z(t + \tau) = \underbrace{e^{\tau \mathcal{L}} P_z(t)}_{\text{linear (Fourier SSM)}} + \underbrace{\int_0^\tau e^{(\tau-s)\mathcal{L}} \mathcal{N}(P_z(t+s)) ds}_{\text{nonlinear correction}}. \quad (3)$$

This is exact. The first term is known in closed form, and it amplifies the noise modes that seed the domain pattern. The second term is the hard part, because it depends on the field at every instant inside the frame.

We replace the unknown path by a single *effective force* that gives the same next frame. Over one frame, the linear dynamics turn a constant force f into $\tau \varphi_1(\tau \mathcal{L}) f$, with $\varphi_1(z) = (e^z - 1)/z$. Since $\varphi_1 > 0$ and $\mathcal{L}$ is symmetric, this operator is invertible, so exactly one constant force reproduces the true next frame:

$$P_z(t + \tau) = e^{\tau \mathcal{L}} P_z(t) + \tau \varphi_1(\tau \mathcal{L}) N_{\text{eff}}(P_z(t)), \quad N_{\text{eff}} = [\tau \varphi_1(\tau \mathcal{L})]^{-1} (P_z(t + \tau) - e^{\tau \mathcal{L}} P_z(t)). \quad (4)$$

N_{eff} depends only on the current field and the material, and it is nearly zero during incubation. If $\mathcal{N}$ hardly changes within a frame, $N_{\text{eff}} \approx \mathcal{N}(P_z(t))$, the classic exponential-Euler scheme (Section B.3). In general it has no closed form, so we learn it.

Let v be the field in material units (Section 3.4), θ the frame interval, and L the Fourier symbol of $\mathcal{L}$, so that $e^{\theta L}$ and $\varphi_1(\theta L)$ act on each Fourier mode of v . A neural operator $N_\psi(v; \mathbf{m})$, which also sees the material numbers $\mathbf{m}$, replaces N_{eff} . One step of Duhamel-NO is then (Figure 1)

$$v_{n+1} = e^{\theta L} v_n + \theta \varphi_1(\theta L) N_\psi(v_n; \mathbf{m}). \quad (5)$$

Each piece has a clear job. The term $e^{\theta L}$ moves every Fourier mode by its exact linear factor, including the fine modes the network cannot resolve, so the growth of the noise is exact from the first frame. The filter φ_1 shrinks the network's output on modes that decay within one frame, as the true integral does. The network's last layer starts at zero, so the untrained model is the exact linear solution (Algorithm 1).

3.3 The Fourier SSM

$\mathcal{L}$ is the same everywhere in the plane of the film, so it acts on each in-plane Fourier mode k separately. The electrostatics couple the ten depth layers, so for each k the ten layer amplitudes $\hat{P}_z(k, t) \in \mathbb{C}^{10}$

obey

$$\partial_t \hat{P}_z(k, t) = L(k) \hat{P}_z(k, t), \quad L(k) = -\Gamma[(\alpha + G_{11}|k|^2)I + C(k)], \quad (6)$$

where $C(k)$ is a 10×10 coupling matrix. A mode grows when $L(k)$ has a positive eigenvalue. The negative α drives growth, the gradient term damps short wavelengths, and the depolarization field damps long ones, so only a band of k grows. This band sets the domain size. Because $L(k)$ is symmetric, $e^{\theta L(k)}$ is a diagonal linear state-space model (SSM) in the eigenbasis of $L(k)$ (Proposition 1, Section B.4).

We measure $B(k) = e^{\theta L(k)}$ from data instead of building it from the continuum formula, because the solver’s discrete operators differ slightly. While the field is small, $\mathcal{N}$ is negligible and consecutive frames obey $\hat{P}_z^{t+1} = B\hat{P}_z^t$. So for each of 46 rings of $|k|$ we stack the eight frame pairs $(t, t+1)$, $t = 0, \dots, 7$, into matrices X and Y and fit B by least squares, as in exact dynamic mode decomposition (DMD) [Schmid, 2010, Tu et al., 2014]:

$$B_b = \arg \min_B \|Y - BX\|_F^2 = YX^H(XX^H + \epsilon I)^{-1}. \quad (7)$$

For a new material, we average $\theta^{-1} \log B_b$ over the eight training materials with the closest α ; β and γ are not needed. Rings that decay within one frame get a learnable, strongly damped rate. This Fourier SSM alone reproduces frame 1 at correlation 0.997 on validation materials.

3.4 The neural operator and its training

Material units. The network sees each material in its own units, with P_z in units of P_S and time in units of the linear growth time, both from closed-form laws in (α, β, γ) (Section D.3). In these units the nonlinear force of every material has the shape $-(1-r)v^3 - rv^5$, up to a prefactor that depends on $\mathbf{m}$ (Section B.5), so β and γ enter through a single shape number r .

Network and loss. N_ψ is a U-shaped neural operator (UNO) [Rahman et al., 2023] with 9.5M parameters (Section E). Its input is the field divided by its root-mean-square amplitude s , together with $\log s$ and the material numbers. Dividing by s lets one network handle fields whose size changes by 10^3 , and $\log s$ tells it how close the field is to saturation. We train on five-frame windows, rolling out $k = 4$ steps with gradients through all steps. The loss is the relative error,

$$\mathcal{L} = \frac{1}{k} \sum_{q=1}^k \frac{\|\tilde{P}^q - P^q\|^2}{\|P^q\|^2}, \quad \tilde{P}^0 = P^0, \quad \tilde{P}^q = \Phi_\theta(\tilde{P}^{q-1}), \quad (8)$$

where Φ_θ is one step of Equation (5). Without the normalization, the tiny early frames would not count.

4 Results

Baselines. We compare against seven neural operators and U-Nets, namely FNO [Li et al., 2021], TFNO [Kossaifi et al., 2024], F-FNO [Tran et al., 2023], UNO, U-Net, CNextU-Net [Gupta and Brandstetter, 2023, Ohana et al., 2024] and CNO [Raonić et al., 2023]. For a fair comparison, every baseline gets the same training as Duhamel-NO and is tuned on validation materials. This setup matters, because with standard training no backbone exceeds a correlation of about 0.84 (Section E). The main tables show four baselines (all in Section F). A fine-tuned foundation model, Poseidon-T [Herde et al., 2024], a transformer, Transolver [Wu et al., 2024], and generative models [Kohl et al., 2023, Lipman et al., 2023, Lippe et al., 2023] did not learn the task.

Zero-shot generalization. On the 256 held-out D1 materials (Table 1), Duhamel-NO reaches correlation 0.982 at frame 400, and its error is 23% lower than that of the strongest baseline, F-FNO. On D2, which is partly outside the training range, its error is 26% lower. Across the five D2 regions, its correlation stays between 0.974 and 0.985. UNO, the same network and training without the Duhamel step, has error 0.219 against 0.129, which isolates the architecture. Figure 2 shows a typical held-out material.

Table 1: **Accuracy on unseen materials and noise** (mean over seeds; **best, second**). Correlation and error at frame 400, and error at frame 1 (D1). All models are in Table 6.

model	D1 corr.	D1 error	D2 error	frame-1 error
Duhamel-NO (ours)	0.982	0.129	0.145	0.082
F-FNO	0.974	0.167	0.195	0.215
U-Net	0.966	0.207	0.242	0.313
UNO	0.962	0.219	0.238	0.315
FNO	0.947	0.270	0.268	0.568
Poseidon-T	0.553	2.917	2.570	4.342

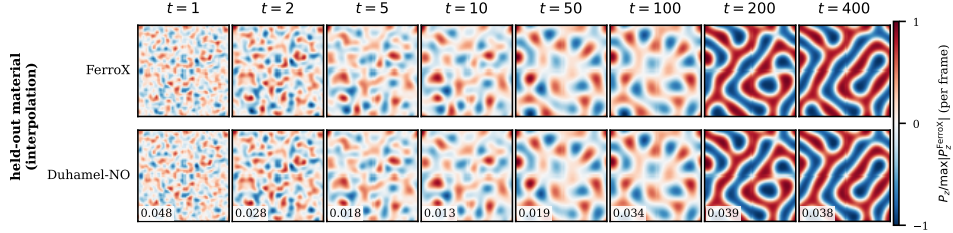


Figure 2: **Duhamel-NO rolled out from unseen noise on a held-out material** (median error), against FERROX (top layer, scaled per frame by the true maximum, with relative errors).

The first frames. The gain is largest where the pattern is decided. At frame 1 the error of Duhamel-NO is 0.082, against 0.215 for F-FNO and 0.568 for FNO (Figure 9), and its correlation, 0.997, is above the cap of 0.853 for any FNO with our mode set (Propositions 5 and 6).

Physical quantities. A low field error does not guarantee correct physics, so we recompute from each prediction the quantities a materials scientist would use, a list fixed before evaluating any model. In paired tests over 12 metrics (Table 2, the field errors at frames 1, 10 and 400, and the wall energy, not shown), Duhamel-NO beats the best baseline on 8 and ties on 4 that describe late-time domain shape. It predicts the nucleation frame exactly for 73% of trajectories, against 62% for the best baseline.

Table 2: **Physical-quantity errors**, D1 held-out (**best, second**, excluding Poseidon-T; one seed for U-Net, Poseidon-T). t_{nuc} in frames, $F(t)$ relative to the energy drop, others relative error at frame 400. All models in Table 8.

model	t_{nuc}	P_S	domain size	wall density	P/P_S hist.	φ	E_z	$F(t)$
Duhamel-NO (ours)	0.28	0.003	0.005	0.007	0.010	0.139	0.135	0.004
F-FNO	0.70	0.006	0.005	0.008	0.010	0.165	0.164	0.007
U-Net	0.61	0.008	0.005	0.008	0.011	0.201	0.200	0.006
UNO	0.54	0.024	0.006	0.009	0.020	0.220	0.218	0.012
FNO	0.68	0.012	0.008	0.011	0.014	0.270	0.267	0.009
Poseidon-T	57.7	2.136	0.246	0.665	1.301	3.035	2.979	1106

5 Conclusion

Duhamel-NO computes the linear growth of the noise exactly and learns only the nonlinear correction. It generalizes zero-shot to unseen noise and materials with lower error and better physics than seven tuned baselines, at $1814\times$ the throughput of FERROX. On Allen–Cahn it transfers to a finer grid but not to a coarser one (Section I). It is built for noise-seeded dynamics, is second to CNO on hysteresis, and is tested against one simulator only (Section J).

Acknowledgements

The authors would like to acknowledge helpful comments and suggestions by Andrew Nonaka from LBNL CCSE. We thank Arnav Shah for meaningful discussions and assistance in revisions. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy User Facility.

References

- Kévin Alhada-Lahbabi, Damien Deleruyelle, and Brice Gautier. Machine learning surrogate model for acceleration of ferroelectric phase-field modeling. *ACS Applied Electronic Materials*, 5(7): 3894–3907, 2023.
- Kévin Alhada-Lahbabi, Damien Deleruyelle, and Brice Gautier. Machine learning surrogate for 3D phase-field modeling of ferroelectric tip-induced electrical switching. *npj Computational Materials*, 10:197, 2024.
- Samuel M. Allen and John W. Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metallurgica*, 27(6):1085–1095, 1979. doi: 10.1016/0001-6160(79)90196-2.
- Christophe Bonneville, Nathan Bieberdorf, Arun Hegde, Mark Asta, Habib N. Najm, Laurent Capolungo, and Cosmin Safta. Accelerating phase field simulations through a hybrid adaptive Fourier neural operator with U-net backbone. *npj Computational Materials*, 11:12, 2025a. arXiv:2406.17119.
- Christophe Bonneville, Nathan Bieberdorf, Pieterjan Robbe, Mark Asta, Habib N. Najm, Laurent Capolungo, and Cosmin Safta. Extrapolating phase-field simulations in space and time with purely convolutional architectures, 2025b. arXiv:2509.20770.
- Johannes Brandstetter, Daniel E. Worrall, and Max Welling. Message passing neural PDE solvers. In *International Conference on Learning Representations (ICLR)*, 2022.
- John W. Cahn. On spinodal decomposition. *Acta Metallurgica*, 9(9):795–801, 1961. doi: 10.1016/0001-6160(61)90182-1.
- John W. Cahn and John E. Hilliard. Free energy of a nonuniform system. I. interfacial free energy. *The Journal of Chemical Physics*, 28(2):258–267, 1958. doi: 10.1063/1.1744102.
- Long-Qing Chen. Phase-field models for microstructure evolution. *Annual Review of Materials Research*, 32:113–140, 2002.
- S. M. Cox and P. C. Matthews. Exponential time differencing for stiff systems. *Journal of Computational Physics*, 176(2):430–455, 2002.
- A. Fantasia, D. Lanzoni, N. Di Eugenio, A. Monteleone, R. Bergamaschini, and F. Montalenti. A parametrically-conditioned deep learning surrogate for coherent spinodal decomposition. *Advanced Theory and Simulations*, 9(2):e02144, 2026. doi: 10.1002/adts.202502144.
- Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations (ICLR)*, 2022a.
- Albert Gu, Ankit Gupta, Karan Goel, and Christopher Ré. On the parameterization and initialization of diagonal state space models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022b.
- Ankit Gupta, Albert Gu, and Jonathan Berant. Diagonal state spaces are as effective as structured state spaces. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- Jayesh K. Gupta and Johannes Brandstetter. Towards multi-spatiotemporal-scale generalized PDE modeling. *Transactions on Machine Learning Research*, 2023.
- Erfan Hamdi and Emma Lejeune. Towards robust surrogate models: Benchmarking machine learning approaches to expediting phase field simulations of brittle fracture, 2025. arXiv:2507.07237.

- Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for PDEs. In *Advances in Neural Information Processing Systems*, 2024.
- Marlis Hochbruck and Alexander Ostermann. Exponential integrators. *Acta Numerica*, 19:209–286, 2010.
- Sayash Kapoor and Arvind Narayanan. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9):100804, 2023.
- Aly-Khan Kassam and Lloyd N. Trefethen. Fourth-order time-stepping for stiff PDEs. *SIAM Journal on Scientific Computing*, 26(4):1214–1233, 2005.
- Georg Kohl, Li-Wei Chen, and Nils Thuerey. Benchmarking autoregressive conditional diffusion models for turbulent flow simulation, 2023. arXiv:2309.01745.
- Jean Kossaifi, Nikola Kovachki, Kamyar Azizzadenesheli, and Anima Anandkumar. Multi-grid tensorized Fourier neural operator for high-resolution PDEs. *Transactions on Machine Learning Research*, 2024.
- Prabhat Kumar, Andrew Nonaka, Revathi Jambunathan, Girish Pahwa, Sayeef Salahuddin, and Zhi Yao. FerroX: A GPU-accelerated, 3D phase-field simulation framework for modeling ferroelectric devices. *Computer Physics Communications*, 290:108757, 2023. doi: 10.1016/j.cpc.2023.108757.
- J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM, 2016.
- Jiayi Li, Zhaonan Wang, and Flora D. Salim. SGNO: Spectral generator neural operators for stable long horizon PDE rollouts, 2026. arXiv:2602.18801.
- Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.
- Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.
- Phillip Lippe, Bas S. Veeling, Paris Perdikaris, Richard E. Turner, and Johannes Brandstetter. PDE-Refiner: Achieving accurate long rollouts with temporal neural PDE solvers. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.
- Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.
- Nick McGreivy and Ammar Hakim. Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nature Machine Intelligence*, 6: 1256–1269, 2024.
- Michael D. McKay, Richard J. Beckman, and William J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.
- David Nualart. *The Malliavin Calculus and Related Topics*. Springer, 2 edition, 2006.
- Ruben Ohana, Michael McCabe, Lucas Meyer, et al. The Well: a large-scale collection of diverse physics simulations for machine learning. In *Advances in Neural Information Processing Systems, Datasets and Benchmarks Track*, 2024.
- Vivek Oommen, Khemraj Shukla, Somdatta Goswami, Rémi Dingreville, and George Em Karniadakis. Learning two-phase microstructure evolution using neural operators and autoencoder architectures. *npj Computational Materials*, 8:190, 2022.

- Antonio Orvieto, Samuel L Smith, Albert Gu, Anushan Fernando, Caglar Gulcehre, Razvan Pascanu, and Soham De. Resurrecting recurrent neural networks for long sequences. In *International Conference on Machine Learning (ICML)*, 2023.
- Md Ashiqur Rahman, Zachary E. Ross, and Kamyar Azizzadenesheli. U-NO: U-shaped neural operators. *Transactions on Machine Learning Research*, 2023.
- Bogdan Raonić, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel de Bézenac. Convolutional neural operators for robust and accurate learning of PDEs. In *Advances in Neural Information Processing Systems*, 2023.
- Cristopher Salvi, Maud Lemerrier, and Andris Gerasimovičs. Neural stochastic PDEs: Resolution-invariant learning of continuous spatiotemporal dynamics. In *Advances in Neural Information Processing Systems*, 2022.
- Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.
- Leslie N. Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, volume 11006 of *Proc. SPIE*, 2019.
- Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.
- Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized Fourier neural operators. In *International Conference on Learning Representations*, 2023.
- Jonathan H Tu, Clarence W Rowley, Dirk M Luchtenburg, Steven L Brunton, and J Nathan Kutz. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1(2):391–421, 2014.
- Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In *International Conference on Machine Learning*, 2024.
- Rui Zhang, Han Wan, Yang Liu, and Hao Sun. Stable spectral neural operator for learning stiff PDE systems from limited data, 2025. arXiv:2512.11686.

A Related work

Exponential integrators and neural operators. Exponential time differencing solves the stiff linear part of an equation exactly and approximates only the nonlinear part [Cox and Matthews, 2002, Kassam and Trefethen, 2005, Hochbruck and Ostermann, 2010]. Our update is the simplest member of this family, the exponential-Euler step. Some neural operators also apply a Fourier symbol exactly. Neural stochastic partial differential equations (SPDEs) use the mild (Duhamel) form of the equation [Salvi et al., 2022]. SSNO and SGNO (stable spectral and spectral generator neural operators) learn a symbol that only damps [Zhang et al., 2025, Li et al., 2026]. Duhamel-NO differs in three ways. First, its linear operator must *grow* on a band of modes, because this growth seeds the pattern. Second, the operator is *measured* from simulations rather than learned (only a few rings that decay within one frame get a learnable, strongly damped rate), and it couples the depth layers of a 3D film. Third, it is interpolated across materials.

State-space models and data-driven linear operators. Deep state-space models became practical once the state transition was made diagonal [Gu et al., 2022a,b, Gupta et al., 2022, Orvieto et al., 2023]. The linear part of TDGL is exactly such a model (Proposition 1). So it can be computed in closed form from pairs of consecutive frames, as in exact dynamic mode decomposition (DMD) and Koopman methods [Schmid, 2010, Tu et al., 2014, Kutz et al., 2016]. The physics of this linear stage is the classical theory of spinodal decomposition [Cahn and Hilliard, 1958, Cahn, 1961, Allen and Cahn, 1979].

Ferroelectric and phase-field surrogates. Earlier ferroelectric surrogates roll the trajectory forward with one learned model, such as an autoregressive network [Alhada-Lahbabi et al., 2024] or a long short-term memory (LSTM) network over principal components [Alhada-Lahbabi et al., 2023]. Surrogates for other phase-field systems also fit one model over the whole trajectory [Bonneville et al., 2025a,b, Oommen et al., 2022], and some condition on material parameters [Fantasia et al., 2026, Hamdi and Lejeune, 2025]. To our knowledge, earlier ferroelectric surrogates are trained and tested at fixed coefficients and on initial noise seen during training. This kind of leakage inflates accuracy [Kapoor and Narayanan, 2023, McGreivy and Hakim, 2024], so every test trajectory here has its own noise.

B Derivations and proofs

This section collects the math behind Section 3. Sections B.1 to B.3 derive the update of Duhamel-NO. Sections B.4 and B.5 show why the linear part is an SSM and why one shape number describes the nonlinear part. Sections B.6 and B.7 explain the behaviour of the error in the first frames.

Notation. $\hat{f}(k)$ is the in-plane Fourier transform of a field f , with $k = (k_x, k_y)$. A field on the $64 \times 64 \times 10$ grid is a vector of ten depth layers at every in-plane point. $\|u\|$ is the L_2 norm over all voxels.

B.1 The split of the TDGL equation

The free energy of Equation (1) is $F = F_2 + F_{46}$, with

$$F_2[P_z] = \int \left[\frac{\alpha}{2} P_z^2 + \frac{G_{11}}{2} \|\nabla P_z\|^2 \right] d\mathbf{r} + U_{\text{elec}}[P_z], \quad F_{46}[P_z] = \int \left[\frac{\beta}{4} P_z^4 + \frac{\gamma}{6} P_z^6 \right] d\mathbf{r}.$$

The electrostatic energy is $U_{\text{elec}} = -\frac{1}{2} \int E_z[P_z] P_z d\mathbf{r}$, where $E_z[P_z]$ is the z -component of the field that solves the Poisson problem of the stack with P_z as its source. The Poisson problem is linear with fixed (grounded) boundary values, so $E_z[P_z]$ is linear in P_z and the map $P_z \mapsto -E_z[P_z]$ is symmetric. The variational derivatives are therefore

$$\frac{\delta F_2}{\delta P_z} = \alpha P_z - G_{11} \nabla^2 P_z - E_z[P_z], \quad \frac{\delta F_{46}}{\delta P_z} = \beta P_z^3 + \gamma P_z^5,$$

and $\partial_t P_z = -\Gamma \delta F / \delta P_z$ is exactly $\partial_t P_z = \mathcal{L} P_z + \mathcal{N}(P_z)$ with $\mathcal{L}$ and $\mathcal{N}$ as in Equation (2). Two facts follow. First, $\mathcal{L}$ is linear and self-adjoint (up to the factor $-\Gamma$), and it holds every coefficient

that still matters when the field is small, namely α , G_{11} and the geometry of the stack. Second, $\mathcal{N}$ holds only β and γ . This is why the Fourier SSM of Duhamel-NO depends on α alone, while β and γ enter only through the correction. (The network also receives α , through the material numbers.)

B.2 Duhamel's principle

The original principle. Duhamel's principle (variation of constants) solves a *linear* equation driven by a *known* source $f(t)$. For $\partial_t u = \mathcal{L}u + f(t)$, the state one frame of length τ later is

$$u(t + \tau) = e^{\tau\mathcal{L}}u(t) + \int_0^\tau e^{(\tau-s)\mathcal{L}}f(t+s)ds. \quad (9)$$

To see this, multiply the equation by $e^{-s\mathcal{L}}$. This gives $\frac{d}{ds}(e^{-s\mathcal{L}}u) = e^{-s\mathcal{L}}f$, and integrating over the frame gives Equation (9). The first term is how the current state evolves on its own. The second term adds up the effect of the source, each instant carried forward to the end of the frame.

Applying it to TDGL. In Equation (2), the nonlinear term plays the role of the source, $f(s) = \mathcal{N}(P_z(t+s))$. Unlike the textbook case, this source is not known in advance, because it depends on the solution itself. Equation (9) then becomes Equation (3). The result is still exact, but it is implicit, because the integral needs the field at every instant inside the frame, and one frame spans many solver steps.

The roles of the two terms. $\mathcal{L}$ is unstable on a band of wavenumbers, so the first term amplifies the noise modes that seed the domain pattern. The force $\mathcal{N}$ comes from the higher-order Landau terms and scales like P_z^3 . It is negligible while the field is small. Once the field grows, it opposes further growth and holds the field at P_S . (For $\beta < 0$ the quartic term pushes the other way, and the sextic term provides the restoring force.) So the integral is the correction that turns exponential growth into domains, just as the higher-order terms of F keep the energy bounded from below. In Duhamel-NO the only approximation is $N_\psi \approx N_{\text{eff}}$. Any error in the measured L is absorbed by the network.

B.3 The effective force, the exponential-Euler step and the filter φ_1

This subsection shows three things. The effective force is well defined, it is close to the true force on modes that are not stiff, and the filter φ_1 makes errors on stiff modes harmless.

The operator $\tau\varphi_1(\tau\mathcal{L}) = \int_0^\tau e^{(\tau-s)\mathcal{L}}ds$ is symmetric and positive definite, because $\mathcal{L}$ is symmetric and $\varphi_1(z) = \int_0^1 e^{zr}dr > 0$. So it is invertible, and the effective force of Equation (4) is well defined. By Equation (3), it satisfies $\tau\varphi_1(\tau\mathcal{L})N_{\text{eff}}(P_z(t)) = \int_0^\tau e^{(\tau-s)\mathcal{L}}\mathcal{N}(P_z(t+s))ds$. We also use that φ_1 is increasing on the real line.

Distance from the frozen force. Freezing the integrand of Equation (3) at $s = 0$ gives

$$P_z(t + \tau) \approx e^{\tau\mathcal{L}}P_z(t) + \tau\varphi_1(\tau\mathcal{L})\mathcal{N}(P_z(t)),$$

the exponential-Euler scheme [Cox and Matthews, 2002, Hochbruck and Ostermann, 2010], so $N_{\text{eff}} = \mathcal{N}(P_z(t))$ if $\mathcal{N}$ is constant over the frame. In general,

$$\tau\varphi_1(\tau\mathcal{L})[N_{\text{eff}}(P_z(t)) - \mathcal{N}(P_z(t))] = \int_0^\tau e^{(\tau-s)\mathcal{L}}[\mathcal{N}(P_z(t+s)) - \mathcal{N}(P_z(t))]ds,$$

and the norm of the right-hand side is at most $\frac{\tau^2}{2} \sup_{s \leq \tau} \|e^{(\tau-s)\mathcal{L}}\| \sup_{s \leq \tau} \|\frac{d}{ds}\mathcal{N}(P_z(t+s))\|$.

