Shearlet, a Novel Operator Learning Model

Júlio Vargas Bruno Jucá Jairson Sami Fábio Santos Alexandre Evsukoff

National Laboratory of Scientific Computing (LNCC)
Federal University of Rio de Janeiro (UFRJ)

May 15, 2026

Preprint: arXiv:2604.25181
https://arxiv.org/abs/2604.25181

Neural operators in a nutshell

Problem statement

Given a parametric PDE

$$\mathcal{N}(u; a) = 0, \qquad a \in \mathcal{A},\ u \in \mathcal{U},$$

learn the solution operator $\mathcal{G}^\dagger : a \mapsto u$ from data, so that new instances of a can be mapped to u without re-running a classical solver.

Goal. Approximate $\mathcal{G}^\dagger$ between function spaces. A typical neural-operator layer acts on a hidden function $v_t(x)$, combining a local linear map W with a non-local integral operator.

$$v_{t+1}(x) = \sigma\Big(W\, v_t(x) \;+\; \big[\mathcal{K}(a)\, v_t\big](x)\Big),$$

$$\big[\mathcal{K}(a)\, v\big](x) \;=\; \int_D \kappa(x, y; a)\, v(y)\, dy.$$

Design question: how do we parameterise / compute $\mathcal{K}$ efficiently?

- Graph kernels (GNO): flexible but costly.
- Low-rank kernels (LNO): efficient but global.
- Spectral parameterisation $\Rightarrow$ FNO and its descendants.

Where FNO struggles — in PDEs and in rocks

The Fourier basis is global, isotropic, and smooth.
That hurts whenever the solution exhibits:

In parametric PDEs

- Anisotropic advective transport
- Shocks and curved fronts (Burgers, conservation laws)
- Sharp interfaces (multi-phase, multi-scale media)

In porous media

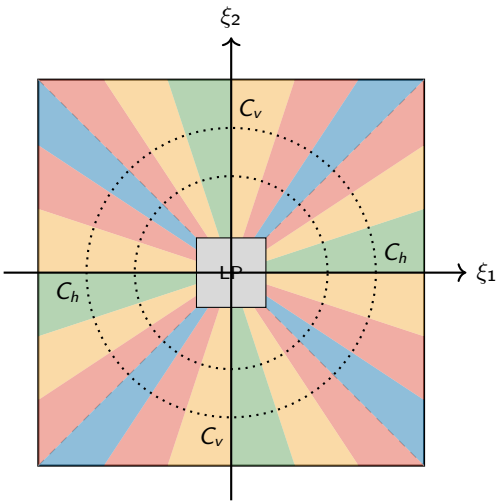
- Oriented pore networks, laminations, fractures
- Invasion fronts, saturation discontinuities
- Rock–fluid and grain boundaries at every voxel

In porous media, FNO inherits exactly these failure modes.
Our proposal: an operator whose spectral basis is matched to the problem geometry.

What that basis should be

Directional, multiscale, and spatially localized: $\Longrightarrow$ **Shearlets**.

Background — Shearlets: the intuition



Frequency-plane tiling.

- Two cones — C_h (horizontal) and C_v (vertical) — plus a central low-pass square.
- Each cone is split into dyadic radial shells indexed by scale j.
- Each shell is partitioned into angular wedges indexed by shear k.
- Finer scales $\Rightarrow$ more wedges $\Rightarrow$ better directional resolution.

Three group operations:

- A_a parabolic scaling
- S_s shearing (orientation)
- T_t translation (locality)

Background — Why shearlets are powerful

Optimal sparse approximation (Guo & Labate)

For the class $\mathcal{E}^2(\mathbb{R}^2)$ of *cartoon-like* functions (piecewise C^2 with C^2 edges), the best N-term shearlet approximation f_N satisfies

$$\|f - f_N\|_2^2 \lesssim N^{-2} (\log N)^3,$$

which is optimal up to a log factor, better than Fourier or wavelets.

Three crucial properties:

- Directional selectivity: shear parameter resolves orientation, ideal for fronts and anisotropy.
- Parabolic scaling $a \times \sqrt{a}$: matches the geometry of curvilinear singularities.
- Spatial localisation: unlike Fourier atoms, each shearlet is concentrated around t.

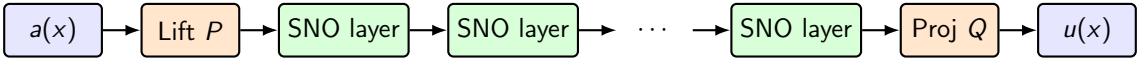
These are exactly the features needed for shock-dominated PDE solutions.

Shearlet Neural Operator (SNO) — architecture

Idea. Replace the Fourier transform in the spectral layer of FNO with the discrete shearlet transform SH.

A single SNO layer applied to a hidden representation v_t:

$$v_{t+1}(x) \;=\; \sigma\Big(W\, v_t(x) \;+\; \mathrm{SH}^{-1}\big(R_\theta \odot \mathrm{SH}(v_t)\big)(x)\Big).$$



Pipeline. Lifting $P \to$ stacked SNO layers $\to$ projection Q to physical output.

Benchmark suite — seven PDE families

We compare SNO against FNO on seven 2D parametric problems grouped by regime:

Diffusion-dominated

- **Multi-orientation texture** — anisotropic diffusion of superposed oriented waves.

