
Recovering the Lévy Noise: Stochastic Pseudodifferential Symbols of Neural Operators

Jose Marie Antonio Miñoza
Center for AI Research PH

Abstract

Training a neural operator on a stochastic dynamical system yields a deterministic map between function spaces, into which the driving noise is absorbed. Existing interpretations of trained operators recover the deterministic component, and that noise is rarely characterized. This work shows that the trained operator still encodes that noise, and that expressing the operator in the Fourier basis recovers it in closed form as a pseudodifferential symbol, the Lévy–Khintchine exponent whose stability index α sets the tail, with $\alpha = 2$ the Gaussian case. The same symbol separates the deterministic operator from the noise, so the noise becomes a quantity one can measure and control. Two consistent methods recover the symbol, from the learned transition symbol of an α -stable process and from the residual of a deterministic fit under a verifiable tail condition. Furthermore, heavy-tailed initialization of a Fourier Neural Operator yields an α -stable operator field whose law is the spectral pair (α, Γ) , with the shared skip coupling its harmonic modes. Across nine PDE families in one and two dimensions, recovery reaches relative L_2 error 7×10^{-7} for the deterministic symbol and about ten percent on the injected stability index.

1 INTRODUCTION

A neural operator trained on a stochastic dynamical system, such as an SDE, a jump process, or an SPDE, learns its evolution as a map between function spaces. That map is deterministic, yet the system it learned is driven by noise whose law is set by a diffusion coefficient, a jump measure, or a stability index. The trained operator encodes that law, and a pseudodifferential symbol recovers it in closed form. That symbol

is the noise’s Lévy–Khintchine exponent, the exponent of its characteristic function. For a stable law the exponent is the stability index α , which sets how heavy the noise tails are, together with a scale. The Gaussian case is $\alpha = 2$, and a smaller α is heavier-tailed.

There is an explicit call for methods that explain, in a generalizable way, what a neural operator has learned (Gao et al., 2025). Existing interpretation is deterministic and case-specific. A learned operator’s apparent randomness is sometimes deterministic structure it failed to capture, and sometimes genuine noise. The two must be separated to control the noise. The symbol separates them. Its Lévy–Khintchine form makes the heavy-tailed, jump-driven regime explicit, where a Gaussian model of the same data misses the tails that carry the rare large events.

A Fourier Neural Operator (FNO) layer $u \mapsto \mathcal{F}^{-1}(R_\theta(k)\mathcal{F}u)$ acts one Fourier mode at a time. A translation-invariant operator multiplies each Fourier mode ξ by a single factor $\hat{p}(\xi)$, and that factor, expressed as a function of ξ , is its symbol. The PDE operators, the isotropic α -stable generator $-(-\Delta)^{\alpha/2}$ with symbol $-|\xi|^\alpha$, and the random limit of a wide operator all act this way. One basis diagonalizes them, so the symbol is recoverable. A differential operator has a polynomial symbol, which represents only local structure. The α -stable generator is nonlocal, its symbol $-|\xi|^\alpha$ is not a polynomial, so the operator is pseudodifferential, and only its symbol carries the jump law. A generic network gives no such representation for each frequency.

Contributions. A trained neural operator’s driving noise is a spatial random *field*, and the operator acts one Fourier mode at a time. This function-space structure is absent in the finite-dimensional trajectory setting. (i) A decomposition of the learned map into a deterministic operator symbol $p(\xi)$ and a noise symbol, both expressed in the one basis that diagonalizes the operator (§3). (ii) The heavy-tailed initialization limit of a wide FNO is an α -stable operator field with *independent modes*, so its symbol is the pair

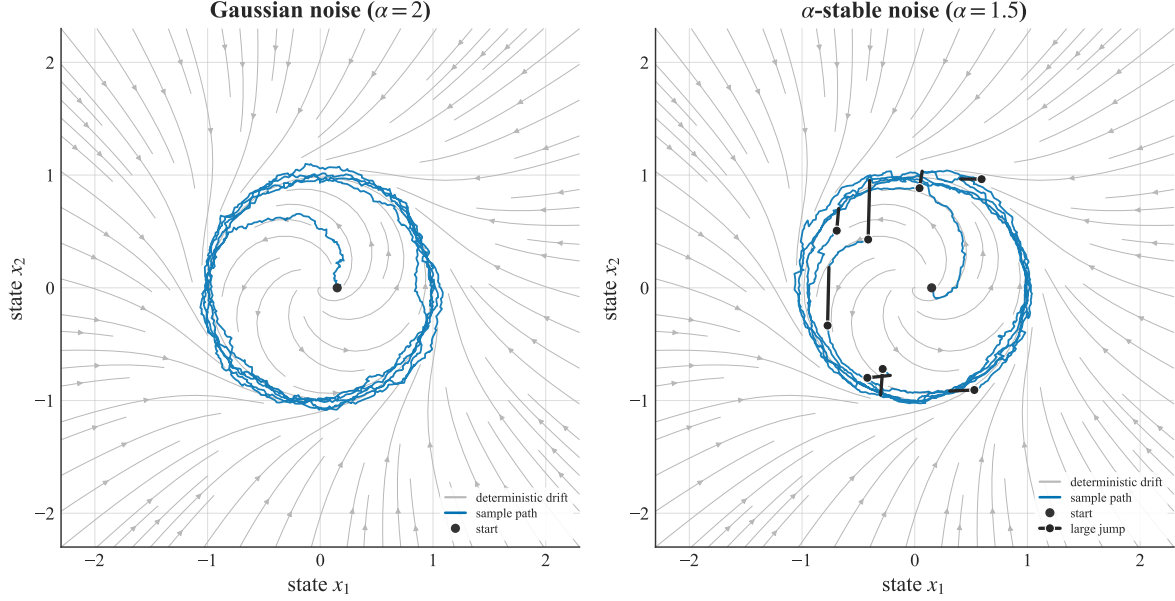


Figure 1: The operator’s symbol separates deterministic structure from genuine noise. Both panels share one *drift field* $b(x)$ (grey streamlines) and one random noise stream. The drift field is the deterministic velocity of a state $x = (x_1, x_2)$. Here it is a Stuart–Landau limit cycle $b(x) = (1 - r^2)x + \omega x^\perp$, with $r^2 = x_1^2 + x_2^2$, $x^\perp = (-x_2, x_1)$, and $\omega = 1$. It draws trajectories onto the unit circle. The blue path solves $dx = b(x) dt + \sigma dL_t^\alpha$. Only the noise law differs: the Gaussian case ($\alpha = 2$, left) hugs the cycle, and the heavy-tailed α -stable case ($\alpha = 1.5$, right) takes occasional large jumps off it (dark segments, landing points marked). The stability index α is the part that changes, and the operator’s symbol recovers it.

(α, Γ) indexed by mode, where Γ is the spectral measure that spreads the jump intensity over the modes, and a depth result, proven at $L = 2$ and conditional beyond it, shows a shared skip connection couples harmonic modes so Γ is non-diagonal (§5). The $\alpha = 2$ corner is the Gaussian operator process of de Souza et al. (2025). (iii) Recovery of the driving noise’s stability index from a trained operator’s residuals, guaranteed by a verifiable tail condition on the fit error, the function-space counterpart of stability-index estimation for finite-dimensional SDEs, alongside recovery of the governing operator’s terms from the deterministic symbol (§4–§6, §7). (iv) Experiments across nine PDE families spanning one and two dimensions, and two physical-parameter ranges (§8). Symbolic accounts of neural operators recover a deterministic symbol (Shin et al., 2022; Boullé et al., 2022), and stability-index estimation is established for finite-dimensional SDEs. The operator field, its explicit symbol indexed by mode, and the heavy-tailed regime are what the function-space setting adds.

2 SETUP

The operators act on the torus, a periodic spatial domain, and are mode-truncated FNOs (Li et al., 2021b;

Kovachki et al., 2023). The results that extract the symbol (§4, §5) need an operator that is diagonal in a fixed basis, so they hold for FNO and for the other spectral operators (for example wavelet-based ones) in their own basis, and not for operators without such a basis. DeepONet (Lu et al., 2021) is one such case, not diagonal in any fixed basis, and no result that extracts the symbol uses it. The residual method (§6) needs only a trained mean fit and applies to any architecture. A pseudodifferential operator acts as $(Au)(x) = \int p(x, \xi) \hat{u}(\xi) e^{ix \cdot \xi} d\xi$, where $\hat{u}$ is the Fourier transform of u and p is the symbol. When p does not depend on x the operator is *translation-invariant*. It commutes with spatial shifts and is diagonal in the Fourier basis. It multiplies each mode k by a single number $\hat{p}(k)$, its *Fourier multiplier*. This is why the symbol can be extracted from an FNO layer. A Lévy process has stationary independent increments, and its law is fixed by the Lévy–Khintchine exponent $\psi(\xi)$ in $\mathbb{E} e^{i\xi X_t} = e^{t\psi(\xi)}$. Its generator is the pseudodifferential operator with symbol ψ . The isotropic α -stable family has $\psi(\xi) = -|\xi|^\alpha$, generator $-(-\Delta)^{\alpha/2}$, and transition semigroup $e^{-t|\xi|^\alpha}$. The stability index $\alpha \in (0, 2]$ sets the tail: $\alpha = 2$ is Gaussian, and $\alpha < 2$ has power-law tails, jumps, and infinite variance. The stochastic claims below describe the operator’s initialization

prior, its law at initialization, before any training. The field’s symbol is the pair (α, Γ) , and convergence is in L^2 .

3 A SYMBOL DECOMPOSITION THAT SEPARATES SIGNAL FROM NOISE

The learned map carries a deterministic operator symbol $p(\xi)$, the Fourier multiplier that fixes what the operator computes, with ξ dual to space. Its randomness is described by two further symbols of different kinds, kept separate. The *noise symbol* is the law of the noise driving the dynamics. It is encoded in the trained operator. In the heavy-tailed regime it is a Lévy–Khintchine characteristic exponent whose variable is dual to the field value, and a Gaussian spectral density at $\alpha = 2$. This is the symbol recovered from a trained operator (§4, §6). The second is the symbol of the random operator itself under heavy-tailed initialization. That operator is an *operator field*: an infinite-width α -stable field with spectral pair (α, Γ) , where Γ is a spectral measure on the retained modes (§5). The first is a property of the learned dynamics, and the second is a property of the model prior. Both are α -stable objects expressed in the basis that diagonalizes the operator, and both meet the Gaussian spectral density at $\alpha = 2$.

4 THE LEARNED TRANSITION SYMBOL

The operator trained on samples of an α -stable process learns the process’s transition semigroup. That semigroup has symbol $e^{-t|\xi|^\alpha}$. A log–log fit of the recovered symbol gives the stability index.

Proposition 1 (symbol recovery at each mode). *A single Fourier-multiplier layer fit by least squares to input/output pairs of a translation-invariant operator with multiplier $\hat{p}(k)$ recovers $R[k] \rightarrow \hat{p}(k)$ on every mode of nonzero input energy. For the target $e^{-t|\xi|^\alpha}$ the stability index is identified by $\log(-\log R[k]) = \log t + \alpha \log |k|$.*

5 THE INITIALIZATION PRIOR AS AN α -STABLE FIELD

The operator prior studied here is distinct from the noise symbol of §4 and §6. The stability index here is that of the weights at initialization, and it describes the prior over operators, not the noise of any dataset. The statement is a property of the initialization and holds independently of the optimizer used later.

