HAAR PARAPRODUCT NEURAL OPERATOR

Anonymous authors
Paper under double-blind review

ABSTRACT

Physical-field prediction involves interactions between regional means and internal variations. Bounding individual coupling coefficients, however, need not control their accumulated effect across a hierarchy. We introduce the Haar Paraproduct Neural Operator (HPNO), which learns these interactions through two input-conditioned bounded functions, called symbols. Their Haar coefficients define a mean-to-detail paraproduct and a detail-to-mean adjoint paraproduct on a weighted binary hierarchy, while a bounded Haar multiplier reweights existing components. Regional orthogonality converts pointwise symbol bounds into joint control of squared transfer coefficients within every subtree. This yields a weighted L^2 bound for each fixed-input kernel action, independent of hierarchy size and depth. We also characterize both directional responses across arbitrary tree cuts and express their exact norms through regional symbol variation. HPNO achieves the lowest reported relative ℓ_2 errors among the compared methods on all six PDE benchmarks spanning regular grids, structured meshes, and point clouds, with competitive results on two unstructured computational fluid dynamics benchmarks. Ablations on Darcy and Elasticity support bidirectional coupling and bounded-symbol construction. Paired Darcy tests further show lower regional response errors than HNO and LinearNO under permeability rearrangements that preserve regional input means.

1 INTRODUCTION

Physical AI and digital twins require models that predict how physical systems respond to changes in geometry, material properties, and operating conditions (Kapteyn et al., 2021). Neural operators address this task by learning mappings from physical inputs to solution fields of partial differential equations (PDEs) (Kovachki et al., 2023). These fields often combine large structures and local variations. In turbulent flows, for example, nonlinear dynamics couple structures across spatial scales (Pope, 2000). Predicting such fields requires learning how components at different scales influence one another, rather than only representing each component separately.

Existing neural operators organize spatial interactions through Fourier kernels (Li et al., 2023a), geometry-aware representations (Li et al., 2023b;c), and learned slice states (Wu et al., 2024; Luo et al., 2025). Wavelet and multiwavelet networks learn localized transformations across resolutions (Fan et al., 2019; Gupta et al., 2021; Tripura & Chakraborty, 2023), while wavelet-domain diffusion combines trajectory generation with multiresolution training (Hu et al., 2025). These approaches already provide mechanisms for multiscale representation and interaction. We focus on the construction of the coupling coefficients: how to learn explicit responses between regional means and internal variations while controlling their accumulated strength across nested regions.

A regional mean does not determine how values are arranged within a region. Under a spatially varying kernel, fields with identical regional means can produce different regional output means. Conversely, a regionally constant incoming field can produce internal output variation. These effects motivate *detail-to-mean* and *mean-to-detail* responses, respectively. Paraproducts provide a classical construction for interactions between coarse and detail components (Bony, 1981; Treil, 2013). Their hierarchical structure also raises a question of control: bounding each transfer coefficient separately need not bound the combined response independently of tree depth. Learning these interactions therefore calls for controlling coefficients jointly across nested regions, rather than only constraining their individual magnitudes.

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.

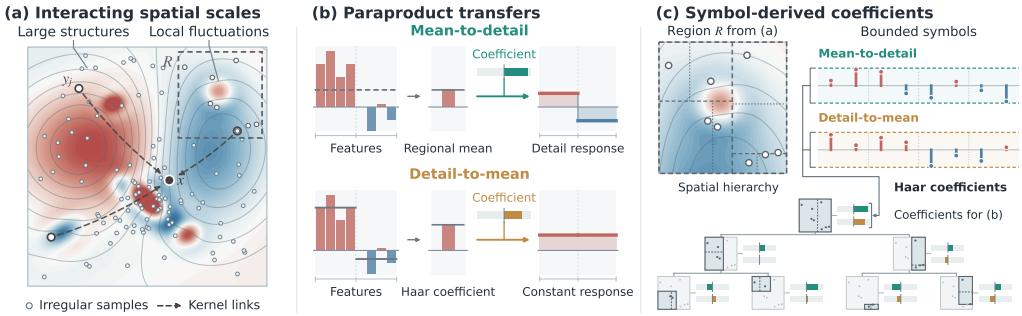

Figure 1: **Haar paraproduct transfers on a spatial hierarchy.** (a) A multiscale field with irregular samples, kernel links from y_j to x, and a highlighted region R. (b) Mean-to-detail and detail-to-mean contributions for equal-mass children. (c) Bounded symbols and their Haar coefficients on the hierarchy of R. Teal and ochre identify the two transfers; red and blue indicate signed values.

We introduce the *Haar Paraproduct Neural Operator (HPNO)* to learn these responses with joint coefficient control. A weighted binary hierarchy organizes the kernel's feature support, and its Haar representation expresses internal contrasts through normalized differences between child-region means. Two input-conditioned bounded functions, called *symbols*, generate the coefficients of a mean-to-detail paraproduct and a detail-to-mean adjoint paraproduct. A learned, bounded Haar multiplier separately reweights existing components. As illustrated in Fig. 1, the directions use separate symbols, while all coefficients within each direction are derived from the same function. Regional Haar orthogonality links the sum of squared coefficients within each subtree to the symbol's regional variance. Pointwise symbol bounds therefore impose joint subtree control, and classical Carleson embedding (Treil, 2013) yields a weighted L^2 bound for each fixed-input kernel action, independent of hierarchy size and depth. Shared upward and downward passes evaluate this action exactly without assembling a dense kernel matrix.

We evaluate HPNO on six PDE benchmarks spanning regular grids, structured meshes, and point clouds, together with two unstructured computational fluid dynamics benchmarks. HPNO achieves the lowest reported relative ℓ_2 errors among the included methods on all six PDE tasks. Ablations on Darcy and Elasticity examine the contributions of directional components and coefficient parameterizations. A paired Darcy experiment further evaluates regional solution changes under permeability rearrangements that preserve regional input means. Our contributions are as follows.

- **Mean–detail kernel construction.** We parameterize separately learned mean-to-detail and detail-to-mean interactions through bounded symbols on a weighted Haar hierarchy.
- **Response characterization and bounds.** We derive exact transfer norms across arbitrary tree cuts and establish a kernel-action bound independent of hierarchy size, depth, and child-mass ratios through joint subtree coefficient control.
- **Field and response evaluation.** We complement benchmark comparisons with learned-kernel diagnostics, architectural ablations, and paired-input evaluation of regional responses to local rearrangements.

2 Related Work

Neural operators and multiresolution learning. Neural operators parameterize mappings between infinite-dimensional function spaces (Kovachki et al., 2023). Fourier Neural Operators (FNO) parameterize convolution kernels in the frequency domain (Li et al., 2023a), with learned deformations and graph embeddings extending spectral processing to irregular geometries (Li et al., 2023b;c). Transolver and Transolver++ structure spatial interactions via learned physical slice states (Wu et al., 2024; Luo et al., 2025), LinearNO linearizes Physics-Attention by bypassing intermediate slice attention (Hu et al., 2026), and HNO introduces Gibbs interaction kernels defined over hyperbolic distances (Pei et al., 2026). Concurrently, multiresolution architectures decompose interactions across scales: BCR-Net leverages nonstandard wavelet forms (Fan et al., 2019), MWT

projects integral kernels onto multiscale polynomial bases (Gupta et al., 2021), and WNO exploits dual spatial-frequency localization (Tripura & Chakraborty, 2023). More recently, WDNO integrates wavelet-domain diffusion for continuous trajectory modeling (Hu et al., 2025), and SAOT couples wavelet subbands with Fourier attention (Zhou et al., 2026). While these architectures provide multiscale representations, HPNO distinguishes itself in its kernel formulation: it couples an exact Haar multiplier with two directional mean–detail transfers whose coefficients are jointly generated by bounded symbols across nested regions.

Paraproduct calculus and operator bounds. Paraproduct calculus decomposes products and nonlinear compositions according to the relative scales of their constituent components (Bony, 1981). In dyadic and martingale analysis, symbol regularity and Carleson coefficient conditions provide the mathematical foundation for establishing paraproduct norm bounds (Beznosova, 2008; Treil, 2013). Recent work generalizes hierarchical paraproducts to finite partition trees (Fasina, 2026) and utilizes tensor paraproducts to regularize and compress kernels defined by nonlinear distance metrics (Fasina & Coifman, 2026). Whereas classical analyses focus on decomposing prescribed analytical kernels, HPNO learns a parametric, input-conditioned kernel family directly from paired data. Its bounded symbols enforce subtree Carleson conditions, allowing classical estimates to yield a weighted $L^2(\mu)$ bound for each fixed-input kernel action, with a constant independent of hierarchy size and depth.

3 Problem Formulation

Setup. Let $\Omega \subset \mathbb{R}^d$ be a spatial domain, and let $u = \mathcal{G}^\dagger_\Omega(a)$ denote the solution field associated with a physical input a. Each instance is observed at $n_\mathcal{X}$ nodes $\mathcal{X} = \{x_j\}_{j=1}^{n_\mathcal{X}} \subset \Omega$ with prescribed positive weights $w = (w_j)$. The available input $a_\mathcal{X}$ contains sampled fields, geometric descriptors, or global conditions broadcast to the nodes. For temporal tasks, it also includes an observed history or a requested time. From input–solution pairs, we learn $\widehat{u}_\mathcal{X} = \mathcal{G}_\theta(\mathcal{X}, w, a_\mathcal{X})$ to predict $u_\mathcal{X} \in \mathbb{R}^{n_\mathcal{X} \times d_u}$, where d_u is the number of output channels. Predictions are evaluated at the supplied nodes, with parameters shared across instances and supported node counts within each task.

Kernel actions. We study input-conditioned kernel actions within $\mathcal{G}_\theta$. Let $\mathcal{S} = \{s_i\}_{i=1}^N$ be the feature support processed by one action, with prescribed positive weights $\omega = (\omega_i)$. For nodal processing, $\mathcal{S} = \mathcal{X}$ and $\omega = w$; otherwise, $\mathcal{S}$ indexes encoded features, so $N = |\mathcal{S}|$ need not equal $n_\mathcal{X}$. For a scalar feature field $f : \mathcal{S} \to \mathbb{R}$, define

$$(\mathcal{K}f)(s) = \sum_{i=1}^{N} \omega_i \, \kappa(s, s_i) \, f(s_i), \qquad s \in \mathcal{S}. \tag{1}$$

The kernel κ depends on $a_\mathcal{X}$ and the learned parameters. We suppress this conditioning in $\mathcal{K}$ and κ, writing $\mathcal{K}[a]$ when comparing conditioning inputs. With the input, parameters, support, and weights fixed, $\mathcal{K}$ is linear in f and acts componentwise on feature channels. The weights define the measure $\mu(\{s_i\}) = \omega_i$ and the norm $\|f\|_\mu^2 = \sum_{i=1}^N \omega_i |f(s_i)|^2$.

Mean–detail responses. Consider a hierarchy of nested regions on $\mathcal{S}$. A cut $\mathcal{C}$ selects disjoint regions that cover $\mathcal{S}$. Let $\mathbb{E}_\mathcal{C}$ replace the field within each selected region by its weighted mean. This orthogonal projection in $L^2(\mu)$ separates the coarse field $\mathbb{E}_\mathcal{C}[f]$ from the detail $(I - \mathbb{E}_\mathcal{C})f$, which has zero weighted mean in each region. The maps $(I - \mathbb{E}_\mathcal{C})\mathcal{K}\mathbb{E}_\mathcal{C}$ and $\mathbb{E}_\mathcal{C}\mathcal{K}(I - \mathbb{E}_\mathcal{C})$ describe mean-to-detail and detail-to-mean responses, respectively. These responses concern the kernel's feature support, which need not coincide with the physical observation nodes.

Research question. How can we construct an input-conditioned kernel action $\mathcal{K}$ with separately learned mean-to-detail and detail-to-mean responses across a hierarchy, while ensuring

$$\|\mathcal{K}f\|_\mu \le B\|f\|_\mu \qquad \text{for all } f, \tag{2}$$

with a constant $B > 0$ independent of the support size N and hierarchy depth?

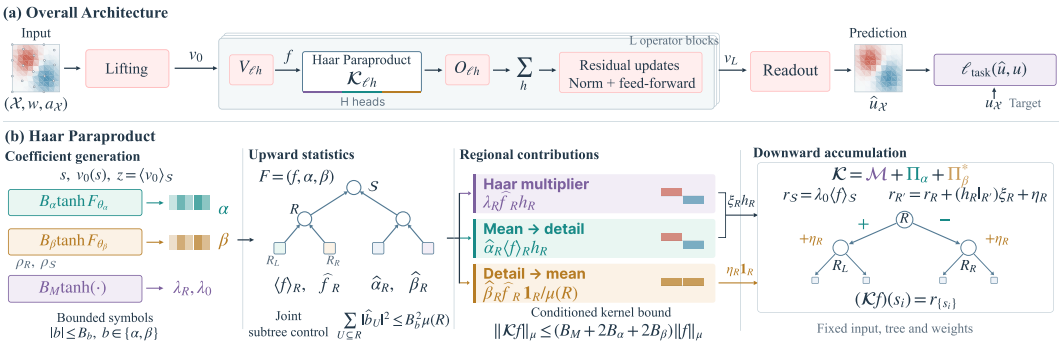

Figure 2: **HPNO architecture and bounded Haar kernel construction.** (a) Input encoding, multi-head operator blocks, readout, and task loss. (b) Input-conditioned symbols and multipliers, shared Haar statistics, parallel regional contributions, and downward accumulation.

4 Haar Paraproduct Neural Operator

4.1 Haar Representation and Paraproducts

Weighted hierarchy. We organize the support $\mathcal{S}$ into a finite binary tree $\mathcal{T}$ with root $\mathcal{S}$ and singleton leaves. Each internal region $R \in \mathcal{T}^\circ$ has two nonempty children R_L, R_R that partition R. The prescribed weights define its mass $\mu(R) = \sum_{s_i \in R} \omega_i$ and mean $\langle f \rangle_R = \mu(R)^{-1} \sum_{s_i \in R} \omega_i f(s_i)$. The same measure defines $\langle f, g \rangle_\mu = \sum_i \omega_i f(s_i) g(s_i)$. Trees use assigned physical, logical, or slot coordinates, according to the task representation.

Haar representation. For $R \in \mathcal{T}^\circ$, let $\nu_R = (\mu(R_L)\mu(R_R)/\mu(R))^{1/2}$, and let $\mathbf{1}_R$ denote the indicator of R. Define

$$h_R = \nu_R \left(\frac{\mathbf{1}_{R_L}}{\mu(R_L)} - \frac{\mathbf{1}_{R_R}}{\mu(R_R)} \right), \qquad \widehat{f}_R = \langle f, h_R \rangle_\mu = \nu_R \big(\langle f \rangle_{R_L} - \langle f \rangle_{R_R} \big). \tag{3}$$

Each h_R has zero weighted mean and unit norm, including when the child masses differ. Together with $\mu(\mathcal{S})^{-1/2}\mathbf{1}_\mathcal{S}$, these functions form an orthonormal basis $\mathcal{H}$, giving $f = \langle f \rangle_\mathcal{S} \mathbf{1}_\mathcal{S} + \sum_{R \in \mathcal{T}^\circ} \widehat{f}_R h_R$. The coefficient $\widehat{f}_R$ records a normalized regional contrast; $\widehat{f}_R h_R$ is its detail component. Means and coefficients apply componentwise to vector features.

Why Haar? Childwise constancy expresses each coefficient through regional mean differences, while zero weighted means yield orthogonality across nested regions. These properties connect directional responses to symbol variation and squared coefficient sums to regional variance. The hierarchy and weights determine the Haar transform; learning enters through the symbols and multipliers below.

Paraproduct and adjoint. A real-valued function b on $\mathcal{S}$, called a *symbol*, supplies the transfer coefficients. The finite-tree paraproduct and its weighted adjoint are (Beznosova, 2008; Treil, 2013)

$$\Pi_b f = \sum_{R \in \mathcal{T}^\circ} \widehat{b}_R \langle f \rangle_R h_R, \qquad \Pi_b^* f = \sum_{R \in \mathcal{T}^\circ} \widehat{b}_R \widehat{f}_R \frac{\mathbf{1}_R}{\mu(R)}. \tag{4}$$

The first sends a regional mean to a Haar detail. The second sends an incoming detail to a response constant on that region. Their normalization gives $\langle \Pi_b f, g \rangle_\mu = \langle f, \Pi_b^* g \rangle_\mu$.

