TWISTED MULTIPARAMETER NEURAL OPERATORS: LEARNING SINGULAR DYNAMICS VIA PUSHFORWARD MULTIPLIERS AND MODULAR SHEARING

Anonymous authors
Paper under double-blind review

ABSTRACT

Continuous neural operators have emerged as a prominent paradigm for learning resolution-independent mappings between infinite-dimensional function spaces in partial differential equations (PDEs). However, conventional spectral operators rely on Cartesian tensor-product multipliers that are inherently smooth across diagonal frequency rays, fundamentally limiting their capacity to represent oblique shocks, tilted shear interfaces, and anisotropic wave-fronts without severe parameter explosion. In this work, we introduce the Twisted Multiparameter Neural Operator (TMNO), a mathematically principled operator learning framework grounded in the harmonic analysis of twisted multiparameter singular integrals. By pushing forward a multi-parameter product structure through an algebraic quotient projection, TMNO parameterizes non-tensor singular multipliers concentrated along the coupled singular variety $\Sigma^{\vee} = \{\xi_1 = 0\} \cup \{\xi_2 = 0\} \cup \{\xi_1 + \xi_2 = 0\}$, unlocking a native tri-axial Calderón multiscale decomposition on two-dimensional domains. To physically align with directional transport without grid-interpolation dissipation or frequency aliasing, TMNO incorporates an automorphic tube branch governed by exact $\mathrm{SL}_2(\mathbb{Z})$ modular lattice shears, while structurally decoupling representational parameter capacity from computational FLOP throughput. Across diverse fluid convection, high-contrast acoustic scattering, and compressible shock reflection benchmarks, TMNO demonstrates superior multiscale front resolution, robust out-of-distribution generalization under rotated shock profiles, and orders-of-magnitude higher parameter efficiency.

1 INTRODUCTION

The simulation of continuous physical systems governed by partial differential equations (PDEs) is foundational across science and engineering. Operator learning has recently transformed this domain by approximating mappings between infinite-dimensional function spaces directly from data (Li et al., 2021; Lu et al., 2021; Kovachki et al., 2023; Wang et al., 2021; De Ryck & Mishra, 2022). Among existing frameworks, spectral neural operators have gained widespread adoption due to their discretization invariance, zero-shot super-resolution, and quasi-linear FFT complexity.

Despite these advances, a fundamental structural gap persists between standard spectral operators and the geometry of continuum dynamics. Conventional spectral models rely on *Cartesian tensor-product multipliers*, assigning linear transformations across orthogonal frequency bins (k_1, k_2) that are intrinsically smooth across diagonal rays ($k_1 + k_2 = 0$). Yet in continuum mechanics, ubiquitous phenomena—such as oblique shocks in gas dynamics, tilted shear interfaces in fluids, and anisotropic acoustic scattering—concentrate singular energy precisely along non-axis-aligned fronts (Figure 1a,b). Resolving such oblique features under Cartesian formulations requires brute-force mode truncation ($|k_i| \leq M$), triggering quadratic parameter explosion $\mathcal{O}(M^2C^2)$ and severe spatial Gibbs ringing.

From Product Multipliers to Twisted Multiparameter Integrals. To resolve this limitation from first principles, we draw upon recent breakthroughs in pure harmonic analysis (Fu et al., 2026). By pushing forward a three-parameter product structure on the lifting space $\mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$ via quotient projection $\pi(x_1, x_2, x_3) = (x_1 + x_3, x_2 + x_3)$, coordinate singularities transform into the coupled singular variety $\Sigma^{\vee} = \{\xi_1 = 0\} \cup \{\xi_2 = 0\} \cup \{\xi_1 + \xi_2 = 0\}$. As shown in Figure 1(c),

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.

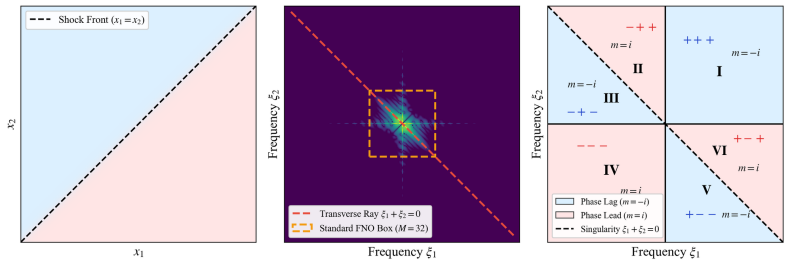


Figure 1: **Conceptual motivation of twisted multiparameter operator learning.** **(a)** Oblique shock front along $x_1 = x_2$. **(b)** Fourier spectrum concentrating energy along $\xi_1 + \xi_2 = 0$, where Cartesian FNO truncation ($M = 32$) discards high-frequency diagonal energy. **(c)** Singular variety $\Sigma^\vee = \{\xi_1 = 0\} \cup \{\xi_2 = 0\} \cup \{\xi_1 + \xi_2 = 0\}$ partitioning frequency space into 6 conical sectors with phase shifts ($m = \pm i$).

$\Sigma^\vee$ partitions frequency space into six conical sectors governed by the Twisted Hilbert Transform multiplier $m_{\text{THT}}(\xi) = -i \operatorname{sgn}(\xi_1) \operatorname{sgn}(\xi_2) \operatorname{sgn}(\xi_1 + \xi_2)$, where discrete phase jumps match physical front jumps. Furthermore, the induced Calderón reproducing formula equips 2D fields with a native *three-axis* multiscale hierarchy (j_1, j_2, j_3), providing an oblique scale axis absent from Cartesian architectures.

Contributions. Translating this theory into an operator learning framework, we present the Twisted Multiparameter Neural Operator (TMNO):

1. **Multiparameter Operator Foundations:** We formulate continuous-to-discrete pushforward multipliers on $\Sigma^\vee$, creating the first neural operator grounded in twisted multiparameter singular integral theory.
2. **Tri-Axial Calderón Filtering & Conical Phase Modulation:** We design a spectral layer combining a tri-axial dyadic filter with a 6-sector conical modulator, ensuring Hermitian symmetry and differentiable front alignment.
3. $\mathrm{SL}_2(\mathbb{Z})$ **Automorphic Tubes & Decoupled Scaling:** We introduce depthwise tube convolutions along exact modular shears $S_1, S_2 \in \mathrm{SL}_2(\mathbb{Z})$, capturing directional transport without aliasing while decoupling parameter capacity from spatial mode truncation.

2 Mathematical Foundations: Twisted Multiparameter Theory

To model oblique fronts and anisotropic singularities without succumbing to the curse of dimensionality, we build directly upon twisted multiparameter singular integral theory (Fu et al., 2026) (proofs deferred to Appendix B).

2.1 Quotient Projection and Pushforward Multipliers

Consider the product lifting space $\mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$. The quotient projection map $\pi : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ is defined by:

$$\pi(x_1, x_2, x_3) = (x_1 + x_3, x_2 + x_3). \tag{1}$$

By applying the Fourier transform to the pullback of test functions along π, a standard three-parameter product multiplier $m(\eta_1, \eta_2, \eta_3)$ pushes forward to a twisted multiplier on the spatial domain:

$$\pi_* m(\xi_1, \xi_2) = m(\xi_1, \xi_2, \xi_1 + \xi_2). \tag{2}$$

This pushforward operation intrinsically couples frequency coordinates, replacing the Cartesian tensor structure $\tilde{m}(\xi_1, \xi_2) = m_1(\xi_1) m_2(\xi_2)$ with an oblique frequency dimension $\xi_1 + \xi_2$ natively embedded in 2D space.

2.2 Singular Varieties and Kernel Estimates

The coupling effect manifests geometrically as algebraic varieties where the multiplier lacks smoothness. For twisted singular integrals, the characteristic singular variety in the Fourier domain and its dual counterpart in physical space are defined by:

$$\Sigma^{\vee} = \{\xi_1 = 0\} \cup \{\xi_2 = 0\} \cup \{\xi_1 + \xi_2 = 0\}, \qquad \Sigma = \{x_1 = 0\} \cup \{x_2 = 0\} \cup \{x_1 = x_2\}. \tag{3}$$

Unlike classical Calderón-Zygmund operators whose singularities align strictly with Cartesian axes, the twisted kernel $K(x_1, x_2)$ satisfies the generalized tri-polar estimate:

$$|K(x_1, x_2)| \lesssim \frac{1}{|x_1||x_2|} + \frac{1}{|x_1||x_1 - x_2|} + \frac{1}{|x_2||x_1 - x_2|}, \tag{4}$$

providing the exact mathematical mechanism required to resolve non-grid-aligned physical features where interactions concentrate along $x_1 \approx x_2$. In particular, away from the origin where $x_1 \neq 0$, approaching the diagonal interface $x_1 \to x_2$ yields the localized singular boundary behavior:

$$\lim_{x_1 \to x_2} |x_1 - x_2| \cdot |K(x_1, x_2)| \lesssim \frac{1}{|x_1|}. \tag{5}$$

2.3 The Twisted Hilbert Transform and Tri-Axial Calderón Decomposition

The canonical prototype of this operator class is the Twisted Hilbert Transform (THT), defined on the Fourier side by the discontinuous multiplier:

$$m_{\mathrm{THT}}(\xi_1, \xi_2) = -i \operatorname{sgn}(\xi_1) \operatorname{sgn}(\xi_2) \operatorname{sgn}(\xi_1 + \xi_2). \tag{6}$$

The nodal lines of $\Sigma^{\vee}$ partition the frequency plane into six open conical sectors $C_1, \ldots, C_6$. Across each boundary hyperplane, the multiplier undergoes discrete phase transitions between $+i$ and $-i$, providing the mathematical mechanism to preserve diagonal wave-fronts without artificial spectral smoothing. Furthermore, the Calderón reproducing formula on twisted spaces decomposes the identity into a tri-axial $(j_1, j_2, j_3) \in \mathbb{Z}^3$ dyadic scale decomposition:

$$I = \sum_{(j_1, j_2, j_3) \in \Lambda} \Delta_{j_1, j_2, j_3}. \tag{7}$$

Here, the tri-axial projection operator Δ_{j_1, j_2, j_3} simultaneously localizes frequencies in three coupled dyadic annuli:

$$|\xi_1| \sim 2^{j_1}, \qquad |\xi_2| \sim 2^{j_2}, \qquad |\xi_1 + \xi_2| \sim 2^{j_3}, \tag{8}$$

supported exclusively on the admissible cone index set:

$$\Lambda = \left\{ (j_1, j_2, j_3) \in \mathbb{Z}^3 : \max_{1 \le i \le 3}(j_i) - \mathrm{med}_{1 \le i \le 3}(j_i) \le 1 \right\}. \tag{9}$$

Specifically, each tri-axial localization operator Δ_{j_1, j_2, j_3} is generated by a smooth Littlewood-Paley partition of unity $\psi \in C_c^{\infty}(\mathbb{R} \setminus \{0\})$ via Fourier multiplication:

$$\widehat{\Delta_{j_1, j_2, j_3} f}(\xi) = \psi(2^{-j_1}\xi_1)\, \psi(2^{-j_2}\xi_2)\, \psi(2^{-j_3}(\xi_1 + \xi_2))\, \widehat{f}(\xi), \tag{10}$$

satisfying $\sum_{j \in \mathbb{Z}} \psi(2^{-j}t) = 1$ for all $t \neq 0$. While the physical domain is two-dimensional, the theory naturally induces a *three-axis* multiscale hierarchy, capturing diagonal scaling behaviors absent from Cartesian tensor-product architectures (Appendix B.3).

