DIRECTION, NOT BANDWIDTH: WHAT LIMITS SPECTRAL NEURAL OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural operators are adopted for accuracy that survives a change of discretization, and Fourier neural operators demonstrate it on turbulent flow, reporting that the network recovers the frequencies it discards. We show that this recovery breaks down on heterogeneous solids, where the operator damps a high frequency no more than a low one. On the elastic Green's operator of homogenization, only the model that keeps every mode is accurate: keeping sixteen modes per axis gives a relative error of 0.44, against 0.026 for keeping all of them. We show this limitation belongs to the architecture rather than to training, and remove it by changing what the spectral weights represent rather than how they are learned. First, we prove that the error of any linear mode-truncating architecture is bounded below by how much the operator varies with direction. The bound is independent of the number of modes, the width and the depth, and can be computed before training. Second, we measure it on the elastic Green's operator, its anisotropic form, and the Stokes projection. Trained nonlinear models sit at the bound, truncated ones at or above it and isotropic ones within 4%, and neither added width, nor longer training, nor a loss reweighted toward high frequencies moves them off it. On a smoothing operator, where the bound is near zero, the same models are accurate and an isotropic parameterization wins instead. The direction dependence is therefore the cause. Third, we replace the table of per-mode weights with an expansion in spherical harmonics of the wavevector direction, applied at every frequency. Six coefficients then reproduce the Stokes projection exactly, at any resolution, where no finite mode table does. Placed inside a standard homogenization solver, and initialized to reproduce that solver exactly, 81 learned coefficients halve its iteration count at a phase contrast of 100 and cut it by more than $2.5\times$ at 1000. The analytic operator at the center of such solvers can therefore be replaced by a learned one that reaches the same accuracy in fewer iterations.

1 INTRODUCTION

The Fourier neural operator (FNO; Li et al., 2021) parameterizes each layer as a learned table of weights on a fixed set of low-frequency Fourier modes plus a pointwise linear map, and discards the table's action above a cutoff K. The justification is a statement about physics rather than about networks. For the equations on which neural operators are usually evaluated—Darcy flow, Navier–Stokes, reaction–diffusion—the solution operator is *smoothing*: its Fourier symbol decays, so high-frequency content in the input is damped on its way to the output, and quantitative approximation rates (Kovachki et al., 2021) exploit exactly this decay.

A large and practically important class of operators does not smooth at all. Write a Fourier multiplier as $\widehat{(Tu)}(k) = a(k)\,\hat{u}(k)$ and consider symbols of the form

$$a(k) = \sigma(n_k), \qquad n_k := k/|k| \in \mathbb{S}^{d-1}, \tag{1}$$

which are *homogeneous of degree zero*: constant along every ray, varying only over the sphere of directions, and never decaying. This class contains the Eshelby–Green operator that FFT-based homogenization applies thousands of times per solve (Moulinec & Suquet, 1998), the Leray projector $P(n) = I - n \otimes n$ of incompressible flow (Chorin, 1968), and the Riesz transforms (Stein, 1970). For these operators the smoothing argument is not merely loose; it is false.

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

Above the cutoff, an FNO layer applies only its pointwise channel mixing W, a matrix that is constant in k and therefore blind to direction. The target above the cutoff is $\sigma(n_k)$, which depends on nothing *but* direction. Raising K moves the cutoff but never supplies angular structure, because every direction realized by some lattice mode is also realized above every cutoff: if k appears, so does jk for all $j \geq 1$. The resulting error floor is the angular variation of σ about its mean.

The floor is large. On the 2D elastic Green's operator an FNO keeping 16 modes per axis reaches a relative error of 0.44, where the same model keeping every mode reaches 0.026 (Table 1). The floor is also not reachable by tuning: it is set by the architecture, so added width, longer training and spectral reweighting of the loss all leave it in place (§5.2). The remedy is to change what the spectral weights represent rather than how they are learned.

Our contributions are as follows.

- We isolate the order-zero class and show why the justification for mode truncation fails on it (§2).
- We prove a lower bound on the error of linear mode-truncating spectral networks that is uniform in the cutoff, width and depth, together with an ensemble identity that predicts the achievable error of each architecture in closed form (§3).
- We show that trained nonlinear models match these parameter-free predictions, which rules out the competing optimization-side explanation of mode saturation (§5.2).
- We introduce a *directional spectral layer*, parameterized by even spherical harmonics of the wavevector direction and applied at every frequency, and prove it is exact at finite angular degree for the canonical order-zero operators (§4); placed inside an FFT homogenization solver, 81 learned coefficients cut its iteration count by 2–3.6$\times$ (§5.6).
- We report two results that do not favor our layer: on a smoothing control the ordering of architectures reverses, and on the full nonlinear homogenization map a large free mode table is more accurate (§5).

2 Order-zero operators

Let $\mathbb{T}^d$ be the periodic domain and T_a the Fourier multiplier with symbol $a : \mathbb{Z}^d \setminus \{0\} \to \mathbb{C}^{c\times c}$. The operators studied in the operator-learning literature have decaying symbols: $a(k) = e^{-\nu t|k|^2}$ for the heat semigroup, $a(k) \sim |k|^{-2}$ for an elliptic solve. Decay is what makes mode truncation safe, since content above a cutoff has already been suppressed by the operator. We study the opposite case (1), the order-zero operators of Calderón–Zygmund theory (Stein, 1970), and write T_σ for the operator with symbol $\sigma(n_k)$. Three instances carry most of the weight in computational practice.

The Eshelby–Green operator. For a homogeneous reference stiffness C^0, let $Z(n)_{ik} := C^0_{ijkh} n_j n_h$ be the acoustic tensor. FFT-based homogenization (Moulinec & Suquet, 1998; Schneider, 2021) solves the Lippmann–Schwinger equation $\varepsilon = \bar{\varepsilon} - \Gamma^0 * \tau$ by repeated application of (Eshelby, 1957)

$$\hat{\Gamma}^0_{ijkh}(n) = \tfrac{1}{4}\Big[Z^{-1}_{ik}(n)\, n_j n_h + Z^{-1}_{jk}(n)\, n_i n_h + Z^{-1}_{ih}(n)\, n_j n_k + Z^{-1}_{jh}(n)\, n_i n_k\Big]. \tag{2}$$

The magnitudes cancel: $Z(k)^{-1} = Z(n_k)^{-1}/|k|^2$ against two factors of k in the numerator. For an isotropic reference $C^0_{ijkh} = \lambda\delta_{ij}\delta_{kh} + \mu(\delta_{ik}\delta_{jh} + \delta_{ih}\delta_{jk})$ one has $Z(n) = \mu I + (\lambda+\mu)\, n \otimes n$, hence

$$Z^{-1}(n) = \tfrac{1}{\mu}\big[I - \beta\, n \otimes n\big], \qquad \beta = \tfrac{\lambda+\mu}{\lambda+2\mu}. \tag{3}$$

For an anisotropic reference no closed form exists, and $Z(n)$ must be inverted numerically at every direction.

The Leray projection $P(n) = I - n \otimes n$ is applied at every step of a projection method for incompressible flow. It annihilates the longitudinal component entirely: these operators can damp a great deal, just not more at higher frequency. The *Riesz transforms* $\widehat{R_j u}(k) = -i\,(n_k)_j\,\hat{u}(k)$ are the canonical order-zero family, of which Γ^0 and P are even-degree combinations.

We do not claim that truncation is useless. Truncation error still tends to zero as $K \to \infty$ on inputs of positive smoothness, since every order-zero operator is bounded on L^2, and matched $\Theta(K^{-2s})$ rates are known (Subedi & Tewari, 2024). Our statement is that truncation buys no *smoothing gain*: the error inherits the input's own decay and nothing from the operator (Theorem 9).

3 WHAT A TRUNCATING SPECTRAL LAYER CANNOT DO

Let $\mathcal{L}_K$ denote the set of single spectral layers S that apply an arbitrary matrix A_k at each mode $|k|_\infty \le K$ and one pointwise matrix W at every mode $|k|_\infty > K$; i.e. the multipliers whose symbol is constant above the cutoff. All proofs are in Appendix C.

Lemma 1 (The realized direction set does not depend on K)**.** *For every $K \in \mathbb{N}$, the set $\mathcal{N}_K := \{n_k : k \in \mathbb{Z}^d,\ |k|_\infty > K\}$ equals $\mathcal{N}_0 = \{q/|q| : q \in \mathbb{Q}^d \setminus \{0\}\}$ and is dense in $\mathbb{S}^{d-1}$.*

Lemma 1 is the whole mechanism: raising the cutoff removes directions from *nowhere*. The band $|k|_\infty > K$ always sees the full sphere.