Definition 1 (HPNO kernel action)**.** *For two symbols α, β and real coefficients $\lambda_0, (\lambda_R)_{R \in \mathcal{T}^\circ}$, define*

$$\mathcal{K} = \mathcal{M} + \Pi_\alpha + \Pi_\beta^*, \qquad \mathcal{M} f = \lambda_0 \langle f \rangle_\mathcal{S} \mathbf{1}_\mathcal{S} + \sum_{R \in \mathcal{T}^\circ} \lambda_R \widehat{f}_R h_R. \tag{5}$$

The Haar multiplier $\mathcal{M}$ reweights existing components without mixing them. Here λ_0 weights the root mean, whereas $\lambda_\mathcal{S}$ weights the root detail when present. All three terms act on the same incoming field and are added, rather than applied sequentially.

4.2 Coefficient Generation

Why joint control is needed. Consider free coefficients in $\Pi_\tau f = \sum_{R\in\mathcal{T}^\circ} \tau_R \langle f\rangle_R h_R$. On a complete binary tree with J internal levels and $\mu(\mathcal{S}) = 1$, choose $\tau_R = \eta B_\alpha \sqrt{\mu(R)}$ for $0 < \eta < 1$. Every coefficient satisfies $|\tau_R| \le B_\alpha \sqrt{\mu(R)}$, yet orthogonality gives $\|\Pi_\tau \mathbf{1}_\mathcal{S}\|_\mu = \eta B_\alpha \sqrt{J}$. Individual coefficient bounds therefore permit accumulation with tree depth. We instead generate coefficients from bounded symbols, which control their squared sums within every subtree.

Input-conditioned symbols. Let $v_0 : \mathcal{S} \to \mathbb{R}^C$ be the initial input encoding on the kernel support, and let $z = \langle v_0\rangle_\mathcal{S}$. Two pointwise networks generate

$$\alpha(s) = B_\alpha \tanh\big(F_{\theta_\alpha}(s, v_0(s), z)\big), \qquad \beta(s) = B_\beta \tanh\big(F_{\theta_\beta}(s, v_0(s), z)\big), \tag{6}$$

where s denotes the assigned coordinates and $B_\alpha, B_\beta > 0$ are fixed amplitudes. Applying Eq. (3) gives $\widehat{\alpha}_R, \widehat{\beta}_R$ at every internal region. The directions are learned separately, but all coefficients within each direction derive from one bounded function. The descriptors use v_0, not the evolving block features; a new physical input therefore generates new coefficients even when the hierarchy is reused.

Bounded multipliers. For each region, form $\rho_R = (c_R, s_R, \langle v_0\rangle_R, z)$, where $c_R = \langle s\rangle_R$ is its weighted centroid and s_R contains its bounding-box side lengths. We set $\lambda_R = B_M \tanh(F_{\theta_M}(\rho_R))$ and $\lambda_0 = B_M \tanh(F_{\theta_0}(\rho_\mathcal{S}))$ with fixed $B_M > 0$. Each layer and head has separate coefficient networks. Within a head, symbol networks share parameters across support elements, and the detail-multiplier network shares parameters across regions. Their scalar outputs are shared across that head's feature channels.

4.3 Operator Architecture

Multi-head mixing. The input encoding initializes L operator blocks and supplies their coefficient descriptors. A block processes current features $v : \mathcal{S} \to \mathbb{R}^C$ through H heads of width $C_h = C/H$:

$$\mathcal{A}_\ell(v) = \sum_{h=1}^{H} O_{\ell h} \mathcal{K}_{\ell h}\big(V_{\ell h} v\big), \qquad \ell = 0, \dots, L-1, \tag{7}$$

where $V_{\ell h} : \mathbb{R}^C \to \mathbb{R}^{C_h}$ and $O_{\ell h} : \mathbb{R}^{C_h} \to \mathbb{R}^C$ are learned pointwise linear maps. Each scalar action $\mathcal{K}_{\ell h}$ applies componentwise. The result enters residual updates with normalization and pointwise feed-forward processing, followed by output reconstruction at $\mathcal{X}$.

Task representations. The representation maps and residual processing are adapted from Pei et al. (2026). Haar leaves are physical nodes for Elasticity, patch features for grid and structured-mesh tasks, and slice states for the two CFD tasks. Slice-based blocks aggregate and reconstruct features around the Haar action; their tree regions index slots rather than physical length scales. Additional local, coarse, and output-refinement paths belong to the predictor, not to $\mathcal{K}$, and remain unchanged in the kernel ablations. Their configurations are specified in Appendix E.3.

End-to-end training. All trainable components are optimized jointly from input–solution pairs. For the n_{tr} training units defined by a task, we minimize

$$\mathcal{L}_{\mathrm{train}}(\theta) = \frac{1}{n_{\mathrm{tr}}} \sum_{i=1}^{n_{\mathrm{tr}}} \ell_{\mathrm{task}}(\widehat{u}_i, u_i). \tag{8}$$

The symbols receive no separate supervision. Gradients pass through both coefficient generation and the features on which the kernels act. The hierarchy, prescribed weights, and amplitude bounds remain fixed during differentiation. Task losses and training settings are given in Appendix E.

4.4 Mean–Detail Responses and Bounds

We now fix the conditioning input, network parameters, hierarchy, and weights. Let $\mathcal{C}$ be a tree cut and $\mathbb{E}_\mathcal{C}$ its averaging projection from § 3. All operator norms below are induced by $\|\cdot\|_\mu$.

Proposition 1 (Mean–detail responses)**.** *For any cut of the fixed weighted tree and any scalar field f, the action in Def. 1 satisfies*

$$\begin{aligned}(I-\mathbb{E}_{\mathcal{C}})\mathcal{K}\big(\mathbb{E}_{\mathcal{C}}[f]\big) &= \big(\alpha-\mathbb{E}_{\mathcal{C}}[\alpha]\big)\mathbb{E}_{\mathcal{C}}[f],\\ \mathbb{E}_{\mathcal{C}}\big[\mathcal{K}((I-\mathbb{E}_{\mathcal{C}})f)\big] &= \mathbb{E}_{\mathcal{C}}\big[(\beta-\mathbb{E}_{\mathcal{C}}[\beta])(I-\mathbb{E}_{\mathcal{C}})f\big].\end{aligned} \tag{9}$$

Products are pointwise. The first response modulates regional means by the centered symbol α. The second averages the product of centered β and incoming detail, giving a regional weighted covariance. The multiplier contributes to neither cross-cut response: it is diagonal in the Haar basis and commutes with $\mathbb{E}_{\mathcal{C}}$.

Transfer strength. Let $\sigma_R(b) = \sqrt{\langle b^2\rangle_R - \langle b\rangle_R^2}$ be the regional weighted standard deviation. The transfer norms are exactly

$$\begin{aligned}\|(I-\mathbb{E}_{\mathcal{C}})\mathcal{K}\mathbb{E}_{\mathcal{C}}\|_{\mu\to\mu} &= \max_{R\in\mathcal{C}}\sigma_R(\alpha) \le B_\alpha,\\ \|\mathbb{E}_{\mathcal{C}}\mathcal{K}(I-\mathbb{E}_{\mathcal{C}})\|_{\mu\to\mu} &= \max_{R\in\mathcal{C}}\sigma_R(\beta) \le B_\beta.\end{aligned} \tag{10}$$

Thus regional symbol variation determines both the response in Eq. (9) and its maximum amplification. A symbol constant within every cut region produces zero transfer in its direction. The response identities and norm equalities follow from regional Haar reconstruction; detailed proofs are provided in the appendix.

Separate directional symbols. Tying $\alpha = \beta$ restricts the fixed scalar kernel to pointwise multiplication plus a Haar multiplier. Separate symbols remove this restriction and motivate the tied-symbol ablation. The corresponding binary-tree identity and derivation are given in Appendix B.4.

Joint subtree control. Bounds at one cut do not alone control contributions accumulated across all levels. Regional Haar orthogonality gives, for every $R \in \mathcal{T}$,

$$\sum_{\substack{U\in\mathcal{T}^\circ\\ U\subseteq R}} |\widehat{\alpha}_U|^2 = \mu(R)\sigma_R(\alpha)^2 \le B_\alpha^2\mu(R). \tag{11}$$

The same statement holds for β with bound B_β. This is a Carleson condition on the transfer coefficients. For finite-tree paraproducts, uniform control of these normalized subtree sums is equivalent, up to absolute constants, to a uniform operator bound (Treil, 2013). The bounded-symbol construction enforces this condition without separately normalizing the coefficient array.

Theorem 1 (Bounded kernel action)**.** *On any finite binary tree with prescribed positive weights, suppose $\|\alpha\|_\infty \le B_\alpha$, $\|\beta\|_\infty \le B_\beta$, and $|\lambda_0|, |\lambda_R| \le B_M$ for all $R \in \mathcal{T}^\circ$. Then*

$$\|\mathcal{K}f\|_\mu \le (B_M + 2B_\alpha + 2B_\beta)\|f\|_\mu \qquad \text{for all } f. \tag{12}$$

The constant is independent of N, tree depth, and child-mass ratios.

The multiplier has norm at most B_M. Carleson embedding applied to Eq. (11) bounds the paraproduct norms by $2B_\alpha$ and $2B_\beta$; the adjoint has the same norm as its paraproduct. The triangle inequality yields Eq. (12) without requiring the three branch outputs to be orthogonal. The embedding argument and proof are given in Appendices C.1 and C.2.

The bound extends to componentwise vector actions. With common amplitude bounds across heads, Eq. (7) acquires the factor $\sum_h \|O_{\ell h}\|_2\|V_{\ell h}\|_2$. This controls the conditioned kernel's feature amplification; changes in input-generated coefficients, representation maps, and nonlinear updates require separate estimates (Appendix C.3).

Kernel Evaluation: shared tree passes. An upward pass computes regional means and Haar coefficients of the packed field $F = (f, \alpha, \beta)$. Leaf means equal the supplied values, and each parent mean is the mass-weighted average of its child means. At an internal region, set $\xi_R = \lambda_R\widehat{f}_R + \widehat{\alpha}_R\langle f\rangle_R$ and $\eta_R = \widehat{\beta}_R\widehat{f}_R/\mu(R)$. The regional contribution is $\xi_R h_R + \eta_R \mathbf{1}_R$. A downward pass starts from $r_{\mathcal{S}} = \lambda_0\langle f\rangle_{\mathcal{S}}$ and updates

$$r_{R'} = r_R + (h_R|_{R'})\xi_R + \eta_R, \qquad R' \in \{R_L, R_R\}. \tag{13}$$

Both children receive the same η_R, while their Haar contributions have opposite signs and the prescribed mass normalization. At a leaf, $(\mathcal{K}f)(s_i) = r_{\{s_i\}}$. These passes evaluate Eq. (5) exactly in real arithmetic without assembling a kernel matrix. Backward derivatives are given in Appendix D.4.

Table 1: **Relative ℓ_2 errors on six PDE benchmarks.** Bold and underlined values denote the lowest and second-lowest reported errors in each column, respectively.

Model	Point cloud	Regular grid		Structured mesh		
	Elasticity	**Navier–Stokes**	**Darcy**	**Plasticity**	**Airfoil**	**Pipe**
FNO (**?**)	0.0229	0.1556	0.0108	0.0074	0.0138	0.0067
MWT (Gupta et al., 2021)	0.0359	0.1541	0.0082	0.0076	0.0075	0.0077
WNO (Tripura & Chakraborty, 2023)	0.0465	0.1614	0.0242	-	0.0188	0.0070
RNO (**?**)	0.0078	0.0894	0.0054	0.0012	0.0054	0.0044
GNOT (Hao et al., 2023)	0.0086	0.1380	0.0105	0.0336	0.0076	0.0047
LNO (Wang & Wang, 2024)	0.0052	0.0845	0.0049	0.0029	0.0051	0.0026
Transolver (Wu et al., 2024)	0.0064	0.0900	0.0058	0.0013	0.0053	0.0033
Transolver++ (Luo et al., 2025)	0.0052	0.0719	0.0049	0.0011	0.0049	0.0027
SAOT (Zhou et al., 2026)	0.0080	0.0688	0.0049	<u>0.0008</u>	<u>0.0048</u>	0.0063
LinearNO (Hu et al., 2026)	0.0050	0.0699	0.0050	0.0011	0.0049	<u>0.0024</u>
HNO (Pei et al., 2026)	<u>0.0037</u>	<u>0.0676</u>	<u>0.0045</u>	0.0009	<u>0.0048</u>	0.0027
HPNO (Ours)	**0.0031**	**0.0617**	**0.0031**	**0.0004**	**0.0043**	**0.0022**

Table 2: **CFD benchmark results.** Volume and surface scores use the task-specific metrics in the text. Bold and underlined values denote the best and second-best scores.

Model	Shape-Net Car				AirfRANS			
	Volume↓	Surface↓	$E_{C_D}\downarrow$	$\rho_D\uparrow$	Volume↓	Surface↓	$E_{C_L}\downarrow$	$\rho_L\uparrow$
GEO-FNO (**?**)	0.1670	0.2378	0.0664	0.8280	0.0361	0.0301	0.6161	0.9257
GNOT (Hao et al., 2023)	0.0329	0.0798	0.0178	0.9833	0.0049	0.0152	0.1992	0.9942
LNO (Wang & Wang, 2024)	0.0269	0.0870	0.0174	0.9781	0.0214	0.0268	0.1480	0.9744
Transolver (Wu et al., 2024)	0.0207	**0.0745**	<u>0.0103</u>	<u>0.9935</u>	0.0037	0.0142	0.1030	0.9978
Transolver++ (Luo et al., 2025)	0.0226	0.0800	0.0132	0.9914	0.0068	0.0159	0.1880	0.9910
LinearNO (Hu et al., 2026)	**0.0194**	0.0754	0.0106	0.9925	<u>0.0011</u>	0.0077	**0.0491**	<u>0.9992</u>
HNO (Pei et al., 2026)	0.0214	0.0764	0.0130	0.9903	0.0045	0.0172	0.1178	0.9977
HPNO (ours)	<u>0.0205</u>	<u>0.0753</u>	**0.0102**	**0.9936**	**0.0009**	**0.0025**	<u>0.0624</u>	**0.9993**

5 Experiments

Experimental setup. We evaluate HPNO on six PDE benchmarks (regular grids, structured meshes, point clouds) alongside ShapeNet Car and AirfRANS (**?**Li et al., 2023b; Wu et al., 2024). The PDE tasks use mean per-case relative ℓ_2 error across full sequences, whereas ShapeNet Car and AirfRANS assess volume/surface errors and aerodynamic drag/lift metrics (relative error and Spearman correlation). We compare HPNO with spectral, multiresolution, and attention-based operators (Tables 1 and 2), evaluating HNO (Pei et al., 2026) and LinearNO (Hu et al., 2026) on identical Darcy inputs. We train a separate HPNO model for each task via Eq. (8), and retrain its ablations with the same task-specific setup.

5.1 Physical-Field Benchmark Evaluation

Predictive performance across benchmarks. HPNO achieves the lowest reported error among the compared methods on all six tasks in Table 1, spanning coefficient-to-solution, geometry-to-field, and temporal forecasting tasks across distinct spatial representations. Extending the evaluation to unstructured aerodynamic fields (Table 2), HPNO obtains the lowest drag error and highest drag correlation on ShapeNet Car alongside second-best volume and surface errors, while on AirfRANS it yields the lowest volume and surface pressure scores with LinearNO achieving the lowest lift error. These varying rankings across field and force metrics highlight that superior field reconstruction does not uniformly guarantee the best prediction for every integrated physical quantity. Qualitative comparisons are shown in Fig. 3.

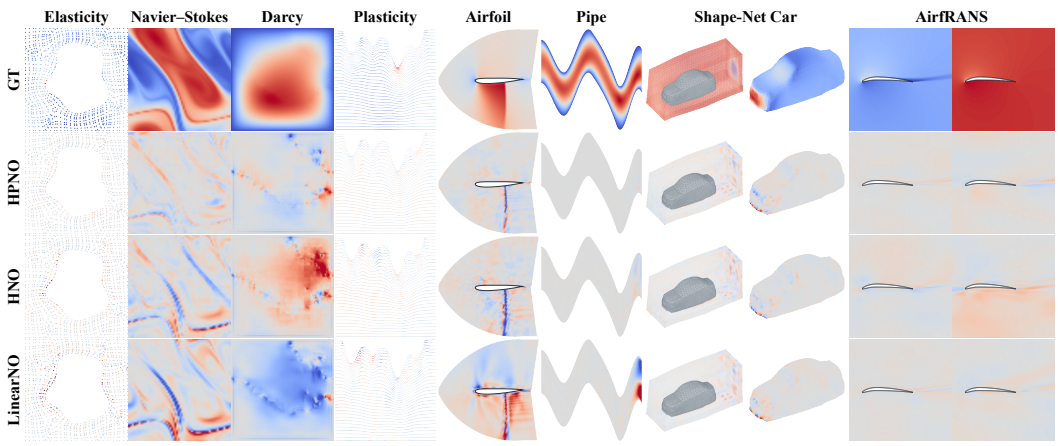

Figure 3: Ground-truth fields and signed prediction errors $\widehat{u} - u$ for HPNO, HNO, and LinearNO across benchmark tasks. ShapeNet Car and AirfRANS show velocity and pressure separately.