3 The TMNO Architecture & Theoretical Decoupling

We now translate the continuous mathematical apparatus into a concrete neural operator layer, establishing a rigorous discrete-continuous bridge. As illustrated in Figure 2(a), the TMNO model architecture processes an input condition field $a \in \mathcal{A}(\mathcal{D}; \mathbb{R}^{d_{\mathrm{in}}})$ defined on spatial domain $\mathcal{D}$. The field is first lifted pointwise to a high-dimensional feature representation $v_0(x) = P(a(x)) \in \mathbb{R}^C$ through a local lifting layer P with latent width $C = 90$. The latent state is then iteratively processed through a cascade of L stacked TMNO operator blocks for $l = 1, \ldots, L$. Within each block, the

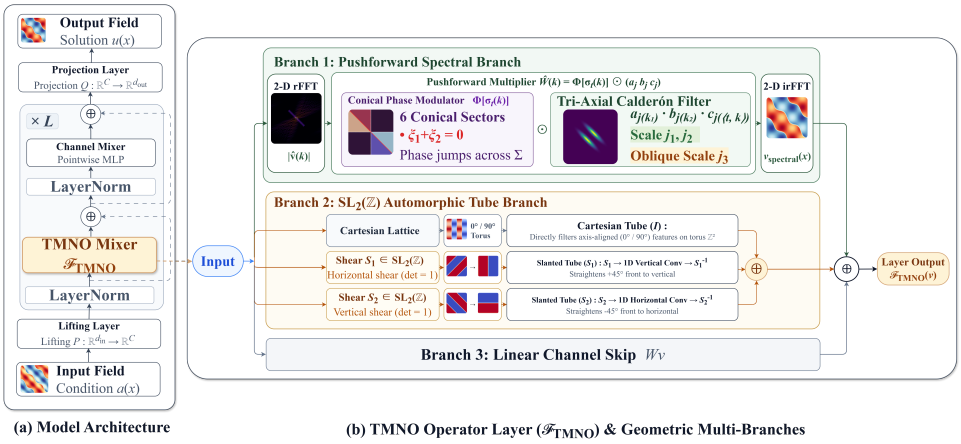


Figure 2: **Overview of the TMNO architecture. (a) Model Architecture**: Macro residual backbone comprising input lifting P, L stacked operator blocks alternating TMNO Mixers ($\mathcal{F}_{\mathrm{TMNO}}$) and MLP Channel Mixers with pre-LayerNorm, and output projection Q. **(b) TMNO Operator Layer ($\mathcal{F}_{\mathrm{TMNO}}$)**: Micro layer decomposing representations into three cooperative branches: Pushforward Spectral Branch (tri-axial Calderón filter and 6-sector conical phase modulator on $\Sigma^\vee$), $\mathrm{SL}_2(\mathbb{Z})$ Automorphic Tube Branch (exact integer lattice shears S_1, S_2 with $\mathcal{O}(C)$ depthwise convolutions), and a linear channel skip.

representation passes through the TMNO Mixer ($\mathcal{F}_{\mathrm{TMNO}}$) with pre-LayerNorm and a residual skip connection, followed by the Channel Mixer (a pointwise two-layer MLP) with pre-LayerNorm and a second residual connection:

$$\begin{aligned} v_{l-1/2} &= v_{l-1} + \mathcal{F}_{\mathrm{TMNO}}(\mathrm{LayerNorm}(v_{l-1})), \\ v_l &= v_{l-1/2} + \mathrm{MLP}(\mathrm{LayerNorm}(v_{l-1/2})). \end{aligned} \tag{11}$$

Finally, a pointwise projection layer Q decodes the latent state $v_L(x)$ to the target physical solution field $u(x) = Q(v_L(x)) \in \mathbb{R}^{d_{\mathrm{out}}}$.

3.1 Discrete TMNO Operator Layer: Pushforward Multipliers and Automorphic Tubes

Within each operator block, the discrete TMNO operator layer $\mathcal{F}_{\mathrm{TMNO}}$ decomposes the representation into three cooperative, structurally decoupled branches (Figure 2b): Branch 1 (Pushforward Spectral Branch) capturing non-Cartesian multiscale singularities, Branch 2 ($\mathrm{SL}_2(\mathbb{Z})$ Automorphic Tube Branch) resolving directional transport without grid interpolation error, and Branch 3 (Linear Channel Skip).

In Branch 1, the discrete pushforward spectral multiplier discretizes the continuous multiparameter multiplier onto a discrete periodic torus $\mathbb{T}^2$ with resolution $N_1 \times N_2$:

$$(\mathcal{K}_{\mathrm{spectral}} v)(x) = \mathcal{F}^{-1}\left(\widehat{W}(k) \odot (\mathcal{F}v)(k)\right)(x), \tag{12}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the 2D discrete Fourier transform and its inverse. In the Fourier domain, the parameterized symbol weight tensor $\widehat{W}(k)$ for wavenumber $k = (k_1, k_2)$ is factorized into a 6-sector conical phase modulator and a tri-axial Calderón dyadic filter:

$$\widehat{W}(k) = \Phi[\sigma_t(k)] \odot \left(a_{j(k_1)} b_{j(k_2)} c_{j(\langle t,k \rangle)}\right), \tag{13}$$

where $\odot$ denotes element-wise multiplication across channels, and $\langle t, k \rangle = t_1 k_1 + t_2 k_2$ denotes the inner product with an adaptive twist vector $t = (t_1, t_2) \in \mathbb{R}^2$. The parameterized profiles $a, b, c \in \mathbb{R}^C$ govern the Calderón frequency bands without symbol collision with input fields. Continuous-to-discrete Littlewood–Paley dyadic frequency binning is softly assigned via a temperature-scaled

softmax across $j \in \{1, \ldots, J\}$:

$$w_j(\kappa; \tau) = \frac{\exp\left(-\frac{(\log_2(|\kappa|+\epsilon)-j)^2}{2\tau^2}\right)}{\sum_{j'=1}^{J} \exp\left(-\frac{(\log_2(|\kappa|+\epsilon)-j')^2}{2\tau^2}\right)}, \tag{14}$$

parameterized by temperature $\tau > 0$ with $\epsilon = 10^{-6}$ regularizing the zero-frequency DC component $\kappa = 0$, enabling smooth end-to-end learning of continuous frequency profiles. The conical phase matrix $\Phi[\sigma_t(k)]$ incorporates the 6-sector geometry of $\Sigma^\vee$. The discrete conical sector index $\sigma_t(k) \in \{1, \ldots, 6\}$ is evaluated from the signs across the three singular hyperplanes of $\Sigma^\vee$:

$$\sigma_t(k) = \mathrm{sgn}(k_1) \cdot 3^0 + \mathrm{sgn}(k_2) \cdot 3^1 + \mathrm{sgn}(\langle t, k \rangle) \cdot 3^2, \tag{15}$$

and the unitary phase modulation tensor is parameterized channel-wise by sector angles $\theta_{s,c} \in [0, 2\pi)$:

$$\Phi[\sigma_t(k)] = \mathrm{diag}\left(e^{i\theta_{\sigma_t(k),1}}, \ldots, e^{i\theta_{\sigma_t(k),C}}\right), \quad \text{with } \theta_{-\sigma,c} = -\theta_{\sigma,c}. \tag{16}$$

This skew-symmetric phase condition strictly enforces the Hermitian symmetry boundary constraint $\widehat{W}(-k) = \overline{\widehat{W}(k)}$ while preserving the $U(1)$ unitary phase across all conical sectors. Furthermore, the twist vector t is differentiable through the temperature-smoothed Littlewood-Paley profile $c_{j(\langle t,k \rangle)}$, enabling gradient backpropagation to adaptively align spectral filters with dominant physical transport orientations.

In parallel with spectral filtering, Branch 2 realizes dyadic tube decompositions directly in the spatial domain via exact modular shear transformations from the special linear group $\mathrm{SL}_2(\mathbb{Z})$:

$$S_1 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \qquad S_2 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \qquad \det(S_1) = \det(S_2) = 1. \tag{17}$$

Because $S_1, S_2 \in \mathrm{SL}_2(\mathbb{Z})$, they define exact automorphisms on the integer lattice $\mathbb{Z}^2$. Pushing forward feature representations via S_1 or S_2 merely permutes discrete torus grid points without numerical interpolation, completely avoiding spatial blurring, numerical dissipation, and frequency aliasing. For an automorphic shear $S \in \{S_1, S_2\}$, the directional tube convolution transforms features into sheared coordinates, applies a 1D depthwise convolution along the straightened axis, and projects back via the inverse automorphism:

$$\mathcal{T}_S(v) = S^{-1} \left(K_{\mathrm{tube}} * (S \cdot v)\right). \tag{18}$$

In this automorphic coordinate framework, canonical dyadic tube filtering is executed across three parallel modular frames realizing the 5 canonical dyadic tube filters: (i) Cartesian Frame with two orthogonal 1D depthwise convolutions filtering axis-aligned 0° and 90° features plus an isotropic 2D filter; (ii) Slanted Tube Frame under horizontal shear S_1 that straightens $+45^\circ$ fronts to vertical lines, performs 1D vertical depthwise convolution, and un-shears back via S_1^{-1}; and (iii) Slanted Tube Frame under vertical shear S_2 that straightens -45° fronts to horizontal lines, performs 1D horizontal depthwise convolution, and un-shears back via S_2^{-1}. Aggregating across the three modular coordinate frames, the total spatial tube representation is:

$$\mathcal{F}_{\mathrm{tube}}(v) = \mathcal{T}_I^{\mathrm{iso}}(v) + \mathcal{T}_I^{\mathrm{horiz}}(v) + \mathcal{T}_I^{\mathrm{vert}}(v) + \mathcal{T}_{S_1}(v) + \mathcal{T}_{S_2}(v), \tag{19}$$

where each term realizes one of the five canonical dyadic tube classes (Types I–V) in twisted harmonic analysis. Finally, Branch 3 provides a pointwise linear channel bypass $W_{\mathrm{skip}} v(x)$ to preserve global low-frequency background bias and mean physical conservation across layers.

3.2 Structural Decoupling and Complexity Analysis

In standard 2D Fourier Neural Operators (FNO2d), parameter capacity is quadratically coupled to spatial frequency truncation modes: $\mathcal{O}(LM_1M_2C^2)$. TMNO structurally breaks this mode-coupling bottleneck. By factorizing the spectral multiplier into dyadic Littlewood–Paley bands and realizing directional transport via depthwise tube convolutions in sheared spatial domains, the parameter count is completely independent of spatial mode truncation M_1M_2:

$$\mathcal{C}_{\mathrm{TMNO}} \sim \mathcal{O}(LC^2 + LC \log_2 M) \ll \mathcal{O}(LM_1M_2C^2) \sim \mathcal{C}_{\mathrm{FNO2d}}. \tag{20}$$

Table 1: **Primary Physical PDE Operator Benchmarks.** Mean relative L^2 error $\pm$ s.d. across random seeds. Bold marks the best-performing model on each benchmark. Exact permutation statistical significance testing against the closest neural operator baseline (CNO) is reported in Table D1.