Non-stiff modes. On a mode with $\tau\lambda \geq -1$, $\tau\varphi_1(\tau\lambda) \geq \tau\varphi_1(-1) = 0.63\tau$, so dividing by it gives $N_{\text{eff}} - \mathcal{N}(P_z(t)) = O(\tau)$. The difference vanishes when $\mathcal{N}$ does not change within a frame, and in particular during incubation, where $\mathcal{N} = O(P_z^3)$.

Stiff modes. On modes with $\tau\lambda \ll -1$, $\tau\varphi_1(\tau\lambda) \approx 1/|\lambda|$ is small. So whatever the effective force is on those modes, it enters the step with a small weight, and the network does not need to get it right.

Duhamel-NO replaces N_{eff} by a learned N_ψ , trained so that the whole step matches the data. It therefore learns the effective force along the path, not the frozen value at $s = 0$.

B.4 The linear propagator is a diagonal Fourier SSM

In words, once each Fourier mode is written in the right basis, every component simply gets multiplied by a fixed number at every frame. This is exactly the form of the diagonal state-space models used in deep learning.

$L(k)$ of Equation (6) is real symmetric (up to the factor $-\Gamma$), so it has real eigenvalues $\lambda_i(k)$ and orthonormal eigenvectors $V(k)$. Integrating Equation (6) over one frame gives $B(k) = e^{\theta L(k)} = V(k) \text{diag}(e^{\theta \lambda_i(k)}) V(k)^\top$. Since α multiplies the identity, changing α only shifts the generator, $L(k; \alpha) = L(k; \alpha') - \Gamma(\alpha - \alpha')I$. And β, γ do not appear in $L(k)$ at all.

Proposition 1 (The linear propagator is a diagonal Fourier SSM). Write $B(k) = V(k)\Lambda(k)V(k)^{-1}$. Then the linear dynamics are exactly the diagonal, linear, time-invariant state-space recurrence

$$h^0(k) = V(k)^{-1} \hat{P}_z^0(k), \quad h^{t+1}(k) = \Lambda(k) h^t(k), \quad \hat{P}_z^t(k) = V(k) h^t(k),$$

with $A = \Lambda$, $C = V$ and $D = 0$, and with the initial noise as an impulse input. This is the form of the diagonal SSMs S4D (diagonal structured state space), DSS (diagonal state spaces) and LRU (linear recurrent unit) [Gu et al., 2022b, Gupta et al., 2022, Orvieto et al., 2023].

Each in-plane mode is an independent copy of this system, so one Fourier SSM step costs one 10×10 matrix–vector product per mode.

Proof of Proposition 1. Let $\Lambda(k)$ be diagonal and set $h^t(k) = V(k)^{-1} \hat{P}_z^t(k)$. Then $h^{t+1} = V^{-1} B \hat{P}_z^t = V^{-1} V \Lambda V^{-1} \hat{P}_z^t = \Lambda h^t$ and $\hat{P}_z^t = V h^t$. This is the state-space recurrence $h^{t+1} = A h^t + B x^t$, $y^t = C h^t + D x^t$ with $A = \Lambda$, $C = V$, $D = 0$ and an input x^t that is the initial noise at $t = 0$ and zero afterwards. Every mode k is an independent copy of this system, so the full linear dynamics are a bank of diagonal, linear, time-invariant SSMs, one per in-plane mode. $\square$

Remark. We also checked the measured (discrete) map numerically. On the reference material, the relative reconstruction error $\|V\Lambda V^{-1} - B\|_\infty / \|B\|_\infty$ is 2.9×10^{-15} , and the condition number of V is at most 6.39. In the continuum the eigenvalues are real (Section B.1). In the fitted map, all 128 growing eigenvalues of the reference material are real to 3.9×10^{-8} . Complex pairs occur only among modes that decay within two frames.

B.5 Material units and the shared nonlinearity

In material units, every material has the same nonlinear force up to one number, r . To see this, write $P_z = P_S v$, measure time in units of λ , and divide $\mathcal{N}$ by $\Gamma |a_{\text{sat}}| P_S$, where $|a_{\text{sat}}| = \beta P_S^2 + \gamma P_S^4$ defines P_S (Section D.3):

$$\frac{\mathcal{N}(P_S v)}{\Gamma |a_{\text{sat}}| P_S} = -\frac{\beta P_S^2}{|a_{\text{sat}}|} v^3 - \frac{\gamma P_S^4}{|a_{\text{sat}}|} v^5 = -(1-r) v^3 - r v^5, \quad r = \frac{\gamma P_S^4}{|a_{\text{sat}}|}.$$

The second equality uses $\beta P_S^2 / |a_{\text{sat}}| = 1 - r$, which follows from the definition of P_S . The matching reduced potential $\phi_r(v) = -v^2/2 + (1-r)v^4/4 + r v^6/6$ has its minima at $v = \pm 1$ for every r (Figure 5). For $r < 1$ the well is second order. At $r = 1$, $\beta = 0$. For $r > 1$ the well is first order ($\beta < 0$), and only the sextic term holds it.

The shape is shared, but the size is not quite. Time is measured in $\lambda = 1/(\Gamma |a_{\text{rate}}|)$, not in $1/(\Gamma |a_{\text{sat}}|)$, so in material units the force is $(|a_{\text{sat}}|/|a_{\text{rate}}|) (-(1-r)v^3 - r v^5)$. The two coefficients have different critical points (Section D.3), so the prefactor $|a_{\text{sat}}|/|a_{\text{rate}}|$ varies across materials. It is a function of α and r , which the network receives through the material numbers $\mathbf{m}$ (α through Π_h), so the network can absorb it.

B.6 Why the error falls during incubation

Errors usually grow in an autoregressive rollout, but in our experiments they *shrink* over the first ten frames (Section F.2). The reason is that most Fourier modes decay during incubation, and they take their errors with them. On the few growing modes, an error grows at the same rate as the field, so its relative size stays fixed. The proposition below makes this precise for the linear stage. Let the true field obey $P_{n+1} = B P_n$ with B diagonalizable, $B = V \Lambda V^{-1}$, and split the eigen-coordinates

into an unstable set $\mathcal{U}$ ($|\lambda_i| > 1$) and a stable set $\mathcal{S}$ ($|\lambda_i| \leq \rho_s < 1$). Let a surrogate produce $\tilde{P}_{n+1} = B\tilde{P}_n + \delta_n$, where δ_n is the error it adds at step n .

Proposition 2 (error propagation in the linear stage). The error $e_n = \tilde{P}_n - P_n$ satisfies $e_n = B^n e_0 + \sum_{j < n} B^{n-1-j} \delta_j$. In eigen-coordinates $c = V^{-1}e$, the stable components of the carried-over error decay geometrically, $|c_i(B^n e_0)| \leq \rho_s^n |c_i(e_0)|$ for $i \in \mathcal{S}$, while the unstable components grow at exactly the rate of the corresponding components of the true field. Consequently, if the injected errors are small after the first step ($\|\delta_j\| \ll \|e_0\|$ for $j \geq 1$), the relative error $\|e_n\|/\|P_n\|$ decreases with n as long as the stable part of e_0 dominates the error and the field is dominated by its unstable part.

Proof. The recursion follows by induction. In eigen-coordinates B^n acts as Λ^n , so component i of $B^n e_0$ is $\lambda_i^n c_i(e_0)$. It shrinks as ρ_s^n on $\mathcal{S}$ and grows as $|\lambda_i|^n$ on $\mathcal{U}$, the same factor that multiplies component i of $P_n = B^n P_0$. The ratio of unstable error to unstable field is therefore constant, while the stable error decays and the stable part of the field decays too, but the field norm is increasingly carried by $\mathcal{U}$. $\square$

For Duhamel-NO, the hypotheses hold with the true rates during the linear stage, because its Fourier SSM is the measured B and its correction is small while the field is small. For the baselines the proposition does not apply directly, because their linear response is learned and is not exactly B . The same argument holds for them only to the extent that their own learned response damps the decaying modes.

B.7 The linear stage, an exact factorization and the truncation cap

The idea in words. At frame 1 the true field is, to a very good approximation, a linear function of the starting noise, with each Fourier mode of the noise is multiplied by its own growth factor. A model can match this only as well as its own linear response matches these factors. An FNO applies one and the same response to every mode above its cut-off, while the true factors vary a lot there. This puts a ceiling on the correlation any FNO can reach at frame 1, however it is trained. The statements below make this precise.

Definitions. Write the one-frame map as $S(u) = Au + R(u)$ with A a Fourier multiplier $g(k)$ (a $Z \times Z$ matrix per wavenumber) and $\|R(u)\| \leq C\|u\|^p$, $p \geq 2$. A field F with its spatial mean removed is written $\tilde{F}$, and $\rho(F, T) = \mathbb{E}\langle \tilde{F}, \tilde{T} \rangle / \sqrt{\mathbb{E}\|\tilde{F}\|^2 \mathbb{E}\|\tilde{T}\|^2}$ is the population pattern correlation. Let $J_1 F$ be the orthogonal projection of the random field F onto the linear span of the starting-noise components, i.e. the part of F that is linear in the noise (for Gaussian noise, the first Wiener chaos [Nualart, 2006]; Gaussianity is not needed). Let $\mathcal{H}(F) = 1 - \mathbb{E}\|J_1 \tilde{F}\|^2 / \mathbb{E}\|\tilde{F}\|^2$ be the share of the output that is not linear in the noise, and $\rho_{\text{lin}}(F, T) = \rho(J_1 F, T)$.

Proposition 3 (exact factorization). Let $T = A\epsilon\xi$ be the true first frame from noise $\epsilon\xi$. For every surrogate output $G = G(\epsilon\xi)$ with finite second moment, $\rho(G, T) = \rho_{\text{lin}}(G, T)\sqrt{1 - \mathcal{H}(G)} \leq \sqrt{1 - \mathcal{H}(G)}$.

Proof. $\tilde{T}$ lies in the linear span of the noise components and $\tilde{G} - J_1 \tilde{G}$ is orthogonal to that span, so $\mathbb{E}\langle \tilde{G}, \tilde{T} \rangle = \mathbb{E}\langle J_1 \tilde{G}, \tilde{T} \rangle$ and $\mathbb{E}\|\tilde{G}\|^2 = \mathbb{E}\|J_1 \tilde{G}\|^2 + \mathbb{E}\|\tilde{G} - J_1 \tilde{G}\|^2$. Hence $\rho = \rho_{\text{lin}} \sqrt{\mathbb{E}\|J_1 \tilde{G}\|^2 / \mathbb{E}\|\tilde{G}\|^2} = \rho_{\text{lin}} \sqrt{1 - \mathcal{H}}$, and $|\rho_{\text{lin}}| \leq 1$ by Cauchy–Schwarz. $\square$

Proposition 4 (three behaviours as the noise shrinks). (a) If G is Lipschitz with a non-zero, spatially varying output at zero input (biases with positional channels), then $\mathcal{H}(G(\epsilon\xi)) \rightarrow 1$ and $\rho \rightarrow 0$ as $\epsilon \rightarrow 0$. (b) If $G(u) = r(u)\Phi(u/r(u), \log r(u))$ with $r = \text{rms}(u)$, then $\mathcal{H}$ depends on ϵ only through the explicit $\log r$ input. (c) If $G(u) = \hat{A}u + N(u)$ with $\|N(u)\| \leq K\|u\|^2$, then $\mathcal{H} = O(\epsilon^2)$ and $\rho \rightarrow \rho_{\text{lin}}(\hat{A}, A)$.

Proof. (a) Let L be the Lipschitz constant, $b = \tilde{G}(0) \neq 0$ and $c = L\epsilon(\mathbb{E}\|\xi\|^2)^{1/2}$. Then $\mathbb{E}\|\tilde{G}(\epsilon\xi) - b\|^2 \leq c^2$, so the mean part satisfies $\|J_0 \tilde{G}\| \geq \|b\| - c$ and $\mathbb{E}\|\tilde{G}\|^2 \leq (\|b\| + c)^2$, giving $\mathcal{H} \geq (\|b\| - c)^2 / (\|b\| + c)^2 \rightarrow 1$; and $\mathbb{E}\|J_1 \tilde{G}\|^2 \leq c^2$ bounds $\rho \leq c / (\|b\| - c) \rightarrow 0$. (b) $\mathcal{H}$ is invariant under deterministic rescaling, so $\mathcal{H}(G(\epsilon\xi)) = \mathcal{H}(r(\xi)\Phi(\xi/r(\xi), \log \epsilon + \log r(\xi)))$. (c) $\hat{A}\epsilon\xi$ is

linear in the noise, so the non-linear share comes only from N , with $\mathbb{E}\|N\|^2 \leq K^2\epsilon^4\mathbb{E}\|\xi\|^4$, while $\mathbb{E}\|\tilde{G}\|^2 \geq (\epsilon\|\hat{A}\tilde{\xi}\| - K\epsilon^2(\mathbb{E}\|\xi\|^4)^{1/2})^2$. Hence $\mathcal{H} = O(\epsilon^2)$, and the linear part of N is $O(\epsilon^2)$, so $\rho_{\text{lin}}(G) \rightarrow \rho_{\text{lin}}(\hat{A})$. $\square$

Proposition 5 (FNO linearization). For an FNO without positional channels, the Jacobian at any spatially homogeneous state is a Fourier multiplier $M(k)$ that equals one fixed matrix M_∞ for every k outside the retained mode set $\mathcal{K}$.