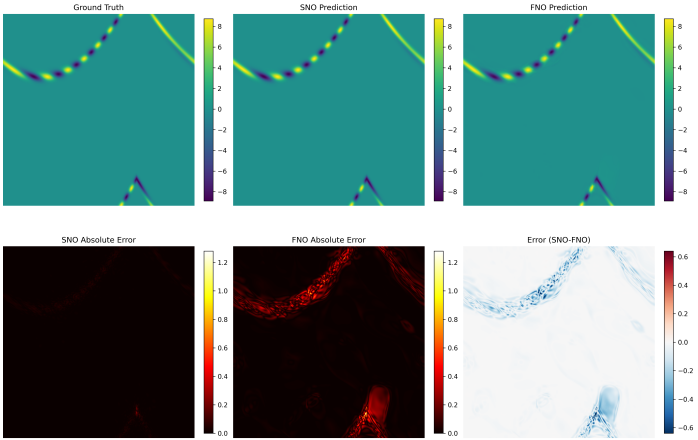
Convection-dominated

- **Bent ridge advection** — parabolically curved front.
- **Anisotropic ridge advection** — extreme aspect-ratio jet ($c_y/c_x = 120$).
- **Sheared Kelvin–Helmholtz stripes** — noisy oriented stripes.
- **Polygonal shock pattern** — 6-fold angular symmetry.

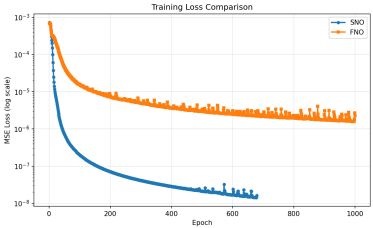
Combined-effect (Burgers, low viscosity)

- **Multi-angle shocks** — two interacting shock fronts.
- **Spiral shock** — Archimedean spiral with continuously varying orientation.

Results — Bent ridge advection



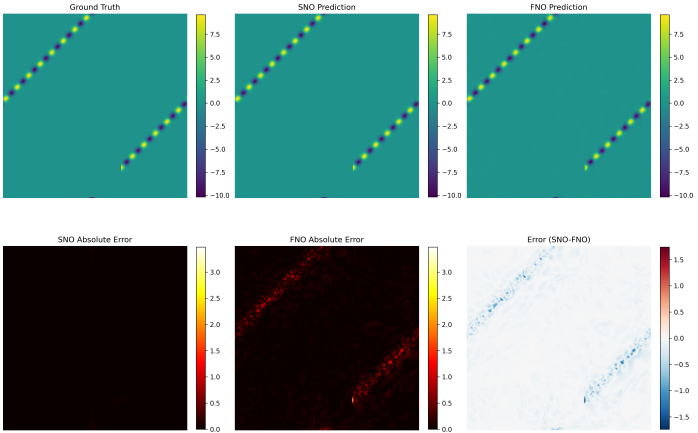
Curved front advected with $(c_x, c_y) = (0.5, 5)$. (N=512, T=150)



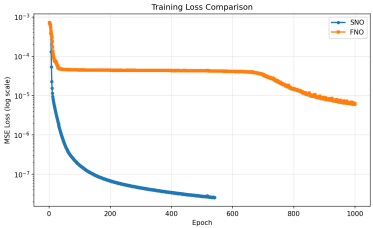
Training loss (MSE, log).

- SNO error near zero along the curved trajectory; FNO smears it over a wide tube.
- L^2 ratio (SNO/FNO) ≈ 0.10 — an order of magnitude better.
- Loss gap is roughly two orders of magnitude throughout training.

Results — Anisotropic ridge advection (jet stream)



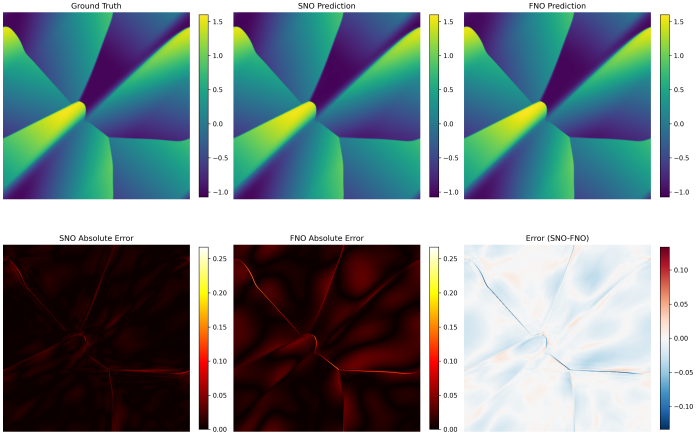
Aspect-ratio 120 transport, $(c_x, c_y) = (0.05, 6.0)$. (N=512, T=150)



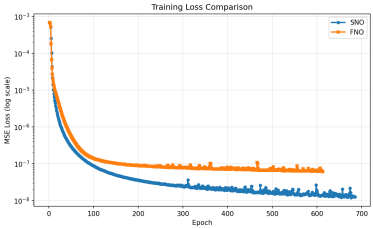
Training loss (MSE, log).

- Extreme directional bias: FNO's isotropic modes cannot align with the jet.
- SNO captures the oriented ridge essentially exactly.
- Largest gap of the suite: L^2 ratio ≈ 0.06 ($\sim$16$\times$); SSIM 0.9999 vs. 0.958.