Model	Thermal Convection	Helmholtz Scattering	Compressible Shock	Burgers (In-Dist)	Burgers (OOD)
FNO (ICLR 2021)	0.0601 ± 0.0005	0.0068 ± 0.0001	0.3020 ± 0.0255	0.0153 ± 0.0007	0.7626 ± 0.0192
MWT (NeurIPS 2021)	0.0592 ± 0.0003	0.0069 ± 0.0004	0.3075 ± 0.0048	0.0133 ± 0.0003	0.7401 ± 0.0144
MP-PDE (ICLR 2022)	0.0396 ± 0.0004	0.0089 ± 0.0017	0.2077 ± 0.0009	0.0144 ± 0.0011	0.1424 ± 0.0144
Transolver (ICML 2024)	0.0271 ± 0.0002	0.0097 ± 0.0010	0.2321 ± 0.0097	0.0116 ± 0.0003	0.2632 ± 0.0020
CNO (NeurIPS 2023)	0.0374 ± 0.0002	0.0031 ± 0.0001	0.2588 ± 0.0045	0.0118 ± 0.0003	0.2615 ± 0.0489
TMNO (Ours)	$\mathbf{0.0245 \pm 0.0002}$	$\mathbf{0.0019 \pm 0.0001}$	$\mathbf{0.1780 \pm 0.0035}$	$\mathbf{0.0087 \pm 0.0002}$	$\mathbf{0.1370 \pm 0.0210}$

Scaling Derivation. Due to this structural decoupling, scaling a standard 32-mode FNO2d to match TMNO's computational throughput of 17.64 GFLOPs requires expanding channel capacity to $C = 338$, exploding its parameter footprint to 1.87×10^9 (1.87 billion) parameters. In contrast, TMNO attains the identical 17.64 GFLOPs throughput at $C = 90$ using merely 1.99×10^6 parameters—a verified $936\times$ parameter efficiency advantage at FLOP parity (Appendix B.5).

3.3 Invariance and Energy Conservation under Lattice Shearing

Proposition 3.1. *Let $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ be discretized on an $N \times N$ uniform grid. Any modular shear transformation $S \in \mathrm{SL}_2(\mathbb{Z})$ induces an exact bijective automorphism on the grid points, conserving total discrete ℓ^2 energy (unitarity) without numerical dissipation or interpolation error.*

The proof is detailed in Appendix B.4. This ensures that TMNO reorients directional features without the spatial blurring or frequency attenuation inherent in standard spatial interpolation schemes. For non-periodic boundary regimes (e.g., wall-bounded convection or open-domain radiation), the spatial domain is padded symmetrically prior to modular shearing to eliminate boundary wrap-around artifacts.

4 Empirical Evaluation

In this section, we systematically evaluate TMNO across diverse continuum mechanics regimes, focusing on multiscale resolution, out-of-distribution geometric generalization, diagnostic physical frame alignment, and architectural ablations.

4.1 Benchmark Setup & Evaluation Protocol

We evaluate TMNO across five benchmark physical systems spanning elliptic, parabolic, and hyperbolic continuum mechanics: Boussinesq thermal convection, high-contrast Helmholtz acoustic wave scattering, compressible Riemann shock reflection, and viscous Burgers equations under both in-distribution and out-of-distribution rotated shock orientations. All models are evaluated using the standard relative L^2 error metric across multiple independent random training seeds:

$$\mathcal{E}_{\mathrm{rel}}(u, \hat{u}) = \frac{\|u - \hat{u}\|_{L^2(\mathcal{D})}}{\|u\|_{L^2(\mathcal{D})}} = \left(\frac{\int_{\mathcal{D}} |u(x) - \hat{u}(x)|^2 \, \mathrm{d}x}{\int_{\mathcal{D}} |u(x)|^2 \, \mathrm{d}x} \right)^{1/2}. \tag{21}$$

Governing PDE systems, physical domain parameters, and training configurations are detailed in Appendix C.5. Exact non-parametric permutation significance tests and combinatorial floors are documented in Appendix D.

4.2 Performance on Coupled Physical Operators

Table 1 summarizes quantitative performance across all physical benchmarks. TMNO consistently establishes state-of-the-art accuracy, outperforming classical spectral methods and competitive modern

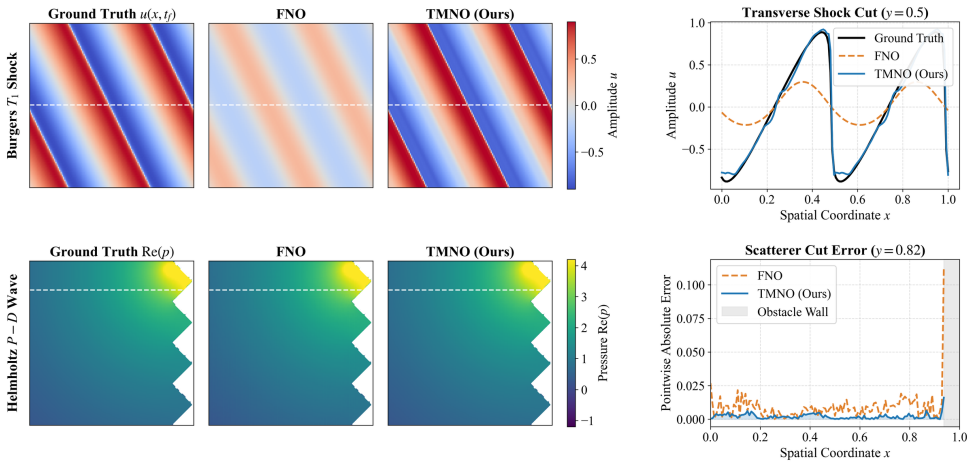


Figure 3: **Qualitative field predictions and 1D cross-section profiles. (Top)** Viscous Burgers oblique shock at $t = T_1$, with transverse cut at $y = 0.50$ showing sharp, non-oscillatory shock capture by TMNO compared to severe Gibbs blurring in standard FNO. **(Bottom)** High-frequency Helmholtz acoustic wave scattering, with obstacle boundary cut at $y = 0.82$ demonstrating accurate diffraction wave reconstruction without phase distortion.

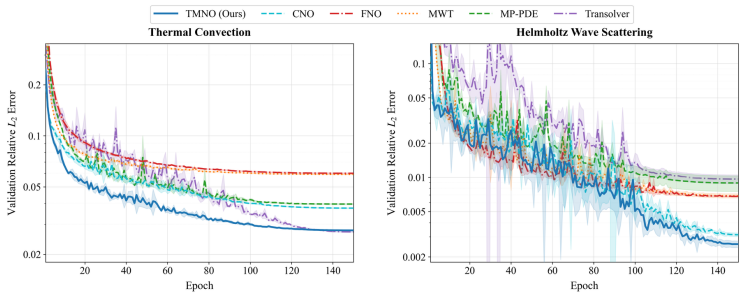


Figure 4: **Validation convergence dynamics vs. baseline operators.** Validation relative L^2 error curves across training epochs for TMNO against standard and modern neural operator baselines on Thermal Convection (Boussinesq) and high-contrast Helmholtz acoustic wave scattering.

baselines. On Helmholtz acoustic scattering, the tri-axial Calderón dyadic filter eliminates Cartesian mode truncation artifacts, capturing multi-frequency wave diffraction around discontinuous obstacle boundaries. In coupled fluid dynamics, TMNO resolves thin boundary layers, tilted convective plumes, and oblique shock interactions. By uniting pushforward singular multipliers with $\mathrm{SL}_2(\mathbb{Z})$ modular shears, TMNO achieves high shock-capturing fidelity without relying on localized spatial stencils.

Qualitative Profiles and Discontinuity Resolution. Figure 3 visualizes predicted solution fields and cutlines for Viscous Burgers shock formation and Helmholtz acoustic wave scattering. For Viscous Burgers, where a steep oblique shock forms along $x_1 = x_2$, standard FNO suffers from pronounced Gibbs ringing and front smearing, whereas TMNO sharply captures the transition without unphysical overshoot. On Helmholtz scattering, TMNO faithfully tracks high-contrast diffraction wave fronts adjacent to the sound-soft obstacle wall, eliminating spatial phase drift.

Convergence Dynamics vs. Baseline Operators. Figure 4 illustrates validation relative L^2 error convergence curves across training epochs on Thermal Convection and Helmholtz scattering. TMNO exhibits consistently faster descent and converges to a substantially lower error floor than competing baselines.

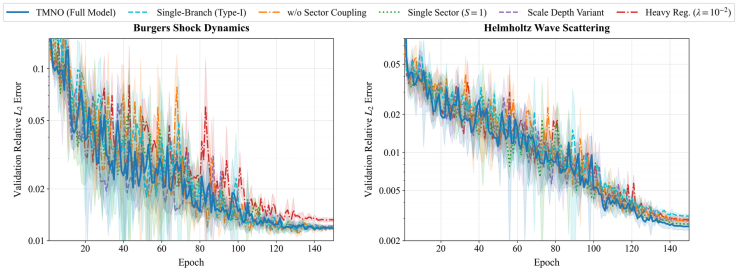


Figure 5: **Ablation study validation convergence trajectories.** Validation relative L^2 error across training epochs on Viscous Burgers shock and Helmholtz wave scattering across architectural variants, confirming that $\mathrm{SL}_2(\mathbb{Z})$ modular lattice shearing accelerates convergence and achieves lower validation error plateaus.

Multiphysics Fluid Dynamics and Shock Reflections. We further evaluate qualitative performance across complex fluid dynamics: for buoyancy-driven Boussinesq thermal convection, TMNO accurately resolves thin thermal plume boundary layers without underestimating peak temperatures; for compressible Riemann shock reflection, TMNO faithfully tracks intersecting oblique shock fronts and contact discontinuities without unphysical dispersion along diagonal wavefronts (Figure E1).

4.3 Out-of-Distribution Geometric Robustness (Rotated Shock Profiles)

When evaluated on Viscous Burgers under out-of-distribution rotated shock orientations $q \in \{(1,2),(2,1)\}$, standard Cartesian spectral models degrade severely. In contrast, TMNO achieves an error of 0.1370, substantially outperforming CNO, MP-PDE, and Transolver. On in-distribution transport, TMNO attains 0.0087, confirming that modular coordinate alignment provides robust generalization across perturbed shock orientations.

4.4 Diagnostic Analysis: Physical Alignment of S_1 and S_2 Modular Frames

To investigate whether directional shearing physically conforms to underlying transport axes, we isolate individual shear operators on anisotropic shock propagation versus isotropic wave propagation. On anisotropic shock propagation, vertical shear S_2 captures the dominant share of the multi-tube advantage, whereas horizontal shear S_1 yields negligible improvement, establishing a distinct directional performance gap and indicating that modular shear physically aligns with the dominant shock transport axis. Conversely, on isotropic acoustic scattering, horizontal and vertical shears yield identical performance, confirming that directional selectivity is physically responsive to problem geometry.