Proof. At a homogeneous state every pre-activation is spatially constant, so layer l has Jacobian $D_l(W_l + R_l(k)\mathbb{1}[k \in \mathcal{K}])$ with $D_l = \text{diag}(\phi'(\cdot))$. Products of multipliers are multipliers, and outside $\mathcal{K}$ only the k -independent W_l terms remain. With coordinate channels the action above the cut-off is multiplication by a fixed smooth field, so the statement holds up to the low-wavenumber content of that field. $\square$

Proposition 6 (truncation cap). If the linear response of G is constant outside $\mathcal{K}$, then for white noise $\rho_{\text{lin}}(G, A^n) \leq \rho_{\mathcal{K}}(n) = (1 - |\mathcal{K}^c| \text{Var}_{k \in \mathcal{K}^c}[g_n(k)] / \sum_{k \neq 0} |g_n(k)|^2)^{1/2}$, and $\rho(G, T_n) \leq \rho_{\mathcal{K}}(n) \sqrt{1 - \mathcal{H}(G)}$.

Proof. For translation-equivariant G and white noise, $J_1 G = \mathcal{F}^{-1}[m(k)\mathcal{F}\xi]$ (linear regression on white noise is a convolution), so ρ_{lin} is the spectral cosine between m and g_n . Maximizing it over m with $m \equiv c$ on $\mathcal{K}^c$ gives $m = g_n$ on $\mathcal{K}$ and c equal to the mean of g_n on $\mathcal{K}^c$; substituting gives $\rho_{\mathcal{K}}(n)$. The second bound follows from Proposition 3. $\square$

What this means for our models. By Proposition 5, an FNO’s linear response is constant outside its mode set, so Proposition 6 applies to it. This is the cap quoted in Section 4. We compute it from the measured Fourier SSM of the D1 held-out materials, for the FNO used here (modes up to 8 in each in-plane direction). The cap is 0.853, 0.882 and 0.926 at frames 1–3 and 0.996 at frame 10. The trained FNO reaches 0.826, 0.877 and 0.923 (Figure 3). Two caveats apply. For rms-normalized networks (class (b)) the cap is an observation, not a proven bound. And it does not apply to architectures with a different mode set (F-FNO) or without spectral truncation (U-Nets, CNO). Duhamel-NO is of class (c) in its linear part. Its Fourier SSM acts on every mode, so no cap applies, and it reaches 0.997 at frame 1.

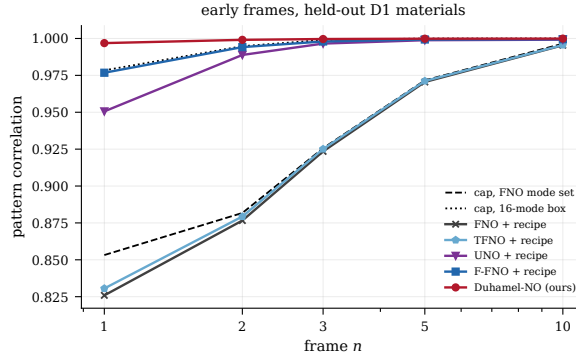


Figure 3: **Truncation cap against trained models** (D1 held-out materials, frames 1–10). Dashed and dotted lines show the best correlation any model with the FNO’s mode set, or with a 16-mode box, can reach (Proposition 6), computed from the measured Fourier SSM. Markers show trained models.

C Algorithms

Algorithm 1 rolls out Duhamel-NO (Section 3) from the initial noise (FFT and IFFT denote the fast Fourier transform and its inverse). The other three procedures measure the Fourier SSM of a training material (Algorithm 2), build it for a new material (Algorithm 3), and train the network (Algorithm 4).

Algorithm 1 Rollout of Duhamel-NO from initial noise

Require: initial noise P^0 ; coefficients (α, β, γ) ; number of frames T ; trained N_ψ

- 1: $P_S, \lambda, r \leftarrow$ closed-form material units (Section D.3); $\theta \leftarrow \Delta t_{\text{frame}}/\lambda$
- 2: $\mathbf{m} \leftarrow (\log \theta, \log P_S, r, \log \Pi_h)$ $\triangleright \Pi_h = \Delta x/\mu$ is the grid spacing in wall widths
- 3: for every in-plane mode k : $E(k) \leftarrow e^{\theta L(k)}$, $\Phi(k) \leftarrow \theta \varphi_1(\theta L(k))$ $\triangleright L(k)$ from Algorithm 3
- 4: $v \leftarrow P^0/P_S$
- 5: **for** $n = 0, \dots, T-1$ **do**
- 6: $s \leftarrow \text{rms}(v)$; $h \leftarrow s \cdot N_\psi(v/s, \log_{10} s, \mathbf{m})$ $\triangleright$ nonlinear correction
- 7: $\hat{v} \leftarrow \text{FFT}(v)$; $\hat{h} \leftarrow \text{FFT}(h)$ $\triangleright$ in-plane transform of the ten layers
- 8: $\hat{v}(k) \leftarrow E(k) \hat{v}(k) + \Phi(k) \hat{h}(k)$ for every mode k $\triangleright$ Equation (5)
- 9: $v \leftarrow \text{bounded IFFT}(\hat{v})$; $P^{n+1} \leftarrow P_S v$
- 10: **return** $P^1, \dots, P^T$

Algorithm 2 Measuring the Fourier SSM of one training material

Require: frames $P^0, \dots, P^8$ of one training simulation (field still small); frame interval θ (in growth times)

- 1: $\hat{P}^t \leftarrow$ in-plane FFT of P^t for $t = 0, \dots, 8$ $\triangleright \hat{P}^t(k) \in \mathbb{C}^{10}$, one vector per mode
- 2: **for** each radial bin b of $|k|$ (46 bins) **do**
- 3: $X \leftarrow [\hat{P}^t(k)]_{k \in b, t < 8}$, $Y \leftarrow [\hat{P}^{t+1}(k)]_{k \in b, t < 8}$ $\triangleright$ snapshot pairs
- 4: $B_b \leftarrow Y X^H (X X^H + \epsilon I)^{-1}$ $\triangleright$ least squares (small ridge), 10×10
- 5: $L_b \leftarrow \theta^{-1} \log B_b$ $\triangleright$ generator, symmetrized
- 6: bins whose modes decay below round-off within one frame get a learnable, strongly damped rate; the uniform mode is damped
- 7: **return** $\{L_b\}$ and $\log |\alpha|$ of this material

Algorithm 3 The Fourier SSM of a new material

Require: α of the new material; generators $\{L_b^{(i)}\}$ of all training materials i

- 1: $\mathcal{I} \leftarrow$ the eight training materials with $\log |\alpha_i|$ closest to $\log |\alpha|$
- 2: $L_b \leftarrow \frac{1}{8} \sum_{i \in \mathcal{I}} L_b^{(i)}$ for every bin b $\triangleright \beta, \gamma$ are not needed, since they do not enter $\mathcal{L}$
- 3: **return** $L(k) = L_{b(|k|)}$, and $e^{\theta L(k)}$, $\theta \varphi_1(\theta L(k))$ by `matrix_exp`

Algorithm 4 Training Duhamel-NO

Require: training trajectories (682 materials, one trajectory each); Fourier SSMs from Algorithm 2

- 1: initialize N_ψ with a zero read-out $\triangleright$ the untrained model is the exact linear solution
- 2: **for** 20 000 steps **do**
- 3: sample 16 windows of $k+1 = 5$ frames; with probability 0.6 the window starts in frames 0–120 (growth phase)
- 4: apply a random exact symmetry (90° rotation, reflection or $P_z \rightarrow -P_z$)
- 5: roll out $k = 4$ steps with Algorithm 1 from the first frame of the window (gradients flow through all steps)
- 6: $\mathcal{L} \leftarrow \frac{1}{k} \sum_{q=1}^k \|\tilde{P}^q - P^q\|^2 / \|P^q\|^2$ $\triangleright$ relative L_2
- 7: AdamW (Adam with decoupled weight decay) step (one-cycle schedule, weight decay 10^{-4} , gradient clipping at 1)

D Datasets and material units

D.1 Simulations

All data are FERROX [Kumar et al., 2023] simulations of a metal–ferroelectric–insulator–metal stack. A 5 nm ferroelectric film ($\varepsilon_{\text{fe}}=24$) sits on a 4 nm dielectric ($\varepsilon_{\text{de}}=10$) between shorted electrodes. The

domain is periodic in-plane, with $G_{11}=10^{-9} \text{ V m}^3/\text{C}$ and $\Gamma=100 (\Omega \text{ m})^{-1}$. At every step the solver takes an explicit-Euler TDGL step and solves the Poisson equation for the whole stack.

Each run is saved for 401 frames. The frame interval is chosen per material so that the frames cover about $1.25 \times$ its time to saturation. Each run starts from a field drawn uniformly from $\pm 2 \times 10^{-3} \text{ C/m}^2$ in every cell. This is 0.5–2% of P_S for the D1 materials and 0.2–4% for D2. The field is set by an integer seed (D1 material i uses seed $2000 + i$, and D2 uses $6000 + i$). We would reject a trajectory with non-finite values or $|P_z| > 1.35P_S$; none was rejected.

Unseen noise. FERROX generates the same initial field for the same seed, whatever the material. A model can therefore memorize shared seeds. In a test with eight shared seeds, a model scored 0.99 on a seen seed but only 0.66–0.82 on an unseen one. We therefore give every trajectory its own seed. The largest correlation between a test and a training initial field is 0.023.

Table 3: **Datasets.** Every trajectory has its own noise seed; no seed is shared between training and test data. Hysteresis is described in Section H.

dataset	grid	frames	varied parameters	training	test
D1	$64^2 \times 10$	401	α, β, γ (Latin hypercube [McKay et al., 1979] in $\log(\alpha , P_S, r)$)	682 (+63 validation)	256 held-out materials (82 interior, 174 edge)
D2	$64^2 \times 10$	401	α, β, γ drawn independently, five regions	–	300 (5×60), no retraining
Hysteresis	$64^2 \times 10$	440–2765	α, β, γ , peak voltage	90 loops (+10 validation)	30 held-out material, 34 new noise (12 evaluated), 10 closed set

Table 3 summarizes the datasets, and Figure 4 shows where the D1 and D2 materials lie in $(|\alpha|, P_S, r)$.

D1. We train on 745 materials (682 for training and 63, every twelfth one, for validation) and hold out 256. Of these, 82 lie inside the sampled box and 174 at its edges, including every well with $r \geq 1.15$. We sample in $(|\alpha|, P_S, r)$ and solve for the coefficients, $\gamma = r|a_{\text{sat}}|/P_S^4$ and $\beta = (1 - r)|a_{\text{sat}}|/P_S^2$, so every sample is a valid double well. Sampling the raw coefficients instead would waste more than 95% of draws, because the valid range of β moves by two orders of magnitude across the range of α .

D2. 300 materials whose α, β, γ are drawn independently (log-uniform), 60 in each of five regions:

- **A** has small P_S .
- **B_{units}** has large P_S , up to $1.8 \times$ the training maximum. This is a pure change of units, because the dimensionless numbers stay in range.
- **B _{α}** has large $|\alpha|$ with small P_S .
- **C** has first-order wells with $r = 1.15$ –2. This changes a dimensionless number, so it is new physics.
- **D** combines all of these, with P_S up to $2.2 \times$ the training maximum.

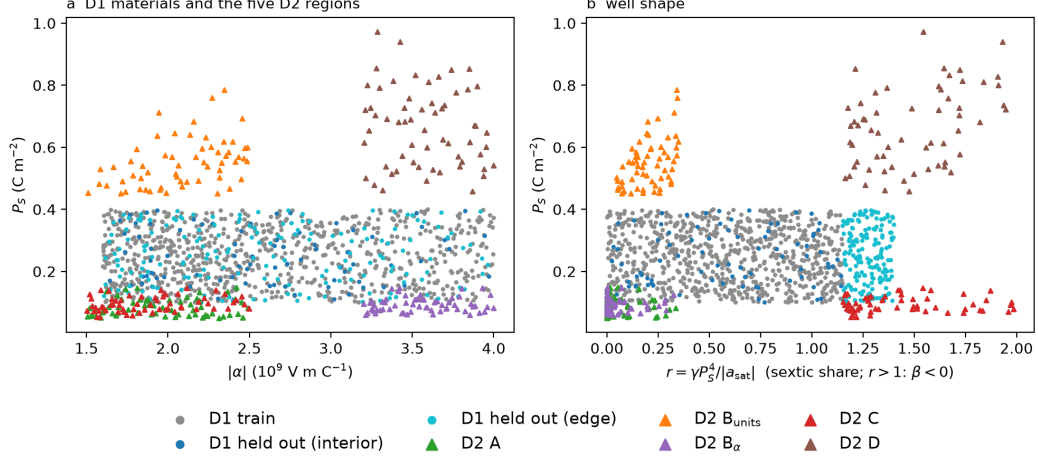


Figure 4: **Where the materials lie.** D1 materials (circles, grey for training and blue for held out) and D2 materials (triangles, coloured by region) in well depth $|\alpha|$, saturation polarization P_S and well shape r . D2 leaves the training range in P_S and in r .