Theorem 1 (single-layer α -stable operator field). *Under heavy-tailed initialization of a mode-truncated single-layer FNO, with i.i.d. symmetric weights whose tails are exactly power-law (regularly varying) of index α , the output field at a fixed input converges in distribution, in $L^2(\mathbb{T})$, to a symmetric α -stable field with independent modes, symbol (α, Γ) , and a scale σ_k^α for each mode, under normalization $(n \log n)^{-1/\alpha}$. The index α is intrinsic to the operator, while Γ depends on that input. As $\alpha \rightarrow 2$ the limit is structurally the Gaussian neural-operator process of de Souza et al. (2025), and the constants coincide for piecewise-linear activations.*

The proof reduces the output field to its coefficient vector on the retained modes and takes a multivariate stable limit of that vector. Each channel contributes a product of two heavy-tailed weights, the lifting weight and the spectral weight. A product of two index- α tails is again index α with a logarithmic correction, which forces the $(n \log n)^{-1/\alpha}$ normalization (Bordino et al., 2023). The heavy factor is shared across the modes of one channel, while the spectral weights are drawn fresh for each mode, so a joint jump of two modes needs two independent large weights and stays negligible against the single-mode jumps. The limiting Lévy measure charges one mode at a time, the modes are independent, and a multivariate stable central limit theorem assembles the field (Jung et al., 2023). Within that measure each mode $k \geq 1$ carries twice the spectral mass of the zero mode, since the complex Fourier weight contributes two independent real components on its plane. The appendix gives the full argument. As $\alpha \rightarrow 2$ the stable exponent turns into a quadratic, and the field converges to the Gaussian operator prior of de Souza et al. (2025) on the same mode-diagonal covariance.

Theorem 2 (depth limit with a skip connection). *Consider a depth- L mode-truncated FNO with a pointwise skip connection. Take the width limit one layer at a time. For $L = 2$ one coupled skip block gives a symmetric α -stable field with symbol (α, Γ_2) . The shared skip weight makes a dominant input mode m_0 deposit Fourier mass on its harmonic family, so harmonic mode pairs jump together, and Γ_2 is non-diagonal, diagonal if and only if the skip vanishes. For general depth L the same sequential limit supports Γ_L on nested harmonic sets, diagonal if and only if every skip vanishes, under a generic nondegeneracy condition. The simultaneous width limit, where every layer widens together, is open, and the result holds along sufficiently fast width paths.*

The coupling comes from the shared skip weight. A single large input coefficient at mode m_0 passes through the rectifying activation and lands Fourier

mass on the harmonics of m_0 at once. The half-wave rectified cosine fixes those weights, $\hat{\rho}_0 = 1/\pi$, $\hat{\rho}_{\pm 1} = 1/4$, and $\hat{\rho}_{\pm 2k} = \frac{1}{\pi} \frac{(-1)^{k+1}}{4k^2 - 1}$, and vanishes on the higher odd harmonics. The skip multiplies that whole spread by one weight, so those modes jump together, and their joint jump is as likely as a single-mode jump. The spectral part draws a fresh weight for each mode and stays diagonal, so the skip alone produces the coupling. Independent modes return when the skip vanishes. Depth compounds the effect, since each layer spreads the previous harmonics onto their own harmonics, so the coupled modes fill nested harmonic sets. The coupling shows that heavy-tailed noise correlates the frequency modes through the operator.

Proposition 2 (identifiability requires an ensemble). *The discrete-spectrum field is non-ergodic. One trained network reveals its realized coefficients only. The law (α, Γ) follows from an ensemble of networks. The marginal scale $\sigma_k^\alpha = \int |s_k|^\alpha d\Gamma$ of each mode is estimatable from an ensemble of m initializations at rate $\sqrt{m}$. The cross-mode coupling requires joint multi-mode statistics.*

6 EXTRACTING THE NOISE FROM RESIDUALS

The data has the form $u = G(a) + \epsilon$, where G is a deterministic operator, and ϵ is a symmetric α -stable term drawn fresh for each sample. A deterministic operator fits the conditional mean G . The residual is the noise, and the residual samples give its symbol.

The tail of the fit error decides whether the residual reveals the noise. An imperfect fit leaves that error in the residual alongside the noise. When the error is lighter-tailed than the noise, the heavy tail that fixes the stability index still dominates, and the estimator extracts it, whatever the error's L_2 size. When the error is as heavy as the noise, the two are inseparable.

Proposition 3 (residual estimator). *The empirical characteristic-function estimator of α from held-out residuals recovers the stability index when the fit error δ is lighter-tailed than the noise, $\mathbb{E}|\delta|^p < \infty$ for some $p \in (\alpha, 2]$. It then extracts the stability index regardless of the L_2 size of δ , and is asymptotically normal at the $\sqrt{m}$ rate around a slope within a controllable tolerance of α (Koutrouvelis, 1980; Koutrouvelis and Bauer, 1982). A fit error as heavy-tailed as the noise breaks the recovery.*

7 ESTIMATION AND RECONSTRUCTION

Each method estimates a stability index from data with infinite variance, so empirical moments give nothing. The empirical characteristic function $\hat{\varphi}(\theta) = \frac{1}{m} \sum_j e^{i\langle \theta, X^{(j)} \rangle}$ stays bounded and has finite variance at every θ (Feuerverger and Mureika, 1977; Paulson et al., 1975). A symmetric α -stable law satisfies $|\hat{\varphi}(\theta)| \approx e^{-\sigma^\alpha |\theta|^\alpha}$. The noise symbol is $\psi(\theta) = -\log |\hat{\varphi}(\theta)|$, and the slope of $\log \psi$ against $\log |\theta|$ gives α . A slope that stays flat across the frequency window certifies the extraction. A slope that bends toward two flags a fit error too heavy to separate. The estimate degrades as $\alpha \rightarrow 2$. The heavy-tail signal that fixes the stability index is carried by the rare large values, and those values grow scarce near the Gaussian boundary. The experiments show this limit (§8).

Figure 2 contrasts the characteristic-function estimator with a moment estimator on the same samples. The empirical variance diverges with sample size for $\alpha < 2$ and settles only at $\alpha = 2$, so a Gaussian or moment fit reports a scale that never converges. The characteristic-function estimator recovers the stability index across the range.

A recovered symbol is useful only if it reproduces the operator it came from. The recovered symbols define a closed-form operator, and its relative L_2 error against the trained operator on held-out data measures the reconstruction. A recovered stability index is checked the same way, by whether the residual law it predicts matches the observed residuals. Table 1 reports this error across the linear registry. The closed-form symbol reproduces every linear operator to machine precision. A linear symbol fit against the nonlinear Burgers operator has error about 0.37, which tells an adequate reconstruction from an inadequate one.

8 EXPERIMENTS

The deterministic and residual methods use 1D operators whose symbols are known in closed form: Poisson, heat, the α -stable transition operator, and the fractional Laplacian. The single Fourier-multiplier layer makes the learned symbol directly recoverable, and a multi-layer FNO carries the residual method on the nonlinear benchmarks. The method from the initialization prior uses a wide FNO block at heavy-tailed initialization, estimated across an ensemble of independent initializations. Every method estimates the stability index from the empirical characteristic function, since the heavy-tailed data carries no usable moments.

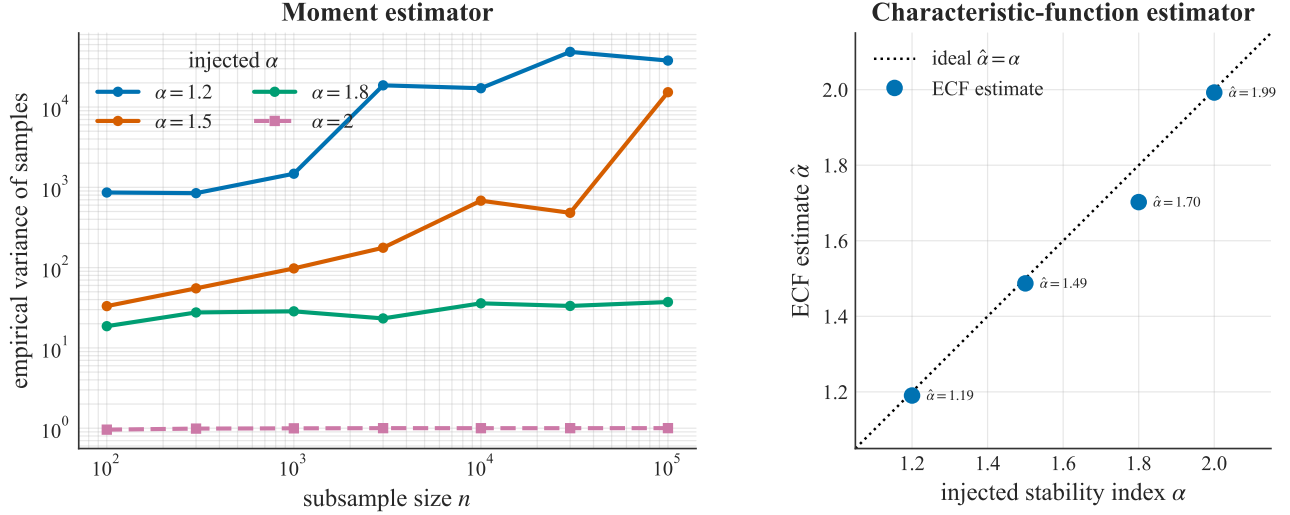


Figure 2: The estimator choice. Left: the empirical variance grows without settling as the sample size grows for $\alpha < 2$, and stays flat at $\alpha = 2$, so a moment estimator has no limit in the heavy-tailed regime. Right: the characteristic-function estimator recovers the injected stability index where the moment estimator fails.

Table 1: Reconstruction error of the recovered closed-form symbol. The closed-form least-squares symbol on each mode is compared with the analytic symbol on held-out data. The five linear operators recover to machine precision. The nonlinear Burgers control, where no linear symbol exists, scores about fifteen orders of magnitude higher.

operator	symbol error	reconstruction rel. L_2
Poisson	3.8×10^{-16}	3.5×10^{-16}
heat	1.9×10^{-15}	2.2×10^{-15}
Helmholtz	2.4×10^{-15}	2.5×10^{-15}
α -stable transition	7.9×10^{-16}	1.5×10^{-15}
fractional Laplacian	4.9×10^{-16}	4.7×10^{-16}
Burgers (nonlinear control)	n/a	3.7×10^{-1}

Recovery reaches near machine precision for the deterministic symbol (Figure 3). From those recovered linear symbols, a dictionary fit extracts the governing operator in closed form: its differential order, its coefficient, and whether it is local or nonlocal. The fit recovers the diffusion viscosity, the Helmholtz wavenumber, and the fractional index, and it separates local from nonlocal operators, at coefficient of determination 1.00 on the linear registry. The stability-index recovery then tracks the injected value both from the heavy-tailed width limit of the initialization prior and from a trained operator’s residuals (Figure 4). The recovery is weakest at very heavy tails near $\alpha = 1.2$, where the infinite-variance noise degrades the mean fit, and it carries a mild downward bias near $\alpha = 2$, where the estimator loses resolution.

The coupling prediction of Theorem 2 also holds on a wide FNO block with α -stable initialization (Figure 5).

Each mode’s index matches the initialization for both conditions. Without the shared skip, harmonic modes are independent, and their tail dependence sits at the independence baseline. With the skip, the modes couple, and the upper-tail dependence, the chance that one harmonic mode is extreme when the other is, rises well above the baseline for every tested α .