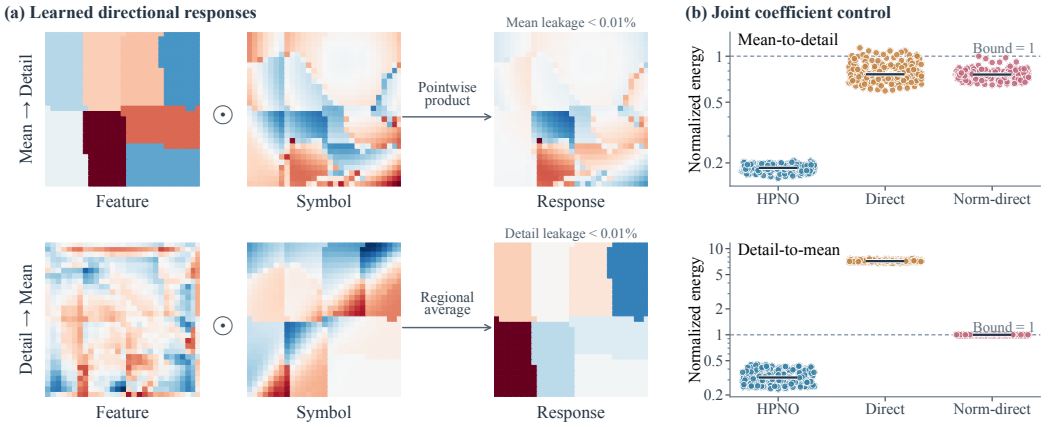

Figure 4: **Kernel responses and joint coefficient control.** (a) Coarse/detail features, centered symbols, and cross-cut responses for a Darcy sample. (b) Case-level maxima of normalized subtree statistics for both directions; points show cases, bars medians, and dashed lines the unit bound.

5.2 Mechanism Analysis and Ablations

Learned responses and joint coefficient control. Fig. 4 examines a Darcy feature at a fixed conditioning input, layer, head, and tree cut. The mean-to-detail response transforms a regionally constant feature into internal variation via centered symbol α. The reverse response averages the pointwise product of the incoming detail and centered β to produce a regional mean. These maps illustrate feature interactions governed by Eq. (9), showing that symbol variation dictates how coarse and detail components couple. Concurrently, the right panel of Fig. 4 evaluates the maximum normalized subtree coefficient sum across regions, layers, and heads for each case and direction against the unit reference in Eq. (11). Both HPNO and Normalized-direct strictly adhere to this bound, whereas Direct exceeds it in the detail-to-mean direction, confirming that pointwise coefficient envelopes fail to control coefficient accumulation across hierarchy levels. This diagnostic highlights the role of joint subtree constraints over isolated amplitude bounds in stabilizing kernel actions; details are in Appendix F.1.

Mean-preserving input changes and regional response. To evaluate whether models capture internal variations indistinguishable by regional input means alone, we rearrange permeability within fixed regions while preserving their weighted means ($P_{\mathcal{X}} a^+ = P_{\mathcal{X}} a^-$) under unchanged domain, forcing, and boundary conditions. Each pair induces a reference change $\Delta u = u^+ - u^-$, and frozen predictors infer $\Delta\widehat{u} = \widehat{u}^+ - \widehat{u}^-$ without response-specific fine-tuning. As visualized in Fig. 5, we

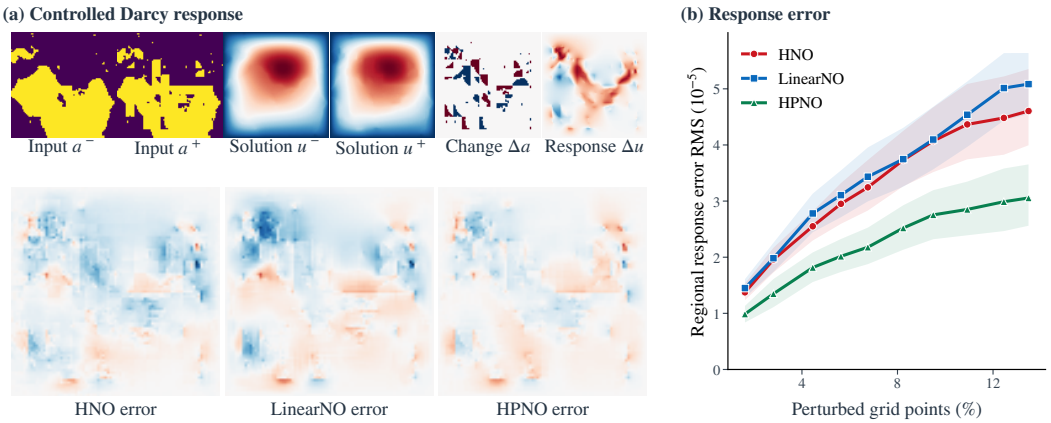

Figure 5: **Paired Darcy responses.** (a) Permeability and reference-solution pairs, their differences, and signed response errors $\Delta\widehat{u} - \Delta u$ for HNO, LinearNO, and HPNO on the original grid. (b) Absolute regional response error against the mean fraction of changed nodes.

quantify performance via the weighted RMS of $P_{\mathcal{X}}(\Delta\widehat{u} - \Delta u)$ across 50 base cases over ten rearrangement levels. HPNO consistently attains lower aggregate error than HNO and LinearNO across all settings; notably, at the most demanding setting (averaging 13.5% altered nodes), HPNO reduces error by 33.6% and 39.9%, respectively. These results confirm HPNO's superior ability to capture regional responses driven by fine-scale spatial rearrangements, complementing our theoretical kernel analysis; full experimental protocols and metrics are detailed in Appendix F.3.

Predictive contribution of architectural designs. Table 3 links these designs to predictive accuracy. The multiplier retains all Haar coordinates; thus, adding either paraproduct alters coordinate interactions rather than supplying missing scales. Adding each direction reduces error on both tasks, with the bidirectional model performing best. Tying $\alpha = \beta$ degrades accuracy, justifying separately parameterized directional pathways. Within direct coefficients, subtree normalization improves performance, yet HPNO outperforms normalized direct prediction. These comparisons demonstrate the benefit of joint subtree control and favor bounded-symbol parameterization over direct coefficient prediction under the same normalized subtree restriction. Implementation details are provided in Appendix F.2.

Table 3: **HPNO ablations.** Relative ℓ_2 errors on Darcy and Elasticity.

Design	Darcy	Elasticity
$\mathcal{M}$	0.00345	0.00364
$\mathcal{M} + \Pi_\alpha$	0.00323	0.00351
$\mathcal{M} + \Pi^*_\beta$	0.00324	0.00347
$\mathcal{M} + \Pi_\alpha + \Pi^*_\beta$	**0.00311**	**0.00314**
$\alpha = \beta$	0.00323	0.00351
c_R^{md}, c_R^{dm}	0.00338	0.00402
$\bar{c}_R^{\mathrm{md}}$, $\bar{c}_R^{\mathrm{dm}}$	0.00327	0.00348

6 Conclusion

We introduced HPNO to learn bidirectional mean–detail interactions with joint control of hierarchical transfer coefficients. The analysis connects directional response strength to regional symbol variation and establishes a weighted L^2 bound for each fixed-input kernel action, independent of hierarchy size and depth. The benchmark results show predictive improvements across different spatial representations. Ablations favor bidirectional coupling over Haar reweighting alone and bounded-symbol generation over direct coefficient parameterizations. Paired Darcy tests further demonstrate more accurate regional responses to mean-preserving permeability rearrangements. Together, these results connect the learning of multiscale interactions with explicit control of their accumulated strength. Extending the analysis to complete predictors under changes in conditioning inputs and hierarchies remains a direction for future work.

AI USE STATEMENT

Generative AI tools were used solely for grammar correction and sentence refinement during manuscript preparation. No generative AI was used for research ideation, theoretical proofs, algorithm development, or data generation. All content was verified by the authors, who take full responsibility for the final manuscript.

ETHICS STATEMENT

This work studies fundamental neural operator architectures evaluated on standard, publicly available scientific benchmarks. It involves no human subjects, sensitive personal data, or direct applications with adverse societal or ethical consequences.

REPRODUCIBILITY STATEMENT

To ensure reproducibility, our complete codebase, training pipelines, and pretrained checkpoints are anonymously available at https://anonymous.4open.science/r/HPNO-57478. Complete theoretical proofs, dataset configurations, network architectures, and implementation parameters are fully documented in the main text and appendix.

REFERENCES

Oleksandra V Beznosova. Linear bound for the dyadic paraproduct on weighted lebesgue space l2 (w). *Journal of Functional Analysis*, 255(4):994–1007, 2008.

Jean-Michel Bony. Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires. In *Annales scientifiques de l'École normale supérieure*, volume 14, pp. 209–246, 1981.

Yuwei Fan, Cindy Orozco Bohorquez, and Lexing Ying. Bcr-net: A neural network based on the nonstandard wavelet form. *Journal of Computational Physics*, 384:1–15, 2019.

Oluwadamilola Fasina. Hierarchical paraproducts. *arXiv preprint arXiv:2602.16644*, 2026.

Oluwadamilola Fasina and Ronald R Coifman. Regularization and compression of integral operator kernels with tensor paraproducts. *arXiv preprint arXiv:2609.15822*, 2026.

Gaurav Gupta, Xiongye Xiao, and Paul Bogdan. Multiwavelet-based operator learning for differential equations. *Advances in neural information processing systems*, 34:24048–24062, 2021.

Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. GNOT: A general neural operator transformer for operator learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 12556–12569. PMLR, 2023.

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 *The Thirteenth International Conference on Learning Representations*, 2025.

Wenjie Hu, Sidun Liu, Peng Qiao, Zhenglun Sun, and Yong Dou. Transolver is a linear transformer: Revisiting physics-attention through the lens of linear attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 408–416, 2026.

Michael G Kapteyn, Jacob VR Pretorius, and Karen E Willcox. A probabilistic graphical model foundation for enabling predictive digital twins at scale. *Nature Computational Science*, 1(5): 337–347, 2021.

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.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023a.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023b.

Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, et al. Geometry-informed neural operator for large-scale 3d pdes. *Advances in Neural Information Processing Systems*, 36:35836–35854, 2023c.

Huakun Luo, Haixu Wu, Hang Zhou, Lanxiang Xing, Yichen Di, Jianmin Wang, and Mingsheng Long. Transolver++: An accurate neural solver for pdes on million-scale geometries. In *International Conference on Machine Learning*, 2025.

Jieyuan Pei, Zhuoxuan Li, Wei Li, Haobo Zhang, jiawei jiang, and Jianwei Zheng. Hyperbolic neural operator. In *International Conference on Machine Learning*, 2026.

Stephen B. Pope. *Turbulent Flows*. Cambridge University Press, 2000.

Sergei Treil. Commutators, paraproducts and bmo in non-homogeneous martingale settings. *Revista Matemática Iberoamericana*, 29(4):1325–1372, 2013.

Tapas Tripura and Souvik Chakraborty. Wavelet neural operator for solving parametric partial differential equations in computational mechanics problems. *Computer Methods in Applied Mechanics and Engineering*, 404:115783, 2023. ISSN 0045-7825.

Tian Wang and Chuang Wang. Latent neural operator for solving forward and inverse pde problems. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

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

Chenhong Zhou, Jie Chen, and Zaifeng Yang. Saot: An enhanced locality-aware spectral transformer for solving pdes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 28928–28936, 2026.

Appendices

A Background on Haar Representations and Paraproducts **13**
A.1 Averages and Haar Details . . . 13
A.2 Averages at Different Resolutions . . . 13
A.3 From Multiscale Products to Paraproducts . . . 14
A.4 Paraproducts as Kernel Actions . . . 14
A.5 Why Coefficients Must Be Controlled Together . . . 14

B Weighted Haar Basis and Mean–Detail Responses **15**
B.1 Basis and Reconstruction . . . 15
B.2 Adjoint, Constants, and Mean Responses . . . 16
B.3 Responses Across an Arbitrary Tree Cut . . . 17
B.4 Tied Symbols and Pointwise Multiplication . . . 18

C Kernel-Action Bounds **19**
C.1 Finite-Tree Carleson Embedding . . . 19
C.2 Proof of the Kernel-Action Bound . . . 20
C.3 Channel Projections and Input Changes . . . 22

D Coefficient Generation and Discrete Computation **22**
D.1 Coefficient Networks and Tensor Interfaces . . . 22
D.2 Explicit Kernel Representation . . . 23
D.3 Two-Pass Evaluation . . . 23
D.4 Backward Computation . . . 24
D.5 Complexity, Preprocessing, and Reuse . . . 25
D.6 Algebraic Consistency Conditions . . . 26

E Experimental Details **26**
E.1 Evaluation Metrics . . . 26
E.2 Baselines . . . 27
E.3 Training Protocol . . . 27

F Additional Analyses and Visualizations **28**
F.1 Kernel Diagnostics . . . 28
F.2 Ablation Definitions . . . 29
F.3 Paired Darcy Protocol and Results . . . 30
F.4 Additional Field Error Maps . . . 31

A Background on Haar Representations and Paraproducts

We use the notation of §§ 3 and 4. All mathematical statements concern finite binary trees with positive leaf weights and real-valued fields. Unless stated otherwise, the conditioning input, network parameters, support, and weights are fixed. The proofs use the same measure in regional averages, Haar coefficients, and operator norms.

Table A.1: Notation used in the appendix.

Notation	Meaning
$a, u, \widehat{u}$	Physical input, reference solution, and prediction.
$\mathcal{X}, w, n_{\mathcal{X}}$	Physical observation nodes, their weights, and $n_{\mathcal{X}} = \lvert\mathcal{X}\rvert$.
$\mathcal{S}, \omega, N$	Kernel feature support, positive weights, and $N = \lvert\mathcal{S}\rvert$.
f, v_0	Incoming kernel feature and initial input encoding.
$\mathcal{T}, \mathcal{T}^\circ$	Binary hierarchy and its internal regions.
$\mu(R), \langle f\rangle_R$	Regional mass and weighted mean.
$h_R, \widehat{f}_R$	Normalized Haar function and coefficient $\langle f, h_R\rangle_\mu$.
α, β, λ	Two spatial symbols and Haar multiplier coefficients.
$\mathcal{M}, \mathcal{K}$	Haar multiplier and complete conditioned kernel action.
$\mathcal{C}, \mathbb{E}_{\mathcal{C}}$	Tree cut and its weighted averaging projection.
$P_{\mathcal{X}}$	Averaging projection on the physical evaluation partition.
B_α, B_β, B_M	Fixed symbol and multiplier amplitude bounds.

A.1 Averages and Haar Details

An average and a difference. For equally weighted values f_1, f_2, set $m = (f_1 + f_2)/2$ and $\delta = (f_1 - f_2)/2$. Then $f_1 = m + \delta$ and $f_2 = m - \delta$. The average records their common level, while the difference records their contrast. Repeating this operation on child-region means gives a hierarchical representation. For an equal-length dyadic split, the usual normalized Haar function is $h_R = \lvert R\rvert^{-1/2}(\mathbf{1}_{R_L} - \mathbf{1}_{R_R})$ (Beznosova, 2008).

Unequal masses. The weighted construction in Eq. (3) uses $\nu_R = (\mu(R_L)\mu(R_R)/\mu(R))^{1/2}$. The parent mean is

$$\langle f\rangle_R = \frac{\mu(R_L)}{\mu(R)}\langle f\rangle_{R_L} + \frac{\mu(R_R)}{\mu(R)}\langle f\rangle_{R_R}. \tag{A.1}$$

The mass factors in h_R ensure zero weighted mean even when $\mu(R_L) \neq \mu(R_R)$. For example, weights $(1/4, 3/4)$ and values $(4, 0)$ give mean 1, Haar function $(\sqrt{3}, -1/\sqrt{3})$, and coefficient $\sqrt{3}$. Their reconstruction is $(4, 0) = (1, 1) + \sqrt{3}(\sqrt{3}, -1/\sqrt{3})$. Using $(1, -1)$ as the contrast would not give zero mean under these weights.