D.2 Free-energy landscapes

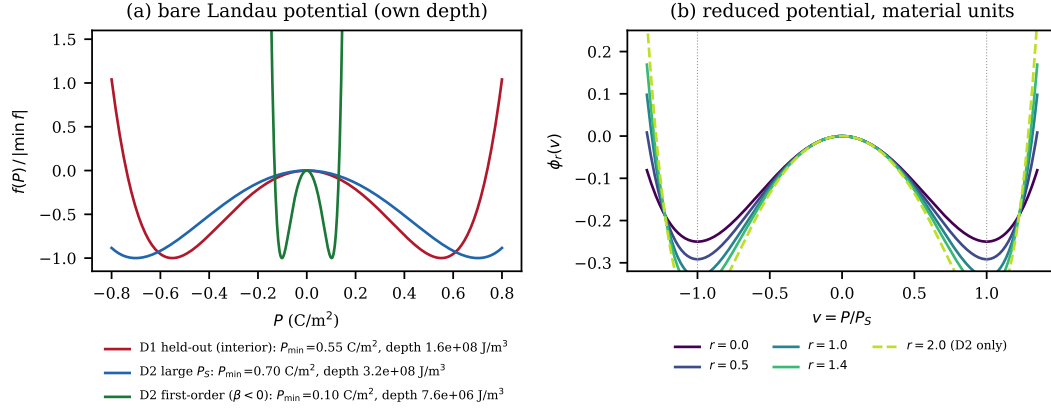


Figure 5: **Free-energy landscapes of the test materials.** (a) Bare Landau potential $f(P) = \alpha P^2/2 + \beta P^4/4 + \gamma P^6/6$ of the two materials of Figure 8 and the D2 large- P_S material, each divided by its own well depth. The wells differ in position by a factor of seven and in depth by a factor of about forty, and the first-order material ($\beta < 0$) has a narrow, steep well. (b) In material units every material has the reduced potential $\phi_r(v)$ of Section B.5, with minima at $v = \pm 1$; only its shape, set by r , differs. D2 reaches $r = 2$, beyond the training range (dashed).

D.3 Material units

The material units are a time unit λ (the linear growth time) and an amplitude unit P_S . Both come from α , but corrected for the depolarization field of the dielectric, which weakens the effective well. This correction differs for the two things α controls, the growth of small fluctuations and the saturated amplitude, so we use two coefficients, both fitted once to simulations:

- a growth coefficient $|a_{\text{rate}}| \propto |\alpha| - 1.37 \times 10^9$, which sets λ ;
- a saturation coefficient $|a_{\text{sat}}| = 1.389 (|\alpha| - 1.31 \times 10^9 (1 - 0.0775 r))$ (units of α , V m/C). P_S then solves $|a_{\text{sat}}| = \beta P_S^2 + \gamma P_S^4$, and $r = \gamma P_S^4 / |a_{\text{sat}}|$.

This law predicts P_S of the D1 materials to 1.03% on average (3.41% at worst). For D2, whose P_S lies up to $2.2\times$ the fit maximum, it is accurate to 1.9% (Figure 6). The value 1.37×10^9 is a

critical point. Below it, the depolarization field outweighs the Landau driving force and the film never nucleates. All D1 and D2 materials lie above it.

The units alone, before any training, make materials look much more alike. Over 120 D1 materials, they reduce the coefficient of variation of the settled amplitude from 0.38 to 0.083, and that of the nucleation time from 0.74 to 0.054.

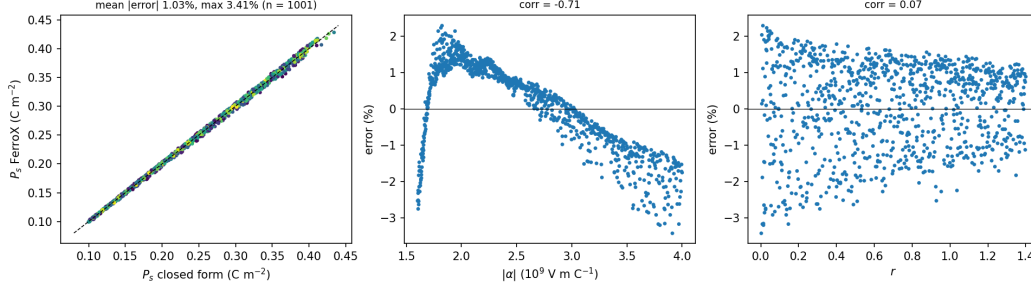


Figure 6: **Closed-form saturation polarization** against the value measured in FERROX (99.5% quantile of $|P|$ over the settled frames) for all D1 materials, coloured by r , with residuals against $|\alpha|$ and r .

E Training and evaluation details

The network. N_ψ is a U-shaped neural operator (UNO) [Rahman et al., 2023], a Fourier neural operator arranged like a U-Net. Each block multiplies the lowest Fourier modes of its input by learned weights, adds a pointwise linear map, applies a Gaussian error linear unit (GELU), and resamples the field to a coarser or finer grid. The ten depth layers are treated as channels.

The input has four parts. They are the field divided by its root-mean-square (rms) amplitude s , a channel holding $\log_{10} s$ (so the network knows how far the field is from saturation), a 16-dimensional embedding of the material numbers (a bias-free two-layer network), and two coordinate channels. A pointwise layer lifts these to 32 channels. Two down-sampling blocks reduce the grid from 64^2 to 32^2 and 16^2 (64 and 128 channels), a middle block works at 16^2 , and two up-sampling blocks return to 64^2 with skip connections. A final block and a two-layer pointwise head give ten output channels, which are multiplied back by s .

The U-shape lets the correction depend on both the local structure of the walls and the larger domain pattern. The network keeps only the lowest Fourier modes (up to 16 per direction). This costs nothing during incubation, because the Fourier SSM in Equation (5) handles every mode.

Why a multi-step loss. Single-step training would let errors compound at inference, so Equation (8) unrolls k steps. Unlike the pushforward trick [Brandstetter et al., 2022], which stops gradients through the model’s own inputs, we backpropagate through all k steps.

Training. Every model uses the same settings, with 20 000 optimizer steps, batch 16, $k = 4$ unrolled steps, AdamW [Loshchilov and Hutter, 2019] (weight decay 10^{-4}) with a one-cycle schedule [Smith and Topin, 2019] (5% warm-up), gradient clipping at 1, 32-bit floating point (fp32), and exact symmetry augmentation. Besides the seven baselines, we trained a transformer, Transolver [Wu et al., 2024], and three generative models on a U-Net backbone. The generative models are trained on their own objectives (one-step denoising, or flow matching on the residual). They are sampled with 20 denoising diffusion implicit model (DDIM) [Song et al., 2021] steps for the autoregressive conditional diffusion model (ACDM) [Kohl et al., 2023], 8 Euler steps for flow matching [Lipman et al., 2023], and 3 refinement steps for PDE-Refiner [Lippe et al., 2023]. Poseidon-T is fine-tuned from its public pretrained weights for 20 000 steps.

Without our training setup. Every baseline is trained with the setup of Section 3.4, which uses material units, an amplitude-aware input, the loss of Equation (8) and symmetry augmentation. In the standard setup instead, a baseline receives the raw material parameters (standardized) as constant input channels, plus periodic coordinate channels. Table 4 shows that every backbone needs the setup.

With standard training, no backbone exceeds a correlation of about 0.84, at a relative error of 1.0–2.9, which is no better than predicting zero.

Table 4: **Standard training against our training setup**, same backbone, same (4000-step) tuning budget, FerroX validation split (new materials, new noise). The best standard configuration is a relative- L_2 loss with per-sample normalization for every backbone except Transolver* (relative- L_2 loss alone).

backbone	best standard configuration		with our training setup	
	corr@400	rel. L2@400	corr@400	rel. L2@400
F-FNO	0.837	2.92	0.895	0.41
U-Net	0.841	2.04	0.909	0.37
UNO	0.842	1.87	0.863	0.47
CNextU-Net	0.742	1.52	0.853	0.49
CNO	0.622	1.00	0.907	0.38
TFNO	0.678	2.18	0.863	0.50
FNO	0.819	1.79	0.849	0.52
Transolver*	0.044	1.76	0.066	1.30

Selection. For each model we choose the learning rate and configuration on the validation split, by correlation minus 0.1 times the relative error (for hysteresis, the loop error). Table 5 lists the choice for every model and task. Variants of U-Net, CNextU-Net and CNO without coordinate channels, and CNO with batch normalization (BatchNorm), are included as candidates. The chosen configuration is retrained with three seeds (one for the slowest models).

Metrics. The pattern correlation of a single frame is $\text{corr}(A, B) = \sum_i (A_i - \bar{A})(B_i - \bar{B}) / [\sum_i (A_i - \bar{A})^2 \sum_i (B_i - \bar{B})^2]^{1/2}$, with the sum over voxels. Every prediction is rolled out from the $t=0$ field, with no solver frame.

Evaluation choices. Predictions are clamped to three times the largest training value of the field. The clamp is scaled per sample by the amplitude unit, so out-of-range materials are not cut off. The neuralop FNO with `n_modes=(16, 16)` keeps modes up to 8 in each direction, and the truncation cap (Proposition 6) uses this true mode set. Timing uses one A100 with nothing else running.

Reproducibility. Every number in the paper and in this appendix is written as a $\LaTeX$ macro or a generated table from the benchmark result files (`make_duhamel_macros.py`, `make_physics_table.py`, `appendix/copy_assets.py`); none is typed by hand. The rollout figures are produced from the evaluated checkpoints by `make_rollout_fig.py` and `appendix/gallery_plot.py`.

Table 5: **Validation-selected configuration, size, number of seeds and training time** (one A100) of every model on every task. MSE is the mean squared error loss.

task	model	params (M)	seeds	train (min)	lr	configuration
FerroX TDGL	Duhamel-NO (ours)	9.49	3	39	0.0003	our setup
	F-FNO	1.19	3	42	0.001	our setup
	U-Net	10.95	1	76	0.0003	our setup
	UNO	9.49	3	27	0.001	our setup
	CNextU-Net	9.93	3	59	0.0003	our setup
	CNO	5.50	3	93	0.001	our setup
	TFNO	1.03	3	27	0.001	our setup
	FNO	2.42	3	27	0.001	our setup
	Transolver	2.19	1	161	0.001	our setup
	Poseidon-T	20.79	1	129	0.0001	rel. L2
	U-Net + ACDM diffusion	10.94	1	16	0.001	global noise scale
	U-Net + flow matching	10.94	1	23	0.001	global noise scale
	U-Net + PDE-Refiner	10.94	1	25	0.001	global noise scale
hysteresis	Duhamel-NO (ours)	9.49	3	81	0.001	our setup
	F-FNO	1.19	1	53	0.001	our setup
	U-Net	10.94	1	52	0.0003	rel. L2
	UNO	9.49	1	22	0.001	our setup
	CNextU-Net	9.93	3	43	0.0003	our setup
	CNO	5.50	3	74	0.0003	our setup
	TFNO	1.03	1	24	0.001	MSE
	FNO	2.42	1	24	0.001	our setup

F Additional results on ferroelectric domain formation

Table 6 lists every model, with standard deviations over seeds. Table 7 and Figure 7 break the results down by test subset. Duhamel-NO is the most accurate model on every subset, inside and outside the training range. Its three training seeds agree to within ± 0.001 in correlation. Its error is 0.120 on held-out materials inside the sampled box and 0.133 on those at its edges, so it does not lose accuracy at the edges.