Results — Polygonal shock pattern



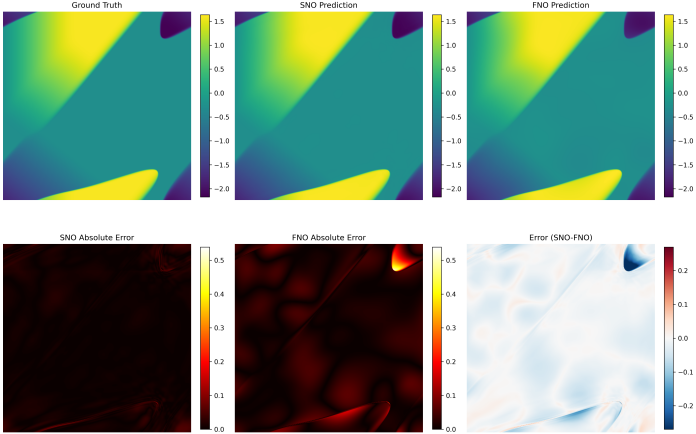
Hexagonal shock with 6-fold symmetry. (N=512, T=220)

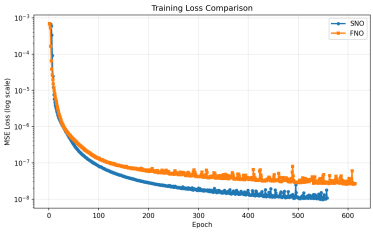


Training loss (MSE, log).

- FNO rounds the polygon's corners; SNO preserves the sharp vertices.
- Error concentrates on the discrete symmetry axes.
- L^2 ratio ≈ 0.43 — gap narrows because structure aligns partly with grid.

Results — Multi-angle shocks (Burgers)





Training loss (MSE, log).

- SNO is visually indistinguishable from ground truth across the domain.
- FNO blurs both fronts and introduces oscillations near intersections.

Two interacting shock fronts at $\arctan(0.8)$ and $\arctan(-1.2)$. (N=512, T=220)

Quantitative comparison — all benchmarks (N=512)

Table: Performance Comparison between SNO and FNO Models for N = 512

Dataset	SNO L2	FNO L2	Ratio	SNO MSE	FNO MSE	SNO MAE	FNO MAE	SNO SSIM	FNO SSIM
anisotropic ridge advect	2.41×10^{-6}	3.97×10^{-5}	0.0608	3.36×10^{-5}	9.08×10^{-3}	2.11×10^{-3}	4.60×10^{-2}	0.9999	0.9583
multi angle shocks	3.93×10^{-6}	1.48×10^{-5}	0.2650	9.08×10^{-5}	1.29×10^{-3}	6.25×10^{-3}	2.17×10^{-2}	0.9990	0.9942
bent ridge advect	2.53×10^{-6}	2.52×10^{-5}	0.1004	3.69×10^{-5}	3.67×10^{-3}	2.43×10^{-3}	2.24×10^{-2}	0.9999	0.9780
multi orientation texture	4.09×10^{-6}	4.21×10^{-6}	0.9699	8.38×10^{-6}	8.91×10^{-6}	2.25×10^{-3}	2.46×10^{-3}	0.9986	0.9988
spiral shock	8.36×10^{-6}	8.49×10^{-6}	0.9852	2.16×10^{-4}	2.22×10^{-4}	5.84×10^{-3}	4.42×10^{-3}	0.9939	0.9942
polygonal shock	3.52×10^{-6}	8.18×10^{-6}	0.4308	4.10×10^{-5}	2.21×10^{-4}	4.19×10^{-3}	1.04×10^{-2}	0.9987	0.9934
sheared kelvin helmotz	4.18×10^{-6}	2.31×10^{-5}	0.1812	3.21×10^{-5}	9.77×10^{-4}	4.44×10^{-3}	2.65×10^{-2}	0.9989	0.9677

- Gap grows with anisotropy and shock complexity; spiral shock and multi-orientation texture are the limiting near-tie cases.

Discussion

What changed? Only the spectral basis — from Fourier exponentials to shearlet atoms. Architecture, training, data, parameter budget were held fixed (SNO actually uses *fewer* parameters: $\sim$12k vs. $\sim$17k for FNO).

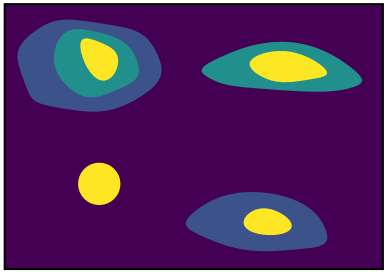
Why such a large effect on hard regimes?

- Fourier needs many modes to approximate localised, anisotropic features.
- Shearlets are near-optimal for cartoon-like functions $\Rightarrow$ a small number of coefficients carries most of the information.
- The neural network only has to learn the residual structure once the basis already captures the geometry.

Limitations.

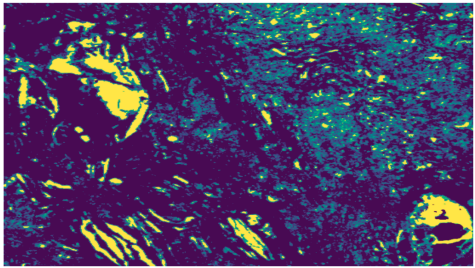
- Shearlet transforms have a higher constant cost than FFT.
- Cases with continuously varying orientation (spiral shock) erode the inductive-bias advantage.
- 3D extensions exist but are heavier to implement and train.