The residual method runs on real PDE benchmarks from PDEBench (Takamoto et al., 2022), across nine families that span one and two dimensions. Table 2 reports eight of them at the full stability-index range, and Navier–Stokes is the ninth, run at $\alpha = 1.5$ only ($\hat{\alpha} = 1.51$). A neural operator is trained on the next-step solution operator, symmetric α -stable noise is added to the targets, and the stability index is extracted from held-out residuals. The injected stability index is known, which fixes a ground truth to score the recovery against, so the real content of the test

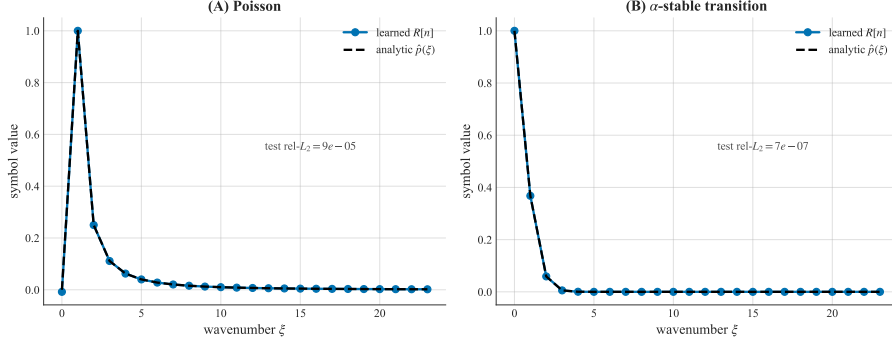


Figure 3: The deterministic method. A single Fourier-multiplier layer recovers the analytic Poisson symbol to test relative L_2 error 9×10^{-5} , and the α -stable transition symbol to 7×10^{-7} .

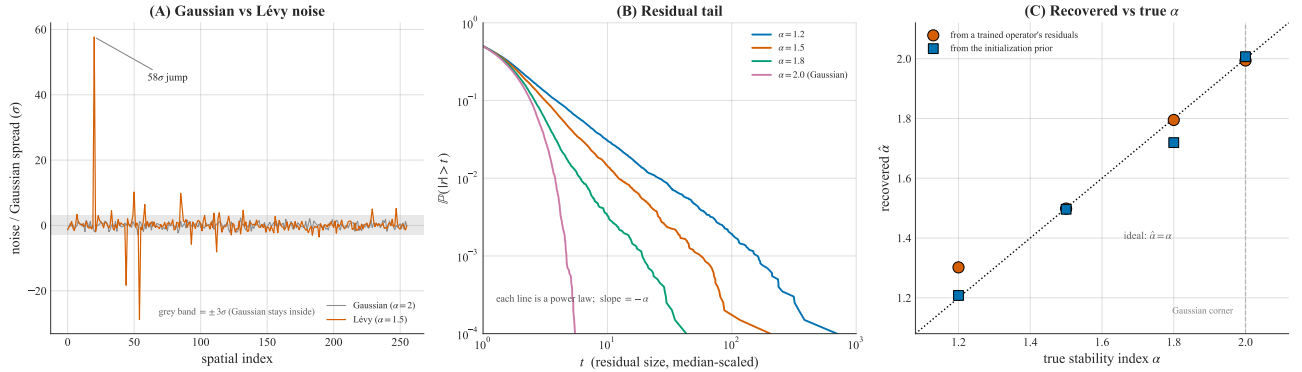


Figure 4: Extracting the stability index by two methods. (A) One realization of a trained operator’s residual: the Gaussian noise ($\alpha = 2$) stays inside the $\pm 3\sigma$ band, while the Lévy noise ($\alpha = 1.5$) carries rare jumps tens of σ tall. (B) The empirical survival function of the residuals is a power law whose log-log slope is $-\alpha$, so the tail exponent is the stability index, and the Gaussian corner ($\alpha = 2$) falls off sharply. (C) Both methods recover the true stability index α . The residual method extracts it from a trained operator’s residuals, separating the deterministic part to relative L_2 error 8×10^{-3} . The method from the initialization prior extracts it from the heavy-tailed width limit. The readout is conservative only near the Gaussian corner.

is the nonlinear dynamics the operator must fit. The recovery is accurate when the residual error is lighter-tailed than the noise, which holds wherever the operator fits the clean dynamics. At $\alpha = 1.5$ and $\alpha = 1.8$ the recovered stability index matches the injected one within about ten percent on every family that fits, with the largest gap on the 2D families near the Gaussian boundary. At $\alpha = 1.2$ the heavy tail degrades the mean fit for most families, so those cells carry no clean split and are left blank. A blank is a fit failure, consistent with Proposition 3, and separate from the estimator.

The recovery is invariant to a family’s physical parameters (Table 3). On Burgers the recovered stability index stays near the injected 1.5 as the viscosity varies over three orders of magnitude, and on advection it stays near 1.5 as the transport coefficient varies. The stability index the operator carries is a property of the

driving noise, and it does not move with the equation’s coefficients.

9 SCOPE AND LIMITATIONS

The depth results hold in the sequential width limit, layer by layer, for symmetric α -stable block weights. The simultaneous width limit, where every layer widens together, is open. Depth L also needs a nondegeneracy assumption. At every layer, for every atom v_a of the input spectral measure, the in-band rectified image $\rho(a)$ loads a mode outside the support of v_a , which holds when the atom’s fundamental satisfies $3m_0 \leq K$ and can fail near the band edge. The condition $\rho(a) \neq 0$ holds automatically. One coupling constant is inherited from the single-layer atom weights. The recovery gives the scalar α and the scale of each mode, and it does not give the full multivari-

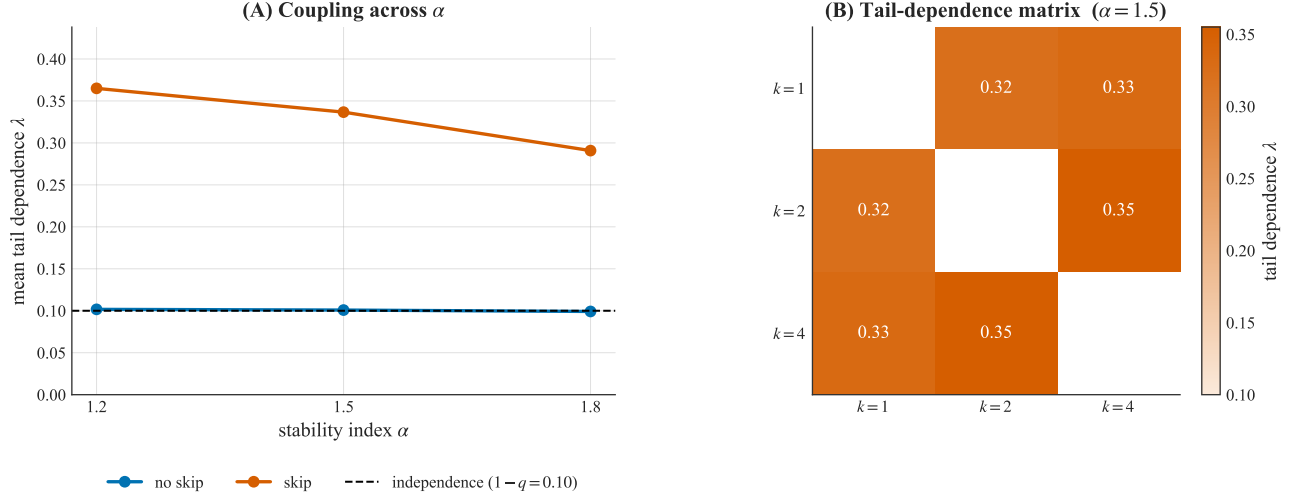


Figure 5: Theorem 2 on a wide FNO block with α -stable initialization, over an ensemble of random initializations. (A) The mean upper-tail dependence λ of harmonic mode pairs versus α : without the shared skip it sits on the independence baseline $1 - q$, and with the skip it rises well above it. (B) The tail-dependence matrix over the harmonic family with the skip, at $\alpha = 1.5$: every pair lifts above the baseline, so one shared weight couples the whole family at once.

ate jump measure. Convergence is in distribution in $L^2(\mathbb{T})$, and the field is stationary only at $\alpha = 2$ or under a translation-invariant input ensemble. Estimation degrades as $\alpha \rightarrow 2$. The PDEBench Navier–Stokes subset used here has four trajectories, so its error on the clean data is measured across time steps of those trajectories. The experiments measure recovery against an injected stability index of known value, and do not test recovery from a native heavy-tailed law. The residual method needs the fit error to be lighter-tailed than the noise. On shock-dominated compressible flow the spectral FNO does not fit, and its residual error is itself heavy-tailed, so the recovered stability index pins to that error and the family is out of scope without a shock-capturing backbone (Appendix B).

10 RELATED WORK

Neural operators map functions to functions. The Fourier Neural Operator applies a learned multiplier on each retained Fourier mode (Li et al., 2021b; Kovachki et al., 2023), and DeepONet learns a branch-trunk representation of the solution map (Lu et al., 2021). These architectures target the deterministic solution operator of a PDE.

Data-driven equation discovery is well developed on the deterministic side. Sparse regression identifies active terms and their coefficients for ODEs and PDEs (Brunton et al., 2016; Rudy et al., 2017). PDNO parameterizes a pseudodifferential symbol as a forward operator (Shin et al., 2022), Green’s-function learning

recovers the kernel of a linear operator (Boullé et al., 2022), and a neural-operator symbolic search recovers operators with fractional and nonlocal terms (Garmaev and Fink, 2025). The recovery of the deterministic terms here (§8) is a scaffold in this line. It extracts the symbol from the trained operator and estimates the fractional index that the symbolic search fixes. None of these recovers a noise distribution.

The Gaussian corner of the field limit is established for operators. de Souza et al. (2025) prove that a wide neural operator converges to a function-valued Gaussian process, and optimal convergence rates for neural operators are known in this finite-variance regime (Nguyen and Mücke, 2024). Both are the $\alpha = 2$ case of the operator field studied here. At $\alpha = 2$ the field has a covariance and a spectral density.

The heavy-tailed case rests on infinite-width stable limits for fully connected networks (Jung et al., 2023; Favaro et al., 2020; Bordino et al., 2023; Favaro et al., 2022). A wide layer with heavy-tailed weights converges to a symmetric α -stable law carrying a spectral measure on the sphere of input coordinates (Samorodnitsky and Taqqu, 1994). The operator setting changes the object: the limit is a random field, it is diagonal in the Fourier basis, so its modes are independent, the spectral measure becomes a symbol indexed by frequency mode, and the depth coupling acts on harmonic modes through the shared skip. Fully connected networks lack that structure, and the function-space regime supplies it.

Table 2: Residual method on PDEBench families across the full stability-index range. Each cell is the recovered stability index $\hat{\alpha}$ for the injected α . A blank marks a cell where the operator’s fit to the clean data stayed above relative L_2 0.35, so the signal and noise cannot be separated. Cells other than Burgers and shallow water are means over five seeds $\{0, 42, 84, 126, 168\}$, with a standard deviation of at most 0.03 across seeds. Navier–Stokes was run at $\alpha = 1.5$ only ($\hat{\alpha} = 1.51$).

PDE family	dim	$\alpha=1.2$	$\alpha=1.5$	$\alpha=1.8$
Burgers	1D	1.27	1.50	1.78
Advection	1D		1.54	1.80
Diffusion–sorption	1D		1.51	1.80
Diffusion–reaction	1D		1.49	1.80
Compressible flow	1D		1.50	1.74
Shallow water	2D	1.38	1.53	1.64
Darcy flow	2D		1.52	1.62
Diffusion–reaction	2D		1.50	1.75

Table 3: Recovery is invariant to a family’s physical parameters, at injected $\alpha = 1.5$. Left: Burgers across viscosity ν . Right: advection across transport coefficient β .

ν	$\hat{\alpha}$	β	$\hat{\alpha}$
0.001	1.53	0.1	1.58
0.01	1.53	0.4	1.54
0.1	1.50	2.0	1.53
1.0	1.50	4.0	1.54

Data-driven recovery of a driving Lévy noise is developed for finite-dimensional systems. Trajectory-based methods estimate the stability index and jump measure of an α -stable SDE from sample paths (Li and Duan, 2020; Li et al., 2021a). The residual method here is the function-space counterpart, extracting a heavy-tailed noise field from a trained operator’s residuals, a regime the trajectory setting does not reach.