Theorem 2 (Angular gap)**.** *Let $\sigma \in C(\mathbb{S}^{d-1}; \mathbb{C}^{c\times c})$. For every $K \in \mathbb{N}$ and every $S \in \mathcal{L}_K$ with pointwise matrix W,*

$$\|T_\sigma - S\|_{L^2\to L^2} \ \ge \ \sup_{n\in\mathbb{S}^{d-1}} \|\sigma(n) - W\|_{\mathrm{op}} \ \ge \ \delta_\infty(\sigma) := \inf_{W'\in\mathbb{C}^{c\times c}} \sup_{n\in\mathbb{S}^{d-1}} \|\sigma(n) - W'\|_{\mathrm{op}}.$$

The bound is independent of K and attained, by $A_k = \sigma(n_k)$ and W a Chebyshev center of $\sigma(\mathbb{S}^{d-1})$. Consequently $\inf_{S\in\mathcal{L}_K} \|T_\sigma - S\| \to 0$ as $K \to \infty$ if and only if σ is constant.

Corollary 3 (Depth and width do not help without nonlinearity)**.** *Let $\mathcal{A}_K$ be the smallest set of operators containing $\mathcal{L}_K$ (for arbitrary channel widths, with constant lifting and projection matrices) and closed under composition, sums and scalar multiples: all* linear *FNOs of arbitrary depth and width with skip connections whose tables retain at most $|k|_\infty \le K$. Every element of $\mathcal{A}_K$ is a multiplier whose symbol is constant on $|k|_\infty > K$, so Theorem 2 holds verbatim for $\mathcal{A}_K$.*

The operator-norm bound is a worst-case statement. Experiments measure an average over an input ensemble, for which the sharp constant is the *angular variance*.

Theorem 4 (Angular gap, ensemble form)**.** *Let the coefficients $\{\hat u(k)\}_{k\neq 0}$ be independent, mean zero, with $\mathbb{E}\,\hat u(k)\hat u(k)^* = \frac{p_k}{c} I_c$ for a summable input spectrum $p \ge 0$. Let $S \in \mathcal{L}_K$ be exact below the cutoff ($A_k = \sigma(n_k)$ for $|k|_\infty \le K$, the best case). Let $p^{>K}$ be p restricted to $|k|_\infty > K$ and normalized to a probability measure, and $\bar\sigma := \mathbb{E}_{p^{>K}}[\sigma(n_k)]$. Then*

$$\frac{\mathbb{E}\|(T_\sigma - S)u\|^2}{\mathbb{E}\|T_\sigma u\|^2} \ = \ \eta_K \cdot \frac{\mathbb{E}_{p^{>K}}\|\sigma(n_k) - W\|_F^2}{\mathbb{E}_{p^{>K}}\|\sigma(n_k)\|_F^2} \ \ge \ \eta_K\, R_\sigma^{>K}, \qquad R_\sigma^{>K} := \frac{\mathbb{E}_{p^{>K}}\|\sigma - \bar\sigma\|_F^2}{\mathbb{E}_{p^{>K}}\|\sigma\|_F^2}, \tag{4}$$

where $\eta_K := \sum_{|k|_\infty>K} p_k\|\sigma(n_k)\|_F^2 \big/ \sum_k p_k\|\sigma(n_k)\|_F^2$ is the fraction of output *energy above the cutoff, with equality iff $W = \bar\sigma$. If the directions weighted by $p^{>K}$ equidistribute on $\mathbb{S}^{d-1}$, then $R_\sigma^{>K} \to R_\sigma$, the normalized angular variance $\|\sigma - \langle\sigma\rangle\|^2_{L^2(\mathbb{S}^{d-1})}/\|\sigma\|^2_{L^2(\mathbb{S}^{d-1})}$.*

Equation (4) is an identity, not an estimate. It predicts the best achievable relative error $\sqrt{\eta_K R_\sigma^{>K}}$ of every architecture in $\mathcal{A}_K$ from the symbol and the input spectrum alone, with no free parameters and no reference to training. A continuous isotropic symbol, which can at best apply the angular mean everywhere, is the case $K = 0$, $\eta_0 = 1$. For the 2D Leray projector, $R_\sigma = 1/2$ exactly; for the 2D isotropic elastic operator with $\lambda = \mu$, $R_\sigma = 4/13$.

Remark 5 (Scope). The class $\mathcal{A}_K$ is linear. A deployed FNO interleaves pointwise nonlinearities, which act in Fourier space as convolutions and *do* manufacture high-frequency, direction-dependent content, so they are not covered by Corollary 3. We claim the lower bound for the linear class only. We do *not* claim that FNOs cannot represent order-zero operators: that is refuted by universality (Kovachki et al., 2021) and by Lanthaler et al. (2023), who approximate operators with discontinuous data using a single retained mode on finite-dimensional input families. Our empirical claim is the separate, measurable one that trained nonlinear models on our tasks sit at the predictions of Theorem 4 (§5.2).

No smoothing gain. Theorem 2 concerns L^2 inputs. On H^s inputs, $s > 0$, the truncation error of an order-zero operator does vanish, but at rate $\Theta(\delta K^{-s})$ with δ the angular variation of σ, whereas an operator of order $-\alpha < 0$ gains an extra factor $K^{-\alpha}$ (Theorem 9, with matching upper and lower bounds). At order zero the operator contributes nothing to the decay.

4 A DIRECTIONAL SPECTRAL LAYER

Theorem 2 identifies the angular variation of the symbol as the binding quantity, so we parameterize exactly that.

Definition 6 (Directional spectral layer)**.** *Fix an even $L \in \mathbb{N}$ and let $\{Y_{\ell m}\}$ be the real spherical harmonics on $\mathbb{S}^2$ ($d = 3$), or $\{1, \cos \ell\varphi, \sin \ell\varphi\}$ on $\mathbb{S}^1$ ($d = 2$). The layer is*

$$\widehat{\mathcal{M}u}(k) = M(n_k)\,\hat{u}(k) \quad (k \neq 0), \qquad \widehat{\mathcal{M}u}(0) = 0, \qquad M(n) = \sum_{\substack{0 \le \ell \le L \\ \ell \text{ even}}} \sum_{|m| \le \ell} A_{\ell m} Y_{\ell m}(n), \tag{5}$$

with $A_{\ell m} \in \mathbb{R}^{c \times c}$, applied at all *frequencies: no cutoff, no radial dependence, no dependence on the grid.*

No cutoff and no grid. The parameters live on $\mathbb{S}^{d-1}$, so (5) defines one continuum operator whose restriction to any grid uses the same function (Theorem 16). A mode table with cutoff K costs $(2K+1)^d c^2$ weights; the layer costs $N_d(L)\, c^2$ with $N_3(L) = (L/2+1)(L+1)$ and $N_2(L) = L+1$, independent of resolution (Proposition 15).

Direction-dependent channel mixing. Each $A_{\ell m}$ is a matrix, so the mixing of components changes with n. A pointwise 1×1 convolution applies the same matrix at every mode; this is precisely the capability whose absence Theorem 2 penalizes.

Even degrees only. $\hat{\Gamma}^0$ and P are built from products of *pairs* of components of n, hence even under $n \mapsto -n$, so odd degrees are physically absent. With real coefficients they are also numerically harmful. A real multiplier requires $a(-k) = \overline{a(k)}$, not $a(-k) = a(k)$; odd harmonics with imaginary coefficients are legitimate and are exactly the Riesz transforms. With real coefficients applied on the half-spectrum, however, the inverse real FFT imposes conjugate symmetry by fiat and realizes a discontinuous even reflection of M (Proposition 13, Remark 14). We measured a relative deviation of 6.28 from the intended full-plane multiplier at $\ell = 1$, against 2.1×10^{-7} at $\ell = 2$. Restricting to even degrees removes this and halves the parameter count.

Proposition 7 (Isotropic Eshelby–Green: exact at $L = 4$, and not before)**.** *Let C^0 be isotropic and positive definite ($\mu > 0$, $d\lambda + 2\mu > 0$). Then*

(a) substituting (3) *into* (2) *gives the Moulinec–Suquet form*

$$\hat{\Gamma}^0_{ijkh}(n) = \tfrac{1}{4\mu}\big[\delta_{ki} n_h n_j + \delta_{hi} n_k n_j + \delta_{kj} n_h n_i + \delta_{hj} n_k n_i\big] - \tfrac{\lambda+\mu}{\mu(\lambda+2\mu)}\, n_i n_j n_k n_h;$$

(b) every entry of $\hat{\Gamma}^0$ is the restriction to $\mathbb{S}^{d-1}$ of an even homogeneous polynomial of degree 4, so its spherical-harmonic content is supported on $\ell \in \{0, 2, 4\}$;

(c) the degree-4 component is nonzero because $\beta \in (0, 1)$. Hence $L = 4$ is necessary and sufficient for exact representation by Definition 6.

The Leray projector $P = I - n \otimes n$ is quadratic and is exact at $L = 2$. For a cubic reference with Zener ratio $A \neq 1$, $\hat{\Gamma}^0$ is rational rather than polynomial in n; its expansion is infinite but converges exponentially (Proposition 12).

In 3D, Proposition 7 means $N_3(4) = 15$ coefficients per channel pair represent the isotropic Eshelby–Green operator exactly and $N_3(2) = 6$ represent the Leray projector, at every resolution. By Theorem 2, no finite mode table represents either.