Bridge — carbonate porosity slices *exhibit* cartoon-like structure



Synthetic cartoon $f \in \mathcal{E}^2(\mathbb{R}^2)$:

piecewise smooth on C^2 regions; sharp jumps on a C^2 curve.



Real porosity field (μCT):

Near-constant interiors, curved pore–solid interfaces.

Why this matters

Similar structure: roughly constant interiors with sharp jumps along curved boundaries. The Guo–Labate $N^{-2}(\log N)^3$ rate (slide 5) is optimal *exactly* for this signal class — a stronger fit than "anisotropic" alone.

Application — Porosity prediction from dry rock samples

Setting. 3D porosity mapping of carbonate plug samples via micro-CT (COPPE/UFRJ × Petrobras).

Classical pipeline (paired dry/wet μCT)

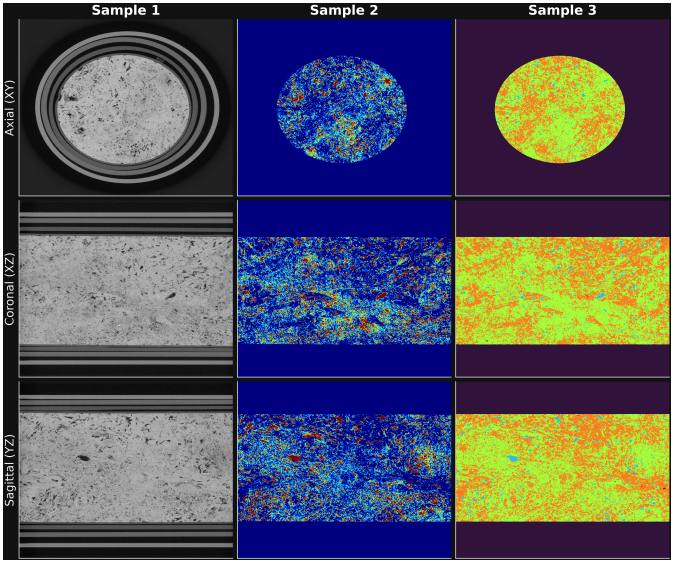
- Two scans of the *same* plug: dry and saturated with 10% NaI at 2,500 psi.
- Calibrated intensity difference gives voxel-wise porosity.
- Accurate but expensive: two scans, chemicals, high-pressure cel.

Goal. Skip the wet scan — predict the porosity map $\phi(x) \in [0,1]$ directly from the dry image alone, at *voxel scale*.

Why is SNO a natural fit here?

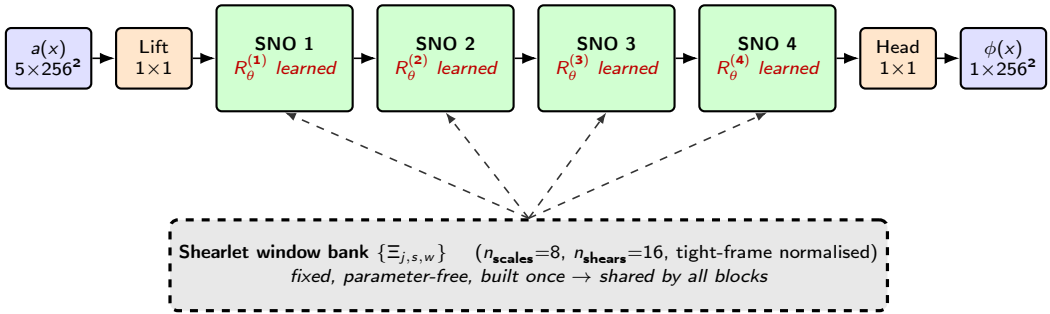
- Carbonate pore networks are strongly anisotropic (vugs, channels, oolitic shells, laminations).
- Pore-solid interfaces are near-cartoon: piecewise-smooth fields with curved C^2 jump sets — the regime where shearlets give near-optimal sparse approximation.
- Dataset: 11 plugs.

Application — Porosity prediction from dry rock samples



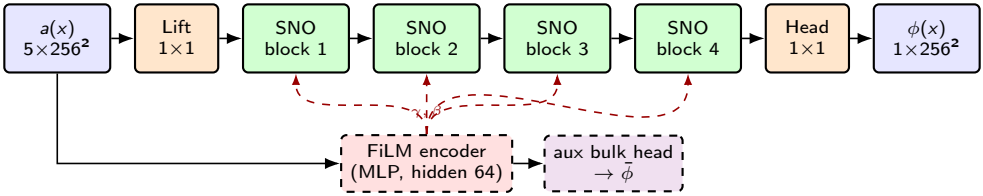
Spectral core — shared bank, per-block learned weights

Task. Input: $5\times256\times256$ (5 neighbouring dry slices). Output: $1\times256\times256$ (central porosity slice). $L=4$ spectral layers between a lift and a head.



Inside each SNO block, *all of these are learned, separately per block*: spectral mixing weights $R_\theta^{(\ell)}$ acting on shearlet coefficients, sigmoid gates that combine the three mixing branches (spectral + depthwise 3×3 + pointwise 1×1), the depthwise/pointwise conv weights themselves, SE channel attention ($r=4$), and the transformer-style FFN ($4\times$ expansion). Channel width $W=16$ throughout. Per-sample GroupNorm ($G=1$); long-range skip from lifted input to head.