A.2 Averages at Different Resolutions

Nested cuts. Let $\mathcal{C}_0 = \{\mathcal{S}\}$. Obtain $\mathcal{C}_{j+1}$ by replacing each internal region of $\mathcal{C}_j$ by its two children and retaining singleton leaves. For a finite tree, some J satisfies $\mathcal{C}_J = \{\{s_i\} : 1 \le i \le N\}$. Define

$$\mathbb{E}_j[f] = \sum_{R\in\mathcal{C}_j} \langle f\rangle_R \mathbf{1}_R, \qquad \Delta_j f = \mathbb{E}_{j+1}[f] - \mathbb{E}_j[f]. \tag{A.2}$$

These are finite conditional expectations and martingale differences (Treil, 2013). Multiplying μ by a common positive constant does not change these averaging operators; in particular, one may normalize it to a probability measure when interpreting the expectation notation.

Reconstruction. For a single split, substitution of the definitions gives

$$\langle f\rangle_{R_L}\mathbf{1}_{R_L} + \langle f\rangle_{R_R}\mathbf{1}_{R_R} - \langle f\rangle_R\mathbf{1}_R = \widehat{f}_R h_R. \tag{A.3}$$

For example, on R_L the left side is $\mu(R_R)(\langle f\rangle_{R_L} - \langle f\rangle_{R_R})/\mu(R)$; the right side has the same value because $\nu_R^2/\mu(R_L) = \mu(R_R)/\mu(R)$. The calculation on R_R is analogous. Since $\mathbb{E}_J[f] = f$,

telescoping yields

$$f = \mathbb{E}_0[f] + \sum_{j=0}^{J-1} \Delta_j f = \langle f \rangle_{\mathcal{S}} \mathbf{1}_{\mathcal{S}} + \sum_{R \in \mathcal{T}^\circ} \widehat{f}_R h_R. \tag{A.4}$$

All details are retained. Equal means at one split do not imply equal values inside its children: splitting $(1, 3, 2, 2)$ into $(1, 3)$ and $(2, 2)$ has zero parent contrast but a nonzero contrast in the first child. Tree regions represent physical extents only when their coordinates have that interpretation; Haar levels are not disjoint Fourier bands.

A.3 From Multiscale Products to Paraproducts

Separating interactions. Paraproduct calculus separates products and nonlinear compositions by the scales of their components (Bony, 1981). Using $\mathbb{E}_{j+1} = \mathbb{E}_j + \Delta_j$, pointwise multiplication gives

$$\begin{aligned}(\mathbb{E}_{j+1}[b])(\mathbb{E}_{j+1}[f]) - (\mathbb{E}_j[b])(\mathbb{E}_j[f]) \\ = (\mathbb{E}_j[f])(\Delta_j b) + (\mathbb{E}_j[b])(\Delta_j f) + (\Delta_j b)(\Delta_j f).\end{aligned} \tag{A.5}$$

Summing over levels and using the singleton cut gives

$$\begin{aligned}bf = \langle b \rangle_{\mathcal{S}} \langle f \rangle_{\mathcal{S}} \mathbf{1}_{\mathcal{S}} + \sum_{j=0}^{J-1} (\mathbb{E}_j[f])(\Delta_j b) \\ + \sum_{j=0}^{J-1} (\mathbb{E}_j[b])(\Delta_j f) + \sum_{j=0}^{J-1} (\Delta_j b)(\Delta_j f).\end{aligned} \tag{A.6}$$

All sums are finite, so no smoothness or limiting assumption is needed.

Fixing a symbol. Within $R \in \mathcal{C}_j$, the first factor in $(\mathbb{E}_j[f])(\Delta_j b)$ equals $\langle f \rangle_R$, while the contribution of that split to $\Delta_j b$ is $\widehat{b}_R h_R$. Thus the first sum in Eq. (A.6) is $\Pi_b f$. For fixed b, it is linear in f. Adding a constant to b leaves this action unchanged. The second sum is $\Pi_f b$, obtained by exchanging the fields, not by taking the adjoint.

The same-level term and the adjoint. Because regions at the same cut are disjoint, the last sum is $\sum_{R \in \mathcal{T}^\circ} \widehat{b}_R \widehat{f}_R h_R^2$. When child masses are equal, $h_R^2 = \mathbf{1}_R / \mu(R)$. For unequal masses, h_R^2 generally differs between the children, although its regional mean is $1/\mu(R)$. Consequently, this same-level product term is not in general $\Pi_b^* f$. The adjoint in Eq. (4) is instead identified from the weighted inner product in Appendix B.2. For two leaves with weights $(1/2, 1/2)$ and $b = (1, -1)$, $\Pi_b(1, 1) = (1, -1)$ and $\Pi_b^*(1, -1) = (1, 1)$, whereas $\Pi_b^*(1, 1) = 0$.

A.4 Paraproducts as Kernel Actions

Substituting the regional mean into the paraproduct gives

$$(\Pi_b f)(s) = \sum_{i=1}^{N} \omega_i \left[\sum_{R \in \mathcal{T}^\circ} \frac{\widehat{b}_R}{\mu(R)} h_R(s) \mathbf{1}_R(s_i) \right] f(s_i). \tag{A.7}$$

Each regional term separates the output and input arguments. The adjoint exchanges these factors. The resulting kernel can be dense even though its action admits a shared tree evaluation. Appendix D.2 gives the complete HPNO kernel. HPNO parameterizes this family from input–solution pairs; it does not assert that every PDE kernel has this form.

A.5 Why Coefficients Must Be Controlled Together

Accumulation across levels. Consider a complete binary tree with J internal levels and $\mu(\mathcal{S}) = 1$. For $0 < \eta < 1$, choose $c_R = \eta B \sqrt{\mu(R)}$ and define $\Pi_c f = \sum_R c_R \langle f \rangle_R h_R$. Although every

coefficient satisfies $|c_R| \le B\sqrt{\mu(R)}$, orthogonality gives

$$\begin{aligned}\|\Pi_c \mathbf{1}_{\mathcal{S}}\|_\mu^2 &= \sum_{R\in\mathcal{T}^\circ} |c_R|^2 \\ &= \eta^2 B^2 \sum_{j=0}^{J-1} \sum_{R\in\mathcal{C}_j} \mu(R) = \eta^2 B^2 J.\end{aligned} \tag{A.8}$$

Each level partitions total mass one, and $\|\mathbf{1}_{\mathcal{S}}\|_\mu = 1$. The response norm therefore grows as $\sqrt{J}$ despite the individual coefficient bounds.

The joint quantity. For a coefficient array c, define

$$\mathfrak{C}(c) = \max_{R\in\mathcal{T}} \frac{1}{\mu(R)} \sum_{\substack{U\in\mathcal{T}^\circ \\ U\subseteq R}} |c_U|^2. \tag{A.9}$$

Empty sums are zero. The finite embedding argument in Appendix C.1 proves

$$\mathfrak{C}(c) \le \|\Pi_c\|_{\mu\to\mu}^2 \le 4\mathfrak{C}(c). \tag{A.10}$$

For symbol coefficients, regional Parseval yields

$$\begin{aligned}\sum_{\substack{U\in\mathcal{T}^\circ \\ U\subseteq R}} |\widehat{b}_U|^2 &= \sum_{s_i\in R} \omega_i |b(s_i) - \langle b\rangle_R|^2 \\ &= \mu(R)\sigma_R(b)^2 \le \|b\|_\infty^2 \mu(R).\end{aligned} \tag{A.11}$$

Thus $\|b\|_\infty \le B$ implies $\mathfrak{C}(\widehat{b}) \le B^2$. The converse need not impose the same pointwise bound.

The measure in the estimate. Here the weights define the Haar construction itself. This differs from retaining an unweighted dyadic basis and studying its action in a separately weighted space $L^2(w)$, as in Beznosova (2008). The common-measure, finite-tree argument below uses only positive masses; no additional Muckenhoupt condition is imposed.

B Weighted Haar Basis and Mean–Detail Responses

B.1 Basis and Reconstruction

Step 1: zero mean and unit norm. The two children partition R, so $\mu(R) = \mu(R_L) + \mu(R_R)$. Using Eq. (3),

$$\begin{aligned}\langle h_R, \mathbf{1}_{\mathcal{S}}\rangle_\mu &= \nu_R \left(\frac{\mu(R_L)}{\mu(R_L)} - \frac{\mu(R_R)}{\mu(R_R)}\right) = 0, \\ \|h_R\|_\mu^2 &= \nu_R^2 \left(\frac{1}{\mu(R_L)} + \frac{1}{\mu(R_R)}\right) \\ &= \frac{\mu(R_L)\mu(R_R)}{\mu(R)} \frac{\mu(R_L)+\mu(R_R)}{\mu(R_L)\mu(R_R)} = 1.\end{aligned} \tag{B.1}$$

Substitution into the inner product also gives

$$\begin{aligned}\langle f, h_R\rangle_\mu &= \nu_R \left(\frac{\sum_{s_i\in R_L} \omega_i f(s_i)}{\mu(R_L)} - \frac{\sum_{s_i\in R_R} \omega_i f(s_i)}{\mu(R_R)}\right) \\ &= \nu_R(\langle f\rangle_{R_L} - \langle f\rangle_{R_R}) = \widehat{f}_R.\end{aligned} \tag{B.2}$$

Step 2: orthogonality. Two distinct tree regions are disjoint or nested. Disjoint regions give disjoint supports. If $U \subsetneq R$, then U lies in one child of R, and h_R is constant on U. Hence

$$\langle h_R, h_U\rangle_\mu = (h_R|_U) \sum_{s_i\in U} \omega_i h_U(s_i) = 0. \tag{B.3}$$

The reverse inclusion follows by symmetry. All Haar functions are also orthogonal to the constant function. Thus overlapping supports across levels do not prevent orthogonality; it follows from childwise constancy and cancellation.

Step 3: completeness and Parseval. If the tree has n_{int} internal nodes, counting its edges gives $2n_{\text{int}} = n_{\text{int}} + N - 1$. Therefore $n_{\text{int}} = N - 1$. Adding $h_0 = \mu(\mathcal{S})^{-1/2}\mathbf{1}_{\mathcal{S}}$ gives N orthonormal vectors in the N-dimensional field space. They form a basis, and $\langle f, h_0\rangle_\mu = \sqrt{\mu(\mathcal{S})}\langle f\rangle_{\mathcal{S}}$. Consequently,

$$\begin{aligned} f &= \langle f\rangle_{\mathcal{S}}\mathbf{1}_{\mathcal{S}} + \sum_{R\in\mathcal{T}^\circ} \widehat{f}_R h_R, \\ \|f\|_\mu^2 &= \mu(\mathcal{S})|\langle f\rangle_{\mathcal{S}}|^2 + \sum_{R\in\mathcal{T}^\circ} |\widehat{f}_R|^2. \end{aligned} \tag{B.4}$$

For $N = 1$, the internal-node sum is empty and the constant basis alone gives the same statements.

Step 4: regional reconstruction. Fix $R \in \mathcal{T}$ and write $f^{[R]} = \mathbf{1}_R(f - \langle f\rangle_R)$. For $U \subseteq R$, its Haar coefficient equals $\widehat{f}_U$ because h_U has zero mean. Coefficients on disjoint regions vanish by support. On strict ancestors, the Haar function is constant on R, so the coefficient vanishes because $f^{[R]}$ integrates to zero. The root-constant coefficient also vanishes. Applying Eq. (B.4) therefore yields

$$\begin{aligned} \mathbf{1}_R(f - \langle f\rangle_R) &= \sum_{\substack{U\in\mathcal{T}^\circ \\ U\subseteq R}} \widehat{f}_U h_U, \\ \sum_{s_i\in R} \omega_i |f(s_i) - \langle f\rangle_R|^2 &= \sum_{\substack{U\in\mathcal{T}^\circ \\ U\subseteq R}} |\widehat{f}_U|^2. \end{aligned} \tag{B.5}$$

Expanding two such centered fields in the same basis gives

$$\sum_{s_i\in R} \omega_i (b(s_i) - \langle b\rangle_R)(f(s_i) - \langle f\rangle_R) = \sum_{\substack{U\in\mathcal{T}^\circ \\ U\subseteq R}} \widehat{b}_U \widehat{f}_U. \tag{B.6}$$

Singleton regions have zero variance and empty sums. For vector features, these identities hold componentwise, and the squared-norm identity follows by summing over channels.

B.2 Adjoint, Constants, and Mean Responses

Weighted adjoint. Since $\langle f\rangle_R = \langle f, \mathbf{1}_R/\mu(R)\rangle_\mu$, finite summation gives

$$\begin{aligned} \langle \Pi_b f, g\rangle_\mu &= \sum_{R\in\mathcal{T}^\circ} \widehat{b}_R \langle f\rangle_R \langle h_R, g\rangle_\mu \\ &= \sum_{R\in\mathcal{T}^\circ} \widehat{b}_R \widehat{g}_R \left\langle f, \frac{\mathbf{1}_R}{\mu(R)} \right\rangle_\mu \\ &= \left\langle f, \sum_{R\in\mathcal{T}^\circ} \widehat{b}_R \widehat{g}_R \frac{\mathbf{1}_R}{\mu(R)} \right\rangle_\mu = \langle f, \Pi_b^* g\rangle_\mu. \end{aligned} \tag{B.7}$$

This proves the adjoint formula in Eq. (4).

Constant probes. For $f = \mathbf{1}_{\mathcal{S}}$, every regional mean is one and every Haar coefficient is zero. Using reconstruction and the real diagonal multiplier gives

$$\begin{aligned} \Pi_b \mathbf{1}_{\mathcal{S}} &= b - \langle b\rangle_{\mathcal{S}}\mathbf{1}_{\mathcal{S}}, & \Pi_b^* \mathbf{1}_{\mathcal{S}} &= 0, \\ \mathcal{K}\mathbf{1}_{\mathcal{S}} &= \lambda_0 \mathbf{1}_{\mathcal{S}} + \alpha - \langle \alpha\rangle_{\mathcal{S}}\mathbf{1}_{\mathcal{S}}, & & \\ \mathcal{K}^*\mathbf{1}_{\mathcal{S}} &= \lambda_0 \mathbf{1}_{\mathcal{S}} + \beta - \langle \beta\rangle_{\mathcal{S}}\mathbf{1}_{\mathcal{S}}. & & \end{aligned} \tag{B.8}$$

These are probes of the incoming feature at fixed coefficients. A constant physical input need not generate constant symbols, because coefficient networks also receive coordinates.

The global mean response. Each output Haar function integrates to zero, so $\langle \Pi_b f \rangle_{\mathcal{S}} = 0$. For the adjoint, integrating each normalized indicator gives

$$\begin{aligned} \mu(\mathcal{S})\langle \Pi_b^* f \rangle_{\mathcal{S}} &= \sum_{R \in \mathcal{T}^\circ} \widehat{b}_R \widehat{f}_R \\ &= \sum_i \omega_i (b(s_i) - \langle b \rangle_{\mathcal{S}})(f(s_i) - \langle f \rangle_{\mathcal{S}}). \end{aligned} \tag{B.9}$$

The second equality is Eq. (B.6) at the root. Therefore

$$\langle \mathcal{K} f \rangle_{\mathcal{S}} = \lambda_0 \langle f \rangle_{\mathcal{S}} + \frac{1}{\mu(\mathcal{S})} \sum_{R \in \mathcal{T}^\circ} \widehat{\beta}_R \widehat{f}_R. \tag{B.10}$$

The covariance describes a feature response, not physical mass or energy conservation.

B.3 Responses Across an Arbitrary Tree Cut

We prove Prop. 1 and Eq. (10) for a cut $\mathcal{C}$ of the fixed weighted tree. Its regions may have different depths.

Step 1: the averaging projection. By definition, $\mathbb{E}_{\mathcal{C}}[f] = \sum_{R \in \mathcal{C}} \langle f \rangle_R \mathbf{1}_R$. Averaging a regionally constant field leaves it unchanged, so $\mathbb{E}_{\mathcal{C}}^2 = \mathbb{E}_{\mathcal{C}}$. For any g constant on the cut regions,

$$\langle f - \mathbb{E}_{\mathcal{C}}[f], g \rangle_\mu = \sum_{R \in \mathcal{C}} g|_R \sum_{s_i \in R} \omega_i (f(s_i) - \langle f \rangle_R) = 0. \tag{B.11}$$

Thus $\mathbb{E}_{\mathcal{C}}$ is the orthogonal projection onto regionally constant fields. Its complement has zero weighted mean in each cut region.