Nyquist modes. On an even grid, a mode with $|k_j| = N/2$ is its own conjugate and represents several continuum wavevectors with genuinely different directions, while an order-zero symbol has full amplitude there. Its value is therefore ill-defined (Remark 20). We zero those modes, as FFT homogenization codes do. Keeping them breaks the projector identity $\Gamma^0 : C^0 : \varepsilon = \varepsilon$ by 5.6%; zeroing them restores it to 2.6×10^{-7}.

Optionally, multiplying $M(n_k)$ by a learned radial profile $R(|k|)$ recovers decaying symbols. This radial factor is unnecessary for pure order-zero targets and is used, where stated, for the smoothing control and the nonlinear homogenization map.

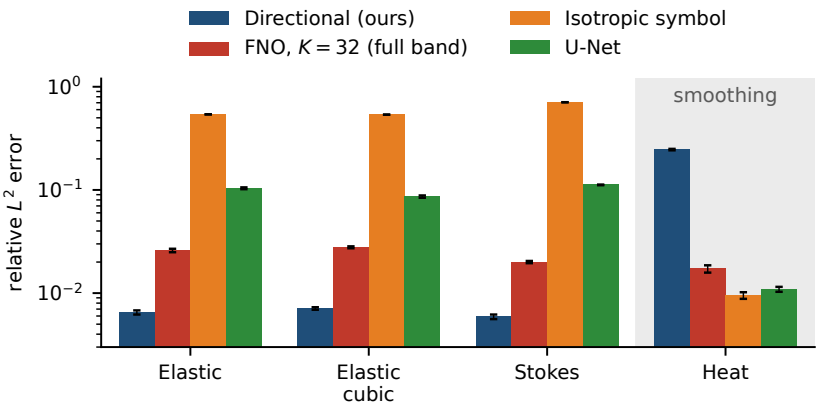

Figure 1: The ordering of architectures inverts between operator classes (Table 1; 64^2, mean $\pm$ s.d. over three seeds). The correct symbol parameterization is set by the operator, not by the architecture.

Table 1: Relative L^2 error in two dimensions, 64^2, mean $\pm$ s.d. over three seeds. The ordering inverts between the order-zero targets and the smoothing control.

	order zero			smoothing
Architecture	Elastic	Elastic cubic	Stokes	Heat
Directional (L=6)	0.0065 $\pm$0.0003	0.0071 $\pm$0.0002	0.0059 $\pm$0.0003	0.2460 $\pm$0.0040
FNO, K=32	0.0259 $\pm$0.0010	0.0278 $\pm$0.0006	0.0200 $\pm$0.0005	0.0172 $\pm$0.0014
FNO, K=16	0.4423 $\pm$0.0044	0.4653 $\pm$0.0033	0.4367 $\pm$0.0090	0.0121 $\pm$0.0002
FNO, K=8	0.4922 $\pm$0.0010	0.4987 $\pm$0.0007	0.6108 $\pm$0.0058	0.1275 $\pm$0.0021
U-Net	0.1038 $\pm$0.0021	0.0862 $\pm$0.0022	0.1119 $\pm$0.0008	0.0109 $\pm$0.0006
Isotropic symbol (EFNO)	0.5402 $\pm$0.0008	0.5364 $\pm$0.0007	0.7077 $\pm$0.0008	0.0095 $\pm$0.0007

5 Experiments

Setup. Inputs are Gaussian random fields with a power-law spectrum whose exponent is a tunable roughness parameter; outputs are obtained by applying the exact target operator. Targets in 2D (64^2) are the isotropic Eshelby–Green operator ($\lambda = \mu$), its cubic-reference form ($C_{11} = 1$, $C_{12} = 0.4$, $A = 3.2$), the Leray projector, and a heat semigroup as a smoothing control; 3D experiments use 32^3. Every target is verified to be degree-zero along rays before training (the control is verified *not* to be), and the Leray projector to be idempotent to 1.3×10^{-7}. We compare the directional layer against FNOs with $K \in \{8, 16, 32\}$ retained modes per axis ($K = 32$ is full band at 64^2), a U-Net (Ronneberger et al., 2015), and a network whose spectral symbol is a continuous isotropic function of $|k|$ in the style of Bosch et al. (2026). The metric is relative L^2 error on held-out inputs. Architectures, parameter counts and training protocol are given in Appendix A.

5.1 The ordering inverts between operator classes

Table 1 and Figure 1 give the central result. On the three order-zero targets the directional layer is 3–4$\times$ more accurate than the full-band FNO, with 38k parameters against 8.4M, and 75–120$\times$ more accurate than the continuous isotropic symbol. On the smoothing control the ordering reverses: the isotropic symbol is best and the purely angular layer worst, by a factor of 26. Neither parameterization is better in general; the homogeneity order of the target dictates which structure is needed, and choosing wrongly costs one to two orders of magnitude in either direction. We regard this reversal, rather than any single win, as the falsifiable content of the claim.

Table 2: Optimization-free floors from (4) versus trained nonlinear models (64^2, three seeds). No fitted constants. The last column is (measured − predicted)/predicted.

Architecture and target	predicted	measured	deviation
FNO $K{=}8$, elastic	0.4813	0.4922	$+2.3\%$
FNO $K{=}16$, elastic	0.3904	0.4423	$+13.3\%$
Isotropic symbol, elastic	0.5536	0.5402	-2.4%
Isotropic symbol, elastic cubic	0.5590	0.5364	-4.0%
Isotropic symbol, Stokes	$1/\sqrt{2} = 0.7071$	0.7077	$+0.1\%$

Table 3: Relative L^2 error in three dimensions, 32^3 (mean over two seeds). The parameter gap widens to three orders of magnitude; the directional layer's weights live on $\mathbb{S}^2$ and are resolution-free.

Target	Architecture	rel. L^2	Params
Elastic (order 0)	Directional ($L{=}4$)	0.0123	12.8k
	FNO, $K{=}16$	0.0447	12584.2k
Elastic cubic (order 0)	Directional ($L{=}8$)	0.0106	35.9k
	FNO, $K{=}16$	0.0496	12584.2k
Stokes (order 0)	Directional ($L{=}2$)	0.0036	5.8k
	FNO, $K{=}16$	0.0292	12584.1k
Heat (smoothing)	Directional + radial	0.0148	18.9k
	FNO, $K{=}16$	0.0172	12584.1k

5.2 Trained models sit at the predicted floors

Theorem 4 predicts, without training, the best achievable error of a truncating model (exact below the cutoff, the angular mean above it) and of an isotropic symbol (the angular mean everywhere). Table 2 compares these predictions with trained models; Table 8 in the appendix shows that the predictions themselves match their closed forms to within 0.5%.

The Stokes entry is worth isolating: the theory predicts $1/\sqrt{2}$ exactly, and the trained model lands within one part in a thousand. The truncated FNOs sit above their floors, as a lower bound requires. Two isotropic entries fall slightly *below* the linear bound, by 2.4% and 4.0%. This does not contradict Theorem 4, which concerns the linear class (Remark 5); the trained arms have pointwise nonlinearities, the theorem's escape hatch. We verified that the gap is not an artifact of an unweighted rather than energy-weighted angular mean (the two differ by 0.05%). The substantive finding is how little the escape hatch buys: at most 4%.

The floors also rule out the optimization account. Mode saturation in spectral operators has been reported before and attributed to optimization: to the neglect of low-amplitude spectral content by the mean-squared objective (Lippe et al., 2023), or to a bias of Fourier parameterizations toward dominant frequencies (Qin et al., 2024; Rahaman et al., 2019). Those accounts predict that a suitable remedy closes the gap. We tested three—a spectrally reweighted loss, an H^s Sobolev loss, and a SpecB-style residual cascade—and none improves the truncated model by more than 9%, while the Sobolev loss makes it worse (Table 7). Table 2 explains why: the baselines are already at their architectural floors, and there is no headroom for an optimizer to reach.

Capacity does not help either. At fixed $K{=}16$, increasing width from 16 to 128 moves the FNO from 0.408 to 0.510, and training eight times longer moves it from 0.414 to 0.622; both go the wrong way, while the directional layer improves in both sweeps ($0.0089 \to 0.0062$ and $0.0098 \to 0.0033$; Figure 2). This is the signature of a floor imposed by the architecture, not the optimizer, and it answers the standard objection that a negative result reflects an under-tuned baseline (Duruisseaux & Kossaifi, 2025).