The estimator is the empirical characteristic function of a stable law (Koutrouvelis, 1980; Koutrouvelis and Bauer, 1982; Nolan, 2020). Heavy tails in neural networks are also studied at the optimizer level, where stochastic-gradient noise is heavy-tailed and weight iterates go heavy-tailed even from light-tailed initialization (Simsekli et al., 2019; Gurbuzbalaban et al., 2021). That phenomenon is a property of the training dynamics, while the operator field’s prior here is fixed at initialization.

11 CONCLUSION

A neural operator trained on stochastic dynamics carries the distribution of its noise, and a pseudodiffer-

ential symbol expresses it in closed form. The light-tailed regime gives a Gaussian spectral density, and the heavy-tailed regime a stable characterization with index α , recovered from the learned transition symbol, the initialization prior, and the residuals. The same symbol separates the deterministic operator from the noise.

The separation makes the noise a quantity one can measure and act on. A recovered stability index certifies whether apparent randomness is deterministic structure the operator missed or genuine heavy-tailed noise. It sets the stability index of a denoiser, and it supplies a controlled Lévy forcing when injection is the goal.

AI use statement

Generative AI tools assisted this work in three roles: editing the manuscript for clarity and consistency, writing supporting code for the experiments and figures, and checking derivations and surveying related work. The theoretical results and their proofs, and all reported measurements, were produced and verified by the authors, with every proof checked line by line and open steps flagged as such in the text. All AI-assisted text, code, and derivation checks were reviewed by the authors, who take responsibility for the final content of this work, including its text, claims, and artifacts.

References

- Alberto Bordino, Stefano Favaro, and Sandra Fortini. Infinitely wide limits for deep stable neural networks: sub-linear, linear and super-linear activation functions. *Transactions on Machine Learning Research*, 2023. arXiv:2304.04008.
- Nicolas Boullé, Seick Kim, Tianyi Shi, and Alex

- Townsend. Learning green's functions associated with time-dependent partial differential equations. *Journal of Machine Learning Research*, 23(218):1–34, 2022. arXiv:2204.12789.
- Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.
- Daren B. H. Cline. Convolution tails, product tails and domains of attraction. *Probability Theory and Related Fields*, 72(4):529–557, 1986. doi: 10.1007/BF00344720.
- Daniel Augusto de Souza, Fangzheng Zhu, John P. Cunningham, Yuri Saporito, Diego Mesquita, and Marc Peter Deisenroth. Infinite neural operators: Gaussian processes on functions. In *Advances in Neural Information Processing Systems 38 (NeurIPS)*, 2025. arXiv:2510.16675.
- Paul Embrechts and Charles M. Goldie. On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society, Series A*, 29(2):243–256, 1980. doi: 10.1017/S1446788700021224.
- Stefano Favaro, Sandra Fortini, and Stefano Peluchetti. Stable behaviour of infinitely wide deep neural networks. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108 of *Proceedings of Machine Learning Research*, pages 1137–1146, 2020. arXiv:2003.00394.
- Stefano Favaro, Sandra Fortini, and Stefano Peluchetti. Large-width asymptotics for ReLU neural networks with α -stable initializations. *arXiv preprint arXiv:2206.08065*, 2022.
- Andrey Feuerverger and Philip McDunnough. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society, Series B*, 43(1):20–27, 1981. doi: 10.1111/j.2517-6161.1981.tb01142.x.
- Andrey Feuerverger and Roman A. Mureika. The empirical characteristic function and its applications. *The Annals of Statistics*, 5(1):88–97, 1977. doi: 10.1214/aos/1176343742.
- Wenhan Gao, Jian Luo, Fang Wan, Ruichen Xu, Xiang Liu, Haipeng Xing, and Yi Liu. Can data-driven dynamics reveal hidden physics? there is a need for interpretable neural operators. *arXiv preprint arXiv:2510.02683*, 2025.
- Sergei Garmayev and Olga Fink. A neural operator-based approach to symbolic discovery of PDEs, 2025.
- Mert Gurbuzbalaban, Umut Simsekli, and Lingjiong Zhu. The heavy-tail phenomenon in SGD. In *ICML*, 2021.
- Paul Jung, Hoil Lee, Jungwoo Lee, and Hongseok Yang. α -stable convergence of heavy-tailed infinitely-wide neural networks. *Advances in Applied Probability*, 55(4):1415–1441, 2023. doi: 10.1017/apr.2023.3. arXiv:2106.11064.
- Ioannis A. Koutrouvelis. Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75(372):918–928, 1980. doi: 10.1080/01621459.1980.10477573.
- Ioannis A. Koutrouvelis and David F. Bauer. Asymptotic distribution of regression-type estimators of parameters of stable laws. *Communications in Statistics - Theory and Methods*, 11(23):2715–2730, 1982.
- Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023. arXiv:2108.08481.
- Yang Li and Jinqiao Duan. A data-driven approach for discovering stochastic dynamical systems with non-gaussian lévy noise, 2020.
- Yang Li, Yubin Lu, Shengyuan Xu, and Jinqiao Duan. Extracting stochastic dynamical systems with α -stable lévy noise from data, 2021a.
- 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)*, 2021b. arXiv:2010.08895.
- Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning non-linear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5. arXiv:1910.03193.
- Mark M. Meerschaert and Hans-Peter Scheffler. *Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice*. John Wiley & Sons, New York, 2001. ISBN 9780471356295.
- Mike Nguyen and Nicole Mücke. Optimal convergence rates for neural operators, 2024.
- John P. Nolan. *Univariate Stable Distributions: Models for Heavy Tailed Data*. Springer Series in Operations Research and Financial Engineering. Springer, 2020. doi: 10.1007/978-3-030-52915-4.

- A. S. Paulson, E. W. Holcomb, and R. A. Leitch. The estimation of the parameters of the stable laws. *Biometrika*, 62(1):163–170, 1975. doi: 10.1093/biomet/62.1.163.
- Sidney I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007. ISBN 9780387242729.
- Samuel H. Rudy, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017.
- E. L. Rvačeva. On domains of attraction of multi-dimensional distributions. *Select. Transl. Math. Statist. Probab.*, 2:183–205, 1962.
- Gennady Samorodnitsky and Murad S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall/CRC, New York, 1994. ISBN 9780412051715.
- Jin Young Shin, Jae Yong Lee, and Hyung Ju Hwang. Pseudo-differential neural operator: Generalized fourier neural operator for learning solution operators of partial differential equations. In *arXiv preprint arXiv:2201.11967*, 2022.
- Umut Simsekli, Levent Sagun, and Mert Gurbuzbalaban. A tail-index analysis of stochastic gradient noise in deep neural networks. In *ICML*, 2019.
- Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Dan MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. PDEBench: An extensive benchmark for scientific machine learning. In *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*, 2022.

CHECKLIST

1. For all models and algorithms presented, check if you include:
 - (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes] Section 2 and the theorem/proposition statements.
 - (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] The recovery is consistent at the $\sqrt{m}$ rate in the sample or ensemble size (Propositions 3, 2).
 - (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes] In the supplementary material.
2. For any theoretical claim, check if you include:
 - (a) Statements of the full set of assumptions of all theoretical results. [Yes] Stated with each result and in Appendix A.
 - (b) Complete proofs of all theoretical results. [Yes] Appendix A; open steps are flagged explicitly.
 - (c) Clear explanations of any assumptions. [Yes]
3. For all figures and tables that present empirical results, check if you include:
 - (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes] Code in the supplementary material; data are the public PDEBench benchmark (Takamoto et al., 2022).
 - (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes] Section 8; the stability index is estimated on held-out residual splits.
 - (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes] Recovered indices are means over five seeds {0, 42, 84, 126, 168} with standard deviation at most 0.03 (Table 2).
 - (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] A single NVIDIA T4 GPU.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
 - (a) Citations of the creator if your work uses existing assets. [Yes] PDEBench (Takamoto et al., 2022).
 - (b) The license information of the assets, if applicable. [Yes] PDEBench is released under an open license.
 - (c) New assets either in the supplemental material or as a URL, if applicable. [Yes] Code in the supplementary material.
 - (d) Information about consent from data providers/curators. [Not Applicable]
 - (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]
5. If you used crowdsourcing or conducted research with human subjects, check if you include:

- (a) The full text of instructions given to participants and screenshots. [Not Applicable]
- (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
- (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Recovering the Lévy Noise: Stochastic Pseudodifferential Symbols of Neural Operators

Supplementary Materials

A PROOFS

This appendix proves the five results of the main text: Theorem 1, Theorem 2, Proposition 1, Proposition 3, and Proposition 2. Appendix A.1 treats the two field-limit theorems. Appendix A.2 treats the three retrieval propositions. The status of each result is stated with it: Theorem 1, Proposition 1, and Proposition 2 are unconditional, Proposition 3 rests on cited estimator theory for the $\sqrt{m}$ limit, with the tail and slope bounds proved here, and Theorem 2 holds at $L = 2$ in the sequential width limit, with its one inherited atom weight now fixed by Lemma 4, and for general depth in the same sequential limit under a nondegeneracy assumption, with the simultaneous width limit open.

A.1 The field limit of the initialization prior

The field limit describes the operator’s initialization prior, its law at initialization. All convergence is in $L^2(\mathbb{T})$ on the torus $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$. The field symbol is the pair (α, Γ) throughout, a stability index and a spectral measure on the mode sphere. It is not a Lévy generator exponent.

Standing notation. The Fourier coefficients are $\hat{f}(k) = \frac{1}{2\pi} \int_{\mathbb{T}} f(x) e^{-ikx} dx$, and Parseval reads $\|f\|_{L^2}^2 = 2\pi \sum_k |\hat{f}(k)|^2$. A mode cutoff K fixes the retained band, with V_K the span of $\{e^{ikx} : |k| \leq K\}$ and $V_K^{\mathbb{R}}$ its real elements, of real dimension $2K + 1$. The activation is the GELU $\phi(z) = z\Phi(z)$ with Φ the standard normal CDF, which is asymptotically linear on the right ($\phi(z)/z \rightarrow 1$ as $z \rightarrow +\infty$, $\phi(z) \rightarrow 0$ super-polynomially as $z \rightarrow -\infty$, and $|\phi(z)| \leq |z|$). For a fixed input u_0 the positive and negative parts are $g_+ = (u_0)_+$ and $g_- = (u_0)_-$, with band coefficients $\gamma_+(k) = \widehat{g_+}(k)$ and $\gamma_-(k) = \widehat{g_-}(k)$.

The FNO with a single spectral layer, of width n , scalar in and scalar out, retaining $|k| \leq K$, is

$$f_n(x) = \sum_{|k| \leq K} \hat{f}_n(k) e^{ikx}, \quad \hat{f}_n(k) = \frac{1}{(n \log n)^{1/\alpha}} \sum_{i=1}^n R_{i,k} Y_{i,k}, \quad Y_{i,k} = \widehat{\phi(a_i u_0)}(k), \quad (1)$$

with a lifting weight a_i for each channel, a spectral weight $R_{i,k}$ for each channel and mode, and the read-out folded into the spectral contraction. Reality of $\phi(a_i u_0)$ and the constraint $R_{i,-k} = \overline{R_{i,k}}$ give $\hat{f}_n(-k) = \overline{\hat{f}_n(k)}$, so $f_n \in V_K^{\mathbb{R}}$ almost surely.

Assumption 1 (input). $u_0 \in L^2(\mathbb{T}; \mathbb{R})$, $u_0 \neq 0$, and nondegenerate on the retained band: $(\gamma_+(k), \gamma_-(k)) \neq (0, 0)$ for every $|k| \leq K$. A mode where both vanish is null in the limit and is dropped from the active band.

Assumption 2 (lifting weights). The weights $a_1, \dots, a_n$ are i.i.d., symmetric, with an exactly regularly varying tail of index $\alpha \in (0, 2)$: $\mathbb{P}(|a_1| > t) \sim c_a t^{-\alpha}$ as $t \rightarrow \infty$, $c_a > 0$.