Step 2: Haar coordinates relative to the cut. Each internal region U either lies within one cut region or strictly above the cut. In the second case its children are unions of cut regions. Consequently,

$$\mathbb{E}_{\mathcal{C}}[h_U] = \begin{cases} 0, & U \subseteq R \text{ for some } R \in \mathcal{C}, \\ h_U, & U \text{ lies strictly above the cut.} \end{cases} \tag{B.12}$$

The first case uses the zero integral of h_U; the second uses its constancy on each cut region. Both $\mathcal{M}$ and $\mathbb{E}_{\mathcal{C}}$ are diagonal in these coordinates. Hence they commute, and both cross-cut blocks of $\mathcal{M}$ vanish.

Step 3: mean-to-detail response. If $U \subseteq R \in \mathcal{C}$, then $\langle \mathbb{E}_{\mathcal{C}}[f] \rangle_U = \langle f \rangle_R$. The complement projection removes output Haar functions above the cut. By regional reconstruction,

$$\begin{aligned} &(I - \mathbb{E}_{\mathcal{C}}) \Pi_\alpha (\mathbb{E}_{\mathcal{C}}[f]) \\ &\quad = \sum_{R \in \mathcal{C}} \langle f \rangle_R \sum_{\substack{U \in \mathcal{T}^\circ \\ U \subseteq R}} \widehat{\alpha}_U h_U \\ &\quad = \sum_{R \in \mathcal{C}} \langle f \rangle_R \mathbf{1}_R (\alpha - \langle \alpha \rangle_R) \\ &\quad = (\alpha - \mathbb{E}_{\mathcal{C}}[\alpha]) \mathbb{E}_{\mathcal{C}}[f]. \end{aligned} \tag{B.13}$$

For $\Pi_\beta^*(\mathbb{E}_{\mathcal{C}}[f])$, the incoming Haar coefficients below the cut vanish. Indicators associated with the remaining regions are constant on each cut region, so the complement projection removes their outputs as well. Combining these facts with Step 2 proves the first response identity.

Step 4: detail-to-mean response. Set $g = (I - \mathbb{E}_{\mathcal{C}}) f$. Its means and Haar coefficients above the cut vanish. Below the cut, $\mathbb{E}_{\mathcal{C}}$ removes the Haar outputs of $\Pi_\alpha g$. For $U \subseteq R \in \mathcal{C}$, $\mathbb{E}_{\mathcal{C}}[\mathbf{1}_U / \mu(U)] =$

$\mathbf{1}_R/\mu(R)$. Therefore

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[\Pi_\beta^* g] &= \sum_{R\in\mathcal{C}} \frac{\mathbf{1}_R}{\mu(R)} \sum_{\substack{U\in\mathcal{T}^\circ \\ U\subseteq R}} \widehat{\beta}_U \widehat{g}_U \\ &= \sum_{R\in\mathcal{C}} \frac{\mathbf{1}_R}{\mu(R)} \sum_{s_i\in R} \omega_i(\beta(s_i) - \langle\beta\rangle_R) g(s_i) \\ &= \mathbb{E}_{\mathcal{C}}[(\beta - \mathbb{E}_{\mathcal{C}}[\beta])(I - \mathbb{E}_{\mathcal{C}})f].\end{aligned} \tag{B.14}$$

The second equality uses Eq. (B.6) and $\langle g\rangle_R = 0$. The multiplier has no cross-cut contribution by Step 2. This proves the second identity in Prop. 1.

Step 5: exact transfer norms. For a symbol b, the forward block has the form $(b - \mathbb{E}_{\mathcal{C}}[b])\mathbb{E}_{\mathcal{C}}$. Disjointness of cut regions gives

$$\begin{aligned}\|(b - \mathbb{E}_{\mathcal{C}}[b])\mathbb{E}_{\mathcal{C}}[f]\|_\mu^2 &= \sum_{R\in\mathcal{C}} \mu(R)\sigma_R(b)^2 |\langle f\rangle_R|^2 \\ &\le \left(\max_{R\in\mathcal{C}} \sigma_R(b)^2\right) \|\mathbb{E}_{\mathcal{C}}[f]\|_\mu^2 \\ &\le \left(\max_{R\in\mathcal{C}} \sigma_R(b)^2\right) \|f\|_\mu^2.\end{aligned} \tag{B.15}$$

The last inequality is contractivity of an orthogonal projection. Choose a maximizing region R_* and the unit input $f = \mathbf{1}_{R_*}/\sqrt{\mu(R_*)}$. The first line then equals $\sigma_{R_*}(b)^2$, proving equality in the operator norm. Since $[(I - \mathbb{E}_{\mathcal{C}})\Pi_b\mathbb{E}_{\mathcal{C}}]^* = \mathbb{E}_{\mathcal{C}}\Pi_b^*(I - \mathbb{E}_{\mathcal{C}})$, the reverse block for the same symbol has the same norm. Applying these facts to α and β gives

$$\begin{aligned}\|(I - \mathbb{E}_{\mathcal{C}})\mathcal{K}\mathbb{E}_{\mathcal{C}}\|_{\mu\to\mu} &= \max_{R\in\mathcal{C}} \sigma_R(\alpha) \le B_\alpha, \\ \|\mathbb{E}_{\mathcal{C}}\mathcal{K}(I - \mathbb{E}_{\mathcal{C}})\|_{\mu\to\mu} &= \max_{R\in\mathcal{C}} \sigma_R(\beta) \le B_\beta.\end{aligned} \tag{B.16}$$

Here $\sigma_R(b)^2 = \langle b^2\rangle_R - \langle b\rangle_R^2 \le \|b\|_\infty^2$. When every cut region is a singleton, both norms are zero. For componentwise vector actions, the upper bounds follow by summing squared scalar bounds, and equality is attained in one channel.

Omitting incoming details. Linearity at fixed coefficients and Eq. (B.14) imply

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[\mathcal{K}f] - \mathbb{E}_{\mathcal{C}}[\mathcal{K}(\mathbb{E}_{\mathcal{C}}[f])] &= \mathbb{E}_{\mathcal{C}}[(\beta - \mathbb{E}_{\mathcal{C}}[\beta])(I - \mathbb{E}_{\mathcal{C}})f], \\ \|\mathbb{E}_{\mathcal{C}}[\mathcal{K}f] - \mathbb{E}_{\mathcal{C}}[\mathcal{K}(\mathbb{E}_{\mathcal{C}}[f])]\|_\mu &\le \max_{R\in\mathcal{C}} \sigma_R(\beta) \|(I - \mathbb{E}_{\mathcal{C}})f\|_\mu.\end{aligned} \tag{B.17}$$

Structure versus parameterization. Any finite coefficient array c defines $b_c = \sum_{R\in\mathcal{T}^\circ} c_R h_R$, with $\widehat{(b_c)}_R = c_R$ by orthogonality. The cut identities therefore also apply to direct-coefficient paraproducts through their reconstructed symbols. HPNO additionally constrains the symbol's pointwise amplitude; the existence of a Haar expansion alone does not impose this constraint.

B.4 Tied Symbols and Pointwise Multiplication

Let $M_b f = bf$ denote pointwise multiplication, distinct from the Haar multiplier $\mathcal{M}$. Its diagonal part in the orthonormal Haar basis is

$$\mathrm{Diag}_{\mathcal{H}}(M_b)f = \sum_{h\in\mathcal{H}} \langle M_b h, h\rangle_\mu \langle f, h\rangle_\mu h. \tag{B.18}$$

We verify

$$\Pi_b + \Pi_b^* = M_b - \mathrm{Diag}_{\mathcal{H}}(M_b) \tag{B.19}$$

by comparing matrix entries in this basis.

Step 1: diagonal entries. In $\langle \Pi_b h_R, h_R \rangle_\mu$, orthogonality selects only $\widehat{b}_R \langle h_R \rangle_R = 0$. The constant diagonal entry is also zero, and the same holds for Π_b^* by adjointness. For multiplication,

$$\begin{aligned} \langle M_b h_0, h_0 \rangle_\mu &= \langle b \rangle_{\mathcal{S}}, \\ \langle M_b h_R, h_R \rangle_\mu &= \frac{\mu(R_R)}{\mu(R)} \langle b \rangle_{R_L} + \frac{\mu(R_L)}{\mu(R)} \langle b \rangle_{R_R}. \end{aligned} \tag{B.20}$$

The second line follows by integrating $b h_R^2$ on the two children. Subtracting the diagonal operator makes the right side of Eq. (B.19) zero on all diagonal entries.

Step 2: off-diagonal entries. Disjoint regions give zero entries for all operators. For $U \subsetneq R$,

$$\langle M_b h_R, h_U \rangle_\mu = (h_R|_U) \widehat{b}_U = \langle \Pi_b h_R, h_U \rangle_\mu. \tag{B.21}$$

In this orientation, the corresponding Π_b^* entry is zero: $\Pi_b h_U$ has no coefficient on the ancestor R. The reverse orientation is supplied by the adjoint. Between the constant basis function and h_R, multiplication has entry $\widehat{b}_R / \sqrt{\mu(\mathcal{S})}$, which is again the appropriate paraproduct entry. These cases exhaust the basis, proving Eq. (B.19).

Consequence for the tied action. When $\alpha = \beta = b$,

$$\mathcal{K} = M_b + [\mathcal{M} - \mathrm{Diag}_{\mathcal{H}}(M_b)]. \tag{B.22}$$

The bracket is another Haar-diagonal operator, but need not have the original multiplier's amplitude bound or network parameterization. This is an identity for the conditioned scalar action, not an equivalence of complete nonlinear predictors.

C Kernel-Action Bounds

C.1 Finite-Tree Carleson Embedding

The following finite L^2 argument is the Carleson embedding estimate used for martingale paraproducts (Treil, 2013). It uses the same measure as the averages and Haar basis.

Lemma 1 (Finite-tree Carleson embedding)**.** *Suppose $\chi_R \geq 0$ for $R \in \mathcal{T}^\circ$ and*

$$\sum_{\substack{R \in \mathcal{T}^\circ \\ R \subseteq U}} \chi_R \leq A \mu(U) \qquad \text{for every } U \in \mathcal{T}, \tag{C.1}$$

where $A \geq 0$. Then

$$\sum_{R \in \mathcal{T}^\circ} \chi_R |\langle f \rangle_R|^2 \leq 4A \|f\|_\mu^2. \tag{C.2}$$

Step 1: maximal regions. Define the tree maximal function and its level sets by

$$M_{\mathcal{T}} f(s) = \max_{\substack{R \in \mathcal{T} \\ s \in R}} \langle |f| \rangle_R, \qquad E_t = \{ s : M_{\mathcal{T}} f(s) > t \}. \tag{C.3}$$

Let $\mathcal{R}_t$ be the inclusion-maximal regions satisfying $\langle |f| \rangle_R > t$. They exist whenever E_t is nonempty because the tree is finite. Intersecting tree regions are nested, so distinct maximal regions are disjoint and cover E_t. It follows that

$$t \mu(E_t) \leq \sum_{R \in \mathcal{R}_t} \mu(R) \langle |f| \rangle_R = \sum_{s_i \in E_t} \omega_i |f(s_i)|. \tag{C.4}$$

Step 2: the maximal estimate. Using $z^2 = 2\int_0^z t\,dt$ for $z \geq 0$ gives

$$\begin{aligned}\|M_{\mathcal{T}} f\|_\mu^2 &= 2\int_0^\infty t\mu(E_t)\,dt \\ &\leq 2\int_0^\infty \sum_{s_i \in E_t} \omega_i |f(s_i)|\,dt \\ &= 2\sum_i \omega_i |f(s_i)| \int_0^\infty \mathbf{1}_{\{t < M_{\mathcal{T}} f(s_i)\}}\,dt \\ &= 2\sum_i \omega_i |f(s_i)| M_{\mathcal{T}} f(s_i) \\ &\leq 2\|f\|_\mu \|M_{\mathcal{T}} f\|_\mu.\end{aligned} \tag{C.5}$$

The inequality on the second line is Eq. (C.4). The next equality exchanges a finite sum and an integral. The last inequality is Cauchy–Schwarz. Dividing when $\|M_{\mathcal{T}} f\|_\mu > 0$ proves $\|M_{\mathcal{T}} f\|_\mu \leq 2\|f\|_\mu$; the zero case is immediate.

Step 3: coefficient level sets. Positive weights imply $|\langle f\rangle_R| \leq \langle |f|\rangle_R$. Every region with $|\langle f\rangle_R| > t$ is therefore contained in some $U \in \mathcal{R}_t$. Applying Eq. (C.1) within each maximal region,

$$\begin{aligned}\sum_{R:|\langle f\rangle_R|>t} \chi_R &\leq \sum_{U\in\mathcal{R}_t} \sum_{\substack{R\in\mathcal{T}^\circ \\ R\subseteq U}} \chi_R \\ &\leq A \sum_{U\in\mathcal{R}_t} \mu(U) = A\mu(E_t).\end{aligned} \tag{C.6}$$

Step 4: integration. Applying the same integral identity to $|\langle f\rangle_R|^2$,

$$\begin{aligned}\sum_{R\in\mathcal{T}^\circ} \chi_R |\langle f\rangle_R|^2 &= 2\int_0^\infty t \sum_{R:|\langle f\rangle_R|>t} \chi_R\,dt \\ &\leq 2A\int_0^\infty t\mu(E_t)\,dt \\ &= A\|M_{\mathcal{T}} f\|_\mu^2 \leq 4A\|f\|_\mu^2.\end{aligned} \tag{C.7}$$

This proves Lem. 1.

Necessary and sufficient subtree control. For $\Pi_c f = \sum_R c_R \langle f\rangle_R h_R$, orthogonality gives $\|\Pi_c f\|_\mu^2 = \sum_R |c_R|^2 |\langle f\rangle_R|^2$. The lemma with $\chi_R = |c_R|^2$ and $A = \mathfrak{C}(c)$ proves $\|\Pi_c\|_{\mu\to\mu}^2 \leq 4\mathfrak{C}(c)$. For the reverse inequality, fix $U \in \mathcal{T}$ and use $f = \mathbf{1}_U$. Every $R \subseteq U$ has $\langle f\rangle_R = 1$, so

$$\sum_{\substack{R\in\mathcal{T}^\circ \\ R\subseteq U}} |c_R|^2 \leq \|\Pi_c \mathbf{1}_U\|_\mu^2 \leq \|\Pi_c\|_{\mu\to\mu}^2 \mu(U). \tag{C.8}$$

The first inequality discards the remaining nonnegative squared Haar coefficients. Dividing by $\mu(U)$ and maximizing over U proves Eq. (A.10). Thus a family of finite-tree paraproducts has uniformly bounded operator norms exactly when its subtree quantities are uniformly bounded, up to the constants in that estimate.

C.2 Proof of the Kernel-Action Bound

We prove Thm. 1 with only the incoming feature f allowed to vary.

Step 1: bounded symbols give subtree control. Fix $R \in \mathcal{T}$ and set $\bar{\alpha}_R = \langle \alpha \rangle_R$. Regional Parseval and expansion of the square give

$$\begin{aligned} \sum_{\substack{U \in \mathcal{T}^\circ \\ U \subseteq R}} |\widehat{\alpha}_U|^2 &= \sum_{s_i \in R} \omega_i |\alpha(s_i) - \bar{\alpha}_R|^2 \\ &= \sum_{s_i \in R} \omega_i |\alpha(s_i)|^2 - 2\bar{\alpha}_R \sum_{s_i \in R} \omega_i \alpha(s_i) + \mu(R)|\bar{\alpha}_R|^2 \\ &= \sum_{s_i \in R} \omega_i |\alpha(s_i)|^2 - \mu(R)|\bar{\alpha}_R|^2 \\ &\le B_\alpha^2 \mu(R). \end{aligned} \tag{C.9}$$

The third line uses $\sum_{s_i \in R} \omega_i \alpha(s_i) = \mu(R)\bar{\alpha}_R$. The last line drops a nonnegative term and uses $|\alpha(s_i)| \le B_\alpha$. The identical argument gives the B_β bound for β.

Step 2: the paraproduct. Its output terms are mutually orthogonal. Using Lem. 1 with $\chi_R = |\widehat{\alpha}_R|^2$ and $A = B_\alpha^2$ gives

$$\|\Pi_\alpha f\|_\mu^2 = \sum_{R \in \mathcal{T}^\circ} |\widehat{\alpha}_R|^2 |\langle f \rangle_R|^2 \le 4B_\alpha^2 \|f\|_\mu^2. \tag{C.10}$$

Taking square roots proves $\|\Pi_\alpha\|_{\mu \to \mu} \le 2B_\alpha$. Likewise $\|\Pi_\beta\|_{\mu \to \mu} \le 2B_\beta$.