Table 6: **All models**, the full version of Table 1 (mean $\pm$ std over seeds; **best**). Pattern correlation and relative L_2 error at frame 400 on D1 (held-out materials) and D2 (out of range), and relative error at frame 1 (D1).

model	family	seeds	D1 held-out (256)		D2 out-of-range (300)		frame 1
			corr@400	rel. L2@400	corr@400	rel. L2@400	rel. L2
Duhamel-NO (ours)	ours	3	0.982$\pm$0.001	0.129$\pm$0.003	0.978$\pm$0.001	0.145$\pm$0.004	0.082$\pm$0.000
F-FNO	spectral NO	3	0.974 $\pm$ 0.001	0.167 $\pm$ 0.003	0.968 $\pm$ 0.000	0.195 $\pm$ 0.004	0.215 $\pm$ 0.001
U-Net	U-Net	1	0.966	0.207	0.954	0.242	0.313
UNO	spectral NO	3	0.962 $\pm$ 0.003	0.219 $\pm$ 0.012	0.959 $\pm$ 0.002	0.238 $\pm$ 0.008	0.315 $\pm$ 0.005
CNextU-Net	U-Net	3	0.959 $\pm$ 0.000	0.229 $\pm$ 0.002	0.949 $\pm$ 0.002	0.261 $\pm$ 0.010	0.347 $\pm$ 0.018
CNO	conv. NO	3	0.951 $\pm$ 0.004	0.256 $\pm$ 0.011	0.937 $\pm$ 0.004	0.292 $\pm$ 0.005	0.451 $\pm$ 0.003
TFNO	spectral NO	3	0.949 $\pm$ 0.005	0.272 $\pm$ 0.017	0.952 $\pm$ 0.003	0.265 $\pm$ 0.012	0.564 $\pm$ 0.004
FNO	spectral NO	3	0.947 $\pm$ 0.004	0.270 $\pm$ 0.015	0.947 $\pm$ 0.003	0.268 $\pm$ 0.012	0.568 $\pm$ 0.003
Poseidon-T	foundation	1	0.553	2.917	0.414	2.570	4.342
U-Net + ACDM diffusion	diffusion	1	0.206	1.293	0.092	2.854	5.209
U-Net + flow matching	flow matching	1	0.169	1.330	0.094	1.895	5.210
U-Net + PDE-Refiner	diffusion	1	0.053	3.578	0.033	2.158	5.351
Transolver	transformer	1	0.003	1.329	0.008	1.313	0.757

Table 7: **Correlation at frame 400 per test subset** (mean $\pm$ std over seeds). D1 is split into interpolation (inside the sampled box) and extrapolation (at its edges). The D2 regions are those of Section D.

model	D1 extrap.	D1 interp.	D2-A	D2-B _{α}	D2-B _{units}	D2-C	D2-D
Duhamel-NO (ours)	0.981 $\pm$ 0.001	0.982 $\pm$ 0.001	0.974 $\pm$ 0.001	0.977 $\pm$ 0.002	0.985 $\pm$ 0.002	0.974 $\pm$ 0.001	0.978 $\pm$ 0.003
F-FNO	0.975 $\pm$ 0.001	0.973 $\pm$ 0.001	0.967 $\pm$ 0.002	0.958 $\pm$ 0.002	0.982 $\pm$ 0.001	0.970 $\pm$ 0.003	0.961 $\pm$ 0.001
U-Net	0.963	0.972	0.967	0.970	0.979	0.941	0.915
UNO	0.959 $\pm$ 0.004	0.970 $\pm$ 0.002	0.958 $\pm$ 0.010	0.949 $\pm$ 0.003	0.976 $\pm$ 0.004	0.951 $\pm$ 0.008	0.960 $\pm$ 0.003
CNextU-Net	0.958 $\pm$ 0.000	0.962 $\pm$ 0.001	0.951 $\pm$ 0.008	0.957 $\pm$ 0.004	0.969 $\pm$ 0.002	0.941 $\pm$ 0.003	0.926 $\pm$ 0.016
CNO	0.949 $\pm$ 0.004	0.955 $\pm$ 0.007	0.917 $\pm$ 0.008	0.952 $\pm$ 0.011	0.959 $\pm$ 0.004	0.923 $\pm$ 0.018	0.934 $\pm$ 0.009
TFNO	0.946 $\pm$ 0.005	0.954 $\pm$ 0.006	0.958 $\pm$ 0.004	0.942 $\pm$ 0.007	0.973 $\pm$ 0.006	0.954 $\pm$ 0.004	0.932 $\pm$ 0.007
FNO	0.944 $\pm$ 0.006	0.954 $\pm$ 0.003	0.952 $\pm$ 0.002	0.927 $\pm$ 0.004	0.965 $\pm$ 0.002	0.949 $\pm$ 0.005	0.941 $\pm$ 0.014

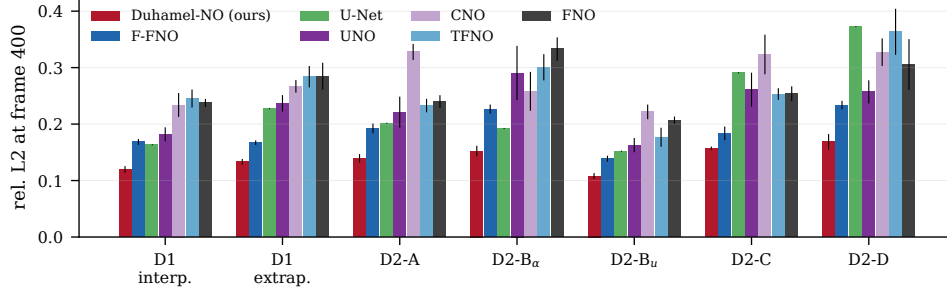


Figure 7: **Relative error at frame 400 per test subset** (bars show the standard deviation over seeds). Duhamel-NO is the most accurate model on every subset, including those outside the training range.

F.1 Rollouts on two test materials

Figure 8 adds a first-order well from D2 to the held-out material of Figure 2. Its well has r beyond the training range, and so a different free-energy landscape (Figure 5).

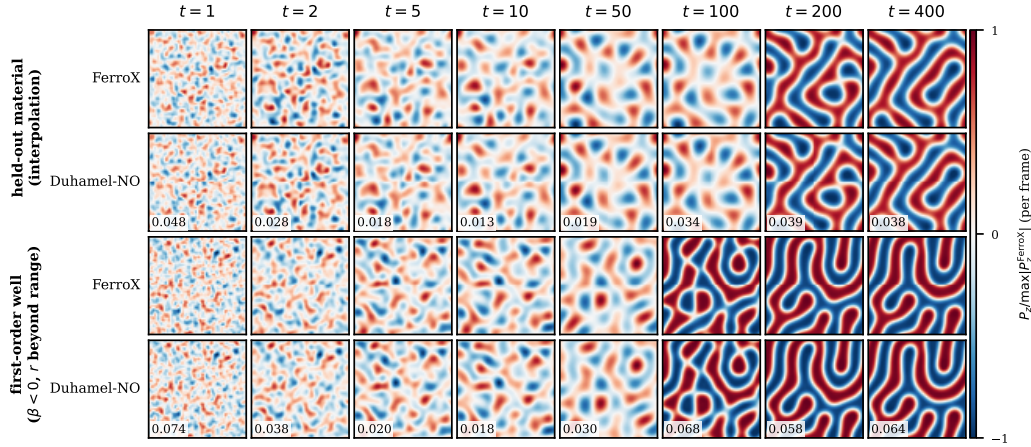


Figure 8: **Duhamel-NO rolled out from unseen noise on two test materials**, against FERROX (top layer; each frame is scaled by the maximum of the true field, and numbers are relative errors). Each material is the median-error trajectory of its group (D1 interior; D2 first-order wells); mean errors are higher because of a tail of harder materials (Table 1).

F.2 Error at every frame

Figure 9a,b shows the error at every frame. For every model, the error falls over the first ten frames and only then rises. This is surprising, because errors in an autoregressive rollout usually compound. For Duhamel-NO the reason is exact (Proposition 2). Its Fourier SSM is the true linear map, so during

incubation an error on a decaying mode decays with it, and on the few growing modes an error grows only as fast as the field, so its relative size stays constant. For the baselines the same explanation is only a heuristic, because their linear response is learned. It is their own learned dynamics, not the true ones, that damp their errors, and they do so to the extent that they approximate the true map. After nucleation this damping is lost, and errors slowly accumulate. UNO, the same network without the Duhamel step, has error 0.315 at frame 1, against 0.082 for Duhamel-NO.

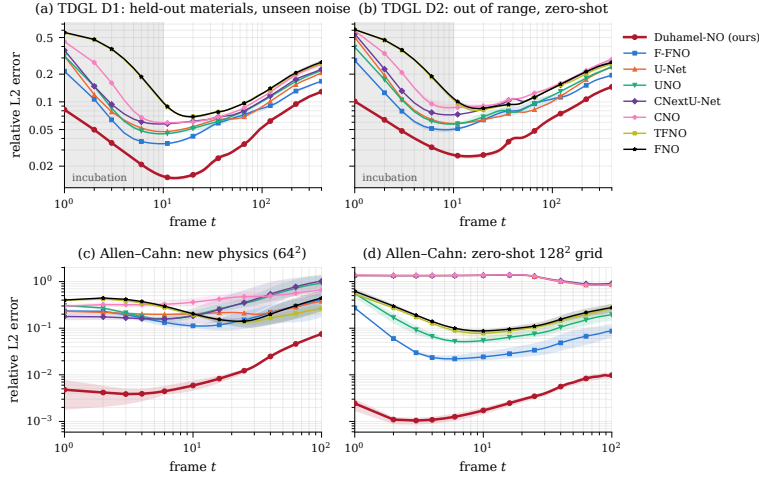


Figure 9: **Relative error at every frame.** (a,b) TDGL, D1 held-out and D2. (c,d) Allen-Cahn, new physics and zero-shot 128^2 grid (trained on 64^2 ; Section I). Lines are means over trajectories, and bands show the range over seeds.

F.3 Physical quantities

Table 8 is the full version of Table 2, with every baseline.

Table 8: **Errors in physical quantities, all models**, D1 held-out materials (mean over trajectories and seeds; **best**, second, not counting Poseidon-T). Physical quantities were computed for one seed of U-Net, CNextU-Net, TFNO and Poseidon-T, although Table 6 reports field errors over three seeds for CNextU-Net and TFNO. t_{nuc} is in frames, $F(t)$ is relative to the total energy drop, and the others are relative errors at frame 400.

model	t_{nuc}	P_S	domain size	wall density	P/P_S hist.	φ	E_z	$F(t)$
Duhamel-NO (ours)	0.28	0.003	0.005	0.007	0.010	0.139	0.135	0.004
F-FNO	0.70	0.006	0.005	<u>0.008</u>	0.010	<u>0.165</u>	<u>0.164</u>	0.007
U-Net	0.61	0.008	0.005	<u>0.008</u>	0.011	0.201	0.200	<u>0.006</u>
UNO	0.54	0.024	<u>0.006</u>	<u>0.009</u>	<u>0.020</u>	0.220	0.218	<u>0.012</u>
CNextU-Net	0.55	<u>0.005</u>	<u>0.006</u>	<u>0.008</u>	0.010	0.224	0.222	0.007
CNO	<u>0.46</u>	0.006	0.008	0.010	<u>0.011</u>	0.256	0.253	0.008
TFNO	<u>0.69</u>	0.011	0.008	0.011	<u>0.013</u>	0.263	0.260	0.008
FNO	0.68	0.012	0.008	0.011	0.014	0.270	0.267	0.009
Poseidon-T	57.7	2.136	0.246	0.665	1.301	3.035	2.979	1106

Figure 10 shows the physical metrics on D1 in more detail.

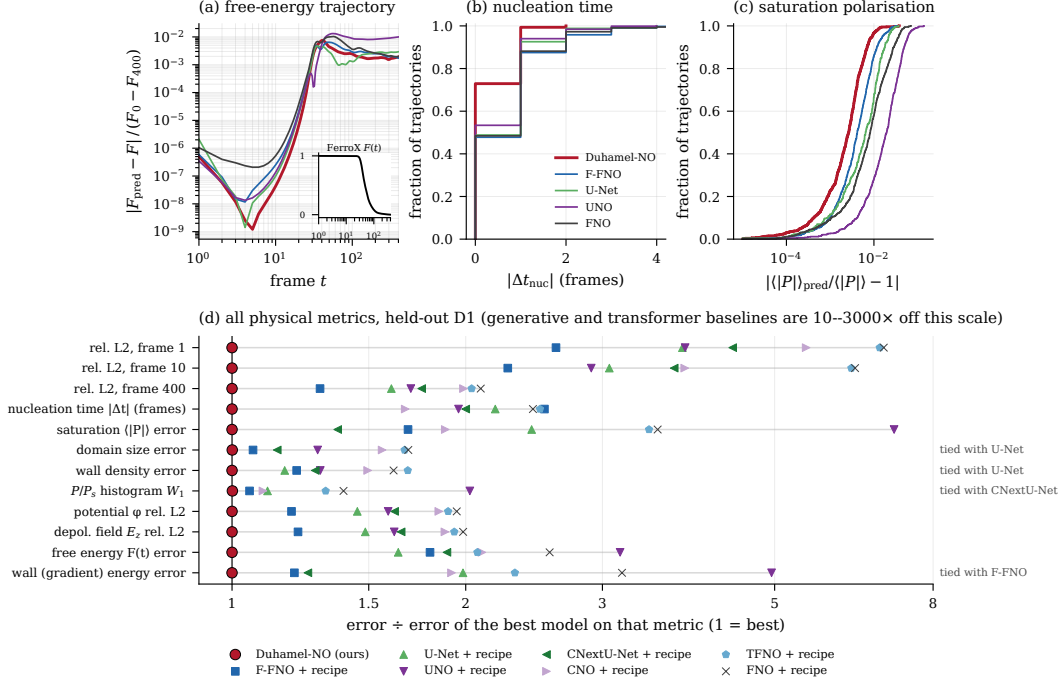


Figure 10: **Physical consistency on the D1 held-out materials.** (a) Error of the free-energy trajectory relative to the total energy drop (the inset shows the true $F(t)$). (b, c) Cumulative distributions of the nucleation-time error and of the saturation-polarization error (further left is better). (d) Every metric as the ratio of each model’s error to that of the best model. In legends, “+ recipe” denotes our training setup.