Finally, input roughness separates the two error sources. Sweeping the input spectral exponent moves the truncated FNO by $5.3\times$ ($0.510 \to 0.097$ as inputs become smoother) while the isotropic symbol stays flat ($0.520 \to 0.581$). Truncation cost is a property of the data, through η_K; isotropy cost is a

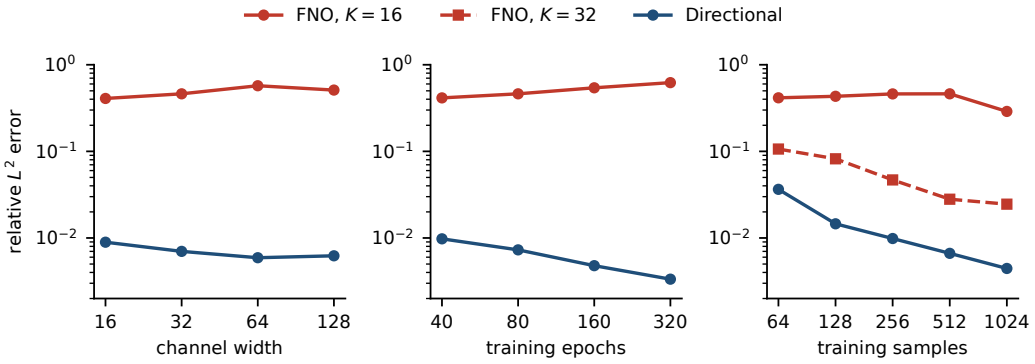

Figure 2: Truncated FNO (K=16) versus the directional layer on the 2D elastic target as width, training length and sample count grow. The FNO does not improve with width or training length and stays far above the full-band FNO (K=32) at every sample count; the directional layer improves throughout. This is the signature of a floor imposed by the architecture, not the optimizer.

Table 4: Zero-shot resolution transfer, $64 \to 128$, on the cubic-reference elastic operator ($A = 3.2$). Inputs are band-limited strictly below the coarse Nyquist frequency so both grids represent the same continuous field; errors therefore differ from Table 1, whose inputs are not band-limited.

Architecture	at 64	at 128	degradation
Directional (L=6)	0.0076	0.0074	1.0×
FNO, K=32	0.0107	0.0138	1.3×
U-Net	0.0822	0.1499	1.8×
FNO, K=16	0.2647	0.2644	1.0×
FNO, K=8	0.4783	0.4783	1.0×
Isotropic symbol (EFNO)	0.5543	0.5755	1.0×

property of the operator, through R_σ, and does not improve on smoother data. Both trends follow from (4).

5.3 Three dimensions and exactness at finite degree

Table 3 repeats the comparison at 32^3. The directional layer is 3.6–8× more accurate than an FNO with K=16 on the order-zero targets, with 350–2000× fewer parameters; on Stokes, 5.8k parameters against 12.6M at an eighth of the error. On the 3D heat control the directional layer *with* a radial factor is competitive with the FNO. Figure 3 verifies Proposition 7 without training, by least-squares projection of each symbol onto real spherical harmonics on the lattice: the residual of the isotropic elastic symbol falls to single-precision round-off (1.5×10^{-6}) at degree 4 and that of the Leray projector (2.9×10^{-7}) at degree 2, while the cubic-reference residual decays steadily without becoming exact (1.7×10^{-2} at $L = 8$).

5.4 Resolution transfer

Theorem 16 shows the directional layer is one continuum operator, so it is resolution-free by construction. A mode table might be expected to fail under refinement on this class; it does not. For a clean test, the coarse input must be a spectral band-limit of the fine one with the band limit strictly below the coarse Nyquist frequency, so that both grids represent the same continuous field and the operator acts identically on each. Protocols that violate this are misleading: with independently generated grids, or a band limit at the coarse Nyquist, 55.7% of the fine field's energy lies above the coarse Nyquist and *every* architecture appears to degrade, including the U-Net, which has no mode truncation at all.

Table 5: Full nonlinear homogenization (microstructure → strain), target obtained by solving the Lippmann–Schwinger fixed point to an equilibrium residual of 1.1×10^{-6}. A full-band FNO is more accurate here, at $88\times$ the parameters; angular structure remains necessary (the isotropic symbol is $15\times$ worse than the best directional model).

Architecture	rel. L^2	Params
FNO, K=32	0.0253	8394.1k
Directional (L=12) + radial	0.0348	95.6k
Directional (L=6) + radial	0.0367	71.0k
Directional (L=12)	0.0481	62.8k
Directional (L=6)	0.0496	38.2k
FNO, K=16	0.1238	2102.6k
U-Net	0.1749	471.0k
Isotropic symbol (EFNO)	0.5138	71.0k

Table 6: Speedup of the Moulinec–Suquet solver with a learned directional Γ^0 (81 parameters): analytic iterations needed to match the learned solver's error after J iterations, 64^2, 3 seeds for J=16 and one seed for J=8. "> 40": the analytic solver does not reach that error within the 40 iterations evaluated.

	phase contrast 100		phase contrast 1000	
	J=8	J=16	J=8	J=16
Matching analytic iterations	18	32–33	29	> 40
Speedup	2.3×	2.0–2.1×	3.6×	$> 2.5\times$

Under the correct protocol (Table 4), the directional layer transfers from 64^2 to 128^2 at $1.0\times$ and the full-band FNO at $1.3\times$, consistent with the zero-shot super-resolution of Li et al. (2021). Truncated FNOs transfer unchanged because their error is already at the floor. The advantage of the directional layer is representational and parametric, not a discretization claim.

5.5 A NEGATIVE RESULT ON THE FULL NONLINEAR MAP

We also learn the full homogenization map—two-phase microstructure to local strain field—with targets obtained by solving the Lippmann–Schwinger fixed point to an equilibrium residual of 1.1×10^{-6}. Here Γ^0 sits *inside* a fixed point that depends on the microstructure, so the map is nonlinear and not itself order zero.

Table 5 reports the outcome, which does not favor us: a full-band FNO is $1.37\times$ more accurate, with $88\times$ the parameters. The result is diagnostic. The map carries radial structure that a purely angular symbol cannot represent, and restoring a radial factor recovers most of the gap ($0.0481 \to 0.0348$ at $L = 12$). Angular structure remains *necessary*: the isotropic symbol is $15\times$ worse. The directional layer is the best structured representation and the best at any comparable parameter budget, but on this task a large unstructured table is more accurate.

5.6 A LEARNED GREEN'S OPERATOR INSIDE AN FFT SOLVER

Regression error is not what an FFT homogenization code needs; it needs fewer iterations. We replace the analytic Γ^0 in the Moulinec–Suquet iteration $\varepsilon^{j+1} = \bar{\varepsilon} - \Gamma^0 * [(C - C^0) : \varepsilon^j]$ by a learned directional symbol ($L = 8$, 81 coefficients), unroll J iterations, and train on anisotropic two-phase microstructures (correlation length stretched $3\times$ along one axis) to match a 3000-iteration reference solve. The learned symbol is *initialized at the analytic* Γ^0 by projection onto the harmonic basis, so at step zero the learned solver is the classical one and any gain is due to training. We report the number of analytic iterations needed to match the error that the learned solver reaches in J (Table 6). A continuous isotropic symbol and a K=16 mode table cannot be initialized to reproduce the solver, as Theorem 2 predicts; after training the isotropic symbol is far slower than the analytic solver, and the mode table (38k parameters) is slower in three of the four settings and reaches only $1.25\times$ in the fourth.

6 Related work

Neural operators and spectral parameterizations. Neural operators (Kovachki et al., 2023; Lu et al., 2021) learn maps between function spaces; the FNO (Li et al., 2021) and its variants (Tran et al., 2023) parameterize a truncated table of Fourier weights. Discretization behavior and aliasing of such models have been analyzed by Bartolucci et al. (2023) and addressed architecturally by Raonić et al. (2023). Bosch et al. (2026) remove the mode cutoff by parameterizing the symbol as a continuous function of the physical wavevector, the infrastructure we build on; their symbol is Gaussian in $|k|^2$ and thus smooth, decaying and isotropic, each of which an order-zero symbol violates, and their experiments are smoothing operators. Shin et al. (2024) learn a symbol without a cutoff but impose no homogeneity or directional structure. Spherical FNOs (Bonev et al., 2023) place the *spatial domain* on a sphere and still truncate in the spectral index; a symbol on the sphere of wavevector *directions* over a flat torus is a different object.

Learning for FFT homogenization. Nguyen & Schneider (2026) hard-code the analytic Eshelby–Green operator at all frequencies and prove a uniform approximation result, noting its Riesz-transform structure. Their construction involves no training, fixes an isotropic reference, and treats linear elasticity only; for an anisotropic reference no closed form exists, which is what a learned directional symbol supplies.

Approximation theory and mode saturation. Universality (Kovachki et al., 2021) and nonlinear reconstruction with very few modes (Lanthaler et al., 2023) show that FNOs *can* represent these operators; the latter operates on finite-dimensional input families where the network effectively performs parameter regression. Subedi & Tewari (2024) give matched $\Theta(K^{-2s})$ truncation rates for bounded Fourier-diagonal operators; our Theorem 9 is its matrix-valued directional analogue, and our contribution is identifying the constant (Remark 10). Mode saturation has been attributed to optimization (Lippe et al., 2023; Qin et al., 2024); §5.2 shows that on order-zero targets it is architectural.