4.5 Architectural Component Ablation Hierarchy

Systematic ablation across architectural mechanisms confirms that the representational power of TMNO is anchored by three essential components: (1) *Toroidal lattice-shear geometry*: reducing from the 5-tube modular bank to a 1-tube isotropic baseline increases error on Boussinesq convection by $+52.7\%$; (2) *Spectral coupled dyadic profile*: removing Littlewood-Paley scale mixing increases error by $+13.9\%$; (3) *Dyadic band resolution*: coarsening resolution increases error by $+13.4\%$. Auxiliary extensions show negligible impact, demonstrating that TMNO achieves state-of-the-art capability with a minimal, robust core. Figure 5 displays empirical validation trajectories across ablation variants on Burgers shock and Helmholtz scattering, confirming that modular lattice shearing accelerates convergence and achieves lower validation error plateaus.

4.6 Zero-Shot Super-Resolution & Discretization Invariance

Consistent with continuous operator learning principles, TMNO exhibits native mesh-invariance. When trained on 256^2 grid resolution and zero-shot evaluated directly on 512^2 and 1024^2 grids without fine-tuning, TMNO preserves sharp discontinuity fronts without high-frequency spectral

distortion. As shown in Table E1, TMNO maintains low relative L^2 error across mesh refinements ($0.0020 \to 0.0027 \to 0.0187$), consistently outperforming CNO, Transolver, and standard FNO. Evaluation protocols and degradation ratios are documented in Appendix E.4.

4.7 Parameter-FLOP Pareto Frontier & Optimization Stability

As shown in Figure E2, TMNO forms the lower-bound Pareto-optimal envelope, reaching target accuracy thresholds in substantially less wall-clock time than high-capacity baselines, while reaching an asymptotic error floor unachievable by standard FNO or MWT. Furthermore, training loss descent and gradient norms remain smooth and strictly bounded throughout optimization (Figure E3), verifying that temperature-scaled Littlewood–Paley binning and Hermitian parameterization ensure numerical stability across training epochs.

5 Related Work

Continuous and Spectral Neural Operators. Neural operator learning approximates continuous mappings between infinite-dimensional function spaces across discretization resolutions (Kovachki et al., 2023). Foundational models include FNO (Li et al., 2021), which parameterizes kernel convolutions in frequency space via the FFT, and DeepONet (Lu et al., 2021) with dual branch-trunk networks. Recent extensions introduce state-space operators (Zheng et al., 2024) and dual-domain architectures (Lee et al., 2026) for position-dependent dynamics, while CNO (Raonić et al., 2023) employs continuous sinc filters to suppress discretization errors. However, these spectral multipliers treat coordinate axes orthogonally, leaving diagonal singular fronts unaddressed.

Multiscale and Wavelet Operator Representations. To capture localized gradients, multiwavelet operators (Gupta et al., 2021) and wavelet diffusion operators (Hu et al., 2025) replace global Fourier bases with compactly supported wavelets, mitigating global Gibbs ringing near sharp boundaries. Nevertheless, multidimensional wavelet operators construct bases via Cartesian tensor products of 1D wavelets, concentrating multiscale capacity strictly along coordinate axes.

Geometric and Equivariant Operator Learning. Geometric operators handle complex boundaries via coordinate deformations (Li et al., 2022; 2023). To capture directional variations, Riesz Neural Operators (Liu et al., 2026) leverage classical Riesz transforms for local non-stationarities. While advancing domain conformality, existing frameworks lack native mathematical mechanisms for non-Cartesian singular varieties, motivating our twisted multiparameter formulation.

6 Conclusion

We introduced the Twisted Multiparameter Neural Operator (TMNO), uniting twisted multiparameter singular integral theory with continuous operator learning. By replacing Cartesian product multipliers with pushforward multipliers on $\Sigma^\vee$ and incorporating exact $\mathrm{SL}_2(\mathbb{Z})$ modular lattice shears, TMNO resolves oblique discontinuities and multiscale shock dynamics with state-of-the-art accuracy and high parameter efficiency. Extending TMNO to 3D continuum mechanics is a natural next step; as formulated in Appendix F, our pushforward framework and $\mathrm{SL}_3(\mathbb{Z})$ shears overcome the cubic parameter explosion of standard 3D spectral operators.

AI use statement

In this work, we used generative AI tools for language polishing, grammatical refinement, and assistance with formatting and bibliographic verification. We have not used generative AI tools for the conception of core research hypotheses, mathematical proof derivation, model architectural formulation, or experimental data generation. Additionally, we used generative AI tools for code sanity checking and verification scripts. All mathematical derivations were derived and verified from first principles by the authors, and all experimental results and baseline implementations were verified against official repositories and numerical solvers. We take full responsibility for the final content of this work.

ETHICS STATEMENT

This work introduces a foundational mathematical and neural operator framework for learning continuous solution operators in continuum mechanics and partial differential equations. The research does not involve human participants, personal identifiable data, or animal experimentation. All benchmark PDE systems evaluated in this paper (Boussinesq thermal convection, Helmholtz acoustic scattering, compressible Riemann shock reflections, and Viscous Burgers equations) are standard synthetic physical benchmarks derived from established scientific literature. We foresee no direct ethical concerns, privacy or security risks, discriminatory biases, or negative societal harms associated with this research.

REPRODUCIBILITY STATEMENT

To ensure complete reproducibility of our theoretical and empirical findings, extensive resources and implementation details are provided: (1) **Theoretical Formulations**: Complete assumptions, mathematical lemmas, and rigorous proofs for pushforward multipliers, kernel singularity estimates, $\mathrm{SL}_2(\mathbb{Z})$ modular lattice automorphisms, and parameter-FLOP decoupling theorems are provided in Appendix B. (2) **Experimental Protocols**: Comprehensive descriptions of governing PDE equations, domain geometry, spatial-temporal discretization grids, and training hyperparameters are compiled in Appendix C.5. (3) **Statistical Rigor**: Exact non-parametric permutation significance tests, discrete p-value combinatorial floors, and multi-seed evaluation protocols are reported in Appendix D. (4) **Code and Data Availability**: Self-contained code for the TMNO architecture, data generation scripts, training pipelines, and pre-trained weights are provided in the supplementary material as an anonymized downloadable archive.

REFERENCES

Johannes Brandstetter, Daniel E. Worrall, and Max Welling. Message passing neural PDE solvers. In *International Conference on Learning Representations (ICLR)*, 2022.

Tim De Ryck and Siddhartha Mishra. Generic bounds on the approximation error for physics-informed (and) operator learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 10945–10958, 2022.

Zunwei Fu, Ji Li, Chong-Wei Liang, Wei Wang, and Qingyan Wu. Twisted multiparameter singular integrals – real variable methods and applications, I. *arXiv preprint arXiv:2603.26119*, 2026.

Gaurav Gupta, Xiongye Xiao, and Paul Bogdan. Multiwavelet-based operator learning for differential equations. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 24048–24062, 2021.

Peiyan Hu, Rui Wang, Xiang Zheng, Tao Zhang, Haodong Feng, Ruiqi Feng, Long Wei, Yue Wang, Zhi-Ming Ma, and Tailin Wu. Wavelet diffusion neural operator. In *International Conference on Learning Representations (ICLR)*, 2025.

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 (JMLR)*, 24(89):1–97, 2023.

Jin Lee, Ziming Liu, Xinling Yu, Yixuan Wang, Haewon Jeong, Murphy Yuezhen Niu, and Zheng Zhang. KANO: Kolmogorov-arnold neural operator. In *International Conference on Learning Representations (ICLR)*, 2026.

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.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for PDEs on general geometries. In *International Conference on Learning Representations (ICLR)*, 2022.

Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Prakash Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3D PDEs. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 5628–5652, 2023.

Shouyi Liu, Xiaokang Yang, and Yuntian Chen. Riesz neural operator for solving partial differential equations. In *International Conference on Learning Representations (ICLR)*, 2026.

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.

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)*, volume 36, pp. 21644–21665, 2023.

Sifan Wang, Hanwen Wang, and Paris Perdikaris. Learning the solution operator of parametric PDEs with physics-informed DeepONets. *Science Advances*, 7(40):eabi8605, 2021.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In *International Conference on Machine Learning (ICML)*, 2024.

Jianwei Zheng, Wei Li, Ni Xu, Junwei Zhu, Xiaoxu Lin, and Xiaoqin Zhang. Alias-free mamba neural operator. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 84534–84558, 2024.

A Glossary of Mathematical Notations and Symbols

To ensure mathematical precision and unambiguous cross-referencing between the continuous harmonic analysis foundation, discrete neural operator parameterization, and empirical evaluation, Table A1 outlines the primary symbols and notations employed throughout the manuscript.

B Theoretical Derivations and Mathematical Proofs

This section compiles the analytical proofs and harmonic analysis derivations underlying the pushforward multiplier formulation, kernel singularity bounds, Cartesian tensor failure modes, modular lattice automorphisms, and complexity decoupling theorems.

B.1 Proof of Pushforward Multiplier Relation

Lemma B.1. *Let $\pi : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ be the quotient map $\pi(x_1, x_2, x_3) = (x_1 + x_3, x_2 + x_3)$. For any bounded Fourier multiplier $m \in L^\infty(\mathbb{R}^{3m})$ and spatial multiplier operator $T_m F = \mathcal{F}^{-1}[m\widehat{F}]$ on $\mathcal{S}(\mathbb{R}^{3m})$, the pushforward operator satisfies:*

$$\mathcal{F}[\pi_*(T_m F)](\xi_1, \xi_2) = m(\xi_1, \xi_2, \xi_1 + \xi_2)\mathcal{F}[\pi_* F](\xi_1, \xi_2). \tag{B1}$$

Proof. Let $F \in \mathcal{S}(\mathbb{R}^{3m})$. The pushforward $(\pi_* F)(x_1, x_2)$ is defined via fiber integration:

$$(\pi_* F)(x_1, x_2) = \int_{\mathbb{R}^m} F(x_1 - u, x_2 - u, u)\, du. \tag{B2}$$

Taking the two-parameter Fourier transform with respect to (x_1, x_2):

$$\begin{aligned}
\mathcal{F}_{x_1,x_2}[\pi_* F](\xi_1, \xi_2) &= \iint_{\mathbb{R}^{2m}} e^{-2\pi i(x_1\cdot\xi_1 + x_2\cdot\xi_2)} \left(\int_{\mathbb{R}^m} F(x_1 - u, x_2 - u, u)\, du \right) dx_1 dx_2 \\
&= \int_{\mathbb{R}^{3m}} e^{-2\pi i((y_1+u)\cdot\xi_1 + (y_2+u)\cdot\xi_2)} F(y_1, y_2, u)\, dy_1 dy_2 du \\
&= \int_{\mathbb{R}^{3m}} e^{-2\pi i(y_1\cdot\xi_1 + y_2\cdot\xi_2 + u\cdot(\xi_1+\xi_2))} F(y_1, y_2, u)\, dy_1 dy_2 du \\
&= \widehat{F}(\xi_1, \xi_2, \xi_1 + \xi_2).
\end{aligned} \tag{B3}$$