SNO for porosity — architecture diagram



Training protocol — optimisation & LOOCV setup

Optimisation

- AdamW, $\beta = (0.9, 0.95)$
- $\eta = 10^{-3}$; weight decay 10^{-4}
- Batch size 32; gradient clip 2.0
- Up to 1000 epochs; early-stop patience 50 on `val_loss`
- GroupNorm-based per-sample input normalisation; no batch statistics

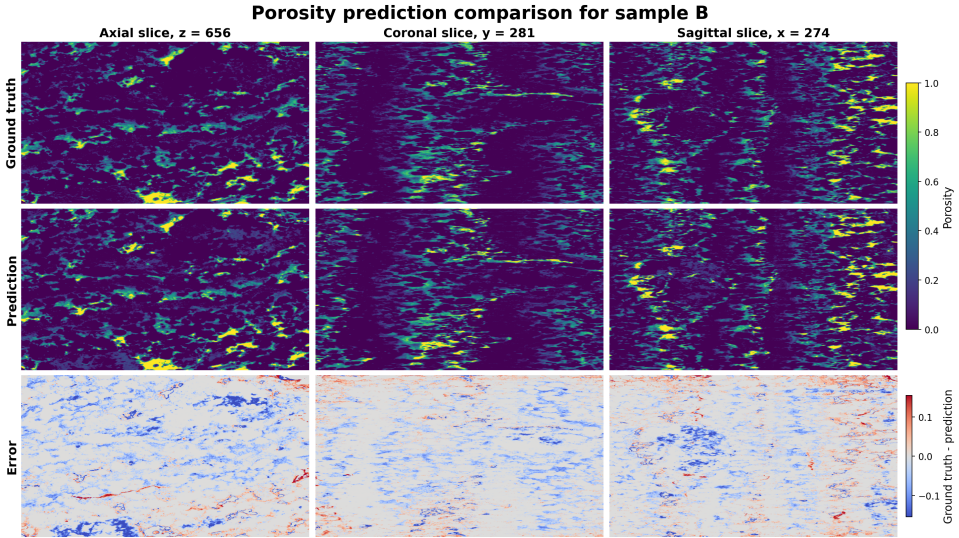
LOOCV setup

- 11 sample codes, one held out per fold ($\times$11 retrains)
- 90 / 10 train / val split within each fold
- Sqrt-weighted sample balancing (per-sample voxel count)
- DDP, multi-GPU; resumes from `last.ckpt`

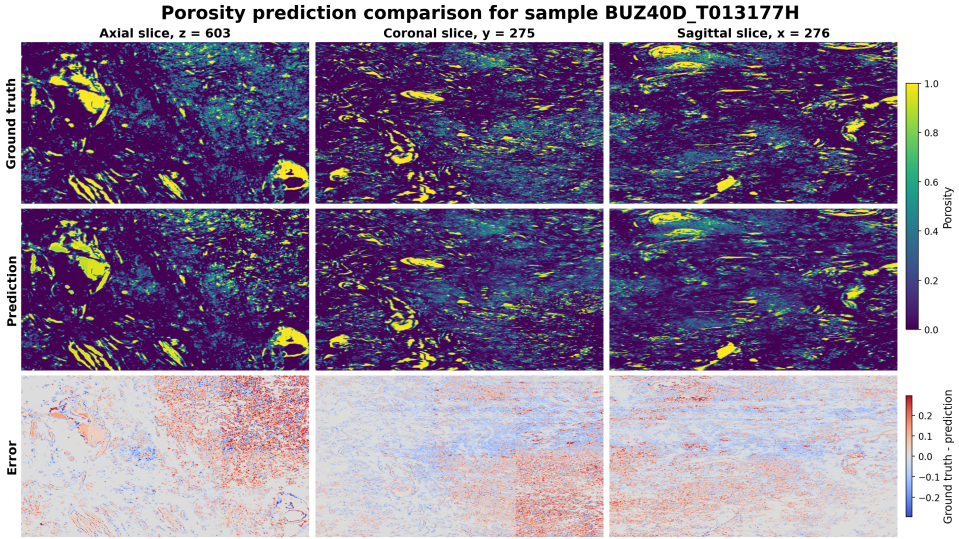
Reporting

Test metrics are aggregated across the 11 held-out folds. Per-fold bulk porosities are then compared against laboratory measurements.