7 Limitations and conclusion

Our theory covers linear spectral networks only. Theorem 2 and Corollary 3 do not apply to FNOs with pointwise nonlinearities, which manufacture high frequencies; that deep nonlinear models sit at the linear floor (§5.2) is a measurement, not a theorem. On the full nonlinear homogenization map a free mode table is $1.37\times$ more accurate than our layer, at $88\times$ the parameters (§5.5). The premise also does not hold for every plausible application: the empirical transfer operator of a crystal-plasticity stress surrogate, measured against a shuffled null control, is 97–99% isotropic wherever the estimate is reliable ($R^2 = 0.93$ and 0.78 at the two shortest horizons), because slowly evolving stress is dominated by the identity map, and a directional parameterization buys little there. Finally, the reversal of §5.1 is cleanest for the purely angular layer; with a radial factor our layer is competitive on smoothing operators too, so the dissociation is a statement about *which component* an operator class requires, not a claim that our layer and existing ones suit disjoint problems.

Mode truncation is safe exactly when the operator smooths. For order-zero operators, which sit at the center of FFT homogenization and incompressible flow solvers, the error of a truncating spectral layer is fixed by the angular variation of the symbol, is predictable in closed form before training, and is not removed by tuning. Parameterizing the symbol on the sphere of wavevector directions removes it with orders of magnitude fewer parameters, and inside an FFT homogenization solver it cuts the iteration count by 2–$3.6\times$. More broadly, the structure of a spectral parameterization should follow the homogeneity order of the target operator.

Reproducibility statement

All tables and figures are generated by scripts that read the raw run outputs (one JSON file per run); no number is entered by hand. Section 2 gives the target symbols in closed form, §4 the layer, Appendix A the data generation, architectures and training protocol, and Appendix C full proofs; the band-limiting protocol for the resolution study is stated in §5.4. Code and the generated datasets will be released upon acceptance.

REFERENCES

Francesca Bartolucci, Emmanuel de Bézenac, Bogdan Raonić, Roberto Molinaro, Siddhartha Mishra, and Rima Alaifari. Representation equivalent neural operators: a framework for alias-free operator learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Boris Bonev, Thorsten Kurth, Christian Hundt, Jaideep Pathak, Maximilian Baust, Karthik Kashinath, and Anima Anandkumar. Spherical Fourier neural operators: Learning stable dynamics on the sphere. In *International Conference on Machine Learning (ICML)*, 2023.

Michael Bosch, Niklas Schmitz, and Michael F. Herbst. Euclidean Fourier neural operators. *arXiv preprint arXiv:2608.28425*, 2026.

Alexandre Joel Chorin. Numerical solution of the Navier–Stokes equations. *Mathematics of Computation*, 22(104):745–762, 1968.

Valentin Duruisseaux and Jean Kossaifi. Fourier neural operators explained: A practical perspective. *arXiv preprint arXiv:2512.01421*, 2025.

John D. Eshelby. The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proceedings of the Royal Society of London. Series A*, 241(1226):376–396, 1957.

Nikola Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On universal approximation and error bounds for Fourier neural operators. *Journal of Machine Learning Research*, 22(290):1–76, 2021. arXiv:2107.07562.

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.

Samuel Lanthaler, Roberto Molinaro, Patrik Hadorn, and Siddhartha Mishra. Nonlinear reconstruction for operator learning of PDEs with discontinuities. In *International Conference on Learning Representations (ICLR)*, 2023. arXiv:2210.01074.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations (ICLR)*, 2021. arXiv:2010.08895.

Phillip Lippe, Bastiaan S. Veeling, Paris Perdikaris, Richard E. Turner, and Johannes Brandstetter. PDE-refiner: Achieving accurate long rollouts with neural PDE solvers. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. arXiv:2308.05732.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

Hervé Moulinec and Pierre Suquet. A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Computer Methods in Applied Mechanics and Engineering*, 157(1-2):69–94, 1998.

Binh Huy Nguyen and Matti Schneider. Universal Fourier neural operators for periodic homogenization problems in linear elasticity. *Journal of the Mechanics and Physics of Solids*, 206:106418, 2026. arXiv:2507.12233.

Shaoxiang Qin et al. Toward a better understanding of Fourier neural operators from a spectral perspective. *arXiv preprint arXiv:2404.07200*, 2024.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning (ICML)*, 2019.

Bogdan Raonić, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel de Bézenac. Convolutional neural operators for robust and accurate learning of PDEs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

Matti Schneider. A review of nonlinear FFT-based computational homogenization methods. *Acta Mechanica*, 232:2051–2100, 2021.

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. *Transactions on Machine Learning Research*, 2024. arXiv:2201.11967.

Elias M. Stein. *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, 1970.

Unique Subedi and Ambuj Tewari. Controlling statistical, discretization, and truncation errors in learning Fourier linear operators. *arXiv preprint arXiv:2408.09004*, 2024.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized Fourier neural operators. In *International Conference on Learning Representations (ICLR)*, 2023.

A Experimental details

Data. Inputs are real Gaussian random fields on $\mathbb{T}^d$ with power spectrum $p_k \propto |k|^{-\gamma}$ (zero mean, normalized to unit variance per sample); $\gamma = 1$ by default, and the roughness sweep of §5.2 uses $\gamma \in \{0.5, 1, 2, 3\}$. Outputs are the exact action of each target symbol, with Nyquist modes zeroed. The heat control is $e^{-\nu t|k|^2}$ with $\nu t = 10^{-4}$, fixed independently of the grid. In 2D we use 512 training and 128 test samples; in 3D, 256 and 64. For the nonlinear map of §5.5, two-phase microstructures with phase contrast 5 are solved by the Moulinec & Suquet (1998) scheme (500 iterations) on 64^2, with 512 training and 128 test samples.

Architectures. All spectral arms share one backbone: a pointwise lifting, 4 layers of width 32 in 2D (3 layers of width 16 in 3D), each computing $\mathrm{GELU}(\mathcal{S}x + Wx)$ with $\mathcal{S}$ the arm's spectral operator and W a 1×1 convolution, and a two-layer pointwise projection. The arms differ only in $\mathcal{S}$: the directional layer of Definition 6 ($L = 6$ in 2D, 38.2k parameters; $L \in \{2, 4, 8\}$ in 3D as in Table 3); an FNO mode table with $K \in \{8, 16, 32\}$ in 2D (0.53M, 2.10M, 8.39M parameters) and $K = 16$ in 3D; or a continuous isotropic symbol expanded in Gaussian radial basis functions of $|k|^2$, in the style of Bosch et al. (2026). The U-Net baseline has 471k parameters. Parameter counts for the nonlinear map are in Table 5.

Training. All arms are trained with AdamW (learning rate 2×10^{-3}, weight decay 10^{-4}, cosine schedule) on the relative L^2 loss, batch size 32 in 2D and 8 in 3D, for 80 epochs (2D operators), 100 epochs (nonlinear map) and 60 epochs (3D). The same settings are used for every arm, with no per-arm tuning; §5.2 and Figure 2 show that the FNO does not improve with more width or longer training. 2D results are mean $\pm$ s.d. over three seeds; 3D results are the mean over two seeds. Each run uses a single AMD MI250X graphics compute die.

B Additional results

Optimization-side remedies. Table 7 reports the three remedies of §5.2 on the 2D isotropic elastic target (one seed each; plain-FNO rows are the three-seed means of Table 1).

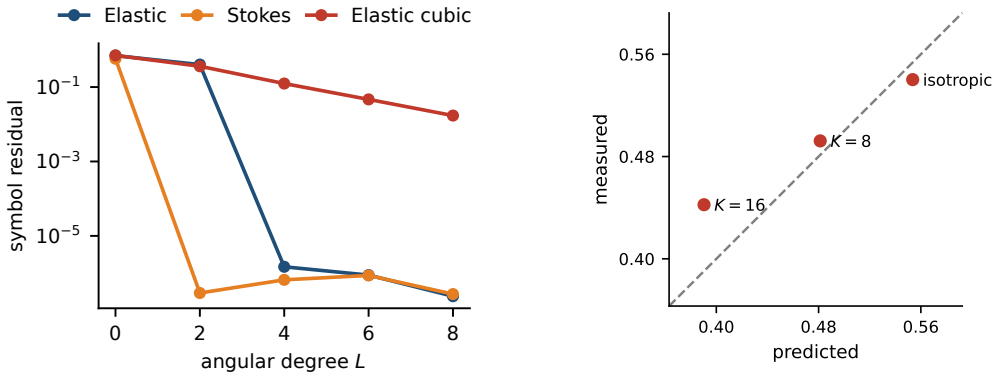

Figure 3: Left: residual of the least-squares projection of each 3D symbol onto real spherical harmonics of degree $\leq L$, with no training. The residual falls to round-off at degree four for the isotropic elastic symbol and at degree two for the Leray projector; the cubic reference decays without becoming exact (Proposition 7). Right: optimization-free floors against trained models (Table 2).

Table 7: Optimization-side remedies for FNOs on the 2D elastic target. None approaches the directional layer (0.0065); the largest improvement of a truncated model is 9%.

Arm	K=8	K=16	K=32
FNO (relative L^2 loss)	0.4922	0.4423	0.0259
+ spectrally reweighted loss	0.4899	0.4044	0.0645
+ H^1 Sobolev loss	0.5053	0.5040	0.0319
+ residual cascade (SpecB-style)	–	0.4325	–

Table 8: The angular-variance identity (4) predicts the achievable floor of each architecture with no fitted quantity. Rows 2–3 combine the measured R_σ with η_K computed from the symbol and the input spectrum; row 1 shows that R_σ is itself predicted by the closed form. *Computed separately from the main sweep.