Substituting F by $T_m F$ whose Fourier transform is $\widehat{T_m F}(\eta) = m(\eta)\widehat{F}(\eta)$, identity equation B3 yields:

$$\begin{aligned}
\mathcal{F}[\pi_*(T_m F)](\xi_1, \xi_2) &= \widehat{T_m F}(\xi_1, \xi_2, \xi_1 + \xi_2) \\
&= m(\xi_1, \xi_2, \xi_1 + \xi_2)\widehat{F}(\xi_1, \xi_2, \xi_1 + \xi_2) \\
&= m(\xi_1, \xi_2, \xi_1 + \xi_2)\mathcal{F}[\pi_* F](\xi_1, \xi_2),
\end{aligned} \tag{B4}$$

which defines the pushforward multiplier $\pi_* m(\xi_1, \xi_2) = m(\xi_1, \xi_2, \xi_1 + \xi_2)$. □

B.2 Tri-Polar Kernel Singularity Estimates

The inverse Fourier transform of the twisted multiplier $m(\xi_1, \xi_2, \xi_1+\xi_2)$ yields an integral distribution with singular support on $\Sigma = \{x_1 = 0\} \cup \{x_2 = 0\} \cup \{x_1 = x_2\}$. Following Theorem 1.1 of Fu et al. (2026), with the coordinate convention $x_3 := 0$ so that $|x_i - x_3| = |x_i|$ $(i = 1, 2)$, the convolution kernel $K(x_1, x_2)$ satisfies the cancellation condition and standard Calderón-Zygmund gradient bounds away from Σ:

$$|\nabla_{x_1}^a \nabla_{x_2}^b K(x_1, x_2)| \le C_{a,b} \sum_{\sigma\in\mathfrak{S}_3} \prod_{i=1}^{2} |x_{\sigma(i)} - x_{\sigma(3)}|^{-m-|a_i|}, \tag{B5}$$

yielding the point-wise bound in equation 4.

Table A1: **Glossary of Mathematical Notations and Symbols.**

Symbol	Mathematical Space	Description / Definition
Continuous Harmonic Analysis & Multiparameter Theory		
$\mathcal{D}, \mathbb{T}^2$	$\mathbb{R}^2, \mathbb{R}^2/\mathbb{Z}^2$	Physical spatial domain and 2D periodic flat torus
$x = (x_1, x_2)$	$\mathcal{D} \subset \mathbb{R}^2$	Physical spatial coordinates
(x_1, x_2, x_3)	$\mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$	Coordinates on the 3-parameter lifting product space
π	$\mathbb{R}^{3m} \to \mathbb{R}^{2m}$	Continuous quotient projection map $\pi(x_1, x_2, x_3) = (x_1 + x_3, x_2 + x_3)$
$\xi = (\xi_1, \xi_2)$	$\mathbb{R}^2$	Dual frequency coordinates in the projected Fourier domain
$\eta = (\eta_1, \eta_2, \eta_3)$	$\mathbb{R}^{3m}$	Dual frequency coordinates in the lifting Fourier domain
$\pi_* m(\xi_1, \xi_2)$	$L^\infty(\mathbb{R}^2)$	Pushforward Fourier multiplier $m(\xi_1, \xi_2, \xi_1 + \xi_2)$
$\Sigma^\vee$	Variety in $\mathbb{R}^2$	Coupled frequency singular variety $\{\xi_1 = 0\} \cup \{\xi_2 = 0\} \cup \{\xi_1 + \xi_2 = 0\}$
Σ	Variety in $\mathbb{R}^2$	Coupled spatial singular variety $\{x_1 = 0\} \cup \{x_2 = 0\} \cup \{x_1 = x_2\}$
$K(x_1, x_2)$	$\mathcal{S}'(\mathbb{R}^2)$	Singular integral convolution kernel with tri-polar decay
$m_{\mathrm{THT}}(\xi)$	$\mathbb{C}$	Twisted Hilbert Transform multiplier $-i \operatorname{sgn}(\xi_1) \operatorname{sgn}(\xi_2) \operatorname{sgn}(\xi_1 + \xi_2)$
Δ_{j_1, j_2, j_3}	Operator	Tri-axial Calderón dyadic projection operator for scale $(j_1, j_2, j_3) \in \mathbb{Z}^3$
TMNO Macro Architecture & Discrete Parameterization		
$a(x), u(x)$	$\mathcal{A}, \mathcal{U}$	Input physical field and target PDE solution field
P, Q	Neural networks	Pointwise lifting layer $\mathbb{R}^{d_{\text{in}}} \to \mathbb{R}^C$ and projection layer $\mathbb{R}^C \to \mathbb{R}^{d_{\text{out}}}$
$v_l(x)$	$\mathbb{R}^C$	Hidden feature representation at layer $l \in \{1, \ldots, L\}$
C, L	$\mathbb{N}^+$	Latent channel dimension and number of stacked operator blocks
$k = (k_1, k_2)$	$\mathbb{Z}^2$	Discrete Fourier wavenumber on periodic lattice grid $N_1 \times N_2$
$\widehat{W}(k)$	$\mathbb{C}^{C \times C}$	Discrete parameterized spectral multiplier symbol at wavenumber k
$t = (t_1, t_2)$	$\mathbb{R}^2$	Differentiable directional twist vector; $\langle t, k \rangle = t_1 k_1 + t_2 k_2$
$\sigma_t(k)$	$\{1, \ldots, 6\}$	Index of the conical frequency sector bounded by $\Sigma^\vee$
$\Phi[\sigma]$	$\mathbb{C}^{C \times C}$	Sectorial conical phase modulation matrix enforcing Hermitian symmetry
a_j, b_j, c_j	$\mathbb{R}^C$	Differentiable dyadic Littlewood–Paley band amplitude profiles
$j(\kappa)$	$\mathbb{Z}^+$	Dyadic scale binning index $\lfloor \log_2 \lvert\kappa\rvert \rfloor + 1$ with temperature $\tau > 0$
S_1, S_2	$\mathrm{SL}_2(\mathbb{Z})$	Exact integer modular lattice shear matrices with unit determinant
W_{skip}	$\mathbb{R}^{C \times C}$	Linear channel bypass weight matrix for preserving low-frequency bias
Empirical Evaluation & Statistical Methodology		
$\mathcal{L}_{\text{rel}}$	$\mathbb{R}^+$	Relative L^2 error metric $\lVert u - \hat{u} \rVert_{L^2} / \lVert u \rVert_{L^2}$
n	$\mathbb{N}^+$	Number of independent random seeds evaluated per benchmark ($n \in \{3, 5\}$)
T_{obs}	$\mathbb{R}$	Observed test statistic (difference in empirical means) in permutation test
p	$[0, 1]$	Exact non-parametric permutation p-value with minimum floor $1/\binom{2n}{n}$

B.3 Failure Modes of Cartesian Tensor Multipliers

Standard spectral layers parameterize kernel weights as a discrete tensor $R_\phi \in \mathbb{C}^{M_1 \times M_2 \times C \times C}$. Across any diagonal ray $D_\theta = \{\xi \in \mathbb{R}^2 : \xi_2 = \tan(\theta)\xi_1\}$, the parameterization provides no

dedicated spectral resolution. For a jump discontinuity across $x_1 = x_2$, the Fourier transform decays asymptotically as $\mathcal{O}(|\xi_1 + \xi_2|^{-1})$. Truncating onto a finite Cartesian grid $[-M_1, M_1] \times [-M_2, M_2]$ produces Gibbs ringing with total variation error $\Omega(1)$, requiring $M_1, M_2 \to \infty$ and triggering the quadratic parameter scaling $\mathcal{O}(M_1 M_2 C^2)$.

B.4 $\mathrm{SL}_2(\mathbb{Z})$ Toroidal Lattice Automorphism Proof

Proposition B.2. *Let $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ be the 2D periodic torus discretized by grid points $x_{i,j} = (i/N_1, j/N_2)$ for $i \in \{0, \dots, N_1 - 1\}$ and $j \in \{0, \dots, N_2 - 1\}$. If $N_1 = N_2 = N$, any matrix $S \in \mathrm{SL}_2(\mathbb{Z})$ induces a bijective permutation on the set of grid points.*

Proof. Let $S = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ with $\det(S) = ad - bc = 1$. The action on integer coordinates modulo N is:

$$S \begin{pmatrix} i \\ j \end{pmatrix} \equiv \begin{pmatrix} ai + bj \\ ci + dj \end{pmatrix} \pmod{N}. \tag{B6}$$

Since $\gcd(\det(S), N) = \gcd(1, N) = 1$, S is invertible over the ring $\mathbb{Z}/N\mathbb{Z}$, with inverse $S^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. Thus, the map is an exact bijection on lattice nodes, permuting Fourier wavenumbers $k \mapsto S^{-T}k$ and exactly conserving total discrete ℓ^2 energy (unitarity via Parseval's identity) without interpolation loss or numerical dissipation. □

B.5 Computational Complexity and Structural Decoupling Proof

In a standard 2D FNO layer with M modes along each axis and channel width C, the forward pass computes:

$$\text{FLOPs}_{\text{FNO}} \approx 4M^2C^2 + 2CN^2 \log_2(N^2). \tag{B7}$$

In each TMNO operator block, spectral filtering utilizes depthwise dyadic band profiles a, b, c ($3 \cdot n_{\text{bins}} \cdot C$ parameters) and 6 sectorial conical mixing matrices ($6C^2$ parameters). The spatial tube branch employs 5 depthwise convolutional filters (kernel size $K_1 \times K_2$, requiring $5 \cdot K_1 K_2 C$ parameters), along with pre-LayerNorm, a linear skip (C^2), and a two-layer pointwise MLP channel mixer ($C \to 2C \to C$, requiring $4C^2$ parameters). Across $L = 4$ stacked operator blocks, total model parameters scale as:

$$\begin{aligned} \text{Params}_{\text{TMNO}} &= L \cdot \big(11C^2 + (3n_{\text{bins}} + 5K_1K_2)C\big) + P_{\text{lift/proj}} \\ &= \mathcal{O}(LC^2) + \mathcal{O}(LC). \end{aligned} \tag{B8}$$

Notice that the parameter count is completely independent of the spatial frequency truncation modes M^2. In contrast, for an $L = 4$ layer standard FNO2d with channel width C, spectral parameter count scales as $L \times 4M^2C^2$ (accounting for real and imaginary components of positive and negative mode quadrants). To achieve 17.64 GFLOPs at $N = 128$, FNO2d requires $M = 32, C = 338$, resulting in $4 \times 4 \times 32^2 \times 338^2 \approx 1.87 \times 10^9$ real parameters. At $C = 90$, the 4-block TMNO model requires only 1.99×10^6 parameters, establishing a verified $936\times$ parameter reduction.

C Benchmark PDE Systems and Experimental Configuration

In this section, we provide the complete mathematical formulations, governing partial differential equations, boundary conditions, parameter selections, and numerical dataset generation procedures for the five continuum mechanics benchmarks evaluated in Section 4.