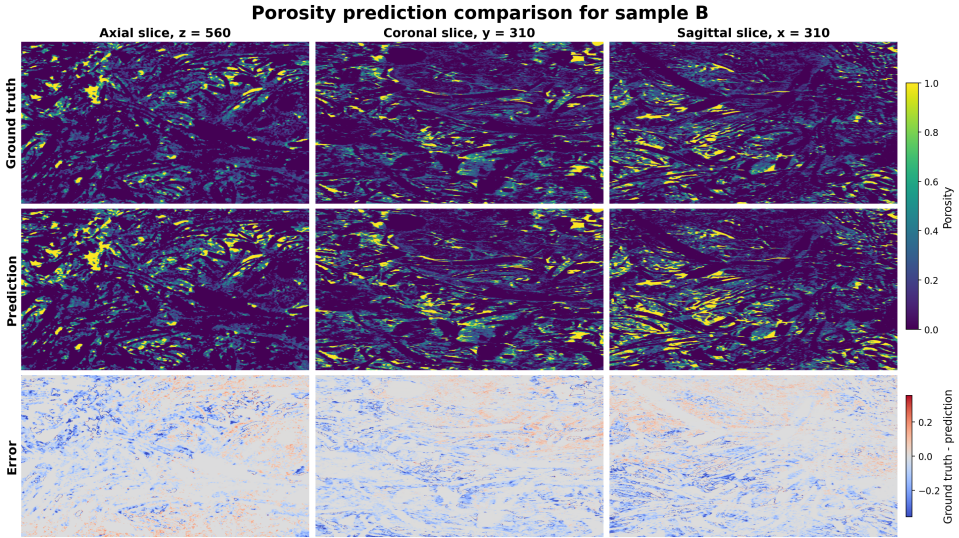
Application results — visual prediction quality



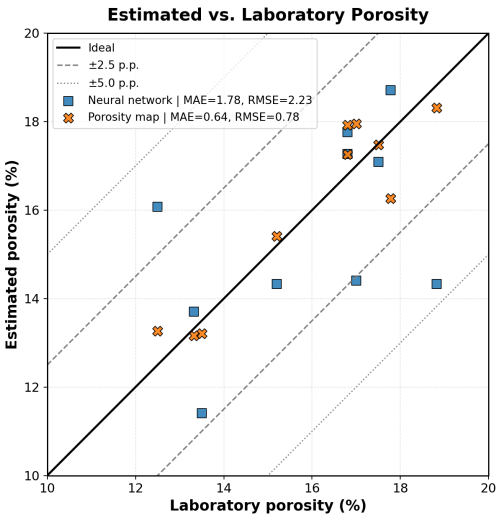
Application results — visual prediction quality



Application results — visual prediction quality



Application results — bulk porosity vs. classical CDF



Predicted vs. lab bulk porosity, 11 plug crops, LOOCV.

Metric (p.p. unless noted)	Porosity map	SNO
MAE	0.64	1.78
RMSE	0.78	2.23
Bias	+0.18	−0.63
MAPE (%)	3.80	10.95
Within ±1 p.p. (%)	72.73	54.55
Within ±2.5 p.p. (%)	100.00	63.64
Within ±5 p.p. (%)	100.00	100.00

- SNO recovers spatial structure of pores per-slice (see prev. slide).
- Bulk calibration trails the porosity map method, which uses *both* dry+wet scans by construction.
- All SNO predictions remain within ±5 p.p. of lab porosity — with the wet scan dropped.

Conclusions, outlook & references

Contributions

- **SNO:** a neural operator built on the discrete shearlet transform — directional, multiscale, spatially-localised spectral layer.
- Consistent improvements over FNO on anisotropic and shock-dominated parametric PDEs across seven benchmark families, with *fewer* parameters.
- Real-world case study: 3D porosity mapping of carbonate rocks from dry-state micro-CT. SNO recovers pore geometry across 11 LOOCV folds; all bulk-porosity predictions within ± 5 p.p., *without* the saturated scan.

Outlook

- 3D shearlets for compressible flow, seismic imaging, and full-volume porosity prediction.
- Hybrid Fourier / shearlet layers; physics-informed losses; adaptive band truncation.
- Tighten bulk-porosity calibration on the rock dataset (mass-conservation loss; larger / more diverse training cores).

Preprint

F. P. dos Santos, J. de C. V. Fernandes, A. M. A. Côrtes. *Shearlet Neural Operators for Anisotropic-Shock-Dominated and Multi-scale Parametric PDEs.* arXiv:2604.25181 — arxiv.org/abs/2604.25181

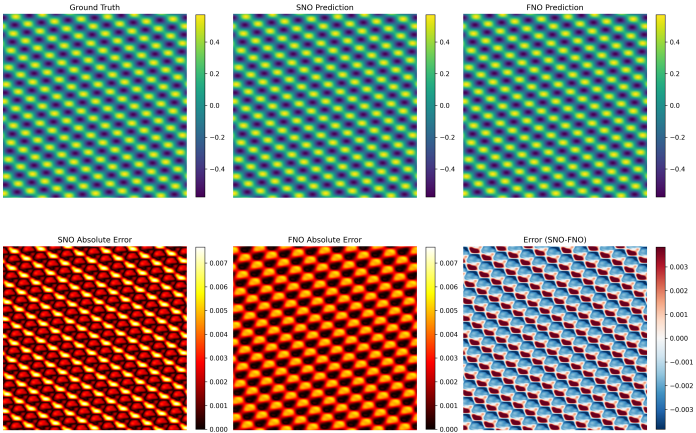
Thank you — questions?