Step 3: the adjoint. The indicator outputs of Π_β^* overlap, so we use duality, not orthogonality:

$$\begin{aligned} \|\Pi_\beta^* f\|_\mu &= \sup_{\|g\|_\mu = 1} |\langle \Pi_\beta^* f, g \rangle_\mu| \\ &= \sup_{\|g\|_\mu = 1} |\langle f, \Pi_\beta g \rangle_\mu| \\ &\le \|f\|_\mu \sup_{\|g\|_\mu = 1} \|\Pi_\beta g\|_\mu \le 2B_\beta \|f\|_\mu. \end{aligned} \tag{C.11}$$

The second line uses Eq. (B.7); the third uses Cauchy–Schwarz and Step 2.

Step 4: the multiplier. The multiplier is diagonal in the orthonormal basis, so

$$\begin{aligned} \|\mathcal{M} f\|_\mu^2 &= |\lambda_0|^2 \mu(\mathcal{S}) |\langle f \rangle_{\mathcal{S}}|^2 + \sum_{R \in \mathcal{T}^\circ} |\lambda_R|^2 |\widehat{f}_R|^2 \\ &\le B_M^2 \left(\mu(\mathcal{S}) |\langle f \rangle_{\mathcal{S}}|^2 + \sum_{R \in \mathcal{T}^\circ} |\widehat{f}_R|^2 \right) = B_M^2 \|f\|_\mu^2. \end{aligned} \tag{C.12}$$

The inequality uses both root and detail multiplier bounds. The last equality is Parseval.

Step 5: the complete action. The three branch outputs need not be mutually orthogonal. The triangle inequality gives

$$\begin{aligned} \|\mathcal{K} f\|_\mu &\le \|\mathcal{M} f\|_\mu + \|\Pi_\alpha f\|_\mu + \|\Pi_\beta^* f\|_\mu \\ &\le (B_M + 2B_\alpha + 2B_\beta) \|f\|_\mu. \end{aligned} \tag{C.13}$$

Every constant arose from an amplitude bound or the universal finite-tree embedding constant. None depends on the number of leaves, tree depth, or child-mass ratios. This proves Thm. 1. For $f = (f_1, \ldots, f_{C_h})$, summing the squared scalar bounds gives $\sum_c \|\mathcal{K} f_c\|_\mu^2 \le (B_M + 2B_\alpha + 2B_\beta)^2 \sum_c \|f_c\|_\mu^2$, which proves the componentwise vector extension.

C.3 Channel Projections and Input Changes

Multi-head action. For a pointwise matrix A, $\|Af\|_\mu^2 = \sum_i \omega_i \|Af(s_i)\|_2^2 \le \|A\|_2^2 \|f\|_\mu^2$. With common amplitude bounds across heads and $\Gamma = B_M + 2B_\alpha + 2B_\beta$, Eq. (7) satisfies

$$\begin{aligned} \|\mathcal{A}_\ell(f)\|_\mu &\le \sum_{h=1}^{H} \|O_{\ell h}\|_2 \|\mathcal{K}_{\ell h}(V_{\ell h} f)\|_\mu \\ &\le \Gamma \left(\sum_{h=1}^{H} \|O_{\ell h}\|_2 \|V_{\ell h}\|_2 \right) \|f\|_\mu. \end{aligned} \tag{C.14}$$

The symbol bounds do not constrain these channel-map norms. For fixed linear aggregation E and reconstruction R between the relevant weighted spaces, the same argument gives $\|R\mathcal{K}E\|_{\mathrm{op}} \le \|R\|_{\mathrm{op}} \Gamma \|E\|_{\mathrm{op}}$. Neither map is assumed to have norm one.

Changing the conditioning input. For two conditioning inputs on the same weighted tree,

$$\mathcal{K}[a]f - \mathcal{K}[a']g = \mathcal{K}[a](f-g) + (\mathcal{K}[a] - \mathcal{K}[a'])g. \tag{C.15}$$

The first term is bounded by $\Gamma \|f-g\|_\mu$. For the second term, define

$$\delta_M = \max\left\{ |\lambda_0[a] - \lambda_0[a']|, \max_{R \in \mathcal{T}^\circ} |\lambda_R[a] - \lambda_R[a']| \right\}, \tag{C.16}$$

with the internal maximum omitted for a singleton tree. The difference of multipliers has norm at most δ_M. Linearity of the paraproduct in its symbol and the preceding bound give

$$\|\mathcal{K}[a] - \mathcal{K}[a']\|_{\mu \to \mu} \le \delta_M + 2\|\alpha[a] - \alpha[a']\|_\infty + 2\|\beta[a] - \beta[a']\|_\infty. \tag{C.17}$$

A bound in terms of the physical input difference additionally requires regularity estimates for the coefficient generators, representation maps, and nonlinear predictor updates. The conditioned feature action $\mathcal{K}[a]$ is not the Jacobian of $\mathcal{G}_\theta$.

Physical norms. The theorem uses the discrete measure μ on kernel features. Interpreting it as an approximation to a physical integral requires appropriate quadrature weights. The supervised losses in Appendix E.3 may use different channels, masks, weights, and reductions.

D Coefficient Generation and Discrete Computation

We describe one Haar head with C_h feature channels. Symbols and multipliers are scalar and shared across those channels. Geometry and prescribed weights remain fixed during differentiation.

D.1 Coefficient Networks and Tensor Interfaces

Descriptors and generated coefficients. The initial encoding $v_0 : \mathcal{S} \to \mathbb{R}^C$ supplies $z = \langle v_0 \rangle_{\mathcal{S}}$ and regional means $\langle v_0 \rangle_R$. For assigned coordinate dimension d_s, $\rho_R = (c_R, s_R, \langle v_0 \rangle_R, z)$ has dimension $2d_s + 2C$. The symbol networks in Eq. (6) take $(s, v_0(s), z)$ of dimension $d_s + 2C$. The multiplier networks are

$$\lambda_R = B_M \tanh F_{\theta_M}(\rho_R), \qquad \lambda_0 = B_M \tanh F_{\theta_0}(\rho_{\mathcal{S}}). \tag{D.1}$$

The root descriptor contains both $\langle v_0 \rangle_{\mathcal{S}}$ and z to follow the same regional input format; they are the same pooled quantity. Networks are separate across layers, heads, and coefficient types, with sharing over support elements or regions within one head. The interfaces are listed in Table D.1.

Input dependence. Descriptors can be reused only while the encoding, support, and weights agree. A changed physical input requires new descriptors, symbols, and multipliers even if the tree is reused. Both the feature path and the coefficient-generation path remain in the training graph.

Table D.1: Tensor interfaces for one Haar head, omitting the batch dimension.

Quantity	Shape	Meaning
f	$N \times C_h$	Incoming feature values.
α, β	$N \times 1$ each	Symbol values.
$\widehat{\alpha}_R, \widehat{\beta}_R$	$(N-1) \times 1$ each	Transfer coefficients.
λ_R, λ_0	$(N-1) \times 1$, scalar	Detail and root-mean multipliers.
$\langle f \rangle_R$	$(2N-1) \times C_h$	Means at all tree nodes.
$\widehat{f}_R, \xi_R, \eta_R$	$(N-1) \times C_h$ each	Internal statistics and contributions.
r_R	$(2N-1) \times C_h$	Accumulated responses.

D.2 Explicit Kernel Representation

Expanding Def. 1 gives

$$(\mathcal{K}f)(s) = \lambda_0 \langle f \rangle_{\mathcal{S}} + \sum_{R \in \mathcal{T}^\circ} [\lambda_R \widehat{f}_R + \widehat{\alpha}_R \langle f \rangle_R] h_R(s) + \sum_{R \in \mathcal{T}^\circ} \widehat{\beta}_R \widehat{f}_R \frac{\mathbf{1}_R(s)}{\mu(R)}. \tag{D.2}$$

Writing each mean and Haar coefficient as a weighted sum yields

$$\kappa(s,t) = \frac{\lambda_0}{\mu(\mathcal{S})} + \sum_{R \in \mathcal{T}^\circ} \lambda_R h_R(s) h_R(t) + \sum_{R \in \mathcal{T}^\circ} \frac{\widehat{\alpha}_R}{\mu(R)} h_R(s) \mathbf{1}_R(t) + \sum_{R \in \mathcal{T}^\circ} \frac{\widehat{\beta}_R}{\mu(R)} \mathbf{1}_R(s) h_R(t). \tag{D.3}$$

Substitution into Eq. (1) recovers Eq. (D.2) term by term. The array matrix is $A_{ij} = \omega_j \kappa(s_i, s_j)$; it differs from the array of kernel values when weights are not all one. The tree evaluation assembles neither array.

D.3 Two-Pass Evaluation

For every internal region, store child indices and childwise Haar values $\phi_{R,L} = \nu_R / \mu(R_L)$ and $\phi_{R,R} = -\nu_R / \mu(R_R)$. Algorithm D.1 evaluates the head once its coefficient networks have produced their outputs.

Step 1: correctness of the upward pass. Leaf means equal the supplied values. If the child means are correct, then

$$\frac{\mu(R_L)\langle F \rangle_{R_L} + \mu(R_R)\langle F \rangle_{R_R}}{\mu(R)} = \frac{1}{\mu(R)} \left(\sum_{s_i \in R_L} \omega_i F(s_i) + \sum_{s_i \in R_R} \omega_i F(s_i) \right) = \langle F \rangle_R. \tag{D.4}$$

The equality uses the disjoint partition of R. Induction proves every regional mean. The Haar coefficients are then exact by Eq. (B.2).

Step 2: correctness of the downward pass. Before processing R, the accumulator satisfies

$$r_R = \lambda_0 \langle f \rangle_{\mathcal{S}} + \sum_{\substack{U \in \mathcal{T}^\circ \\ U \supsetneq R}} [(h_U|_R)\xi_U + \eta_U]. \tag{D.5}$$

At the root, the sum is empty. Passing from R to a child adds precisely the contribution of the new strict ancestor R; all earlier ancestor functions are constant on that child. Induction proves the

Algorithm D.1 One HPNO Haar head on a prepared tree

Require: Tree $\mathcal{T}$, child indices, leaf order, masses, ν_R, and $\phi_{R,L}, \phi_{R,R}$.
Require: Features f, symbols α, β, and multipliers $\lambda_0, \{\lambda_R\}_{R\in\mathcal{T}^\circ}$.
Ensure: $y(s_i) = (\mathcal{K}f)(s_i)$ in the original support order.

$F \leftarrow (f, \alpha, \beta)$
$\langle F\rangle_{\{s_i\}} \leftarrow F(s_i)$ for each leaf
for all $R \in \mathcal{T}^\circ$ in child-before-parent order **do**
  $\langle F\rangle_R \leftarrow [\mu(R_L)\langle F\rangle_{R_L} + \mu(R_R)\langle F\rangle_{R_R}]/\mu(R)$
  $\widehat{F}_R \leftarrow \nu_R[\langle F\rangle_{R_L} - \langle F\rangle_{R_R}]$
end for
$r_{\mathcal{S}} \leftarrow \lambda_0\langle f\rangle_{\mathcal{S}}$
for all $R \in \mathcal{T}^\circ$ in parent-before-child order **do**
  $\xi_R \leftarrow \lambda_R\widehat{f}_R + \widehat{\alpha}_R\langle f\rangle_R$
  $\eta_R \leftarrow \widehat{\beta}_R\widehat{f}_R/\mu(R)$
  $r_{R_L} \leftarrow r_R + \phi_{R,L}\xi_R + \eta_R$
  $r_{R_R} \leftarrow r_R + \phi_{R,R}\xi_R + \eta_R$
end for
$y(s_i) \leftarrow r_{\{s_i\}}$ for each leaf
return y in the original support order

invariant at every node. At a leaf it gives

$$\begin{aligned} r_{\{s_i\}} &= \lambda_0\langle f\rangle_{\mathcal{S}} + \sum_{\substack{R\in\mathcal{T}^\circ \\ s_i\in R}} [h_R(s_i)\xi_R + \eta_R] \\ &= (\mathcal{K}f)(s_i), \end{aligned} \tag{D.6}$$

where the last equality is Eq. (D.2). For $N = 1$, both loops are empty and the root initialization is the output. The identity holds in real arithmetic; finite precision introduces roundoff but no truncation of the specified action.

Ordering and batching. Any storage permutation is inverted before comparison with targets or physical-grid differencing. Nodes with resolved dependencies can be processed together. Different cases keep separate statistics and trees, and padding is masked rather than introduced as a positive-mass leaf.

D.4 Backward Computation

Weighted upstream derivative. Let $y = \mathcal{K}f$ and let $J(y)$ be a scalar objective. Define g by $\delta J = \langle g, \delta y\rangle_\mu$, with the inner product summing over channels. For fixed coefficients, $\delta y = \mathcal{K}(\delta f)$, so

$$\nabla_f^\mu J = \mathcal{K}^* g = \mathcal{M}g + \Pi_\alpha^* g + \Pi_\beta g. \tag{D.7}$$

The multiplier is self-adjoint because its coefficients are real. The remaining terms follow from Eq. (B.7).

Transfer-coefficient derivatives. Varying $\widehat{\alpha}_R$ alone changes the output by $\delta y(s) = \langle f\rangle_R h_R(s)\delta\widehat{\alpha}_R$. Pairing with g gives $\delta J = \langle f\rangle_R^\top \widehat{g}_R\,\delta\widehat{\alpha}_R$. Similarly, varying $\widehat{\beta}_R$ gives $\delta y(s) = \widehat{f}_R \mathbf{1}_R(s)\delta\widehat{\beta}_R/\mu(R)$, whose pairing is $\widehat{f}_R^\top\langle g\rangle_R\,\delta\widehat{\beta}_R$. The multiplier derivatives follow by the same calculation:

$$\begin{aligned} \frac{\partial J}{\partial\widehat{\alpha}_R} &= \langle f\rangle_R^\top\widehat{g}_R, & \frac{\partial J}{\partial\widehat{\beta}_R} &= \widehat{f}_R^\top\langle g\rangle_R, \\ \frac{\partial J}{\partial\lambda_R} &= \widehat{f}_R^\top\widehat{g}_R, & \frac{\partial J}{\partial\lambda_0} &= \mu(\mathcal{S})\langle f\rangle_{\mathcal{S}}^\top\langle g\rangle_{\mathcal{S}}. \end{aligned} \tag{D.8}$$

Dot products sum feature channels that share a scalar coefficient. The root-mean derivative includes $\mu(\mathcal{S})$ because the constant field is stored as its mean rather than its normalized constant-basis coefficient.

Derivatives of symbol values. With the tree and weights fixed, $\partial\widehat{\alpha}_R/\partial\alpha(s_i) = \omega_i h_R(s_i)$. The chain rule therefore gives ordinary array derivatives

$$\begin{aligned}\frac{\partial J}{\partial \alpha(s_i)} &= \omega_i \sum_{R\in\mathcal{T}^\circ} \frac{\partial J}{\partial \widehat{\alpha}_R} h_R(s_i),\\ \frac{\partial J}{\partial \beta(s_i)} &= \omega_i \sum_{R\in\mathcal{T}^\circ} \frac{\partial J}{\partial \widehat{\beta}_R} h_R(s_i).\end{aligned} \tag{D.9}$$

These sums are Haar reconstructions followed by leaf weighting. For a symbol preactivation $z_{\alpha,i}$, $\partial\alpha(s_i)/\partial z_{\alpha,i} = B_\alpha(1-\tanh^2 z_{\alpha,i})$. The other bounded outputs have the same derivative form with their respective amplitudes.

Ordinary feature-array gradients. If an autodifferentiation system supplies $\widetilde{g}_i$ through $\delta J = \sum_i \widetilde{g}_i^\top \delta y(s_i)$, then $g(s_i) = \widetilde{g}_i/\omega_i$. Let $W_\mu = \operatorname{diag}(\omega_1,\dots,\omega_N)$ and let A be the array matrix of $\mathcal{K}$. Its weighted adjoint is represented by $W_\mu^{-1}A^\top W_\mu$, so the ordinary feature gradient is

$$(\nabla_f^{\mathrm{E}} J)_i = \omega_i(\mathcal{K}^* g)(s_i), \qquad \nabla_f^{\mathrm{E}} J = A^\top \widetilde{g}. \tag{D.10}$$

Differentiating the actual weighted arithmetic in Algorithm D.1 already includes these factors. They are not applied again afterward.