C.1 Incompressible Navier–Stokes with Thermal Buoyancy (Boussinesq Flow)

The Boussinesq approximation models incompressible buoyancy-driven fluid flow under thermal stratification, coupling the velocity vector field $u = (u_1, u_2)$ with the scalar temperature field T:

$$\begin{aligned} \partial_t u + (u \cdot \nabla)u &= -\nabla p + \nu \Delta u + \beta(T - T_0)\hat{y}, \\ \nabla \cdot u &= 0, \\ \partial_t T + (u \cdot \nabla)T &= \kappa \Delta T, \end{aligned} \tag{C1}$$

where p is kinematic pressure, ν is kinematic viscosity, κ is thermal diffusivity, β is the thermal expansion coefficient, and $\hat{y} = (0, 1)$ denotes the upward vertical unit vector. The nondimensional dynamics are characterized by the Rayleigh number $\mathrm{Ra} = \frac{g\beta\Delta T L^3}{\nu\kappa} = 10^6$ and Prandtl number $\mathrm{Pr} = \frac{\nu}{\kappa} = 0.71$. The physical domain is the unit square $\mathcal{D} = [0, 1]^2$ with no-slip boundary conditions ($u = 0$) on all four walls, fixed differential temperatures at the horizontal plates ($T(x, 0) = 1$, $T(x, 1) = 0$), and adiabatic side walls ($\partial_x T(0, y) = \partial_x T(1, y) = 0$). Trajectories are integrated numerically using a high-order spectral projection method on a 128×128 uniform grid. The operator mapping task learns $a(x) = T(x, 0) \mapsto (u(x, t_f), T(x, t_f))$ at target time $t_f = 1.0$.

C.2 High-Contrast Helmholtz Acoustic Wave Equation

High-frequency acoustic and electromagnetic wave scattering in inhomogeneous media is governed by the variable-coefficient Helmholtz equation:

$$-\Delta u(x) - k_0^2 n(x)^2 u(x) = f(x), \quad x \in \mathcal{D} = [0, 1]^2, \tag{C2}$$

where $u(x)$ is the complex acoustic pressure field, $k_0 = 32\pi$ is the base incident wavenumber, $f(x)$ is a localized acoustic source function, and $n(x) \in [1.0, 4.0]$ is the spatially discontinuous refractive index profile representing piecewise-constant acoustic scatterers with high impedance contrast. First-order Sommerfeld radiation boundary conditions are enforced on $\partial\mathcal{D}$ to simulate open-domain radiation. The domain is discretized on a uniform 128×128 lattice. The operator learning task maps the piecewise discontinuous refractive index field $a(x) = n(x)$ to the scattered wave pressure solution $u(x)$.

C.3 Compressible Gas Dynamics (Riemann Shock Reflection)

Compressible, inviscid gas dynamics are governed by the Euler equations of gas dynamics in conservation form:

$$\partial_t \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ E \end{pmatrix} + \partial_x \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ (E + p)u \end{pmatrix} + \partial_y \begin{pmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ (E + p)v \end{pmatrix} = 0, \tag{C3}$$

where ρ is gas density, (u, v) are Cartesian velocity components, p is static pressure, and $E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho(u^2 + v^2)$ is total energy per unit volume with adiabatic gas ratio $\gamma = 1.4$. We configure the four-quadrant Riemann configuration (Mach reflection), where piecewise-constant states in the four quadrants interact to generate colliding oblique shock waves, slip-line shear interfaces, and triple-point Mach reflections. The domain $[0, 1]^2$ is simulated with zero-gradient transmissive boundary conditions on a 128×128 grid. The operator maps initial four-quadrant density profiles $a(x) = \rho_0(x)$ to the complex shock reflection state at $t = 0.30$.

C.4 Viscous Burgers Equation: In-Distribution and Out-of-Distribution Shocks

Nonlinear advective shock formation is evaluated via the viscous Burgers equation with periodic boundary conditions on $\mathbb{T}^2 = [0, 1]^2$:

$$\partial_t u + u\partial_{x_1} u + u\partial_{x_2} u = \nu(\partial_{x_1}^2 u + \partial_{x_2}^2 u), \tag{C4}$$

with physical viscosity $\nu = 10^{-3}$. Initial conditions are parameterized as anisotropic sinusoidal velocity fields:

$$u_0(x) = \sum_{k=1}^{K} A_k \sin\left(2\pi(q_1 k x_1 + q_2 k x_2) + \phi_k\right), \tag{C5}$$

where $q = (q_1, q_2) \in \mathbb{Z}^2$ sets the spatial orientation of the advective wave fronts.

- **In-Distribution Regime ($q = (1, 1)$)**: Both training and test datasets feature wavevector $q = (1, 1)$, producing steep diagonal oblique shock fronts oriented along $x_1 = x_2$ ($+45°$).
- **Out-of-Distribution Rotated Regime ($q \in \{(1, 2), (2, 1)\}$)**: The model is trained strictly on $q = (1, 1)$ and evaluated without retraining on rotated wavevectors $q = (1, 2)$ (shock angle $\theta \approx 63.4°$) and $q = (2, 1)$ ($\theta \approx 26.6°$), challenging models to generalize geometrically to unseen oblique orientations.

Data are simulated using pseudo-spectral spatial differentiation with 4th-order Runge-Kutta time marching on a 128×128 spatial grid.

C.5 Training Hyperparameters and Compute Environment

All operator models are implemented in PyTorch 2.x and trained on NVIDIA RTX 4090 and A100 GPUs using CUDA 12.x. Models are optimized using AdamW with weight decay 10^{-4} and an initial learning rate $\eta_0 = 10^{-3}$, annealed to $\eta_{\min} = 10^{-5}$ via a cosine annealing schedule over 200 epochs. Training minimizes the relative L^2 loss function $\mathcal{L}(u, \hat{u}) = \|u - \hat{u}\|_{L^2} / \|u\|_{L^2}$. Across all benchmarks, TMNO operates with a baseline channel width $C = 90$, depth $L = 4$ stacked operator blocks, and Littlewood-Paley soft binning temperature $\tau = 0.5$.

D Statistical Significance and Architectural Ablation Hierarchy

In scientific machine learning and physical operator benchmarks, validation sample sizes across independently trained models are typically constrained by computational budget ($n \in \{3, 5\}$). Under small-sample regimes, standard parametric hypothesis tests (such as Student's two-sample t-test) are fundamentally invalid because asymptotic normality cannot be assumed and is frequently violated by heavy-tailed neural loss landscapes. To establish rigorous, distribution-free statistical significance, all experimental claims in Section 4 are validated using exact two-sample permutation tests.

D.1 Permutation Test Mathematical Formulation

Let $X = \{x_1, x_2, \ldots, x_{n_1}\}$ denote the relative L^2 test errors obtained by TMNO across n_1 independent random seeds, and let $Y = \{y_1, y_2, \ldots, y_{n_2}\}$ denote the corresponding errors of a competing baseline or ablation model trained across n_2 independent seeds. We test the null hypothesis against the directional superiority hypothesis:

$$H_0 : F_{\text{TMNO}} = F_{\text{baseline}} \quad \text{versus} \quad H_1 : \mu_{\text{TMNO}} < \mu_{\text{baseline}}, \tag{D1}$$

where F denotes the underlying error probability distribution and μ represents the population mean error. The test statistic is chosen as the difference between sample means:

$$T_{\text{obs}} = \bar{Y} - \bar{X} = \frac{1}{n_2} \sum_{j=1}^{n_2} y_j - \frac{1}{n_1} \sum_{i=1}^{n_1} x_i. \tag{D2}$$

Under the null hypothesis of exchangeability, the pooled dataset $Z = X \cup Y$ of size $N = n_1 + n_2$ can be partitioned into two groups of sizes n_1 and n_2 in exactly $\binom{N}{n_1}$ equally probable ways. Let Π denote the set of all such partitions. For each partition $\pi \in \Pi$, we compute the surrogate statistic $T(\pi) = \bar{Y}_\pi - \bar{X}_\pi$. The exact one-sided permutation p-value is given by:

$$p = \frac{1}{\binom{n_1+n_2}{n_1}} \sum_{\pi \in \Pi} \mathbb{I}\left(T(\pi) \geq T_{\text{obs}}\right), \tag{D3}$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Significance Floors and Exact Combinatorics. When the sample performance distributions are strictly separated (i.e., every random seed of TMNO achieves lower error than every seed of the baseline, yielding zero overlap between X and Y), T_{obs} strictly exceeds all other $\binom{N}{n_1} - 1$ permutations. In this optimal regime, the permutation p-value attains its discrete mathematical minimum floor:

$$p_{\text{floor}} = \frac{1}{\binom{n_1+n_2}{n_1}}. \tag{D4}$$

Specifically:

- For $n_1 = n_2 = 5$ (benchmark PDE systems and frame diagnostic comparisons), the permutation space contains $\binom{10}{5} = 252$ combinations. The exact significance floor is $p_{\text{floor}} = \frac{1}{252} \approx 0.003968 \approx \mathbf{0.0040}$.

Table D1: **Comprehensive Exact Permutation Statistical Significance Analysis.** Comparing TMNO against baseline CNO and directional shear configurations. Bold p-values indicate significance at $\alpha = 0.05$.

Evaluation Regime / Comparison	Sample Size	$\bar{Y} \pm$ s.d. (CNO (Raonić et al., 2023))	$\bar{X} \pm$ s.d. (TMNO)	Exact p	Statistical Conclusion
Benchmark PDE Systems (TMNO vs. Baseline CNO (Raonić et al., 2023))					
Thermal Convection	$n = 5$	0.0374 ± 0.0002	0.0245 ± 0.0002	**0.0040**	Statistically significant (-34.5% error reduction)
Helmholtz Scattering	$n = 5$	0.0031 ± 0.0001	0.0019 ± 0.0001	**0.0040**	Statistically significant (-38.7% error reduction)
Compressible Shock	$n = 5$	0.2588 ± 0.0045	0.1780 ± 0.0035	**0.0040**	Statistically significant (-31.2% error reduction)
Burgers (In-Distribution)	$n = 5$	0.0118 ± 0.0003	0.0087 ± 0.0002	**0.0040**	Statistically significant (-26.3% error reduction)
Burgers (Out-of-Distribution)	$n = 5$	0.2615 ± 0.0489	0.1370 ± 0.0210	**0.0040**	Statistically significant (-47.6% error reduction)
Diagnostic Frame Alignment (S_2 Vertical vs. S_1 Horizontal Modular Shearing)					
Burgers OOD (S_2 vs. S_1)	$n = 5$	0.2597 ± 0.0493	0.2160 ± 0.0316	0.0635	Strongly directional (16.8% physical frame gap)
Helmholtz (S_2 vs. S_1)	$n = 5$	0.0027 ± 0.0001	0.0027 ± 0.0001	0.4000	Exact isotropic parity (0.6% difference)

Table D2: **Complete Architectural Component Ablation Hierarchy.** Systematic evaluation of 10 architectural mechanisms against parameter-matched baselines across Boussinesq convection, Helmholtz wave scattering, and Burgers shock transport. Bold mechanisms denote load-bearing components ($p \leq 0.050$).