Target ($d = 2, 64^2$)	Quantity	Theory	Measured	Dev.
Isotropic Eshelby	$R_\sigma^{1/2}$, continuum (λ=μ): $\sqrt{4/13}$	0.5547	0.5536	0.2%
Isotropic Eshelby	$\sqrt{R_\sigma\,\eta_8}$, $\eta_8 = 0.7488$	0.4791	0.4813	0.5%
Isotropic Eshelby	$\sqrt{R_\sigma\,\eta_{16}}$, $\eta_{16} = 0.4926$	0.3885	0.3904	0.5%
Leray projector	$R_\sigma^{1/2} = 1/\sqrt{2}$ (exact)	0.7071	0.7064*	0.1%
Cubic Eshelby, A=3.2	$R_\sigma^{1/2}$ (lattice-weighted)	0.5592	0.5590	0.04%

C PROOFS AND SUPPORTING RESULTS

C.1 PROOFS OF THE MAIN-TEXT RESULTS

Proof of Lemma 1. If $n = k/|k|$ with $k \in \mathbb{Z}^d \setminus \{0\}$ then $n = (jk)/|jk|$ for every $j \in \mathbb{N}$, and $|jk|_\infty = j|k|_\infty \to \infty$; hence every direction realized by some lattice point is realized by lattice points of arbitrarily large ℓ^∞ norm, giving $\mathcal{N}_K = \mathcal{N}_0$. Density: given $n \in \mathbb{S}^{d-1}$ and $\varepsilon \in (0, 1]$, pick $q \in \mathbb{Q}^d$ with $|q - n| < \varepsilon/2$; then $|q| \geq 1 - \varepsilon/2$ and $\big|q/|q| - n\big| \leq 2|q - n|/|q| \leq 2\varepsilon$. Clearing denominators makes q integral without changing its direction. □

Remark 8 (Real-valued fields). The proof of Theorem 2 uses complex exponentials. For real u one uses $\hat{u}(k) = v$, $\hat{u}(-k) = \bar{v}$. If σ satisfies the reality condition $\sigma(-n) = \overline{\sigma(n)}$ (Proposition 13) and W is real, both coefficients of the error are conjugate and the same bound follows with no change of constant.

Proof of Theorem 2. Fix $S \in \mathcal{L}_K$ with pointwise matrix W. Let $|k|_\infty > K$ and $v \in \mathbb{C}^c$, $|v| = 1$. Take u with $\hat{u}(k) = v$ and $\hat{u} = 0$ elsewhere (for real-valued fields see Remark 8). Then $\|u\| = 1$ and $(T_\sigma - S)u$ has the single coefficient $(\sigma(n_k) - W)v$, so $\|T_\sigma - S\| \geq |(\sigma(n_k) - W)v|$. Taking the supremum over v and over admissible k,

$$\|T_\sigma - S\| \;\geq\; \sup_{n \in \mathcal{N}_K} \|\sigma(n) - W\|_{\mathrm{op}} \;=\; \sup_{n \in \mathbb{S}^{d-1}} \|\sigma(n) - W\|_{\mathrm{op}},$$

the equality by Lemma 1 and continuity of $n \mapsto \|\sigma(n) - W\|_{\mathrm{op}}$. Minimizing over W gives the second inequality. For attainment, choose $A_k = \sigma(n_k)$ on $|k|_\infty \leq K$; the error operator is then the multiplier $(\sigma(n_k) - W)\mathbf{1}_{|k|_\infty > K}$, whose norm is $\sup_{n \in \mathcal{N}_K} \|\sigma(n) - W\|_{\mathrm{op}} = \sup_{\mathbb{S}^{d-1}} \|\sigma - W\|_{\mathrm{op}}$. Finally $\delta_\infty(\sigma) = 0$ iff $\sigma \equiv W'$ for some W'. □

Proof of Corollary 3. Sums and products of multipliers are multipliers, and sums and products of matrices that are constant in k are constant in k; in particular $QW_J \cdots W_1 P$ is a single constant matrix. The high-frequency symbol of any element of $\mathcal{A}_K$ is therefore constant, and the proof of Theorem 2 used nothing else. □

Proof of Theorem 4. By independence and $\mathbb{E}\hat{u}(k)\hat{u}(k)^* = \frac{p_k}{c} I$, for any matrix M, $\mathbb{E}|M\hat{u}(k)|^2 = \mathrm{tr}\big(M\,\mathbb{E}[\hat{u}\hat{u}^*]M^*\big) = \frac{p_k}{c}\|M\|_F^2$. Summing over k, and using that the error vanishes for $|k|_\infty \leq K$,

$$\mathbb{E}\|(T_\sigma - S)u\|^2 = \tfrac{1}{c} \sum_{|k|_\infty > K} p_k \|\sigma(n_k) - W\|_F^2, \qquad \mathbb{E}\|T_\sigma u\|^2 = \tfrac{1}{c} \sum_{k \neq 0} p_k \|\sigma(n_k)\|_F^2,$$

which is the equality in (4). The map $W \mapsto \mathbb{E}_{p^{>K}}\|\sigma - W\|_F^2$ is a strictly convex quadratic on $\mathbb{C}^{c \times c}$ minimized at the mean $\bar{\sigma}$ (orthogonal projection onto the constants in $L^2(p^{>K}; \mathbb{C}^{c \times c})$), with minimum value the variance; this gives the inequality and the equality case. The last claim is weak convergence of the empirical direction measure against the continuous integrand $\|\sigma - \cdot\|_F^2$. □

Proof of Proposition 7. (a) Expanding (3) in (2), the four δ-terms give $\frac{1}{4\mu}[\delta_{ik}n_jn_h + \delta_{ih}n_jn_k + \delta_{jk}n_in_h + \delta_{jh}n_in_k]$ and the four β-terms each give $-\frac{\beta}{4\mu}n_in_jn_kn_h$, summing to $-\frac{\beta}{\mu}n_in_jn_kn_h = -\frac{\lambda+\mu}{\mu(\lambda+2\mu)}n_in_jn_kn_h$ (verified symbolically against the standard form). (b) The first bracket is a homogeneous polynomial of degree 2, the second of degree 4; both are even. Any homogeneous polynomial of degree p decomposes as $\sum_{j \geq 0} |n|^{2j} H_{p-2j}(n)$ with H_r harmonic of degree r; restricted to $|n| = 1$ this is a sum of spherical harmonics of degrees $p, p-2, \ldots$, all of the parity of p. Hence degree 2 contributes $\ell \in \{0, 2\}$ and degree 4 contributes $\ell \in \{0, 2, 4\}$. (c) The degree-2 bracket has no $\ell = 4$ component, so the $\ell = 4$ part of $\hat{\Gamma}^0$ is $-\frac{\beta}{\mu}$ times the harmonic projection of $n^{\otimes 4}$, which is nonzero. Positive definiteness gives $\mu > 0$ and $d\lambda + 2\mu > 0$, hence $\lambda + \mu > \mu(1 - 2/d) \geq 0$ and $\lambda + 2\mu > 0$, so $\beta \in (0, 1)$. Sufficiency is (b) together with completeness of $\{Y_{\ell m}\}_{\ell \leq 4}$ on the space of even quartics restricted to the sphere. The statement for P follows in the same way from its degree-2 homogeneity, and the statement for the cubic reference is Proposition 12. □

C.2 SUPPORTING RESULTS

Theorem 9 (Truncation on H^s: order 0 versus order $-\alpha$)**.** *Let $\tau(k) = |k|^{-\alpha}\sigma(n_k)$ with $\alpha \geq 0$ and $\sigma \in C(\mathbb{S}^{d-1}; \mathbb{C}^{c\times c})$, and let $S \in \mathcal{L}_K$ be exact below the cutoff.*

(i) (Upper bound.) *With $W = 0$, for every $u \in H^s$, $s \geq 0$,*

$$\|(T_\tau - S)u\|_{L^2} \leq \Big(\sup_{|k|_\infty > K} \|\tau(k)\|_{\mathrm{op}}\Big)\|u_{>K}\|_{L^2} \leq \|\sigma\|_{L^\infty} K^{-\alpha}(1+K^2)^{-s/2}\|u\|_{H^s},$$

For $\alpha = 0$ the optimal W replaces $\|\sigma\|_{L^\infty}$ by $\delta_\infty(\sigma)$, giving $O(\delta_\infty(\sigma)K^{-s})$.