Backup

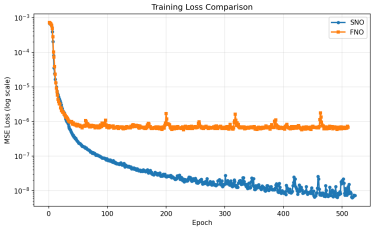
Appendix

- Additional PDE results
- Loss explanation
- Architecture details

Results — Multi-orientation texture (diffusion)



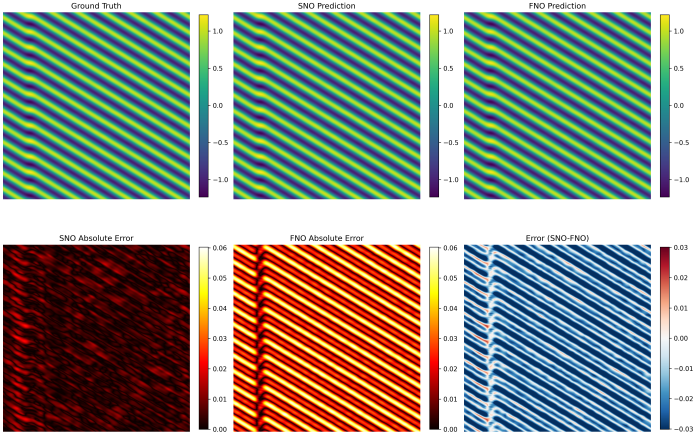
Anisotropic diffusion of superposed oriented waves. (N=512, T=150)



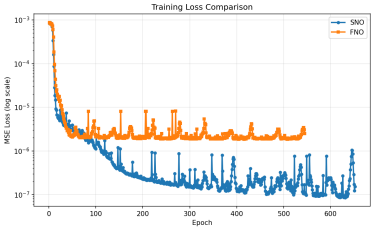
Training loss (MSE, log).

- FNO error visibly larger inside the oriented bands; SNO recovers them sharper.
- Both operators capture the multi-scale texture; L^2 ratio ≈ 0.97 — near-tie at this resolution.
- FNO loss plateaus earlier; SNO converges faster in the early epochs.

Results — Sheared Kelvin–Helmholtz stripes



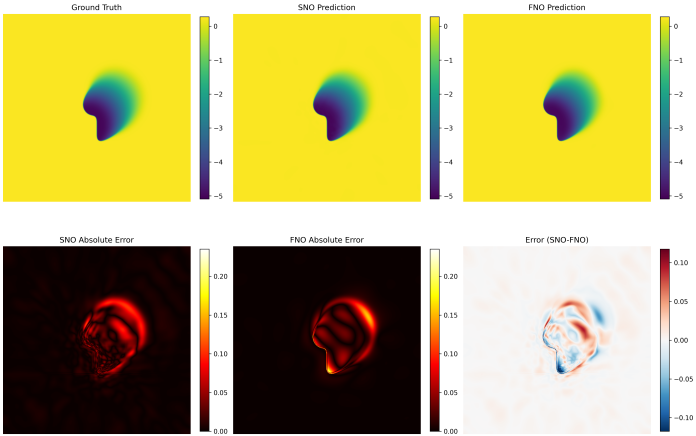
Sheared stripes with additive Gaussian noise. (N=512, T=100)



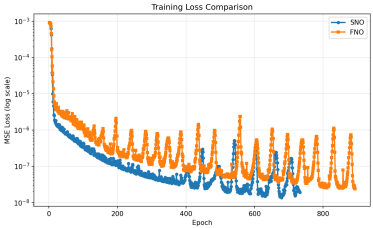
Training loss (MSE, log).

- Noise stresses both operators, but SNO retains stripe orientation and contrast.
- FNO loses stripe sharpness; SSIM drops to 0.968 vs. 0.999 for SNO.
- L^2 ratio ≈ 0.18 — consistent 5–6$\times$ improvement.

Results — Spiral shock pattern (Burgers)



Archimedean spiral with continuously varying orientation. (N=512, T=100)



Training loss (MSE, log).

- Continuously rotating front: the hardest case for a finite shear discretisation.
- SNO still wins overall; FNO slightly better in smooth interior regions (red/blue mix).
- L^2 ratio ≈ 0.99 — essentially tied; limiting case for SNO's bias.

Training protocol — composite porosity loss

$$\mathcal{L} \;=\; \underbrace{w_c\,\mathcal{L}^{\mathrm{ms}}_{\mathrm{Charb}}}_{1.0} + \underbrace{w_s\,\mathcal{L}^{H^s}_{\mathrm{spec}}}_{0.5} + \underbrace{w_g\,\mathcal{L}^{\mathrm{ms}}_{\mathrm{Sobel}}}_{0.15} + \underbrace{w_d\,\mathcal{L}^{\mathrm{Huber}}_{\mathrm{sort}}}_{0.6} + \underbrace{w_m\,\mathcal{L}^{\mathrm{Huber}}_{\mathrm{bulk}}}_{0.5} + \underbrace{w_m\,\mathcal{L}_{\mathrm{aux}}}_{0.5}$$

Building blocks. Charbonnier $\phi_\varepsilon(x)=\sqrt{x^2+\varepsilon^2}$, $\varepsilon=10^{-3}$ (smooth L^1). Huber $H_\delta(x)=\frac{1}{2}x^2$ for $|x|<\delta$, $\delta(|x|-\delta/2)$ otherwise; $\delta=0.05$.