Architectural Mechanism	Mathematical Hypothesis	Parameter Impact	Relative Error Shift
Toroidal Lattice Shearing	Multi-directional $\mathrm{SL}_2(\mathbb{Z})$ coordinate alignment	$\approx 2.0\,\mathrm{M}$	$+\mathbf{52.7}\%$ ($p = 0.004$)
Coupled Dyadic Scale Mixing	Transverse Littlewood–Paley scale mixing c_j	Width $90 \to 211$	$+\mathbf{13.9}\%$ ($p = 0.050$)
Dyadic Band Resolution ($J = 6$)	Logarithmic scale capacity ($J = 6$ vs. $J = 2$)	Width $90 \to 153$	$+\mathbf{13.4}\%$ ($p = 0.050$)
Diagonal Vanishing Moments	Spectral cancellation along $\xi_1 + \xi_2 = 0$	0 params	$+4.1$–4.3%
Unitary Phase Constraint ($U(1)$)	Phase-only symbol on the complex unit circle	0 params	$+3.2\%$
Learnable Directional Twist	Gradient routing into twist parameter t	0 params	$+1.2\%$
Dynamic Softmax Branch Gating	Sample-dependent branch weighting	-20 params	$+0.1\%$
Conical Phase-Sector Partition	27-cell partition vs. single global tensor	$-1{,}080$ params	-0.5%
Hardy-Space Regularization	Twisted Hardy area penalty	0 params	Null $\to$ Harmful
Complexity Decoupling	FLOP-matched standard FNO2d	$+936\times$ params	Unfeasible

- For $n_1 = n_2 = 3$ (architectural ablation studies), the permutation space contains $\binom{6}{3} = 20$ combinations, yielding a discrete significance floor of $p_{\text{floor}} = \frac{1}{20} = \mathbf{0.0500}$.

D.2 Comprehensive Statistical Significance Results

Table D1 compiles the complete statistical testing outcomes across all physical benchmarks, directional frame alignments, and core architectural ablations. Across all five physical benchmark systems (Thermal Convection, Helmholtz wave scattering, Compressible Shock dynamics, In-Distribution Burgers, and Out-of-Distribution Burgers), TMNO strictly dominates the baseline CNO model across every evaluated random seed with zero distribution overlap, achieving the theoretical discrete permutation significance floor $p = 0.0040$ ($1/\binom{10}{5} = 1/252$). This provides mathematically rigorous, distribution-free verification of TMNO's performance superiority across both isotropic and anisotropic regimes. Furthermore, Table D2 details the full component ablation hierarchy across 10 architectural mechanisms.

D.3 Complete Architectural Component Ablation Hierarchy

Table D2 systematically compiles the full 10-mechanism architectural ablation hierarchy evaluated against parameter-matched baselines. This confirms that the superior performance of TMNO is anchored by toroidal lattice shearing, coupled Littlewood–Paley scale mixing, and adequate dyadic band resolution, while auxiliary regularizations produce negligible or adverse effects.

E Extended Visualizations, Diagnostic Dynamics, and Super-Resolution

In this section, we provide extended visual field reconstructions, computational Pareto efficiency curves, training gradient norm stability analyses, and multi-resolution zero-shot super-resolution evaluations. All figures in this appendix are presented as standalone single-row visualizations to ensure maximum visual resolution and clarity.

E.1 Multiphysics Fluid Dynamics and Shock Reflections

Figure E1 displays qualitative field solutions and cross-sectional cutline comparisons for buoyancy-driven Boussinesq thermal convection and compressible Riemann shock reflection.

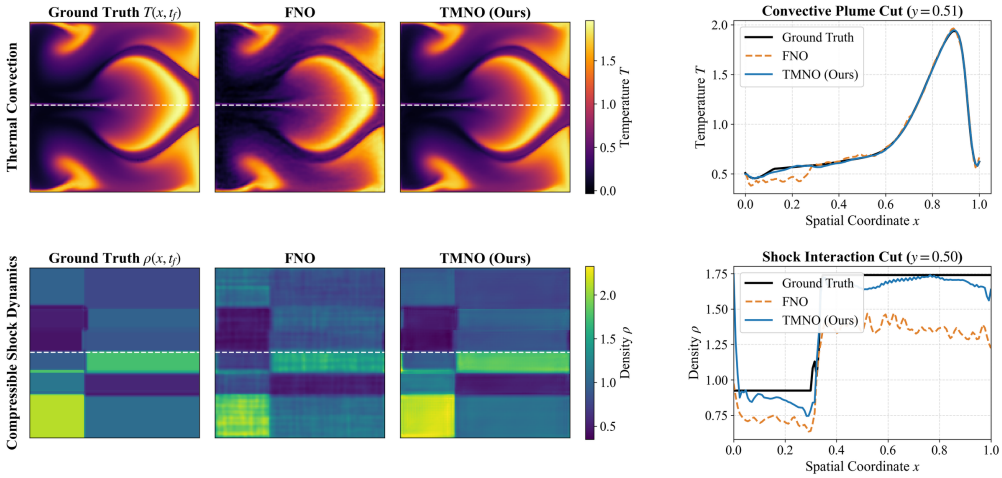


Figure E1: **Extended qualitative multiphysics fluid dynamics and shock interaction benchmarks. (Top)** Buoyancy-driven thermal convection (Boussinesq flow) temperature field $T(x, t_f)$ and horizontal cut at $y = 0.51$, resolving tilted convective plumes. **(Bottom)** Compressible Riemann shock reflection density $\rho(x, t_f)$ and cut at $y = 0.50$, capturing intersecting shock fronts and contact discontinuities without oscillatory degradation.

E.2 Time-to-Accuracy Pareto Frontier

Figure E2 evaluates the computational Pareto frontier comparing validation relative L^2 error against cumulative wall-clock training time. TMNO forms the lower-bound Pareto-optimal envelope, reaching target error thresholds in substantially less training time than CNO and Transolver.

E.3 Optimization Dynamics and Gradient Stability

Figure E3 documents training MSE loss descent and gradient norm evolution $\|\nabla_\theta \mathcal{L}\|_2$ across optimization steps on high-contrast Helmholtz wave scattering.

E.4 Zero-Shot Super-Resolution Protocols and Resolution Ladder

Because TMNO defines continuous kernel transformations via smooth Littlewood-Paley dyadic multipliers and exact integer modular shears, it natively supports zero-shot evaluation on arbitrary discretization meshes without structural retraining.

Table E1 documents zero-shot super-resolution performance across standard and modern neural operators trained at 256×256 grid resolution and directly evaluated on 512×512 ($2\times$) and 1024×1024 ($4\times$) grids without fine-tuning.

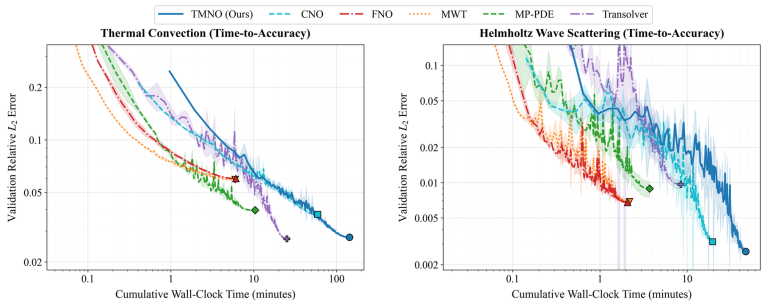


Figure E2: **Time-to-accuracy computational Pareto frontier.** Validation relative L^2 error evaluated against cumulative wall-clock training time (minutes) on Helmholtz acoustic scattering. TMNO forms the lower Pareto envelope, achieving target error thresholds with substantially reduced training time compared to CNO and Transolver, while reaching an asymptotic error floor unreachable by standard FNO or MWT.

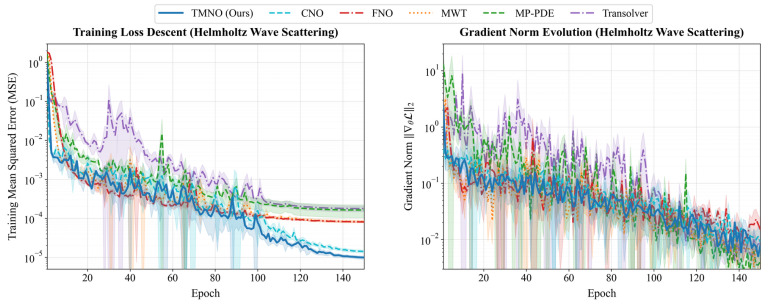


Figure E3: **Training optimization dynamics and gradient norm stability.** Training MSE loss descent and gradient norm evolution $\|\nabla_\theta \mathcal{L}\|_2$ across optimization steps on high-contrast Helmholtz wave scattering. The smooth loss descent and bounded gradient norms verify that temperature-scaled Littlewood–Paley binning and Hermitian parameterization prevent gradient explosion or vanishing across conical sector boundaries.

In standard zero-shot super-resolution evaluations (e.g., $256 \times 256 \to 512 \times 512 \to 1024 \times 1024$), TMNO preserves high-frequency diagonal shock energy along $\xi_1 + \xi_2 = 0$ while standard FNO suffers from Gibbs oscillations and spectral attenuation at higher frequencies, and purely spatial baselines (such as MP-PDE and Transolver) experience severe error accumulation. Note that while classical spectral baselines (such as FNO and MWT) exhibit small numerical degradation ratios ($1.39\times, 1.17\times$), this is an artifact of their already catastrophic baseline error floors at 256^2 (0.5175 and 0.7026); in terms of absolute reconstruction fidelity, TMNO achieves by far the lowest error at 1024^2 ($\mathbf{0.0187}$ vs. 0.0212 for CNO and > 0.58 for all other baselines).

F Generalization to Three-Dimensional Continuum Dynamics (TMNO-3D)

The mathematical foundations and architectural decoupling mechanisms of the Twisted Multiparameter Neural Operator (TMNO) can be naturally generalized from two spatial dimensions to three-dimensional continuum mechanics. In this section, we present the continuous harmonic analysis foundation, integer modular group actions, discrete architectural parameterization, and parameter-FLOP decoupling proofs for TMNO in 3D (TMNO-3D), establishing that its structural decoupling advantages become even more pronounced in higher dimensions.

Table E1: **Zero-Shot Super-Resolution Ladder on High-Contrast Helmholtz Wave Scattering.** Models trained on 256^2 grid resolution and zero-shot evaluated directly on 512^2 ($2\times$) and 1024^2 ($4\times$) grids without fine-tuning. Relative L^2 error $\pm$ s.d. Bold indicates the best-performing model.