Assumption 3 (spectral weights). For $0 \leq k \leq K$ the weights $R_{i,k}$ have real and imaginary parts i.i.d. symmetric with $\mathbb{P}(|\operatorname{Re} R_{1,k}| > t) \sim c_R t^{-\alpha}$, the same index α , with $R_{i,0}$ real and $R_{i,-k} = \overline{R_{i,k}}$. The families $\{a_i\}$ and $\{R_{i,k}\}$ are independent, i.i.d. across the channel index i , and independent across the mode index k . Cross-mode independence of the fresh spectral weights keeps the single-layer modes decoupled.

The exact tails fix the normalization to $(n \log n)^{-1/\alpha}$. A heavier slowly varying factor would change the power of the logarithm. This is the normal domain of attraction with the slowly varying part pinned.

A.1.1 Proof of Theorem 1

Theorem 1. *Under heavy-tailed initialization of a mode-truncated single-layer FNO, with i.i.d. symmetric weights of exactly regularly varying tail index $\alpha \in (0, 2)$, the width limit is, in distribution in $L^2(\mathbb{T})$, a symmetric α -stable operator field with independent modes, symbol (α, Γ) , and a scale σ_k^α for each mode, under normalization $(n \log n)^{-1/\alpha}$. As $\alpha \rightarrow 2$ the limit is structurally the Gaussian neural-operator process of de Souza et al. (2025), and the constants coincide for piecewise-linear activations.*

The channel summand is $T_{i,k} = R_{i,k}Y_{i,k}$, collected as the real vector $\mathbf{T}_i = (T_{i,k})_{|k| \leq K}$ of dimension $2K + 1$ through the coordinates $(T_{i,0}, \operatorname{Re} T_{i,k}, \operatorname{Im} T_{i,k})$, with norming $a_n = (n \log n)^{1/\alpha}$. Four lemmas and an assembly give the limit.

Step 0: reduction to a vector limit. The map $J : \mathbb{R}^{2K+1} \rightarrow V_K^\mathbb{R}$ sending a coordinate vector to its trigonometric polynomial is a linear bijection, and by Parseval the $L^2(\mathbb{T})$ norm is equivalent to the Euclidean norm, so J and J^{-1} are continuous. With $\mathbf{S}_n = a_n^{-1} \sum_i \mathbf{T}_i$ one has $f_n = J(\mathbf{S}_n)$, so by the continuous mapping theorem f_n converges in $L^2(\mathbb{T})$ if and only if $\mathbf{S}_n$ converges in $\mathbb{R}^{2K+1}$. Tightness is automatic in finite dimension, so the task is the multivariate stable limit of $\mathbf{S}_n$.

Lemma 1 (activation tail). *For each $|k| \leq K$ the map $a \mapsto Y(a) = \widehat{\phi(au_0)}(k)$ satisfies $Y(a)/a \rightarrow \gamma_+(k)$ as $a \rightarrow +\infty$ and $Y(a)/|a| \rightarrow \gamma_-(k)$ as $a \rightarrow -\infty$. The random vector $Y_{1,k}$ is then regularly varying of index α , with $\mathbb{P}(|Y_{1,k}| > t) \sim \frac{c_a}{2} t^{-\alpha} (|\gamma_+(k)|^\alpha + |\gamma_-(k)|^\alpha)$, and its angular limit is the asymmetric two-atom measure placing the positive tail of a_1 on the ray $\gamma_+(k)$ and the negative tail on the ray $\gamma_-(k)$.*

Proof. For $a \rightarrow +\infty$ the ratio $\phi(au_0(x))/a = u_0(x)\Phi(au_0(x)) \rightarrow u_0(x)\mathbf{1}_{\{u_0(x)>0\}} = g_+(x)$ pointwise, dominated by $|u_0| \in L^1(\mathbb{T})$ because $|\phi(z)| \leq |z|$, so dominated convergence gives $Y(a)/a \rightarrow \gamma_+(k)$. The branch $a \rightarrow -\infty$ gives $Y(a)/|a| \rightarrow \gamma_-(k)$ by the same argument on $\{u_0 < 0\}$. For large $|a|$ the coefficient $Y(a)$ is therefore a fixed complex slope times the index- α symmetric variable a_1 , and regular variation of a_1 transfers through this asymptotically linear map to the two rays with the stated masses (Resnick, 2007). Nondegeneracy (Assumption 1) makes the limit nonzero. $\square$

Lemma 2 (product tail and the $(n \log n)^{-1/\alpha}$ norming). *For each $|k| \leq K$ the marginal tail is $\mathbb{P}(|T_{1,k}| > t) \sim C_k t^{-\alpha} \log t$, a regularly varying tail of index α with slowly varying part $\log t$, where $C_k = \alpha c_R c_a (|\gamma_+(k)|^\alpha + |\gamma_-(k)|^\alpha)$ for $1 \leq k \leq K$ and $C_0 = \alpha c_R \frac{c_a}{2} (|\gamma_+(0)|^\alpha + |\gamma_-(0)|^\alpha)$. The sequence $a_n = (n \log n)^{1/\alpha}$ then satisfies $n \mathbb{P}(|T_{1,k}| > \kappa_k a_n) \rightarrow 1$ for a constant κ_k .*

Proof. The factor $T_{1,k} = R_{1,k}Y_{1,k}$ is a product of two independent variables, each of exact tail index α by Assumption 3 and Lemma 1. For independent nonnegative X, Y with $\mathbb{P}(X > t) \sim c_X t^{-\alpha}$ and $\mathbb{P}(Y > t) \sim c_Y t^{-\alpha}$,

$$\mathbb{P}(XY > t) = \int_0^\infty \mathbb{P}(X > t/y) dF_Y(y) \sim c_X t^{-\alpha} \int_1^t y^\alpha dF_Y(y) \sim c_X t^{-\alpha} \cdot \alpha c_Y \log t,$$

where $\int_1^s y^\alpha dF_Y(y) = \alpha c_Y \log s + O(1)$ by Karamata's theorem. The logarithm is produced by the intermediate splits $y \in [1, t]$ across all log-scales, while the two endpoint regimes each contribute a vanishing $O(t^{-\alpha})$ fraction. The complex modulus $|R_{1,k}|$ has $c_X = 2c_R$ for $k \geq 1$ from its two i.i.d. index- α parts and $c_X = c_R$ for the real mode $k = 0$, and c_Y comes from Lemma 1, which yields C_k (Embrechts and Goldie, 1980; Cline, 1986). The balance $n C_k a_n^{-\alpha} \log a_n \rightarrow 1$ with $\log a_n \sim \alpha^{-1} \log n$ gives $a_n \asymp (n \log n)^{1/\alpha}$, and the fixed choice $a_n = (n \log n)^{1/\alpha}$ absorbs κ_k into the limiting scale. $\square$

The product structure is the reason the norming carries a logarithmic correction. Each channel summand multiplies two independent exact-index- α tails, and such a product is index α with a slowly varying log factor, so the classical $n^{-1/\alpha}$ norming would leave the sum divergent. This effect is specific to asymptotically linear activations (Bordino et al., 2023).

Lemma 3 (the Lévy measure sits on single-mode subspaces). *For distinct modes $k \neq k'$ in $\{0, \dots, K\}$ and any $\varepsilon > 0$, $n \mathbb{P}(|T_{1,k}| > \varepsilon a_n \text{ and } |T_{1,k'}| > \varepsilon a_n) \rightarrow 0$. The vague limit $\nu = \lim_n n \mathbb{P}(\mathbf{T}_1/a_n \in \cdot)$ therefore charges no direction with two modes simultaneously nonzero.*

Proof. Both $T_{1,k}$ and $T_{1,k'}$ share the single heavy factor a_1 through $Y_{1,k}, Y_{1,k'}$, while $R_{1,k}, R_{1,k'}$ are independent index- α . After conditioning on a_1 , for large $|a_1|$ the moduli satisfy $|Y_{1,k}| \lesssim |a_1|$, so by independence of $R_{1,k}, R_{1,k'}$ the conditional joint exceedance is of order $(|a_1|/\varepsilon a_n)^\alpha \wedge 1$ in each factor. The expectation over a_1 , split at $|a_1| = a_n$, is $O(a_n^{-\alpha})$ from both the tail region and the bulk region $\int_1^{a_n} (y/a_n)^{2\alpha} d(y^{-\alpha}) \asymp a_n^{-\alpha}$. The joint exceedance is thus $O(a_n^{-\alpha})$ with no logarithm, so $n\mathbb{P}(\cdots) = O(na_n^{-\alpha}) = O(1/\log n) \rightarrow 0$. The joint two-mode jumps are negligible against the marginals that set the norming and carry a logarithm, which is the multivariate-regular-variation criterion for asymptotic independence (Resnick, 2007; Samorodnitsky and Taqqu, 1994). $\square$

Lemma 4 (angular measure within a mode). *Fix $k \geq 1$. Inside the mode- k plane $\mathbb{R}^2 \cong \mathbb{C}$ the vague limit $\nu_k = \lim_n n\mathbb{P}(T_{1,k}/a_n \in \cdot)$ exists and equals a radial $\alpha r^{-\alpha-1} dr$ times an atomic angular measure Λ_k supported on the directions $\{\pm r_+(k), \pm r_-(k), \pm i r_+(k), \pm i r_-(k)\}$ with $r_\pm(k) = \gamma_\pm(k)/|\gamma_\pm(k)|$, the mass on the $\gamma_\pm$ -rays proportional to $c_R |\gamma_\pm(k)|^\alpha$. For $k = 0$ the coefficient is real and Λ_0 is symmetric on $\{\pm 1\}$.*

Proof. By Lemma 2 the dominant large- $|T_{1,k}|$ event is an intermediate split in which a_1 and exactly one of $\text{Re } R_{1,k}, \text{Im } R_{1,k}$ are large, the two spectral parts being independent index- α . On the event that $\text{Re } R_{1,k}$ is large, $T_{1,k} \approx \text{Re } R_{1,k} \cdot |a_1| \gamma_{\text{sgn } a_1}(k)$ is a large real scalar along the fixed ray $\mathbb{R} \cdot \gamma_\pm(k)$, an atom at $\pm r_\pm(k)$. The event that $\text{Im } R_{1,k}$ is large gives the i -rotated atoms, and symmetry of the weights supplies both signs. The support and the $c_R |\gamma_\pm(k)|^\alpha$ proportionality are thereby fixed, and the split integral apportions the plane mass in closed form. Because $\text{Re } R_{1,k}$ and $\text{Im } R_{1,k}$ are i.i.d. of index α , the real ray $\hat{\gamma}_\pm(k) = \gamma_\pm(k)/|\gamma_\pm(k)|$ and the i -rotated ray carry equal mass. Averaging over the sign of a_1 places atoms along $\pm \hat{\gamma}_+(k), \pm i \hat{\gamma}_+(k)$ from the $a_1 > 0$ branch and along $\pm \hat{\gamma}_-(k), \pm i \hat{\gamma}_-(k)$ from the $a_1 < 0$ branch, the two branches in mass ratio $|\gamma_+(k)|^\alpha : |\gamma_-(k)|^\alpha$. Each of the four atoms in a branch carries one quarter of that branch's mass, and the two branches sum to the total angular mass $\Lambda_k(S^1) = C_k/\alpha$, where $C_k = \alpha c_R c_\alpha (|\gamma_+(k)|^\alpha + |\gamma_-(k)|^\alpha)$ is the marginal tail constant of Lemma 2. A ray with a vanishing coefficient carries no mass and is dropped. The marginal scale $\sigma_k^\alpha = \int |s_k|^\alpha \Lambda_k(ds)$ depends on this total alone, so the apportionment changes neither the mode scales nor their independence. $\square$

Assembly. By Lemmas 3 and 4 the vague limit is $\nu = \sum_{k=0}^K \nu_k$, each ν_k on the mode- k subspace, so $\mathbf{T}_1$ is multivariate regularly varying of index α with Lévy measure ν . Symmetry of the weights removes drift and skewness for all $\alpha \in (0, 2)$, and the truncated-second-moment condition holds by Karamata since $\alpha < 2$, so the multivariate generalized central limit theorem gives $\mathbf{S}_n = a_n^{-1} \sum_i \mathbf{T}_i \Rightarrow Z$, a symmetric α -stable vector whose spectral measure $\Gamma = \sum_k \Lambda_k$ is the angular part of ν (Rvačeva, 1962; Meerschaert and Scheffler, 2001). Because Γ charges no direction with two modes at once, the axis-support criterion for stable vectors gives mutual independence of $Z_0, \dots, Z_K$ (Samorodnitsky and Taqqu, 1994), which is part (i). The characteristic function $\mathbb{E} e^{i\langle \theta, Z \rangle} = \exp(-\int |\langle \theta, s \rangle|^\alpha \Gamma(ds))$ with $\Gamma = \sum_k \Lambda_k$ gives parts (ii) and (iii), the scale of each mode being $\sigma_k^\alpha = \int |s_k|^\alpha \Gamma(ds)$. The form of this width limit adapted to the network, with the $(n \log n)^{-1/\alpha}$ norming for asymptotically linear activations, matches Jung et al. (2023); Bordino et al. (2023). Transport by J carries the convergence to $f_n \Rightarrow f = J(Z)$ in $L^2(\mathbb{T})$.