F.4 Error spectra

Figure 11 shows at which wavenumbers each model makes its errors. At frame 1, where the truncation cap (Proposition 6) binds, every baseline’s error is spread over all wavenumbers, so the cap is not the only source of error. By frame 10 the FNO’s error is concentrated above its mode cut-off, as its constant response there predicts, but by then the cap itself is no longer restrictive (0.996).

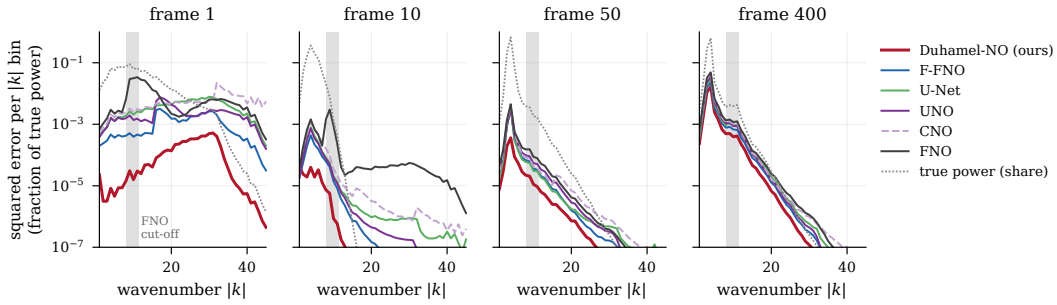


Figure 11: **Error spectra** on D1 held-out materials, as the squared error per radial wavenumber bin, as a fraction of the total true power, at four frames. The grey dotted line is the true power spectrum, and the shaded region is the FNO’s mode cut-off. At frame 1 every baseline’s error is spread over all wavenumbers; by frame 10 the FNO’s error is concentrated above its cut-off.

F.5 Ablations

Table 9 removes or replaces the parts of Duhamel-NO one at a time. Table 10 isolates the effect of the material units.

Table 9: **Ablations of Duhamel-NO** (D1 held-out; mean $\pm$ std over seeds), all with the same 9.5M-parameter UNO. Adding the Duhamel step with the measured Fourier SSM gives the gain. A learned, conditioned symbol, which cannot couple the depth layers, is worse than no Duhamel step. A strictly $O(v^2)$ correction keeps frame 1 accurate but underfits late coarsening.

FerroX TDGL (same 9.5M-parameter UNO backbone)	corr@400 (D1)	rel. L2@400 (D1)	rel. L2, frame 1
UNO, no Duhamel step	0.962 $\pm$ 0.003	0.219 $\pm$ 0.012	0.315 $\pm$ 0.005
+ Duhamel step, measured operator (Duhamel-NO)	0.982 $\pm$ 0.001	0.129 $\pm$ 0.003	0.082 $\pm$ 0.000
Duhamel-NO with an $O(v^2)$ remainder instead of UNO	0.934 $\pm$ 0.013	0.287 $\pm$ 0.042	0.085 $\pm$ 0.000
+ Duhamel step, learned conditioned symbol	0.950 $\pm$ 0.001	0.256 $\pm$ 0.003	0.399 $\pm$ 0.026
Duhamel-NO, uniform mode kept instead of damped	0.980	0.133	0.093
Duhamel, learned generator + small $O(v^2)$ network	0.812 $\pm$ 0.002	0.578 $\pm$ 0.003	0.219 $\pm$ 0.001

Table 10: **The change of units** (mean $\pm$ std over three seeds) for FNO and UNO with our training setup, with and without material units. B_{units} are D2 materials with P_S up to $1.8\times$ the training maximum. The units carry the amplitude out of range ($0.24 \rightarrow 0.02$ for FNO, $0.26 \rightarrow 0.02$ for UNO); they do not improve the pattern in range.

backbone	change of units	D1 rel. L_2	D1 amplitude err.	D2 B_{units} rel. L_2	D2 B_{units} amplitude err.
FNO	yes	0.270 $\pm$ 0.015	0.012 $\pm$ 0.004	0.207 $\pm$ 0.006	0.015 $\pm$ 0.005
FNO	no	0.245 $\pm$ 0.008	0.018 $\pm$ 0.000	0.335 $\pm$ 0.009	0.243 $\pm$ 0.007
UNO	yes	0.219 $\pm$ 0.012	0.025 $\pm$ 0.007	0.163 $\pm$ 0.013	0.019 $\pm$ 0.002
UNO	no	0.199 $\pm$ 0.004	0.049 $\pm$ 0.002	0.330 $\pm$ 0.006	0.258 $\pm$ 0.002

E.6 Cost and training data

Figure 12 relates accuracy to inference cost and to the number of training materials. On one dedicated A100 GPU, FERROX needs a median of 196 s per 400-frame D1 trajectory (range 147–318 s). Duhamel-NO needs 1.66 s for a single trajectory, about $118\times$ faster, and 0.108 s per trajectory in batches of 16, about $1814\times$ faster. This is about the speed of the neural baselines. Simulating the 682 training trajectories with FERROX takes about 37 GPU-hours, and training Duhamel-NO takes about 39 minutes (Table 5). Together they pay for themselves after about 700 predicted trajectories.

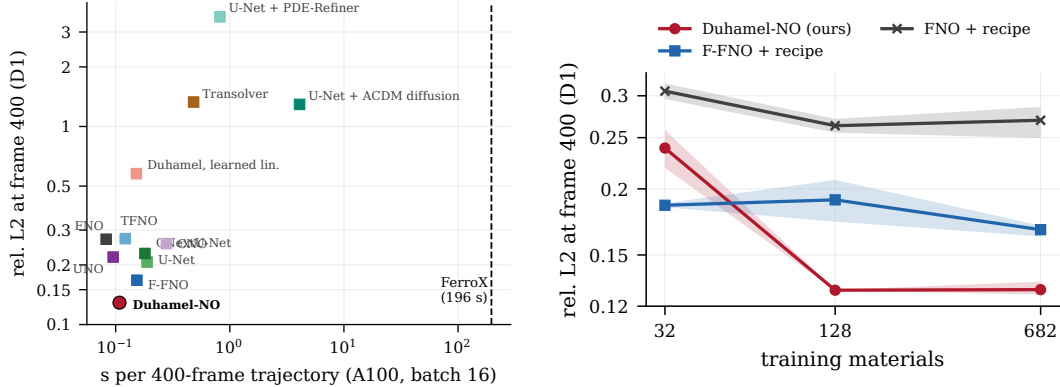


Figure 12: **Left:** error at frame 400 against wall-clock cost per 400-frame trajectory (A100, batch 16), with the FERROX cost. **Right:** error against the number of training materials (the Fourier SSM is rebuilt from the same subset). With 128 materials Duhamel-NO is unchanged (0.129); with 32 it degrades to 0.239, where F-FNO is better (0.186).

E.7 Rollouts of every model

Figures 13 and 14 show one D1 held-out and one D2 trajectory for every model, from the same unseen noise. Each is the trajectory with the median Duhamel-NO error among the saved examples.

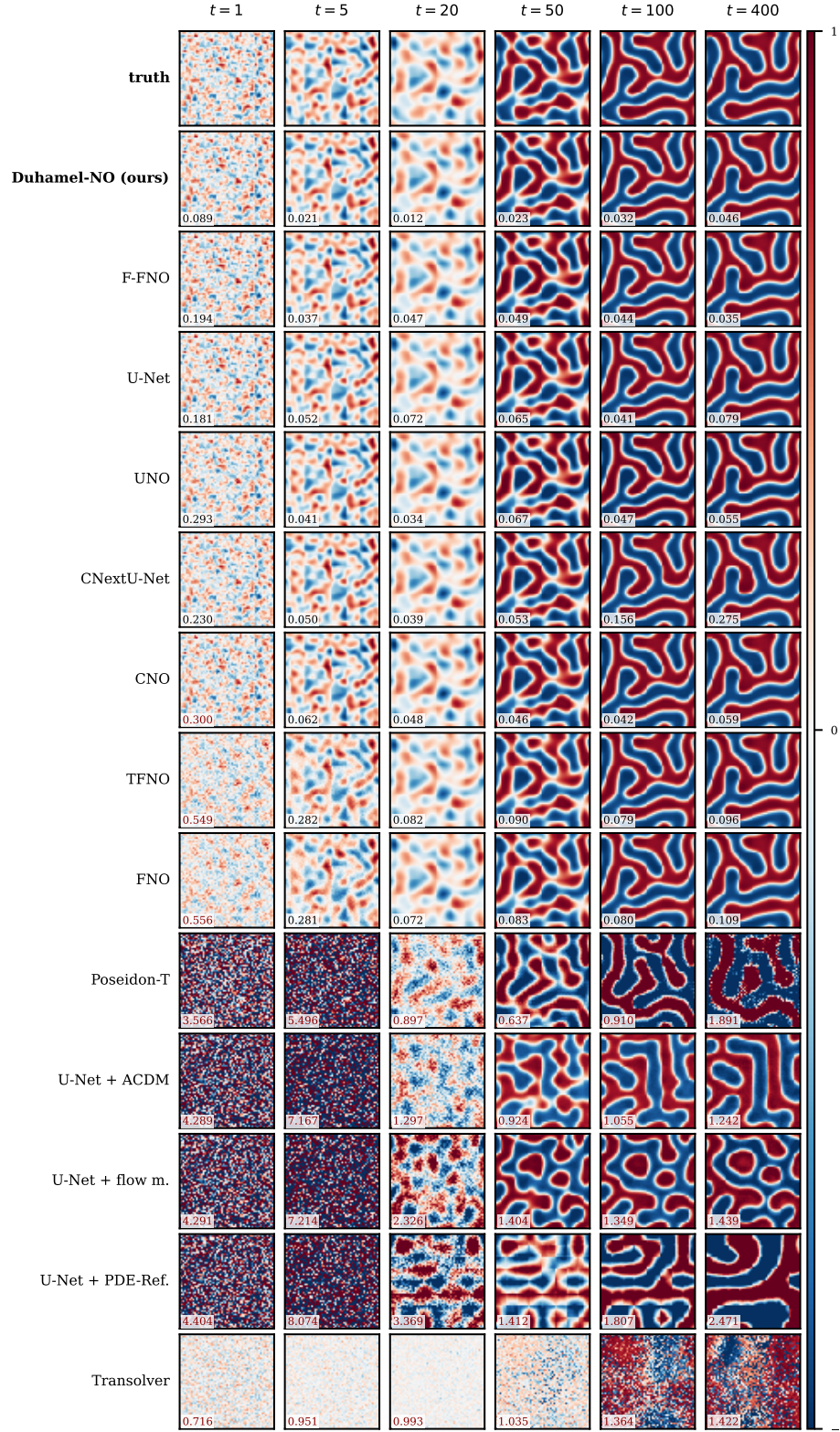


Figure 13: **D1 held-out material, every model** (top ferroelectric layer; field divided by the maximum of the truth at that frame; numbers are relative errors of the frame, red above 0.3).

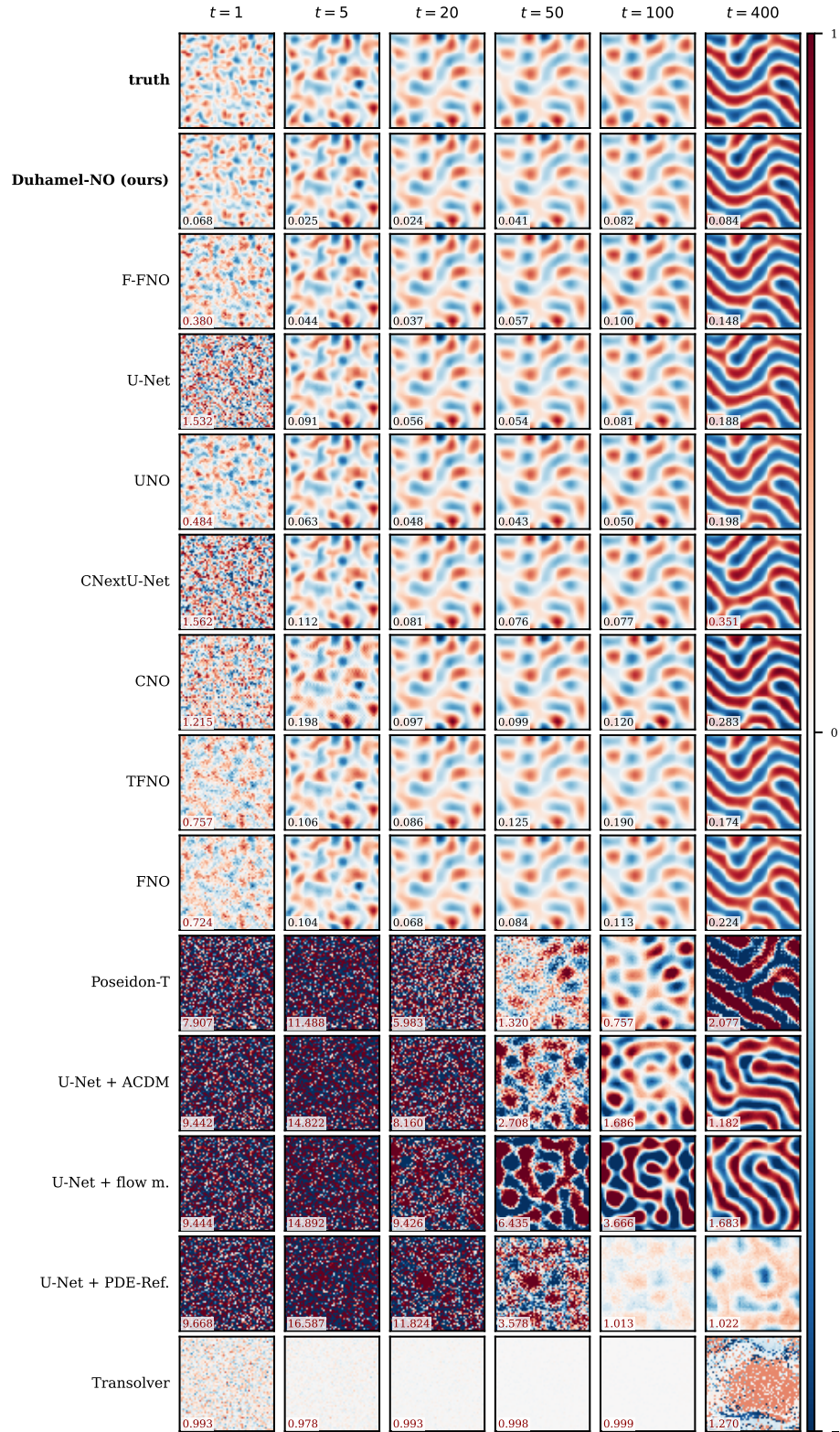


Figure 14: D2 material (out of range, no retraining), every model, as Figure 13.