Reverse accumulation. Let $\bar{m}_R$ and $\bar{d}_R$ denote ordinary upstream derivatives with respect to $\langle F\rangle_R$ and $\widehat{F}_R$. Differentiating the two upward formulas gives the updates

$$\begin{aligned}\bar{m}_{R_L} &\mathrel{+}= \frac{\mu(R_L)}{\mu(R)}\bar{m}_R + \nu_R\bar{d}_R,\\ \bar{m}_{R_R} &\mathrel{+}= \frac{\mu(R_R)}{\mu(R)}\bar{m}_R - \nu_R\bar{d}_R.\end{aligned} \tag{D.11}$$

Parent-before-child traversal propagates these additions to the leaves. Mean-only pooling has $\bar{d}_R = 0$. Gradients from regional and global coefficient descriptors accumulate in the same initial feature encoding.

Total parameter dependence. Features and coefficients both depend on trainable parameters. At fixed tree and weights, the complete variation is

$$\delta y = \mathcal{K}(\delta f) + (\delta\mathcal{K})f, \tag{D.12}$$

where

$$\begin{aligned}((\delta\mathcal{K})f)(s) = {}& \delta\lambda_0\langle f\rangle_{\mathcal{S}}\\ &+ \sum_{R\in\mathcal{T}^\circ}[\delta\lambda_R\widehat{f}_R + \delta\widehat{\alpha}_R\langle f\rangle_R]h_R(s)\\ &+ \sum_{R\in\mathcal{T}^\circ}\delta\widehat{\beta}_R\widehat{f}_R\frac{\mathbf{1}_R(s)}{\mu(R)}.\end{aligned} \tag{D.13}$$

The feature adjoint handles the first term of Eq. (D.12), and coefficient-network derivatives handle the second. These formulas do not differentiate the discrete tree or prescribed weights; learned weights would require additional derivatives.

D.5 Complexity, Preprocessing, and Reuse

Prepared-tree work. A binary tree with N leaves has $N-1$ internal regions. For one head, the upward and downward passes require $\mathcal{O}(N(C_h+2))$ and $\mathcal{O}(NC_h)$ work. Across H heads with $C = HC_h$, this is $\mathcal{O}(N(C+H))$. At fixed head count, coordinate dimension, and network depth, pooling uses $\mathcal{O}(NC)$ work, while coefficient networks with widths proportional to C and channel projections use $\mathcal{O}(NC^2)$ work per block. These counts apply to the prepared Haar support, where $N = |\mathcal{S}|$.

Hierarchy construction. The recursive-sorting construction splits a region at the median of its longest bounding-box axis with deterministic tie handling. Sorting at every node of the balanced recursion costs $\mathcal{O}(N \log^2 N)$ in total. Prescribed logical or slot trees instead use their assigned topology. Given connectivity, masses, centroids, bounding boxes, and Haar values are computed bottom-up in $\mathcal{O}(N)$ additional work at fixed coordinate dimension.

Storage and reuse. Tree geometry uses $\mathcal{O}(N)$ storage, and feature statistics, responses, and fixed-depth network activations use $\mathcal{O}(NC)$ per block at fixed head count. Reverse differentiation has the same asymptotic tree work as the forward pass. Connectivity is reusable while coordinates and partition rules are unchanged; masses and Haar values additionally require unchanged weights. New physical inputs still require new encodings and coefficients. Representation maps and auxiliary processing are outside this tree computation; dense aggregation and reconstruction between $n_{\mathcal{X}}$ nodes and N slots contribute $\mathcal{O}(n_{\mathcal{X}} NC)$ work.

D.6 Algebraic Consistency Conditions

Setting $\alpha = \beta = 0$ and all multipliers to one yields $\mathcal{K} f = f$ by Eq. (B.4). The all-singleton cut has zero cross-cut responses, and a singleton tree requires only root multiplication. The explicit matrix and the two-pass algorithm represent the same linear map by Eq. (D.6). For unequal weights, its induced norm is

$$\|\mathcal{K}\|_{\mu\to\mu} = \|W_\mu^{1/2} A W_\mu^{-1/2}\|_2, \tag{D.14}$$

because $\|f\|_\mu = \|W_\mu^{1/2} f\|_2$. Thus an unweighted matrix spectral norm is not generally the norm used in Thm. 1. These identities provide consistency conditions for an implementation, separate from physical-field accuracy.

E Experimental Details

This section specifies the evaluation metrics, baseline comparisons, and training protocol used in § 5.

E.1 Evaluation Metrics

Six PDE benchmarks. For a prediction $\widehat{u}_i$ and reference u_i with nonzero norm, we use

$$\ell_{\mathrm{rel}}(\widehat{u}_i, u_i) = \frac{\|\operatorname{vec}(\widehat{u}_i - u_i)\|_2}{\|\operatorname{vec}(u_i)\|_2}, \qquad E_{\mathrm{rel}} = \frac{1}{n_{\mathrm{te}}} \sum_{i=1}^{n_{\mathrm{te}}} \ell_{\mathrm{rel}}(\widehat{u}_i, u_i). \tag{E.1}$$

Vectorization includes the evaluated spatial entries and output channels. Predictions and references are decoded to the original target scale before each ratio is computed. Temporal scores include the complete predicted sequence within each case, rather than averaging framewise ratios. This physical-output metric is distinct from the feature norm $\|\cdot\|_\mu$ in Thm. 1.

ShapeNet Car. Let V_i and S_i^{surf} be the volume and surface masks, and let $\|\cdot\|_{2,A}$ restrict a Euclidean norm to mask A. The volume-velocity and surface-pressure scores are

$$E_U = \frac{1}{n_{\mathrm{te}}} \sum_i \frac{\|\widehat{U}_i - U_i\|_{2,V_i}}{\|U_i\|_{2,V_i}}, \qquad E_p = \frac{1}{n_{\mathrm{te}}} \sum_i \frac{\|\widehat{p}_i - p_i\|_{2,S_i^{\mathrm{surf}}}}{\|p_i\|_{2,S_i^{\mathrm{surf}}}}. \tag{E.2}$$

All velocity components are included in E_U. Both fields are decoded before evaluation.

AirfRANS. The volume and surface pressure scores use normalized MSE. With $R_{i,v} = V_i$ and $R_{i,s} = S_i^{\mathrm{surf}}$,

$$E_{r,p} = \frac{1}{n_{\mathrm{te}}} \sum_i \frac{1}{|R_{i,r}|} \sum_{j \in R_{i,r}} |\widehat{\widetilde{p}}_{ij} - \widetilde{p}_{ij}|^2, \qquad r \in \{v, s\}. \tag{E.3}$$

Tildes denote normalization by training statistics. These scores contain pressure entries only.

Table E.1: Benchmark interfaces and test-set sizes. CFD node counts describe the sampled interfaces.

Task	Spatial interface	Input and reference target	Test
Darcy	85×85 grid	Permeability to pressure	200
Navier–Stokes	64×64 grid	Ten observed states to ten future states	200
Elasticity	972 points	Geometry to scalar stress	200
Plasticity	101×31 mesh	Die geometry to twenty deformation states	80
Airfoil	221×51 mesh	Airfoil geometry to Mach number	200
Pipe	129×129 mesh	Pipe geometry to horizontal velocity	200
ShapeNet Car	About 32k points	Geometry to velocity and surface pressure	100
AirfRANS	32k sampled points	Airfoil and flow conditions to flow fields	200

Aerodynamic coefficients. For nonzero reference drag or lift coefficients C_i, report

$$E_C = \frac{1}{n_{\text{te}}} \sum_i \frac{|\widehat{C}_i - C_i|}{|C_i|}, \qquad \rho_C = \operatorname{corr}\big(\operatorname{rank}(\widehat{C}), \operatorname{rank}(C)\big). \tag{E.4}$$

The tables use (E_{C_D}, ρ_D) for ShapeNet Car and (E_{C_L}, ρ_L) for AirfRANS. Predicted and reference coefficients use the same force-integration and reference-flow normalization conventions. Regional-response metrics are defined separately in Appendix F.3.

E.2 Baselines

Comparison set. Tables 1 and 2 include spectral operators, wavelet and multiresolution models, and attention-based operators. The comparison covers Fourier and geometry-aware constructions (**?**Li et al., 2023b;c), wavelet-based operators (Gupta et al., 2021; Tripura & Chakraborty, 2023), and slice, linear, and hyperbolic interaction models (Wu et al., 2024; Luo et al., 2025; Hu et al., 2026; Pei et al., 2026). The method citations in the main tables identify the reported benchmark entries; the tables do not represent a common retraining of every baseline.

Direct comparisons. The paired Darcy evaluation applies HNO, LinearNO, and HPNO to identical constructed inputs and reference solutions. The ablation study separately trains the HPNO variants in Table 3, retaining the same task representations, auxiliary paths, supervision, and evaluation. Appendix F.2 specifies their coefficient constructions.

E.3 Training Protocol

Data and preprocessing. Darcy and Navier–Stokes follow the benchmark interfaces of **?**; the geometry-based tasks follow Li et al. (2023b). The CFD tasks use the comparison interfaces of Wu et al. (2024). Table E.1 summarizes the spatial inputs and test sets. Normalization statistics use training data, and arrays, masks, and targets retain a common node ordering through reconstruction. Darcy subsamples the released 421×421 arrays to 85×85; Navier–Stokes uses periodic flow with viscosity 10^{-5}. Airfoil predicts Mach number on deformed NACA-0012 geometries and is distinct from AirfRANS. ShapeNet Car geometry inputs include coordinates, signed distance, and normals.

Predictor configuration. A separate predictor is trained end-to-end for each task using Eq. (8). The representation maps and residual processing are adapted from Pei et al. (2026). Table E.2 lists the kernel supports, objectives, and additional processing. Airfoil and Pipe hierarchies use logical coordinates, Elasticity uses physical coordinates, and the CFD hierarchies use fixed slot coordinates. Positive Haar weights are prescribed, separately from learned slice assignments and supervised-loss weights. The coefficient-network interfaces are specified in Appendix D.1.

Task losses. Darcy adds supervised derivative matching,

$$\ell_{\text{grad}}(\widehat{u}, u) = \sum_{r=1}^{2} \frac{\|D_r^{\text{pred}}\widehat{u} - D_r^{\text{target}}u\|_2}{\|D_r^{\text{target}}u\|_2}. \tag{E.5}$$

Table E.2: Task configurations. The patch-size parameter is p; $\ell_{\mathrm{rel},t}$ denotes next-frame relative loss. The listed processing is retained across kernel ablations.

Task	Haar support	Objective	Additional processing
Darcy	Patches, $p = 5$	$\ell_{\mathrm{rel}} + 0.1\ell_{\mathrm{grad}}$	Patch reconstruction and local convolution
Navier–Stokes	Patches, $p = 4$	$\sum_{t=11}^{20} \ell_{\mathrm{rel},t}$	History channels, residual output, periodic handling
Plasticity	Patches, $p = 3$	ℓ_{rel} at requested times	Time embedding and displacement reconstruction
Airfoil	Patches, $p = 5$	ℓ_{rel}	Full-resolution refinement and dilated local processing
Pipe	Patches, $p = 3$	ℓ_{rel}	Patch embedding and reconstruction
Elasticity	Physical nodes	ℓ_{rel}	Residual pointwise processing and a coarse-mean path
AirfRANS	64 slice states	$\gamma_v \ell_v + \gamma_s \ell_s$	Slicing and deslicing
ShapeNet Car	64 slice states	$\ell_v + 0.5\ell_s$	Slicing, deslicing, and pressure correction

The central-difference implementation sets predicted boundary entries to zero before differencing and retains reference boundary entries; the operators include this preprocessing. The original grid order is restored first. This term is excluded from the reported field error. For CFD field groups $r \in \{v, s\}$ with supervised node-channel set $\mathcal{E}_r$,

$$\ell_r = \sum_{(j,c)\in\mathcal{E}_r} \vartheta_{jc}^{(r)} |\widehat{u}_{jc}^{\mathrm{tr}} - u_{jc}^{\mathrm{tr}}|^2, \tag{E.6}$$

where the superscript denotes the training scale and fixed nonnegative weights specify the reduction; ordinary MSE uses $\vartheta_{jc}^{(r)} = 1/|\mathcal{E}_r|$. Training across field groups is distinct from the pressure-only AirfRANS reporting metric.

Temporal protocol. For each Navier–Stokes training trajectory $(u_1, \ldots, u_{20})$, next-frame predictions use observed histories $(u_{t-10}, \ldots, u_{t-1})$ for $t = 11, \ldots, 20$. The trajectory objective sums these ten relative losses, and minibatches average trajectory objectives. Evaluation instead rolls out

$$\widehat{u}_t = \mathcal{G}_\theta(\mathcal{X}, w, [\widetilde{u}_{t-10}, \ldots, \widetilde{u}_{t-1}]), \qquad \widetilde{u}_t = \begin{cases} u_t, & t \le 10, \\ \widehat{u}_t, & t > 10, \end{cases} \tag{E.7}$$

for $t = 11, \ldots, 20$, with brackets concatenating history channels. For Plasticity, each requested time is predicted independently; channel selection and displacement reconstruction precede sequence scoring.

F Additional Analyses and Visualizations

F.1 Kernel Diagnostics

Directional responses. For a selected Darcy input, layer, head, feature channel, and cut, we hold the generated coefficients and weighted hierarchy fixed. The coarse and detail features and centered symbols are

$$\begin{aligned} f_{\mathrm{c}} &= \mathbb{E}_{\mathcal{C}}[f], & f_{\mathrm{d}} &= (I - \mathbb{E}_{\mathcal{C}})f, \\ \alpha^\circ &= \alpha - \mathbb{E}_{\mathcal{C}}[\alpha], & \beta^\circ &= \beta - \mathbb{E}_{\mathcal{C}}[\beta]. \end{aligned} \tag{F.1}$$

The responses displayed in Fig. 4 are

$$r_{\mathrm{md}} = (I - \mathbb{E}_{\mathcal{C}})\mathcal{K} f_{\mathrm{c}} = \alpha^\circ f_{\mathrm{c}}, \qquad r_{\mathrm{dm}} = \mathbb{E}_{\mathcal{C}}[\mathcal{K} f_{\mathrm{d}}] = \mathbb{E}_{\mathcal{C}}[\beta^\circ f_{\mathrm{d}}]. \tag{F.2}$$

The reverse response includes regional averaging, so it is constant within each cut region. These maps show the learned feature action on its kernel support; Appendix B.3 gives the identities and exact operator norms. A response-to-input norm ratio for an arbitrary feature is not generally the operator norm.

Normalized subtree statistics. Let $t_R^{(\alpha)}$ be the mean-to-detail coefficient for one layer and head. Define

$$S_R^{(\alpha)} = \frac{\sum_{U\in\mathcal{T}^\circ:\,U\subseteq R} |t_U^{(\alpha)}|^2}{B_\alpha^2 \mu(R)}, \qquad R \in \mathcal{T}, \tag{F.3}$$

with the analogous definition for β. For HPNO, $t_R^{(\alpha)} = \widehat{\alpha}_R$; for Direct and Normalized-direct, it is c_R^{md} and $\bar{c}_R^{\mathrm{md}}$, respectively. For case i and direction $d \in \{\alpha, \beta\}$, the plotted value is

$$S_{i,\max}^{(d)} = \max_{\ell,h} \max_{R\in\mathcal{T}_{i\ell h}} S_{R,i\ell h}^{(d)}. \tag{F.4}$$

Thus points represent cases and short bars their medians. Eq. (11) gives the unit reference for HPNO; Normalized-direct enforces the same subtree restriction. The plotted Direct model exceeds this reference in the detail-to-mean direction. The statistic describes coefficient accumulation; the field-error comparison is in Table 3.

F.2 Ablation Definitions

Directional and tied-symbol controls. The four directional variants separately train $\mathcal{M}$, $\mathcal{M}+\Pi_\alpha$, $\mathcal{M}+\Pi_\beta^*$, and $\mathcal{M}+\Pi_\alpha+\Pi_\beta^*$. Unused coefficient networks are removed, while every Haar coordinate is retained. The remaining predictor and supervision are unchanged, including Darcy's derivative loss and Elasticity's coarse-mean path. The tied variant uses one symbol network with $B_\alpha = B_\beta = B$, giving $\alpha = \beta$; its untied comparison uses the same directional amplitudes. Appendix B.4 characterizes this fixed-kernel restriction.