Lemma 3 is the delicate step. The shared lifting weight a_i induces no cross-mode dependence in the limit because all factors share the same index α , so joint two-mode jumps are $O(a_n^{-\alpha})$ against a marginal of order $a_n^{-\alpha} \log n$ with its extra logarithm, and the ratio $O(1/\log n)$ vanishes. A heavier shared factor, or a Breiman regime with one factor of finite α -moment, would couple the modes and drop the logarithm. The convergence to a symmetric α -stable field on $V_K^{\mathbb{R}}$, the $(n \log n)^{-1/\alpha}$ rate, and mode independence are the robust conclusions. The exact angular measure within each mode is given in closed form by Lemma 4, and the mode scales depend only on its total mass.

A.1.2 Proof of Corollary 1

Corollary 1 (Gaussian operator field at $\alpha = 2$). *Regard $\alpha \in (0, 2)$ as a parameter and let $f^{(\alpha)}$ have spectral measure $\Gamma^{(\alpha)} = \sum_k \Lambda_k^{(\alpha)}$. The atom locations of $\Lambda_k^{(\alpha)}$ depend on u_0 only, through $\gamma_\pm(k)$, and the atom weights $\propto c_R |\gamma_\pm(k)|^\alpha$ vary continuously in α , so $\Gamma^{(\alpha)} \rightarrow \Gamma^{(2)}$ weakly and $f^{(\alpha)} \Rightarrow f^{(2)}$, a mean-zero Gaussian field on $V_K^{\mathbb{R}}$ with independent modes and block-diagonal covariance $\Sigma_{ab} = 2 \int s_a s_b \Gamma^{(2)}(ds)$. This is the same field class as the infinite-width FNO prior of de Souza et al. (2025). The correspondence is structural. Their activation enters through the arc-cosine dual kernel of the whole nonlinearity, whereas the limit here enters through the one-sided slopes $\gamma_\pm(k)$, so the variance spectrum of each mode coincides exactly only for a piecewise-linear activation, where the dual kernel is fixed by those slopes.*

Proof. For fixed θ the integrand $|\langle \theta, s \rangle|^\alpha$ is bounded on the sphere, and $x \mapsto x^\alpha$ converges to $x \mapsto x^2$ uniformly on the bounded range, so with $\Gamma^{(\alpha)} \rightarrow \Gamma^{(2)}$ weakly the stable exponent converges to $\theta^\top M \theta$ with $M_{ab} = \int s_a s_b \Gamma^{(2)}(ds)$. The characteristic functions then converge pointwise to $\exp(-\theta^\top M \theta)$, continuous at the origin, so Lévy continuity gives a Gaussian limit with $\Sigma = 2M$. Independence of modes passes to the limit because $\Gamma^{(2)}$ is again supported on single-mode subspaces, giving block-diagonal Σ . Transport by J gives the field statement. The order of limits is $n \rightarrow \infty$ at fixed α , then $\alpha \rightarrow 2$, so no interchange is used. A direct construction at $\alpha = 2$ carries the boundary tail $t^{-2} \log t$ and the norming $\sqrt{n} \log n$, corrected by a logarithm and specific to the asymptotically linear product structure here, and returns a Gaussian field of the same class as the infinite-width FNO prior of de Souza et al. (2025) under the standard $\sqrt{n}$ norming. $\square$

A.1.3 Proof of Theorem 2

The depth argument keeps the setting of the initialization prior and the symbol, the pair (α, Γ) . One block applies a spectral convolution with fresh heavy-tailed weights for each mode, a pointwise linear skip with a single heavy-tailed weight b shared across all modes, and a pointwise GELU. The decisive asymmetry is that the spectral multiplier draws a new weight for each mode while the skip multiplies the whole field by one scalar. The depth statement is proven for exactly symmetric α -stable block weights, for which the conditional spectral measure of each block is exact. The exactly regularly varying class of Theorem 1 follows the same point-process limit, outlined here.

Theorem 2. *Consider a depth- L mode-truncated FNO with a pointwise skip connection. Take the width limit one layer at a time. For $L = 2$ one coupled skip block gives a symmetric α -stable field with symbol (α, Γ_2) . The shared skip weight makes a dominant input mode m_0 deposit Fourier mass on its harmonic family, so harmonic mode pairs jump together, and Γ_2 is non-diagonal, diagonal if and only if the skip vanishes. For general depth L the same sequential limit supports Γ_L on nested harmonic sets, diagonal if and only if every skip vanishes, under a generic nondegeneracy condition. The simultaneous width limit is open, and the result holds along sufficiently fast width paths.*

The block acts on a discrete-spectrum SaS input field, and the coefficient of output mode k is $A_k = c_n \sum_i [R_i(k) + b_i] \hat{g}_i(k)$ with $\hat{g}_i = \widehat{\phi(z_i)}$ and normalization $c_n = (n \log n)^{-1/\alpha}$. In the sequential width limit the block-2 input is the single-layer field of Theorem 1, so two of the earlier hypotheses become lemmas. The mean-field width limit holds because the block channels are independent copies of that field. The one-large-jump tail transfer holds by a uniform rectification bound and the mapping theorem for regularly varying measures. Only the finite retained band stays structural. The joint width limit, where both widths grow together, would instead need the deep-network result of Favaro et al. (2020), which holds under a sublinear activation envelope and so leaves GELU open.

Lemma 5 (index preservation for asymptotically linear activations). *Let the input value z be SaS with scale σ and let the weights be fresh SaS(1). For a GELU-type activation with $\phi(t)/t \rightarrow 1$ as $t \rightarrow +\infty$ and $\phi(t)/t \rightarrow 0$ as $t \rightarrow -\infty$, the pointwise width limit $c_n \sum_i w_i \phi(z_i)$ under $c_n = (n \log n)^{-1/\alpha}$ is SaS($K^{1/\alpha}$) with $K = c_\alpha^+ \sigma^\alpha$. The index α is preserved and the norming carries the extra logarithm.*

Proof. Conditional on $\{z_i\}$ the sum $\sum_i w_i \phi(z_i)$ is SaS with scale $S_n^{1/\alpha}$, $S_n = \sum_i |\phi(z_i)|^\alpha$, so the conditional characteristic function is $\exp(-c_n^\alpha S_n |\theta|^\alpha)$ and the limit is fixed by that of $c_n^\alpha S_n$. Asymptotic linearity gives $u^\alpha \mathbb{P}(|\phi(z)| > u) \rightarrow c_\alpha^+ \sigma^\alpha$, so $Y = |\phi(z)|^\alpha$ has tail $\mathbb{P}(Y > t) \sim K t^{-1}$ of index 1 with infinite mean. The Feller weak law at index 1 then applies with truncation at $b_n = n \log n$: the maximum term exceeds b_n with probability $O(1/\log n) \rightarrow 0$, the truncated mean is $m(b_n) \sim K \log n$, and Chebyshev gives $S_n/(n \log n) \rightarrow K$ in probability. Hence $c_n^\alpha S_n \rightarrow K$ and the limit is SaS($K^{1/\alpha}$). A bounded or strictly sublinear activation instead has $\mathbb{E}Y < \infty$ and the classical $n^{-1/\alpha}$ norming, and a superlinear $|\phi(t)| \sim c|t|^\beta$ degrades the index to α/β , so the logarithm is the specific signature of asymptotic linearity (Bordino et al., 2023). $\square$

Lemma 6 (exact conditional spectral measure). *Condition on $\{g_i\}$. The output vector $\mathbf{A} = (A_k)_k$ is SaS with spectral measure*

$$\Gamma_n^{\text{cond}} = \underbrace{c_n^\alpha \sum_i \sum_k m_k |\hat{g}_i(k)|^\alpha \lambda_{i,k}}_{\text{spectral term, one plane for each mode}} + \underbrace{c_n^\alpha \sum_i \|\hat{\mathbf{g}}_i\|_\alpha^\alpha \frac{1}{2} (\delta_{u_i} + \delta_{-u_i})}_{\text{skip term, } u_i = \hat{\mathbf{g}}_i / \|\hat{\mathbf{g}}_i\|_\alpha},$$

where $\lambda_{i,k}$ is the angular probability measure within the plane of the fresh complex weight, the plane mass is $m_0 = 1$ and $m_k = 2$ for $k \geq 1$ because the complex weight contributes two independent real components on the mode- k plane, and $\|\hat{\mathbf{g}}_i\|_\alpha^\alpha = \sum_k |\hat{g}_i(k)|^\alpha$. Every spectral atom lies in one mode- k plane, so the spectral term couples no two modes. The skip atom at $\pm u_i$ lies off every single-mode plane whenever $\hat{\mathbf{g}}_i$ loads two or more modes.

Proof. Independence of the fresh weights across (i, k) splits $\mathbf{A}$ into two independent pieces. The spectral piece uses a distinct weight for each mode, so each term $R_i(k)\hat{g}_i(k)$ lives in the mode- k plane, real for $k = 0$ and along the orthogonal real and imaginary pair for $k \geq 1$. The real and imaginary parts of the complex weight are two independent SoS components, so the plane carries the mass $m_k|\hat{g}_i(k)|^\alpha$ of the statement, giving the mode-diagonal term. The skip piece multiplies the whole vector $\hat{\mathbf{g}}_i$ by the single scalar $b_i \sim \text{SoS}(1)$, so $b_i\hat{\mathbf{g}}_i$ is SoS on the line through $\hat{\mathbf{g}}_i$ with spectral measure $\|\hat{\mathbf{g}}_i\|_\alpha^\alpha \frac{1}{2}(\delta_{u_i} + \delta_{-u_i})$. Summing over channels and applying c_n gives the display. The within-plane masses are $m_0 = 1$ and $m_k = 2$ for $k \geq 1$, from the two independent real and imaginary parts of the complex weight. $\square$

Proposition 4 (a dominant mode lands mass on a harmonic family). *Under the one-large-jump assumption and a half-wave rectifying activation ($a_+ = 1, a_- = 0$), condition on a dominant input coefficient B_{im_0} . Then $z_i(x) = 2|B_{im_0}|\cos(m_0x + \psi_i) + O(1)$ and $g_i(x) = 2|B_{im_0}|\rho(m_0x + \psi_i) + o(|B_{im_0}|)$ with $\rho(\theta) = \max(\cos\theta, 0)$ the half-wave rectified cosine, whose coefficients are $\hat{\rho}_0 = \frac{1}{\pi}$, $\hat{\rho}_{\pm 1} = \frac{1}{4}$, $\hat{\rho}_{\pm 2k} = \frac{1}{\pi} \frac{(-1)^{k+1}}{4k^2-1}$, and zero on odd harmonics above one. A single dominant mode m_0 therefore deposits Fourier mass on the whole in-band harmonic family $\mathcal{H}(m_0) = \{km_0 : k \in \{0, 1, 2, 4, 6, \dots\}, |km_0| \leq N\}$ at once, in the fixed proportion $(\hat{\rho}_k)$, so the direction u_i concentrates on the fixed off-axis direction $v^{(m_0)}$ supported on $\mathcal{H}(m_0)$.*