(ii) (Matching lower bound.) *Assume there exist $q_1, q_2 \in \mathbb{Z}^d \setminus \{0\}$ with $|q_1| = |q_2|$ and $\sigma(q_1/|q_1|) \neq \sigma(q_2/|q_2|)$, and put $2c_0 := \|\sigma(q_1/|q_1|) - \sigma(q_2/|q_2|)\|_{\mathrm{op}} > 0$. Then for every K and every $S \in \mathcal{L}_K$ there is $u \in H^s$, $u \neq 0$, with*

$$\frac{\|(T_\tau - S)u\|_{L^2}}{\|u\|_{H^s}} \;\geq\; c_0\, \rho^{-\alpha}(1+\rho^2)^{-s/2}, \qquad \rho := |q_1|\Big\lceil \tfrac{K+1}{\min_i |q_i|_\infty}\Big\rceil \;\leq\; C_q\,(K+1),$$

so the rate $K^{-(s+\alpha)}$ of (i) is sharp, with a constant proportional to the angular variation of σ.

The bound in (i) is $O(K^{-(s+\alpha)})$. Consequently the truncation error of an order-zero operator tends to 0 on H^s inputs with $s > 0$, but inherits only the input's own decay K^{-s}; an order-$(-\alpha)$ operator gains an extra factor $K^{-\alpha}$. There is no smoothing gain at order zero.

Proof. (i) The error multiplier vanishes for $|k|_\infty \leq K$ and equals $\tau(k)$ otherwise, so $\|(T_\tau - S)u\|^2 = \sum_{|k|_\infty > K} \|\tau(k)\hat{u}(k)\|^2 \leq \sup_{|k|_\infty > K} \|\tau(k)\|_{\mathrm{op}}^2 \|u_{>K}\|^2$. For $|k|_\infty > K$ we have $|k| \geq |k|_\infty > K$, hence $\|\tau(k)\|_{\mathrm{op}} \leq \|\sigma\|_{L^\infty} K^{-\alpha}$ and $(1+|k|^2)^{-s/2} \leq (1+K^2)^{-s/2}$, giving $\|u_{>K}\| \leq (1+K^2)^{-s/2}\|u\|_{H^s}$. For $\alpha = 0$ replace $\tau(k)$ by $\tau(k) - W$ and optimize W.

(ii) Let $j := \lceil (K+1)/\min_i |q_i|_\infty \rceil$ and $k_i := j\, q_i$, so $|k_i|_\infty \geq K+1$, $n_{k_i} = q_i/|q_i|$, and $|k_1| = |k_2| = j|q_1| = \rho$. Write $\tilde{W} := \rho^\alpha W$. Pick v with $|v| = 1$ and $|(\sigma(n_{k_1}) - \sigma(n_{k_2}))v| = 2c_0$. Then $\max_i |(\sigma(n_{k_i}) - \tilde{W})v| \geq c_0$ by the triangle inequality; let i^* attain the maximum and take $\hat{u}(k_{i^*}) = v$, $\hat{u} = 0$ elsewhere. The error has the single coefficient $\rho^{-\alpha}(\sigma(n_{k_{i^*}}) - \tilde{W})v$ of modulus $\geq c_0\rho^{-\alpha}$, while $\|u\|_{H^s} = (1+\rho^2)^{s/2}$. Finally $\rho \leq |q_1|\big((K+1)/\min_i |q_i|_\infty + 1\big) \leq C_q(K+1)$. □

Remark 10 (Consistency with Subedi & Tewari (2024)). Subedi & Tewari (2024) analyze a DFT least-squares estimator of a Fourier-diagonal operator with bounded multipliers and no decay assumption, over inputs in $\mathcal{H}^s(\mathbb{T}^d)$ with $s > d/2$, and obtain matched bounds whose truncation term is $\Theta(K^{-2s})$ in the squared risk. Theorem 9 with $\alpha = 0$ is the matrix-valued, directional version of that term. Our contribution is not a better rate, since there is none, but the identification of the *constant*: it is the angular variation of σ, and by Theorem 2 it does not decrease with K at $s = 0$. We therefore say "no smoothing gain", never "irreducible".

Proposition 11 (Exactness in the limit)**.** *Let $\sigma \in L^\infty(\mathbb{S}^{d-1}; \mathbb{R}^{c\times c})$ be even, and let M_L be its spherical-harmonic partial sum of degree L.*

(a) In the ensemble metric of Theorem 4 with equidistributed directions, the relative squared error of $\mathcal{M}_L$ against T_σ equals the tail $\sum_{\ell > L} \|\hat{\sigma}_\ell\|_F^2 / \|\sigma\|^2_{L^2(\mathbb{S}^{d-1})}$, which tends to 0 as $L \to \infty$.

(b) In operator norm, $\|T_\sigma - \mathcal{M}_L\|_{L^2\to L^2} = \sup_n \|\sigma(n) - M_L(n)\|_{\mathrm{op}}$. This tends to 0 if $\sigma \in H^t(\mathbb{S}^{d-1})$ with $t > (d-1)/2$, at rate $O(L^{-(t-t')})$ for any $(d-1)/2 < t' < t$; it does not follow from $\sigma \in L^2$ alone. If σ is real-analytic the convergence is exponential in L.

Proof. (a) is (4) with $K = 0$ and W replaced by the direction-dependent $M_L(n)$, together with Parseval on the sphere and completeness of $\{Y_{\ell m}\}$ in $L^2(\mathbb{S}^{d-1})$; only even degrees appear because σ is even. (b) The norm of a matrix multiplier is the supremum of the matrix norms of its symbol over the dense set $\mathcal{N}_0$ (Lemma 1), which by continuity equals the supremum over $\mathbb{S}^{d-1}$. The rate is Sobolev embedding on the sphere applied to the tail: $\|\sigma - M_L\|_{L^\infty} \leq C\|\sigma - M_L\|_{H^{t'}} \leq CL^{-(t-t')}\|\sigma\|_{H^t}$.

Non-convergence in the merely continuous case is the classical failure of uniform convergence of orthogonal expansions. □

Proposition 12 (When finite L suffices)**.** *Let C^0 be a positive-definite reference. If $\det Z(n)$ is constant on $\mathbb{S}^{d-1}$, equivalently if $Z^{-1} = \operatorname{adj} Z/\det Z$ is polynomial in n on the sphere, then $\hat{\Gamma}^0$ is a spherical polynomial of degree 4 and $L = 4$ suffices. This holds for every isotropic reference, where $\det Z = \mu^{d-1}(\lambda + 2\mu)$. It fails for a cubic reference with Zener ratio $A = 2C_{44}/(C_{11} - C_{12}) \neq 1$: there $\det Z$ is a non-constant homogeneous polynomial of degree $2d$, $\hat{\Gamma}^0$ is rational in n, and its harmonic expansion is infinite. It is however real-analytic on $\mathbb{S}^{d-1}$ (as $\det Z > 0$ there), so by Proposition 11(b) the expansion converges exponentially. For the $d = 2$ cubic reference with $C_{11} = 1$, $C_{12} = 0.4$, $A = 3.2$, the fraction of angular energy above degree L is 1.3×10^{-1} ($L = 2$), 8.7×10^{-3} ($L = 4$), 3.3×10^{-3} ($L = 6$), 8.8×10^{-5} ($L = 10$), 1.6×10^{-7} ($L = 16$).*

Proof. If $\det Z \equiv c$ on the sphere then $Z^{-1} = c^{-1} \operatorname{adj} Z$ has entries that are homogeneous polynomials of degree $2(d-1)$, and (2) multiplies them by quadratics; on the sphere the result reduces modulo $|n|^2 = 1$ to an even polynomial of degree 4, and Proposition 7(b) applies. For the isotropic case Z has eigenvalues $\lambda + 2\mu$ (eigenvector n) and μ (multiplicity $d-1$), both independent of n. For the cubic case in $d = 2$, $Z_{11} = C_{11}n_1^2 + C_{44}n_2^2$, $Z_{22} = C_{44}n_1^2 + C_{11}n_2^2$, $Z_{12} = (C_{12} + C_{44})n_1n_2$, so $\det Z = C_{11}C_{44}(n_1^4 + n_2^4) + \big(C_{11}^2 + C_{44}^2 - (C_{12} + C_{44})^2\big)n_1^2n_2^2$. Since $n_1^4 + n_2^4 = 1 - 2n_1^2n_2^2$ on $\mathbb{S}^1$, this is constant iff $C_{11}^2 + C_{44}^2 - (C_{12} + C_{44})^2 = 2C_{11}C_{44}$, i.e. $(C_{11} - C_{44})^2 = (C_{12} + C_{44})^2$, which for a positive-definite cubic tensor holds iff $C_{11} - C_{12} = 2C_{44}$, i.e. $A = 1$. Real-analyticity follows since $\det Z$ is a nonvanishing polynomial on the compact sphere. □

Proposition 13 (Parity, realness, and odd degrees)**.** *Let $a : \mathbb{Z}^d \setminus \{0\} \to \mathbb{C}^{c\times c}$ be a multiplier.*

- *(a) T_a maps real fields to real fields iff $a(-k) = \overline{a(k)}$ for all k.*
- *(b) Consequently a* real-valued *order-zero symbol must be even, $a(-n) = a(n)$. Since $Y_{\ell m}(-n) = (-1)^\ell Y_{\ell m}(n)$, Definition 6 with real coefficients is real-valued iff all odd-degree coefficients vanish.*
- *(c) $\hat{\Gamma}^0(-n) = \hat{\Gamma}^0(n)$ and $P(-n) = P(n)$, because both are built from $n \otimes n$; the same holds for any product of an even number of Riesz transforms.*
- *(d) Odd degrees are* not *inadmissible in general. If $A_{\ell m}$ is purely imaginary for odd ℓ and real for even ℓ, then $a(-n) = \overline{a(n)}$ and the operator is real. This is the class containing the Riesz transforms, which are order-zero, real, and of odd angular degree.*