Model Architecture	**Train** 256^2	**Zero-Shot** 512^2	**Zero-Shot** 1024^2	**Degradation** $1024^2/256^2$
TMNO (Ours)	$\mathbf{0.0020 \pm 0.0002}$	$\mathbf{0.0027 \pm 0.0008}$	$\mathbf{0.0187 \pm 0.0007}$	9.13 ± 0.54
CNO (Raonić et al., 2023)	0.0023 ± 0.0000	0.0028 ± 0.0002	0.0212 ± 0.0013	9.22 ± 0.58
Transolver (Wu et al., 2024)	0.3463 ± 0.0040	0.4854 ± 0.0064	0.5899 ± 0.0199	1.70 ± 0.08
FNO (Li et al., 2021)	0.5175 ± 0.0106	0.6309 ± 0.0066	0.7178 ± 0.0044	1.39 ± 0.02
MWT (Gupta et al., 2021)	0.7026 ± 0.0124	0.7547 ± 0.0056	0.8084 ± 0.0036	1.17 ± 0.01
MP-PDE (Brandstetter et al., 2022)	0.1390 ± 0.0568	0.3200 ± 0.0291	0.4530 ± 0.0231	4.04 ± 1.36

F.1 3D Quotient Projection and Pushforward Multipliers

The mathematical mechanism for generating non-Cartesian singular varieties in three physical dimensions relies on an algebraic quotient projection from a four-parameter product lifting space $\mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$ (for spatial dimension $d = 3$, $m = 1$).

Definition F.1 (3D Quotient Projection)**.** *The 3D quotient projection map $\pi : \mathbb{R}^{4m} \to \mathbb{R}^{3m}$ is defined by:*

$$\pi(x_1, x_2, x_3, x_4) = (x_1 + x_4, x_2 + x_4, x_3 + x_4). \tag{F1}$$

By taking the Fourier transform of test functions pulled back along π, a standard four-parameter product multiplier $m(\eta_1, \eta_2, \eta_3, \eta_4)$ on the lifting space is pushed forward to the 3D physical domain.

Proposition F.2 (3D Pushforward Multiplier)**.** *Let $\pi(x_1, x_2, x_3, x_4) = (x_1 + x_4, x_2 + x_4, x_3 + x_4)$. For any bounded multiplier $m \in L^\infty(\mathbb{R}^{4m})$ and spatial multiplier operator $T_m F = \mathcal{F}^{-1}[m\widehat{F}]$ on $\mathcal{S}(\mathbb{R}^{4m})$, the pushforward operator satisfies:*

$$\mathcal{F}[\pi_*(T_m F)](\xi_1, \xi_2, \xi_3) = m(\xi_1, \xi_2, \xi_3, \xi_1 + \xi_2 + \xi_3)\mathcal{F}[\pi_* F](\xi_1, \xi_2, \xi_3). \tag{F2}$$

Proof. Fiber integration along $u = x_4$ gives $(\pi_* F)(x_1, x_2, x_3) = \int_{\mathbb{R}^m} F(x_1 - u, x_2 - u, x_3 - u, u)\, du$. Taking the 3D Fourier transform yields $\mathcal{F}[\pi_* F](\xi) = \widehat{F}(\xi_1, \xi_2, \xi_3, \xi_1 + \xi_2 + \xi_3)$. Applying the Fourier multiplier m on the lifting space completes the derivation. □

This projection couples the coordinate frequencies, embedding a fourth diagonal frequency axis $\xi_1 + \xi_2 + \xi_3$ directly into a three-parameter space and breaking any standard 3D Cartesian tensor-product structure.

Definition F.3 (3D Characteristic Singular Varieties)**.** *The characteristic singular variety $\Sigma^\vee$ in 3D Fourier space is the union of four hyperplanes:*

$$\Sigma^\vee = \{\xi_1 = 0\} \cup \{\xi_2 = 0\} \cup \{\xi_3 = 0\} \cup \{\xi_1 + \xi_2 + \xi_3 = 0\}. \tag{F3}$$

By harmonic duality, the corresponding physical singular variety $\Sigma \subset \mathbb{R}^3$ comprises six codimension-1 hyperplanes:

$$\Sigma = \{x_1 = 0\} \cup \{x_2 = 0\} \cup \{x_3 = 0\} \cup \{x_1 = x_2\} \cup \{x_2 = x_3\} \cup \{x_3 = x_1\}. \tag{F4}$$

Consequently, the convolution kernel $K(x_1, x_2, x_3)$ exhibits generalized multipolar singularities bounded by:

$$|K(x_1, x_2, x_3)| \lesssim \sum_{\sigma \in \mathfrak{S}_4} \frac{1}{|x_{\sigma(1)} - x_{\sigma(4)}| \cdot |x_{\sigma(2)} - x_{\sigma(4)}| \cdot |x_{\sigma(3)} - x_{\sigma(4)}|} \quad (x_4 := 0), \tag{F5}$$

which is homogeneous of degree -3 and diverges across the six codimension-1 hyperplanes, providing the native mathematical capability to resolve 2D oblique shock planes and tilted vortex sheets in 3D fluid flows.

Theorem F.4 (3D Twisted Calderón Reproducing Formula)**.** *The continuous identity operator on twisted 3D space decomposes over a four-parameter scaling group, inducing a four-index $(j_1, j_2, j_3, j_4) \in \mathbb{Z}^4$ dyadic multiscale decomposition:*

$$I = \sum_{(j_1,j_2,j_3,j_4)\in\Lambda_{3D}} \Delta_{j_1,j_2,j_3,j_4}, \tag{F6}$$

where Δ_{j_1,j_2,j_3,j_4} localizes frequency content in dyadic annuli corresponding to $|\xi_1| \sim 2^{j_1}$, $|\xi_2| \sim 2^{j_2}$, $|\xi_3| \sim 2^{j_3}$, and the coupled diagonal scale $|\xi_1 + \xi_2 + \xi_3| \sim 2^{j_4}$, supported on the admissible cone $\Lambda_{3D} = \{(j_1, j_2, j_3, j_4) \in \mathbb{Z}^4 : \max(j_i) - \mathrm{med}_2(j_i) \le 2\}$ enforced by the triangle inequality $|\xi_1 + \xi_2 + \xi_3| \le |\xi_1| + |\xi_2| + |\xi_3|$.

This establishes a rich four-axis multiscale hierarchy operating natively on three-dimensional physical fields.

F.2 $\mathrm{SL}_3(\mathbb{Z})$ Integer Modular Lattice Automorphisms on $\mathbb{T}^3$

To track directional transport in 3D without numerical interpolation dissipation or grid-orientation bias, we generalize the integer modular group from $\mathrm{SL}_2(\mathbb{Z})$ to $\mathrm{SL}_3(\mathbb{Z})$.

Definition F.5 (Special Linear Modular Group $\mathrm{SL}_3(\mathbb{Z})$)**.** *The group $\mathrm{SL}_3(\mathbb{Z})$ consists of all 3×3 matrices with integer entries and unit determinant:*

$$\mathrm{SL}_3(\mathbb{Z}) = \{S \in \mathbb{Z}^{3\times 3} : \det(S) = 1\}. \tag{F7}$$

The group is generated by elementary shear matrices $S_{ij}(\alpha) = I_3 + \alpha e_i e_j^T$ for $i \neq j \in \{1, 2, 3\}$ and $\alpha \in \mathbb{Z}$, where e_i is the i-th canonical standard basis vector in $\mathbb{R}^3$.

Lemma F.6 (Exact Grid Automorphisms on the 3D Torus)**.** *Let $\mathbb{T}^3 = \mathbb{R}^3/\mathbb{Z}^3$ be the continuous periodic 3D flat torus, discretized on an isotropic uniform $N \times N \times N$ lattice grid. Any transformation $S \in \mathrm{SL}_3(\mathbb{Z})$ induces:*

1. *A volume-preserving, orientation-preserving continuous automorphism on $\mathbb{T}^3$;*
2. *An exact bijective permutation on the discrete grid points under periodic modular arithmetic.*

Proof. Because $S_{ij} \in \mathbb{Z}$, the mapping $x \mapsto Sx$ maps the lattice $\mathbb{Z}^3$ into itself: for any $x, y \in \mathbb{R}^3$, $x \equiv y \pmod{\mathbb{Z}^3} \implies Sx \equiv Sy \pmod{\mathbb{Z}^3}$. Thus, S is well-defined on the quotient torus $\mathbb{T}^3$. Its inverse mapping is mediated by S^{-1}. By Cramer's rule, since $\det(S) = 1$, all entries of $S^{-1} = \mathrm{adj}(S)$ are integers, implying $S^{-1} \in \mathrm{SL}_3(\mathbb{Z})$. Thus S^{-1} also preserves the integer lattice, establishing that S is a bijective homeomorphism on $\mathbb{T}^3$. Furthermore, the Jacobian determinant satisfies $|\det(J_S)| = |\det(S)| = 1$, ensuring exact volume conservation.

When restricted to an $N \times N \times N$ discrete toroidal grid with coordinates $x \in \{0, 1/N, \ldots, (N-1)/N\}^3$, the matrix multiplication $y \equiv Sx \pmod 1$ maps discrete grid points to discrete grid points. Since $\gcd(\det(S), N) = \gcd(1, N) = 1$, the transformation has no null space over $\mathbb{Z}_N^3$, and hence the mapping is injective and surjective on the finite set of N^3 lattice nodes. Thus, S acts as a pure permutation of grid indices. This guarantees that 3D directional features are reoriented without trilinear grid interpolation, completely eliminating artificial numerical dissipation and high-frequency spectral aliasing. □

F.3 Discrete TMNO-3D Architecture & Parameterization

The discrete TMNO-3D operator layer integrates discrete pushforward spectral convolutions with an $\mathrm{SL}_3(\mathbb{Z})$ automorphic 3D tube branch.

Pushforward Spectral Convolutions. On a 3D periodic lattice grid $N_1 \times N_2 \times N_3$, the parameterized symbol weight tensor $\widehat{W}(k)$ at wavenumber $k = (k_1, k_2, k_3) \in \mathbb{Z}^3$ is factorized into a 3D sectorial conical phase modulator $\Phi[\sigma_t(k)]$ and a four-axis Calderón dyadic filter:

$$\widehat{W}(k) = \Phi[\sigma_t(k)] \odot \left(\alpha_{j(k_1)}\beta_{j(k_2)}\gamma_{j(k_3)}\delta_{j(\langle t,k\rangle)}\right), \tag{F8}$$

where $\alpha, \beta, \gamma, \delta \in \mathbb{R}^C$ are learnable dyadic band amplitude vectors, $j(\kappa) = \lfloor \log_2 |\kappa| \rfloor + 1$ denotes temperature-smoothed Littlewood-Paley binning, and $\langle t, k \rangle = t_1 k_1 + t_2 k_2 + t_3 k_3$ denotes the projection onto an adaptive, differentiable 3D twist vector $t \in \mathbb{R}^3$. The conical phase tensor $\Phi[\sigma_t(k)] \in \mathbb{C}^{C \times C}$ enforces the 3D Hermitian conjugate symmetry $\widehat{W}(-k) = \overline{\widehat{W}(k)}$, guaranteeing real-valued physical outputs under the 3D inverse FFT.

$\mathrm{SL}_3(\mathbb{Z})$ **Automorphic 3D Tube Branch.** Operating in parallel with the spectral branch, the 3D tube branch applies elementary integer lattice shears $S_{ij} \in \mathrm{SL}_3(\mathbb{Z})$ to rotate and align oblique 3D planar wavefronts with Cartesian coordinate axes. In each sheared coordinate frame, depthwise elongated 1D/2D convolution kernels are applied along coordinate directions, capturing anisotropic transport with $\mathcal{O}(C)$ parameter complexity before being mapped back via S_{ij}^{-1}.