Proof. Under dominance the other modes contribute an $O(1)$ field, so the leading term is a large-amplitude cosine. Asymptotic linearity with $a_+ = 1, a_- = 0$ gives $\phi(t) = \max(t, 0) + o(|t|)$, and substitution of the cosine gives $g_i = 2|B_{im_0}|\rho(m_0x + \psi_i) + o(|B_{im_0}|)$. The listed coefficients are the classical half-wave rectified-cosine series, computed from $\hat{\rho}_k = \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} \cos\theta e^{-ik\theta} d\theta$. The coefficient at mode km_0 is $\hat{\rho}_k e^{ik\psi_i}$, nonzero exactly on the stated family, so the limiting direction follows. $\square$

Lemma 7 (the case with no skip is diagonal, by a logarithmic margin). *With $b \equiv 0$ the conditional measure of Lemma 6 is supported on the single-mode planes, so the limit symbol Γ is diagonal and the modes are independent. Diagonality holds only by a logarithmic margin. Under the one-large-jump assumption one large input coefficient makes both harmonic output coordinates large at once, so the spectral joint jump of a pair from the fresh weights is of order $t^{-\alpha}$ with no logarithm, seeded by that single input jump, while the mechanism needing two independent large weights is the strictly subdominant $t^{-2\alpha}$. The marginal carries a logarithm, at $t^{-\alpha} \log t$, so $n\mathbb{P}(\text{spectral joint} > \varepsilon a_n) \asymp \varepsilon^{-\alpha} / \log n \rightarrow 0$ while the marginal count stays positive.*

Proof. The fresh weights for each mode randomize the joint direction of the spectral pair, so its off-axis mass is killed by the single factor $1/\log n$ between the joint order $a_n^{-\alpha}$ and the marginal $a_n^{-\alpha} \log n$, with its logarithm, that fixes the norming. The count with the shared factor is the one in Lemma 3. The remaining step is the marked weak law, the index-1 weak law of Lemma 5 applied to the marks $V_i h(u_i)$ with $V_i = \|\hat{\mathbf{g}}_i\|_\alpha^\alpha$ and h continuous on the sphere, whose index-1 tail follows from the regular variation of $\hat{\mathbf{g}}_i$, and bounded convergence of the exact conditional characteristic function closes the limit. $\square$

The case $L = 2$. Composition of the single spectral layer of Theorem 1 with one coupled skip block, on the diagonal independent-coefficient input that meets the one-large-jump assumption, gives a symmetric α -stable depth-2 field with symbol (α, Γ_2) . The spectral term is mode-diagonal by Lemma 6, placing mass $m_k K_k$ on mode k with $m_0 = 1$ and $m_k = 2$ for $k \geq 1$, where $c_n^\alpha \sum_i |\hat{g}_i(k)|^\alpha \rightarrow K_k$ by Lemma 5. The skip term is off-diagonal. Because b_i is shared across modes, one input jump along mode m_0 makes the whole vector $\hat{\mathbf{g}}_i$ jump along the fixed direction $v^{(m_0)}$ of Proposition 4, and this contribution is the product $b_i \cdot A$ of two independent index- α tails, so it carries a logarithm, at $t^{-\alpha} \log t$, and survives the same norming, whereas the spectral joint from the fresh weights dies by Lemma 7. A marked weak law then places off-axis atoms $w_{m_0} \Theta_{m_0}$ on the harmonic families $\mathcal{H}(m_0)$, so Γ_2 charges directions off the coordinate axes. In the sequential width limit the passage from the exact conditional measure to Γ_2 is a marked weak law over the independent channel copies, so the result is unconditional modulo the finite band. The off-axis atom mass is $w_{m_0} = 2^\alpha C_\alpha \lambda \|(\hat{\rho}_k)\|_\alpha^\alpha$ with

$C_\alpha = (1 - \alpha)/(\Gamma(2 - \alpha) \cos(\pi\alpha/2))$ for $\alpha \neq 1$, removable at $\alpha = 1$ with value $C_1 = 2/\pi$, and the single-layer atom weight λ is fixed in closed form by Lemma 4, the total mass of mode m_0 . The simultaneous width limit, where both widths grow together, stays open.

Dependence of the coupled modes follows from Γ_2 charging off-axis directions. Spectral-measure uniqueness for $\alpha < 2$ and the axis-support criterion make any set of modes loaded by a common off-axis atom dependent (Samorodnitsky and Taqqu, 1994). A covariation is not used for this. For $1 < \alpha < 2$ the covariation of a coupled pair can vanish under a uniform-phase input, since the phase average $\int_0^{2\pi} e^{i\psi} e^{2i\psi} d\psi = 0$ annihilates the fundamental-to-second-harmonic pair, so a null covariation carries no information about independence here. The scale of each mode is $\sigma_k^\alpha = \int |s_k|^\alpha \Gamma_2(ds)$ over all of Γ_2 , off-axis atoms included, and equals the axis mass K_k only in the diagonal case $b \equiv 0$.

General depth. For $L \geq 3$ each block sees a coupled input. In the fully sequential width limit the layer- ℓ input is jointly SaS with symbol $(\alpha, \Gamma^{(\ell)})$ and a finite atomic $\Gamma^{(\ell)}$, so its LePage series represents it as independent stable scalars along the atoms. Its large fluctuations then follow a one-large-jump along a single atom v_a carrying a real band-limited waveform P_a , so the general-symbol hypothesis is a theorem here, not an assumption. The following lemma is the deterministic core of the induction.

Lemma 8 (rectification seeds an in-band nested harmonic set). *For a GELU-type activation and a real band-limited waveform P_a with support S_a , uniformly as $A \rightarrow +\infty$, $\phi(AP_a(x)) = AP_a^+(x) + o(A)$ with $P_a^+ = \frac{1}{2}(P_a + |P_a|)$, so $\hat{g}_i(m) = A \widehat{P_a^+}(m) + o(A)$. Because $|P_a|$ has corners, $\widehat{P_a^+}$ is supported on a set $S'_a \supseteq S_a$ that is generically strictly larger, and its in-band part $\rho(a) = \widehat{P_a^+}|_{\mathcal{N}}$ is a fixed off-axis direction. The single-harmonic case recovers Proposition 4.*

Proof. Pointwise $\phi(AP_a(x))/A \rightarrow \max(P_a(x), 0) = P_a^+(x)$ off the finite zero set of the trigonometric polynomial P_a , dominated by $|P_a|$, so the convergence is in $L^2(\mathbb{T})$ and the coefficients converge to $\widehat{P_a^+}$. The identity $P_a^+ = \frac{1}{2}(P_a + |P_a|)$ and the Lipschitz corner of $|P_a|$ give a Fourier tail of order m^{-2} , hence infinitely many nonzero harmonics unless a symmetry cancels them, and the retained band keeps the in-band part. $\square$

The depth- L width limit is then a symmetric α -stable field with symbol (α, Γ_L) , $\Gamma_L = \mathcal{R}^{L-1}(\Gamma_1)$, where the layer map $\mathcal{R}$ adds one off-axis atom at $\rho(a)$ for each input atom v_a through the shared skip, at the order that carries a logarithm, and keeps the fresh spectral term mode-diagonal. Iteration supports Γ_L on the nested harmonic sets inside the fixed band, and Γ_L is diagonal if and only if every skip vanishes, since $b \equiv 0$ at every layer deletes the skip term of $\mathcal{R}$. The compounding mechanism of Lemma 8, its atom locations and orders, and the diagonal criterion are proven. In the fully sequential width limit the passage from single-channel tail orders to $\mathcal{R}(\Gamma^{(\ell)})$ is the same marked weak law as at $L = 2$, invoked once for each layer, and its i.i.d.-channel hypothesis holds at every layer because the output channels of a block share only the LePage atom scalar of their common input, whose joint jump across two channels integrates two independent fresh-weight tails against the exact index- α atom law and so carries no logarithm, dying by the $1/\log n$ margin while the log-enhanced marginal survives. The depth- L statement therefore holds in the sequential limit, conditional only on a nondegeneracy assumption: at every layer, for every atom v_a of $\Gamma^{(\ell)}$, the in-band rectified image $\rho(a)$ is nonzero and loads a mode outside the support of v_a . The nonvanishing $\rho(a) \neq 0$ is automatic, since each atom pair keeps at least one sign branch with a rectified waveform of positive mean, so only the off-axis loading is assumed, and the family enlargement is not needed for the diagonal criterion. Rectification preserves the $2\pi/m_0$ period of a harmonic-family atom, so its new in-band modes are the odd multiples of m_0 . These are present when $3m_0 \leq K$ and absent for an atom near the band edge, so Γ_L loads the nested harmonic sets only for atoms with in-band odd harmonics. Weak convergence on $L^2(\mathbb{T})$ is metrizable, so the sequential limit extends along any width path that grows fast enough. With the deeper widths fixed the earlier width converges, and a diagonal choice of widths carries every sufficiently fast path to the same field. The simultaneous width limit stays open.

A.2 Retrieval consistency

The three retrieval propositions work on the discretized torus with N grid points, retained band $\mathcal{K} = \{-K, \dots, K\}$, and DFT $\hat{u}(k) = (\mathcal{F}u)(k)$. A translation-invariant operator acts diagonally as $\widehat{Au}(k) = \hat{p}(k)\hat{u}(k)$ with multiplier $\hat{p}$. The single Fourier-multiplier layer $u \mapsto \mathcal{F}^{-1}(R[\cdot]\mathcal{F}u)$ carries a learnable weight $R[k]$ for

each mode, so R is the model symbol. A symmetric α -stable law $\text{SaS}(\sigma)$ has $|\varphi(\theta)| = \exp(-\sigma^\alpha |\theta|^\alpha)$, so $-\log |\varphi(\theta)| = \sigma^\alpha |\theta|^\alpha$.