G The released benchmark

We release the three ferroelectric datasets of Table 3 as an open benchmark under the CC BY 4.0 license. The data are hosted on Hugging Face ([doi:10.57967/hf/11169](https://doi.org/10.57967/hf/11169)), and the code repository links to them. Each dataset is a folder with one file per trajectory and a metadata table (CSV). The table has one row per trajectory, with the Landau coefficients (α, β, γ) , the derived coordinates (P_S, r) , the solver time step and frame interval, the noise seed, and the split used in this paper.

- **D1** (1001 trajectories) and **D2** (300) are arrays of shape $401 \times 10 \times 64 \times 64$ (frames, depth layers, in-plane grid), float32, in C/m^2 . D1 carries the splits train, validation, held-out interior and held-out edge; D2 carries its region (A, B_{units} , B_α , C, D).
- **Hysteresis** (174 loops) holds the field at every saved solver step, with the step index and the applied voltage, and the material, ε_{de} , V_{max} , seed and split (train, validation, new noise, held-out material, closed set).

The ferroelectric film occupies depth layers 8–17 of the simulated stack. The eight dielectric layers carry no polarization (they are exactly zero in every saved frame), so we omit them. A small Python module loads a trajectory and computes the pattern correlation and relative error of Section 2, so results on the benchmark are directly comparable with Table 1. The FERROX input files and job lists reproduce every trajectory from its seed.

H Voltage-driven hysteresis

Setup. We use the capacitor of Section D, but with $\varepsilon_{\text{de}} = 100$ (with 10 the loop is too narrow to resolve). Starting from the unpoled random field, we sweep the voltage slowly, $0 \rightarrow +V_{\text{max}} \rightarrow -V_{\text{max}} \rightarrow +V_{\text{max}}$. The top electrode potential is stepped in 0.125 V increments, and each step is held until the electrostatic solve converges. We use $V_{\text{max}} = 1.0, 1.5$ or 2.0 V and save a frame every 20 solver steps, so a loop takes 440–2765 frames. Every model is retrained on 90 closed loops. It then rolls out the whole sweep from the first frame, with the voltage as an input. The loop error is the mean absolute deviation of the polarization–voltage curve, in units of P_S .

Results. We test on 30 loops of held-out materials, on new noise for training materials, and on an independent set of ten closed loops. CNO is the best model on held-out materials (0.078; Table 11). Duhamel-NO is second (0.138), with $1.76\times$ the loop error of CNO, even though its first-frame error is $4.34\times$ lower than that of the best other model. There are two reasons. First, its Fourier SSM is measured around the unpoled state, but during a sweep the film spends most of its time near saturation, where the linear dynamics are different. Second, with 20 solver steps per frame only some wavenumber bins can be fitted, and the uniform mode, which carries the switched polarization, is not among them. Voltage-driven switching is therefore the task on which Duhamel-NO is least competitive. Figure 15 shows the error of the predicted polarization along the sweep, and Figure 16 the predicted loops of held-out materials.

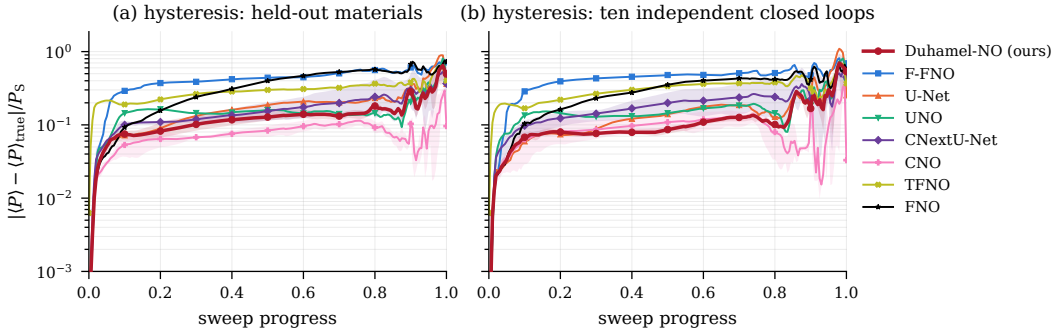


Figure 15: **Hysteresis error along the sweep.** Deviation of the mean polarization from FERROX, in units of P_S , on held-out materials (a) and on ten independent closed loops (b); its time mean is the loop error. Mean over loops, and bands show the range over training seeds.

Table 11: **Hysteresis, closed loops:** loop error (mean $\pm$ std over seeds) on held-out materials, on new noise for training materials, and on an independent set of ten closed loops; field correlation on held-out materials.

model	seeds	held-out material	new noise	extra closed set	field corr. (held-out mat.)
CNO	3	0.078 ± 0.004	0.094 ± 0.012	0.092 ± 0.019	0.722 ± 0.007
Duhamel-NO (ours)	3	0.138 ± 0.014	0.128 ± 0.003	0.121 ± 0.012	0.617 ± 0.042
UNO	1	0.164	0.159	0.172	0.189
CNextU-Net	3	0.171 ± 0.052	0.184 ± 0.055	0.191 ± 0.064	0.455 ± 0.120
U-Net	1	0.194	0.183	0.165	0.174
TFNO	1	0.295	0.310	0.309	0.074
FNO	1	0.361	0.245	0.310	0.397
F-FNO	1	0.431	0.409	0.439	0.264

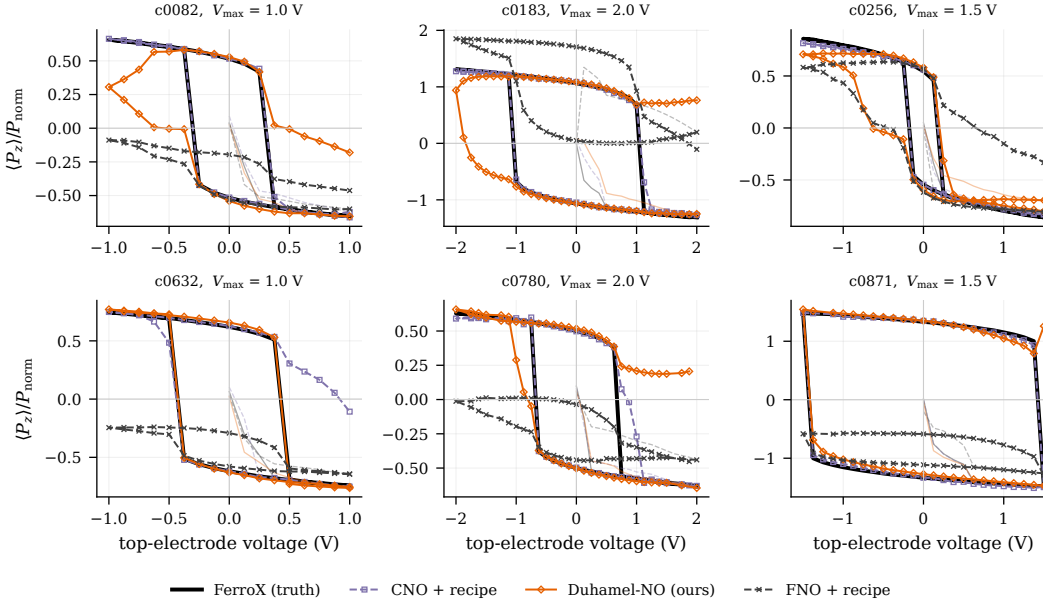


Figure 16: **Closed polarization–voltage loops of held-out materials**, predicted by rollout of the whole sweep from the unpoled random field (virgin branch drawn lightly).

I Allen–Cahn

To test transfer to new grids and wall widths, we also use the 2D Allen–Cahn equation $\partial_t u = -M(au + bu^3 - e\nabla^2 u)$ with $a < 0$. Like TDGL, its dynamics start from small noise that grows and then saturates. We generated these data ourselves, on a periodic 64^2 grid from small random noise, with an exponential-integrator solver. Note that such a solver favours an exponential-Euler model such as Duhamel-NO.

Setup. Each split has 900 training, 100 validation and 128 test trajectories. For training we sample the amplitude unit $U = \sqrt{|a|/b} \in [0.5, 2]$, $|a| \in [0.5, 5]$, and the grid spacing in wall widths $\Pi_h = \Delta x/\mu \in [0.30, 0.70]$, where $\mu = \sqrt{e/|a|}$ is the wall width. Two test splits leave this range, *new units* ($U \in [4, 10]$, $|a| \in [20, 100]$, same Π_h) and *new physics* ($\Pi_h \in [0.85, 1.20]$). We also simulate the in-range test materials on 32^2 and 128^2 grids of the same domain, and roll out every model trained on 64^2 on them unchanged. Table 12 gives the final-frame errors on every split and grid.

Results. Here Duhamel-NO learns its linear symbol, as a polynomial in the dimensionless wavenumber $k\mu$, which does not depend on the grid. A learned symbol works here but not in TDGL (Table 9) because Allen–Cahn has no depth layers to couple. Its symbol is one scalar per wavenumber, while

the TDGL operator is a 10×10 matrix per wavenumber, set by the electrostatics, which a learned conditioned symbol does not capture. In range, its final error is 0.012, against 0.043 for the best baseline, F-FNO. On new physics (Figure 9c) it reaches 0.075, $3.5\times$ lower than the best baseline (0.265, TFNO). Trained on 64^2 , it runs unchanged on 128^2 (Figure 9d) with error 0.010, while F-FNO and FNO degrade to 0.086 and 0.276. The reason is that its linear step depends on the physical wavenumber $k\mu$, not on the grid index. So on the finer grid every mode advances exactly, even modes the training grid cannot represent. The coarser 32^2 grid is different. There Duhamel-NO is fourth (0.124), behind F-FNO (0.049), TFNO and FNO (Table 12).

Table 12: **Allen–Cahn**: relative L_2 error at the final frame (mean over seeds; **best**, second). New units have amplitude and time units outside the training range. New physics has the wall width relative to the grid outside it. The zero-shot grids hold the in-range test materials on 32^2 and 128^2 , with models trained on 64^2 .

model	trained grid (64^2)			zero-shot grid (in range)	
	in range	new units	new physics	32^2	128^2
Duhamel-NO (ours)	0.012	0.010	0.075	0.124	0.010
F-FNO	<u>0.043</u>	<u>0.042</u>	0.416	0.049	<u>0.086</u>
CNextU-Net	0.057	0.051	1.023	0.879	0.909
U-Net	0.088	0.066	0.378	0.857	0.854
CNO	0.097	0.090	0.667	0.884	0.859
UNO	0.092	0.090	0.920	0.717	0.194
TFNO	0.120	0.102	<u>0.265</u>	<u>0.094</u>	0.271
FNO	0.150	0.148	0.443	0.115	0.276

J Limitations

Duhamel-NO is built for *noise-seeded dynamics*, in which a small random field is amplified by an unstable linear stage that decides the pattern. Where the linear stage matters less, its advantage is smaller. On voltage-driven hysteresis it is second to CNO (Section H). It also needs an accurate linear operator, which it measures from the first frames of the training simulations. With only 32 training materials, the interpolated Fourier SSM is too coarse in α , and F-FNO is more accurate (0.186 vs. 0.239; Figure 12). Finally, all ferroelectric data come from one simulator, a uniaxial model without elastic coupling, with one electrode and dielectric configuration. We compare with simulation, not with experiment.