Proof. (a) For real u, $\hat{u}(-k) = \overline{\hat{u}(k)}$; $T_a u$ is real iff $a(-k)\overline{\hat{u}(k)} = \overline{a(k)\hat{u}(k)}$ for all k and all admissible $\hat{u}$, i.e. iff $a(-k) = \overline{a(k)}$. (b) If a is real-valued this reads $a(-k) = a(k)$; applied to $\sum_{\ell m} A_{\ell m}Y_{\ell m}$ with real $A_{\ell m}$ it forces $A_{\ell m} = 0$ for odd ℓ by linear independence of the $Y_{\ell m}$. (c) Immediate from (2)–(3) and $P = I - n \otimes n$. (d) For odd ℓ with $A_{\ell m} = iB_{\ell m}$, $B_{\ell m}$ real, the term $iB_{\ell m}Y_{\ell m}(-n) = -iB_{\ell m}Y_{\ell m}(n) = \overline{iB_{\ell m}Y_{\ell m}(n)}$, and even terms are real and even, so $a(-n) = \overline{a(n)}$. □

Remark 14 (Implementation warning). An implementation that stores only the half-spectrum (`rfft`) and applies real coefficients at odd ℓ does not realize the intended multiplier: the inverse real transform imposes $a(-k) = \overline{a(k)}$ by fiat, so the realized symbol is the *even reflection* of M across the retained half-space. It is still order zero, but discontinuous across the boundary of the half-space, which forfeits the smoothness used in Proposition 11(b).

Proposition 15 (Parameter count)**.** *The layer of Definition 6 has $N_d(L)\,c^2$ real parameters, independent of the grid, with $N_3(L) = \sum_{\ell\le L,\ \ell\ \text{even}}(2\ell+1) = (\frac{L}{2}+1)(L+1)$ and $N_2(L) = L + 1$. In particular $N_3(4) = 15$ and $N_2(4) = 5$. A mode table on $|k|_\infty \le K$ has $(2K+1)^d c^2$ parameters. For $d = 3$, $c = 6$ (Mandel basis) the exactly sufficient directional layer uses $15 \times 36 = 540$ parameters, against $33^3 \times 36 \approx 1.3 \times 10^6$ for a $K = 16$ table that is still, by Theorem 2, bounded away from the target.*

Proof. $\dim\{Y_{\ell m}\}_{|m|\le\ell} = 2\ell+1$ on $\mathbb{S}^2$; summing over $\ell = 2j$, $j \le L/2$, gives $\sum_{j=0}^{L/2}(4j+1) = (\frac{L}{2}+1)(L+1)$. On $\mathbb{S}^1$ the even harmonics are $\{1\} \cup \{\cos\ell\varphi, \sin\ell\varphi\}_{\ell \text{ even}, 2\le\ell\le L}$, of cardinality $1+2(L/2) = L+1$. The count is grid-independent because the parameters index functions on $\mathbb{S}^{d-1}$, not lattice sites. □

Theorem 16 (The directional layer is one continuum operator)**.** *Let Π_N be the Fourier projection onto the grid-N band $|k|_\infty < N/2$ (Nyquist modes excluded) and $\mathcal{M}_N := \Pi_N \mathcal{M} \Pi_N$ the realization of Definition 6 on a grid of size N. Then*

- *(a) with the* same *parameters, $\mathcal{M}_N = \mathcal{M}\Pi_N$ for every N: the discrete operators at all resolutions are restrictions of a single bounded operator on $L^2(\mathbb{T}^d)$;*
- *(b) $\|\mathcal{M}_N\|_{L^2\to L^2} \le \sup_n \|M(n)\|_{\mathrm{op}}$ uniformly in N, and $\mathcal{M}_N$ and $\mathcal{M}_{N'}$ agree exactly on the modes they share;*
- *(c) for $u \in H^s$, $s > 0$, $\|\mathcal{M}_N u - \mathcal{M}u\|_{L^2} \le C\|u\|_{H^s} N^{-s} \to 0$.*

The mechanism is that the mode index $k \in \mathbb{Z}^d$ fixes the direction n_k independently of N; refining the grid only samples the same function M on $\mathbb{S}^{d-1}$ at additional points.

Proof. (a) The layer's symbol at mode k is $M(n_k)$, a function of k alone with no reference to N; multipliers commute with Π_N, so $\Pi_N \mathcal{M} \Pi_N = \mathcal{M}\Pi_N$. (b) The norm of a multiplier restricted to a set of modes is the supremum of the symbol norms there, which is at most the supremum over $\mathbb{S}^{d-1}$; agreement on shared modes is immediate. (c) $\mathcal{M}u - \mathcal{M}_N u = \mathcal{M}(I - \Pi_N)u$, bounded by $\sup_n \|M(n)\|_{\mathrm{op}} \|(I-\Pi_N)u\| \le C(1+(N/2)^2)^{-s/2}\|u\|_{H^s}$. □

Proposition 17 (When a resolution-tied mode table must fail)**.** *Suppose a table is trained at resolution N with $K = N/2$ and evaluated at $N' > N$. The realized operator lies in $\mathcal{L}_K$: it applies the learned A_k on $|k|_\infty \le K$ and the pointwise W on the newly resolved band $K < |k|_\infty \le N'/2$. By Theorem 4 its relative error at the new resolution satisfies*

$$\mathcal{E}(N') \;\ge\; \sqrt{R_\sigma^{>K}\, \eta_K(N')}\,, \tag{6}$$

with $\eta_K(N')$ the fraction of output energy in the new band, however well the model was trained at resolution N. For an operator of order $-\alpha < 0$, $\eta_K(N')$ carries an extra factor $K^{-2\alpha}$, so the penalty is specific to order zero. Under the band-limited protocol of §5.4, the input has no energy in the new band, $\eta_K(N') = 0$, and the bound is vacuous; this is why mode tables transfer there. The bound becomes active only when the finer grid resolves genuinely new input content.

Proof. Apply (4) at cutoff K and resolution N', using that the model is at best exact below K (which only weakens the bound) and that its symbol above K is the constant W. □

Remark 18 (3D constants). For $d = 3$ and an isotropic reference with $\lambda = \mu$, the harmonic energy of $\hat{\Gamma}^0$ in the Mandel basis splits as $\ell = 0$: 0.509, $\ell = 2$: 0.325, $\ell = 4$: 0.166, and $\ell \ge 6$: $< 10^{-29}$, verifying Proposition 7 to machine precision. The angular variance ratio is $R_\sigma = 0.491$, so the floor of Theorem 4 on high-band-dominated inputs is $\sqrt{0.491} = 0.70$. For the 3D Leray projector, $R_\sigma = 1/3$ and the floor is $1/\sqrt{3} = 0.577$; its content is $\ell \in \{0, 2\}$ and $L = 2$ is exactly sufficient.

Remark 19 (What an L-sweep does and does not test). Proposition 7 concerns a *single* layer. A network stacking J directional layers composes their symbols, and a product of two spherical polynomials of degree L has degree up to $2L$; a depth-J stack therefore reaches angular degree JL even before nonlinearities are counted. An empirical saturation of the error in L is consistent with Proposition 7 but is not a sharp test of it; exactness should be verified by fitting the symbol of a single layer, as in Figure 3.

Remark 20 (Order-zero symbols are ill-defined at Nyquist). On a real grid of even size N per axis, a mode index k with $k_j = N/2$ for some j is aliased: the grid samples of $e^{i(N/2)x_j}$ and $e^{-i(N/2)x_j}$ coincide. A grid function with r such components is the restriction of 2^r distinct continuum plane waves whose wavevectors differ by sign flips in the Nyquist coordinates. Unless all remaining coordinates vanish, those wavevectors point in genuinely different directions: for $d = 2$, $N = 64$, the index $(32, 3)$ represents both $k = (32, 3)$ and $k = (-32, 3)$, whose directions differ by $\approx 10.7°$. An

order-zero symbol has full magnitude there, so its value is ambiguous; a symbol of order $-\alpha < 0$ has magnitude $O(N^{-\alpha})$ and is unaffected. Equivalently, the reality constraint forces the self-conjugate Nyquist coefficient to be real, while the compatible strain $\hat{\varepsilon}(k) = \mathrm{sym}(i\,k \otimes \hat{u}(k))$ is purely imaginary there. Consequently the projector identity $\Gamma^0 : C^0 : \varepsilon = \varepsilon$ fails at the Nyquist row: retaining those modes breaks it by 5.6% (an independent 2D reproduction gives 7.6%), and zeroing them restores it to 2.6×10^{-7} (1.9×10^{-7} in the reproduction). We therefore zero the Nyquist row throughout, which is standard practice in FFT homogenization (Moulinec & Suquet, 1998).