A.2.1 Proof of Proposition 1

Proposition 1. *A single Fourier-multiplier layer fit by least squares to input/output pairs of a translation-invariant operator with multiplier $\hat{p}(k)$ recovers $R[k] \rightarrow \hat{p}(k)$ on every mode of nonzero input energy. For the target $e^{-t|\xi|^\alpha}$ the stability index is identified by $\log(-\log R[k]) = \log t + \alpha \log |k|$.*

Proof. Let the pairs be $\{(u^{(i)}, v^{(i)})\}_{i=1}^m$ with $\hat{v}^{(i)}(k) = \hat{p}(k)\hat{u}^{(i)}(k) + \eta^{(i)}(k)$ and $\eta^{(i)}(k)$ zero-mean output noise independent of the inputs, the noiseless case being $\eta \equiv 0$. By Parseval the least-squares loss decouples across modes into $\sum_k L_k(R[k])$ with $L_k(r) = \frac{1}{m} \sum_i |r\hat{u}^{(i)}(k) - \hat{v}^{(i)}(k)|^2$. Each L_k is a strictly convex quadratic in the single complex variable r whenever the mode energy $E_m(k) = \sum_i |\hat{u}^{(i)}(k)|^2$ is positive, and the Wirtinger normal equation $\partial_{\bar{r}} L_k = 0$ gives the closed form

$$R^*[k] = \frac{\sum_i \overline{\hat{u}^{(i)}(k)} \hat{v}^{(i)}(k)}{\sum_i |\hat{u}^{(i)}(k)|^2} = \hat{p}(k) + \frac{\sum_i \overline{\hat{u}^{(i)}(k)} \eta^{(i)}(k)}{E_m(k)}.$$

For $\eta \equiv 0$ this is exact, $R^*[k] = \hat{p}(k)$ whenever $E_m(k) > 0$. For zero-mean output noise of bounded variance with $\frac{1}{m} E_m(k) \rightarrow c_k > 0$, the estimator is unbiased and $\text{Var } R^*[k] \leq s^2/E_m(k) = O(1/m)$, so $R^*[k] \rightarrow \hat{p}(k)$ in L^2 . This L^2 consistency clause covers light-tailed output noise only. The infinite-variance α -stable case of the main text is handled by the residual characteristic-function estimator of Proposition 3, which is built for heavy tails. For the isotropic α -stable transition target $\hat{p}(k) = e^{-t|k|^\alpha}$ with $k \neq 0$, exactness gives $-\log R^*[k] = t|k|^\alpha > 0$, so a second logarithm gives $\log(-\log R^*[k]) = \log t + \alpha \log |k|$, an affine relation whose ordinary least squares slope is α and intercept $\log t$, exactly in the noiseless case and consistently under output noise. The mode $k = 0$ is excluded because $R^*[0] = 1$. Training uses gradient descent on the convex separable loss and converges to R^* , and a symbol of large dynamic range such as $1/k^2$ makes the curvature $E_m(k)$ of each mode span a wide range, so the problem is ill-conditioned but convex. $\square$

At large $|k|$ the multiplier $e^{-t|k|^\alpha}$ underflows, so $R^*[k]$ carries large relative error and $-\log R^*[k]$ is unstable, and a practical fit restricts to a band where $R^*[k]$ is bounded away from 0 and 1. Within that band the recovery reaches the reported relative L_2 error near 7×10^{-7} for the stable symbol.

A.2.2 Proof of Proposition 2

The method from the initialization prior extracts the noise from the initialization prior of the random operator, in the isotropic stationary-input idealization where the width-limit field with discrete spectrum is $X(x) = \sum_{k \in \mathcal{K}} R_k e^{ikx}$ with $R_{-k} = \overline{R_k}$ and R_k independent $\text{SaS}(\sigma_k)$. The representation is the law of the initialization prior of the random operator, indexed by spatial frequency, and its symbol is the pair (α, Γ) .

Proposition 2. *The discrete-spectrum field is non-ergodic. One trained network reveals its realized coefficients only. The law (α, Γ) follows from an ensemble of networks. The marginal scale $\sigma_k^\alpha = \int |s_k|^\alpha d\Gamma$ of each mode is estimable from an ensemble of m initializations at rate $\sqrt{m}$. The cross-mode coupling requires joint multi-mode statistics.*

Proof. Under the spatial shift $T_h : X(\cdot) \mapsto X(\cdot + h)$ the coefficients transform as $R_k \mapsto R_k e^{ikh}$, so each modulus $|R_k|$ is shift-invariant. A positive scale σ_k makes $|R_k|$ non-constant, so the invariant σ -field carries a non-degenerate random variable, which is the failure of ergodicity for a measure-preserving action. A single realization $X(\cdot; \omega)$ is a finite trigonometric sum with distinct frequencies, so its coefficients $R_k(\omega)$ are recovered exactly by projection, but they are one joint draw from $\prod_k \text{SaS}(\sigma_k)$, that is one sample from each mode, so no functional of one realization is a consistent estimator of $(\alpha, \{\sigma_k^\alpha\})$. An ensemble of m independent initializations makes each mode's sample $\{R_k^{(j)}\}_{j=1}^m$ into m i.i.d. $\text{SaS}(\sigma_k)$ draws, and the characteristic-function estimator of Proposition 3, applied to each mode with the shared stability index pooled across modes, returns $\hat{\alpha} \rightarrow \alpha$ and $\hat{\sigma}_k^\alpha \rightarrow \sigma_k^\alpha$ at the $\sqrt{m}$ rate. For $\alpha < 2$ the atomic spectral measure, and hence $(\alpha, \{\sigma_k^\alpha\})$, is uniquely determined by the law, so the target is well posed (Samorodnitsky and Taqqu, 1994). At $\alpha = 2$ the spectral measure is not unique, but the scale of each mode remains identifiable as a variance of the diagonal covariance, and only the finer angular structure within a mode is lost. $\square$

Non-ergodicity does not rest on isotropy. The obstruction is that each $|R_k|$ is a non-degenerate shift-invariant random variable, which holds for any non-degenerate atomic component, anisotropic ones included, so the isotropic idealization is used only to make T_h measure-preserving and to reduce each mode's marginal to a single scalar. At depth the propagated law is (α, Γ) with Γ non-diagonal, so the scales $\sigma_k^\alpha = \int |s_k|^\alpha d\Gamma$ are the marginal scales of the individual modes in the joint law, the integral running over all of Γ . These marginals stay estimable one mode at a time from the ensemble by the estimator for each mode. The cross-mode coupling is not a function of each mode's marginals, so its recovery requires joint multi-mode characteristic-function statistics on vector arguments that excite several modes at once, and only scalar α and the scale of each mode carry a $\sqrt{m}$ guarantee here. The full multivariate jump measure is a hard nonparametric problem and is not claimed.

A.2.3 Proof of Proposition 3

The residual method fits a mean map and extracts α from the residuals. The data are $u^{(i)} = G(a^{(i)}) + \epsilon^{(i)}$ with G the true deterministic operator and $\epsilon^{(i)} \sim \text{S}\alpha\text{S}(\sigma)$ of index $0 < \alpha < 2$ drawn fresh for each sample. The fitted operator $\hat{G}$ gives residuals $r^{(i)} = \epsilon^{(i)} + \delta^{(i)}$ with fit error $\delta^{(i)} = G(a^{(i)}) - \hat{G}(a^{(i)})$. The empirical characteristic function is $\hat{\varphi}(\theta) = \frac{1}{m} \sum_i e^{i\theta r^{(i)}}$, and its double-log regression on $\log |\theta|$ estimates α from the slope.

Proposition 3. *The empirical characteristic-function estimator of α from held-out residuals recovers the stability index when the fit error δ is lighter-tailed than the noise, $\mathbb{E}|\delta|^p < \infty$ for some $p \in (\alpha, 2]$. It then extracts the stability index regardless of the L_2 size of δ , and is asymptotically normal at the $\sqrt{m}$ rate around a slope within a controllable tolerance of α (Koutrouvelis, 1980; Koutrouvelis and Bauer, 1982). A fit error as heavy-tailed as the noise breaks the recovery.*

Proof. The held-out split makes δ and ϵ independent, so $|\varphi_r| = |\varphi_\delta| |\varphi_\epsilon|$ with $|\varphi_\epsilon(\theta)| = e^{-\sigma^\alpha |\theta|^\alpha}$, and $-\log |\varphi_r(\theta)| = \sigma^\alpha |\theta|^\alpha - \log |\varphi_\delta(\theta)|$ with $-\log |\varphi_\delta| \geq 0$. The noise term is a lower bound, so the misfit can only raise the exponent. Let $\delta_s = \delta - \delta'$ for an independent copy, so $|\varphi_\delta|^2 = \varphi_{\delta_s}$. The inequality $1 - \cos u \leq 2^{1-p} |u|^p$ for $p \in (0, 2]$ gives $1 - \varphi_{\delta_s}(\theta) = \mathbb{E}[1 - \cos \theta \delta_s] \leq 4 \mathbb{E}|\delta|^p |\theta|^p$, so on the range $4 \mathbb{E}|\delta|^p |\theta|^p \leq \frac{1}{2}$,

$$\sigma^\alpha |\theta|^\alpha \leq -\log |\varphi_r(\theta)| \leq \sigma^\alpha |\theta|^\alpha + 4 \mathbb{E}|\delta|^p |\theta|^p.$$

The double-log response is $\alpha \log |\theta| + \alpha \log \sigma + \eta(\theta)$ with $0 \leq \eta(\theta) \leq (4 \mathbb{E}|\delta|^p / \sigma^\alpha) |\theta|^{p-\alpha}$. The ordinary least-squares slope on the grid $x_\ell = \log |\theta_\ell|$ is the linear functional $\sum_\ell w_\ell y_\ell$ of the responses, with $w_\ell = (x_\ell - \bar{x}) / \sum_j (x_j - \bar{x})^2$, $\sum_\ell w_\ell = 0$, and $\sum_\ell w_\ell x_\ell = 1$, so on $y_\ell = \alpha x_\ell + \alpha \log \sigma + \eta(\theta_\ell)$ it returns $\alpha + \sum_\ell w_\ell \eta(\theta_\ell)$. Hence $|\text{slope} - \alpha| \leq (\sum_\ell |w_\ell|) \max_\ell \eta(\theta_\ell)$, a grid constant times the largest deviation. A small enough $\theta_{\max}$ puts the slope within any tolerance of α for a fit error of any L_2 size. For the sampling law the empirical characteristic function converges uniformly on compact sets (Feuerverger and Mureika, 1977; Paulson et al., 1975), and its finite-grid central limit theorem (Feuerverger and McDunnough, 1981) carries through the map $\varphi \mapsto \log(-\log |\varphi|)$ on the band where $|\varphi_r|$ is bounded away from 0 and 1, then through least squares, so $\sqrt{m}(\hat{\alpha} - \alpha_{\text{pop}})$ is asymptotically normal about the population slope (Koutrouvelis, 1980; Koutrouvelis and Bauer, 1982). The modulus removes the first-order phase term $i\theta \mathbb{E}\delta$, so the fit error enters at second order, and the exact $\sqrt{m}$ rate about α holds when $(\mathbb{E}|\delta|^p / \sigma^\alpha) \theta_{\max}^{p-\alpha} = o(m^{-1/2})$. If δ has a regularly varying tail of index $\alpha_\delta < \alpha$ the slope converges to α_δ and recovery fails, and at $\alpha_\delta = \alpha$ the stability index survives while the scale is biased up. $\square$

In-sample residuals violate independence, since $\hat{G}$ has partially absorbed the noise, which biases $\hat{\sigma}$ downward, so a held-out split is required. Recovery needs only the held-out fit error to be lighter-tailed than the noise, not small, so a large but light-tailed misfit still returns the stability index. The finite- m variance inflates as $\alpha \rightarrow 2$, because the heavy-tail signal that separates $\alpha < 2$ from the Gaussian lives in the rare large values, which grow scarce near the boundary. This is the mild downward bias seen near $\alpha = 2$. At $\alpha = 1.2$ the infinite-variance noise degrades the mean fit until the error is itself heavy-tailed, the tail condition fails, and the recovery is biased and unstable, so those cells carry no clean split.

B COMPRESSIBLE-FLOW BENCHMARKS

The compressible-flow families were run across the full stability-index range in two and three dimensions. The spectral FNO does not fit their shock-forming dynamics. A Fourier basis rings on the discontinuities, so the fit to the clean data stays well above the 0.35 gate (Table 4). The residual method then cannot separate the signal

from the noise to extract a stability index. The failure is visible in the recovered stability index. Across the six cells it pins to $\hat{\alpha} \approx 1.2$ – 1.35 whether the injected stability index is 1.2, 1.5, or 1.8, so the estimator extracts the heavy-tailed fit error. A shock-capturing backbone, a flux-form or finite-volume head, would be needed to bring the fit error below the noise tail for the residual method to apply.

Table 4: Fit to the clean data (relative L_2) of the FNO on the compressible-flow families. All values are above the 0.35 fit gate, and the recovered stability index pins to the heavy-tailed fit error, not to the injected noise (see text).

family	$\alpha=1.2$	$\alpha=1.5$	$\alpha=1.8$
Compressible flow 2D	1.03	0.78	0.77
Compressible flow 3D	1.46	0.73	0.72