- **Multi-scale Charbonnier** (pixel, peak-region mask W_s, scales $s\in\{0,1,2\}$, $\omega_s=(1, 0.5, 0.25)$):

$$\mathcal{L}^{\mathrm{ms}}_{\mathrm{Charb}} = \sum_s \omega_s \left\langle W_s\, \phi_\varepsilon\big(\hat{u}^{(s)} - u^{(s)}\big)\right\rangle$$

- **Spectral** high-pass Sobolev–Charbonnier (radial weight $w(k) \propto |k|^\alpha$, $\alpha=1.5$):

$$\mathcal{L}^{H^s}_{\mathrm{spec}} = \left\langle \sqrt{\big(w(k)\,|\widehat{u}_k - \widehat{u}_k|\big)^2 + \varepsilon^2}\right\rangle_k$$

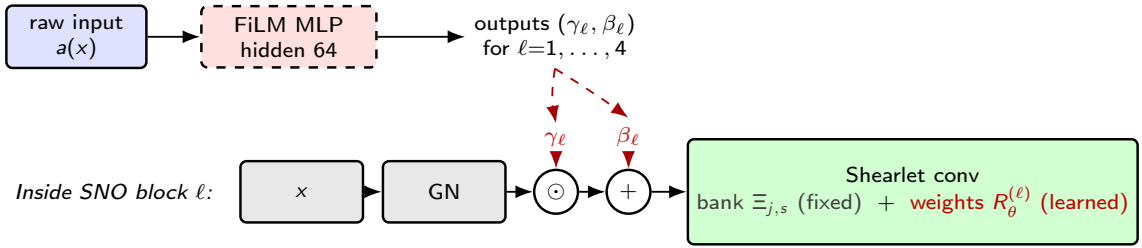
- **Sobel-gradient** multi-scale Charbonnier (boundary sharpness, $s\in\{0,1\}$):

$$\mathcal{L}^{\mathrm{ms}}_{\mathrm{Sobel}} = \sum_s \omega_s \Big(\langle \phi_\varepsilon(\partial_x \Delta u^{(s)})\rangle + \langle \phi_\varepsilon(\partial_y \Delta u^{(s)})\rangle \Big)$$

- **Sorted-pixel Huber** (1-D Wasserstein on sorted pixel values, per sample): $\mathcal{L}^{\mathrm{Huber}}_{\mathrm{sort}} = H_\delta\big(\mathrm{sort}(\hat{u}), \mathrm{sort}(u)\big)$
- **Bulk-mean Huber** (calibration) and **auxiliary FiLM-head Huber**: $\mathcal{L}^{\mathrm{Huber}}_{\mathrm{bulk}} = H_\delta(\bar{\hat{u}}, \bar{u})$, $\mathcal{L}_{\mathrm{aux}} = H_\delta(\phi_{\mathrm{aux}}, \bar{u})$

FiLM conditioning — how the *fixed* bank adapts per sample

Feature-wise Linear Modulation. A small MLP reads the raw input a and emits sample-specific scale/shift pairs $\mathrm{FiLM}(x;a) = \gamma(a) \odot x + \beta(a)$, with $(\gamma_\ell, \beta_\ell) \in \mathbb{R}^W \times \mathbb{R}^W$ for each of the L=4 blocks (W=16 values each).



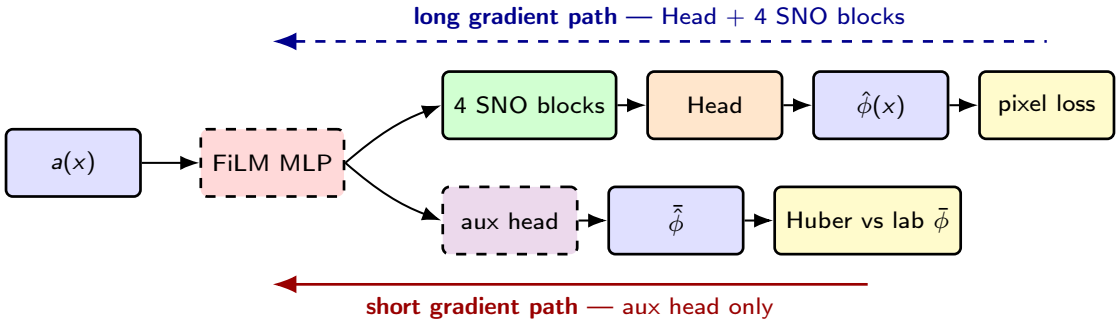
Shared across samples (set once during training):

- the shearlet bank $\Xi_{j,s,w}$ (fixed mathematics, parameter-free);
- every learned weight: $R_\theta^{(\ell)}$, FiLM-MLP, sigmoid gates, depthwise/pointwise convs, SE, FFN.

Sample-specific (depend on *this* $a(x)$):

- only the values $(\gamma_\ell, \beta_\ell)$;
- γ_ℓ rescales each channel before the shearlet conv $\Rightarrow$ the fixed bank Ξ acts as an effectively different filter on each input, without any weight changing.
- Zero-init (γ=1, β=0 at step 0): training starts as if FiLM is absent.

Auxiliary bulk-porosity head — a shortcut for global calibration



What the aux head does. A small MLP off the FiLM encoder predicts *one number* per sample — the bulk porosity:

$$\bar{\hat{\phi}}(a) \;=\; \mathrm{AuxHead}\big(\text{FiLM-MLP}(a)\big),$$

trained against the lab value with a Huber loss (weight 0.5).

Why this matters.

- Bulk porosity is a sample-level property; via pixel loss alone, calibration must back-propagate through 4 spectral blocks + head before reaching FiLM.
- Aux loss at the FiLM encoder gives a one-step gradient path ⇒ FiLM learns bulk *early*, so the spectral core can specialise on *spatial structure*.