Direct coefficients. For the regional descriptor ρ_R in Appendix D.1, set $\zeta_R = (\rho_R, \langle v_0\rangle_{R_L}, \langle v_0\rangle_{R_R})$. Two regional networks generate

$$\begin{aligned} c_R^{\mathrm{md}} &= B_\alpha \sqrt{\mu(R)}\tanh G_{\theta_\alpha}^{\mathrm{dir}}(\zeta_R), \\ c_R^{\mathrm{dm}} &= B_\beta \sqrt{\mu(R)}\tanh G_{\theta_\beta}^{\mathrm{dir}}(\zeta_R), \end{aligned} \tag{F.5}$$

separately for each layer and head, with sharing across regions. The resulting action is

$$\mathcal{K}^{\mathrm{dir}} f = \mathcal{M} f + \sum_{R\in\mathcal{T}^\circ} c_R^{\mathrm{md}} \langle f\rangle_R h_R + \sum_{R\in\mathcal{T}^\circ} c_R^{\mathrm{dm}} \widehat{f}_R \frac{\mathbf{1}_R}{\mu(R)}. \tag{F.6}$$

The multiplier and remaining predictor are unchanged. For HPNO, the bounded child means give

$$|\widehat{\alpha}_R| \le 2B_\alpha \nu_R \le B_\alpha \sqrt{\mu(R)}, \tag{F.7}$$

where the second inequality follows from $4\mu(R_L)\mu(R_R) \le \mu(R)^2$. Direct satisfies the latter envelope, but does not impose a subtree bound. The envelope can be weaker for unequal child masses. The comparison also changes the coefficient networks' conditioning inputs, as specified by ζ_R.

Normalized-direct. Using the subtree quantity $\mathfrak{C}$ in Eq. (A.9), set

$$\bar{c}_R^{\mathrm{md}} = \frac{c_R^{\mathrm{md}}}{\max\{1, \sqrt{\mathfrak{C}(c^{\mathrm{md}})}/B_\alpha\}}, \qquad \bar{c}_R^{\mathrm{dm}} = \frac{c_R^{\mathrm{dm}}}{\max\{1, \sqrt{\mathfrak{C}(c^{\mathrm{dm}})}/B_\beta\}}. \tag{F.8}$$

The homogeneity $\mathfrak{C}(tc) = t^2\mathfrak{C}(c)$ implies

$$\mathfrak{C}(\bar{c}^{\mathrm{md}}) = \frac{\mathfrak{C}(c^{\mathrm{md}})}{\max\{1, \mathfrak{C}(c^{\mathrm{md}})/B_\alpha^2\}} \le B_\alpha^2, \tag{F.9}$$

and similarly for the reverse direction. Normalization is applied during training and evaluation. It leaves coefficients unchanged when inactive; gradients pass through the scaling on the active branch. This imposes HPNO's subtree restriction without imposing its pointwise bound on a reconstructed symbol. Field errors for the trained variants are reported in Table 3.

F.3 Paired Darcy Protocol and Results

Physical partition. All models use the same physical grid, forcing, boundary conditions, and partition $\mathcal{P}_{\mathcal{X}}$. For $x_j \in R \in \mathcal{P}_{\mathcal{X}}$ and positive weights w_j,

$$(P_{\mathcal{X}} a)(x_j) = \frac{\sum_{x_k \in R} w_k a(x_k)}{\sum_{x_k \in R} w_k}. \tag{F.10}$$

This projection acts on physical inputs and outputs, separately from $\mathbb{E}_{\mathcal{C}}$ on kernel features.

Mean-preserving rearrangement. Within each region, let π_R permute node indices with $w_{\pi_R(j)} = w_j$, and set $a^+(x_j) = a^-(x_{\pi_R(j)})$. Changing the summation index gives

$$\begin{aligned} \sum_{x_j \in R} w_j a^+(x_j) &= \sum_{x_k \in R} w_{\pi_R^{-1}(k)} a^-(x_k) \\ &= \sum_{x_k \in R} w_k a^-(x_k). \end{aligned} \tag{F.11}$$

Therefore $P_{\mathcal{X}} a^+ = P_{\mathcal{X}} a^-$, and positive permeability is preserved. Within-region tile exchanges on the equally weighted grid have this property. The evaluation uses 50 held-out base cases and one pair per case at each tile-pair ratio $r \in \{0.1, 0.2, \ldots, 1.0\}$. Pairs are identical across models at each setting.

Extent and reference responses. The tile-pair ratio controls the rearrangement, not grid resolution. The measured fraction of changed nodes is

$$q_i(r) = \frac{1}{n_{\mathcal{X}}} \sum_{j=1}^{n_{\mathcal{X}}} \mathbf{1}_{\{a_i^+(x_j;r) \neq a_i^-(x_j)\}}, \qquad \bar{q}(r) = \frac{1}{n_{\text{pair}}} \sum_i q_i(r). \tag{F.12}$$

Each modified input has its own reference solution; an unchanged base-input solution can be shared across ratios. Frozen predictors evaluate both inputs without response-specific training, giving $\Delta u_i = u_i^+ - u_i^-$ and $\Delta \widehat{u}_i = \widehat{u}_i^+ - \widehat{u}_i^-$. The reference quantity is the solution difference, whose accuracy is distinct from the accuracy of the individual solves.

Response metrics. Let $W = w(\mathcal{X}) = \sum_j w_j$ and $\|z\|_w^2 = \sum_j w_j |z(x_j)|^2$. At each setting, the absolute and relative regional errors are

$$\begin{aligned} E_{\text{resp,abs}}(r) &= \left[\frac{1}{n_{\text{pair}} W} \sum_i \|P_{\mathcal{X}}(\Delta \widehat{u}_i - \Delta u_i)\|_w^2 \right]^{1/2}, \\ E_{\text{resp}}(r) &= \left[\frac{\sum_i \|P_{\mathcal{X}}(\Delta \widehat{u}_i - \Delta u_i)\|_w^2}{\sum_i \|P_{\mathcal{X}} \Delta u_i\|_w^2} \right]^{1/2}. \end{aligned} \tag{F.13}$$

The relative score is a ratio of aggregate energies, not a mean of pairwise ratios. It equals one for a zero-response predictor when the reference energy is positive; otherwise only the absolute score is defined. For reference-response RMS

$$G(r) = \left[\frac{1}{n_{\text{pair}} W} \sum_i \|P_{\mathcal{X}} \Delta u_i\|_w^2 \right]^{1/2}, \qquad E_{\text{resp}}(r) = \frac{E_{\text{resp,abs}}(r)}{G(r)}, \tag{F.14}$$

the last equality holds whenever $G(r) > 0$.

Results. HPNO has the lowest absolute regional error at every setting in Table F.1. At $r = 1$, the corresponding relative errors are 0.1281, 0.1930, and 0.2131 for HPNO, HNO, and LinearNO. All three relative errors decrease from $r = 0.9$ to $r = 1$ because $G(r)$ increases faster than their absolute errors, as expressed by Eq. (F.14).

Relation to the maps. Fig. 5 displays full-resolution Δu and $\Delta \widehat{u} - \Delta u$, whereas the scores above apply $P_{\mathcal{X}}$ before aggregation. Writing $e_i^{\pm} = \widehat{u}_i^{\pm} - u_i^{\pm}$ gives regional response error $P_{\mathcal{X}}(e_i^+ - e_i^-)$; a shared field bias can therefore cancel. This evaluation measures changes in the complete predictor, including its input-generated coefficients.

Table F.1: Paired Darcy results over 50 pairs at each tile-pair ratio. Reference RMS G and absolute regional errors are in units of 10^{-5}; changed-node percentages are averaged over pairs.

Ratio	Changed (%)	Reference G	HNO	LinearNO	HPNO
0.1	1.61	5.5311	1.3742	1.4505	0.9871
0.2	2.81	7.7780	1.9570	1.9831	1.3485
0.3	4.45	10.9869	2.5513	2.7799	1.8162
0.4	5.63	12.8290	2.9513	3.1066	2.0161
0.5	6.76	13.8938	3.2450	3.4340	2.1785
0.6	8.24	16.2834	3.7307	3.7478	2.5202
0.7	9.50	18.0043	4.0703	4.0968	2.7544
0.8	10.92	20.3364	4.3662	4.5377	2.8514
0.9	12.45	22.7296	4.4817	5.0138	2.9885
1.0	13.49	23.8540	4.6046	5.0841	3.0556

F.4 Additional Field Error Maps

This section presents extended qualitative evaluations across representative test cases for all evaluated PDE and CFD benchmarks. Each comparison displays the ground-truth reference field u alongside signed prediction errors $\widehat{u} - u$ for HPNO, HNO, and LinearNO under identical domains, coordinate views, and physical output channels. Reference fields and error maps use separate colormaps, with a common symmetric error scale enforced across all models within each case. For aerodynamic benchmarks (ShapeNet Car and AirfRANS), velocity components and surface/volume pressure distributions are visualized separately to evaluate boundary-layer and wake resolution.

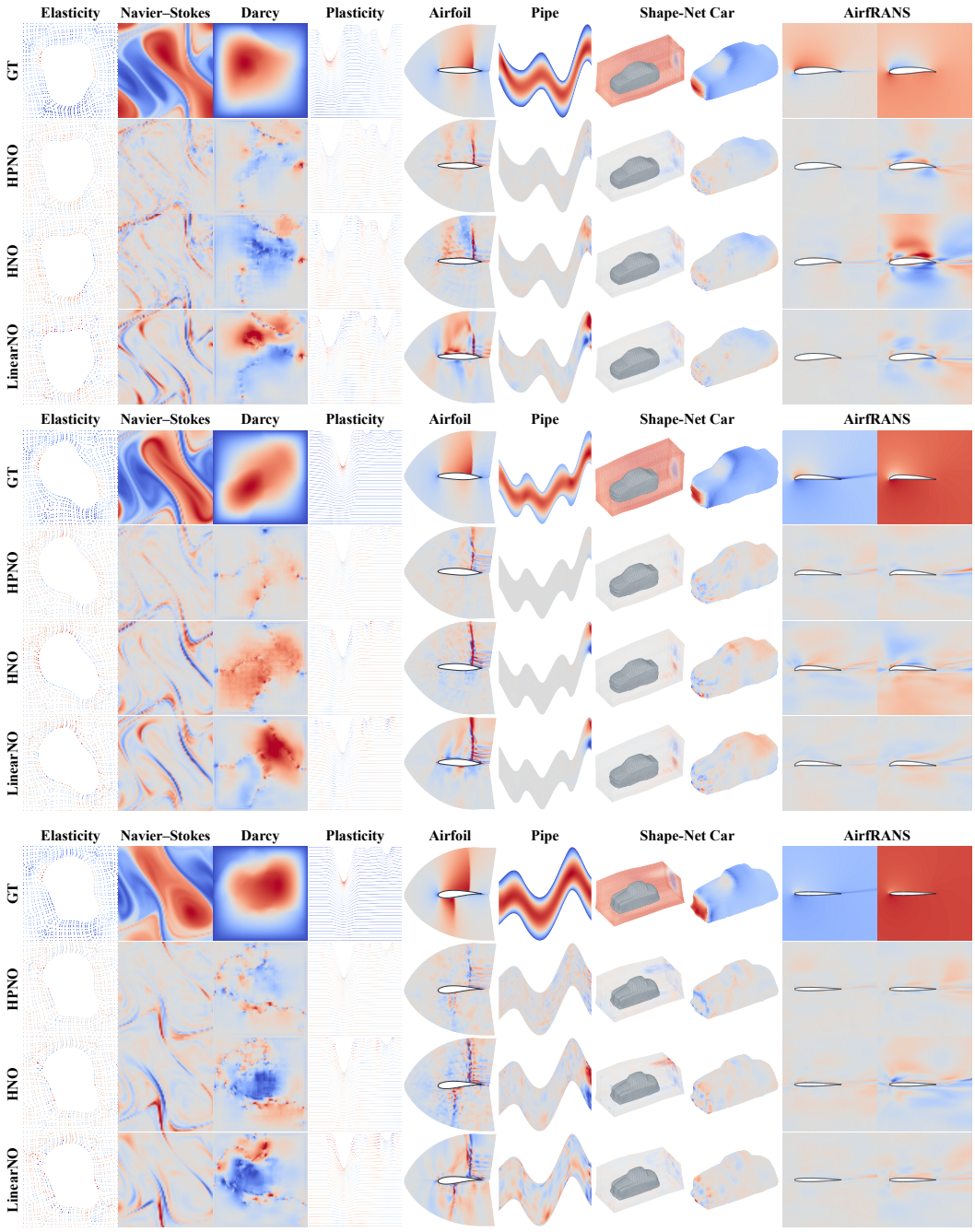

Figure F.1: Ground-truth fields and signed prediction errors (continued).

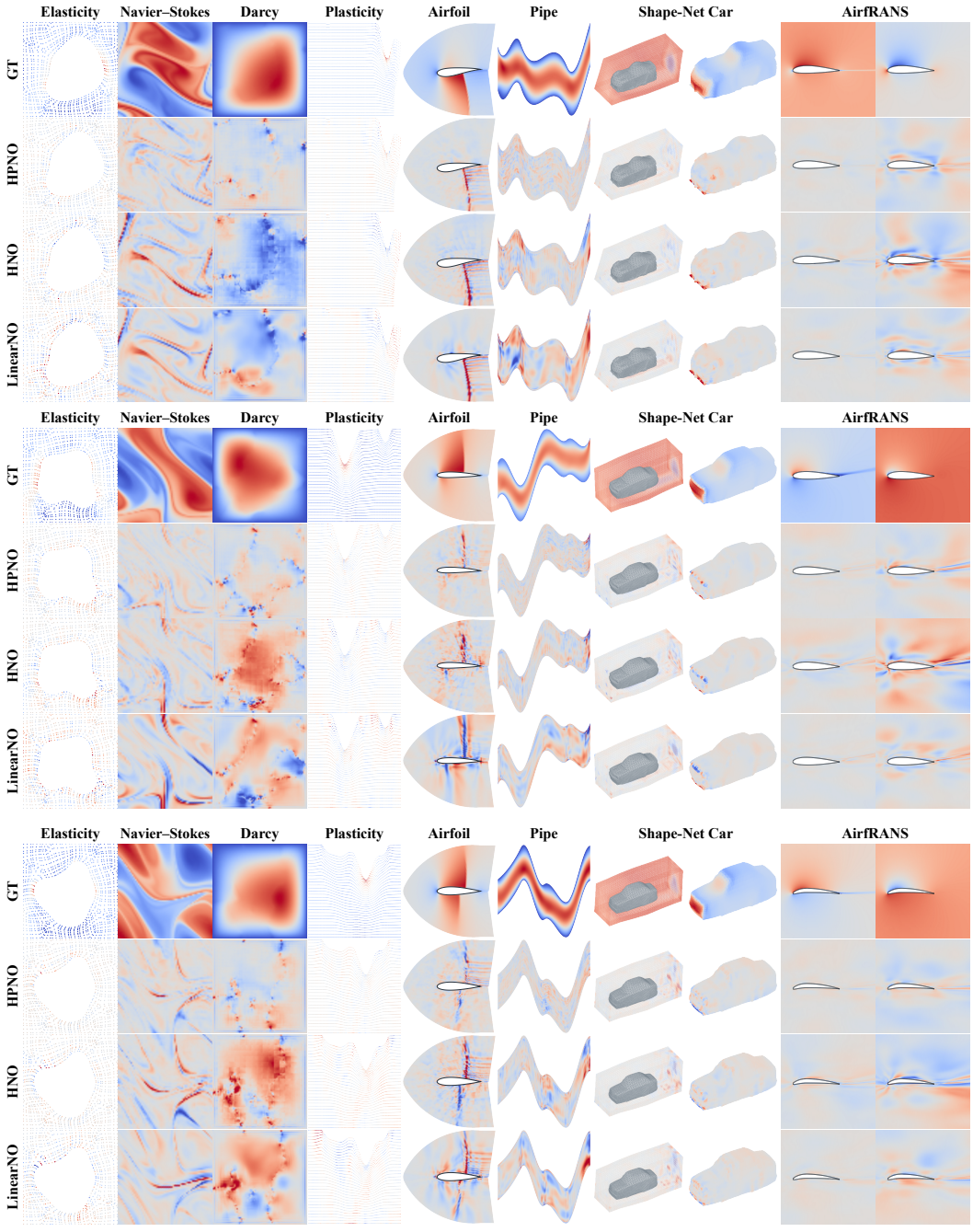

Figure F.1: Ground-truth fields and signed prediction errors (continued).

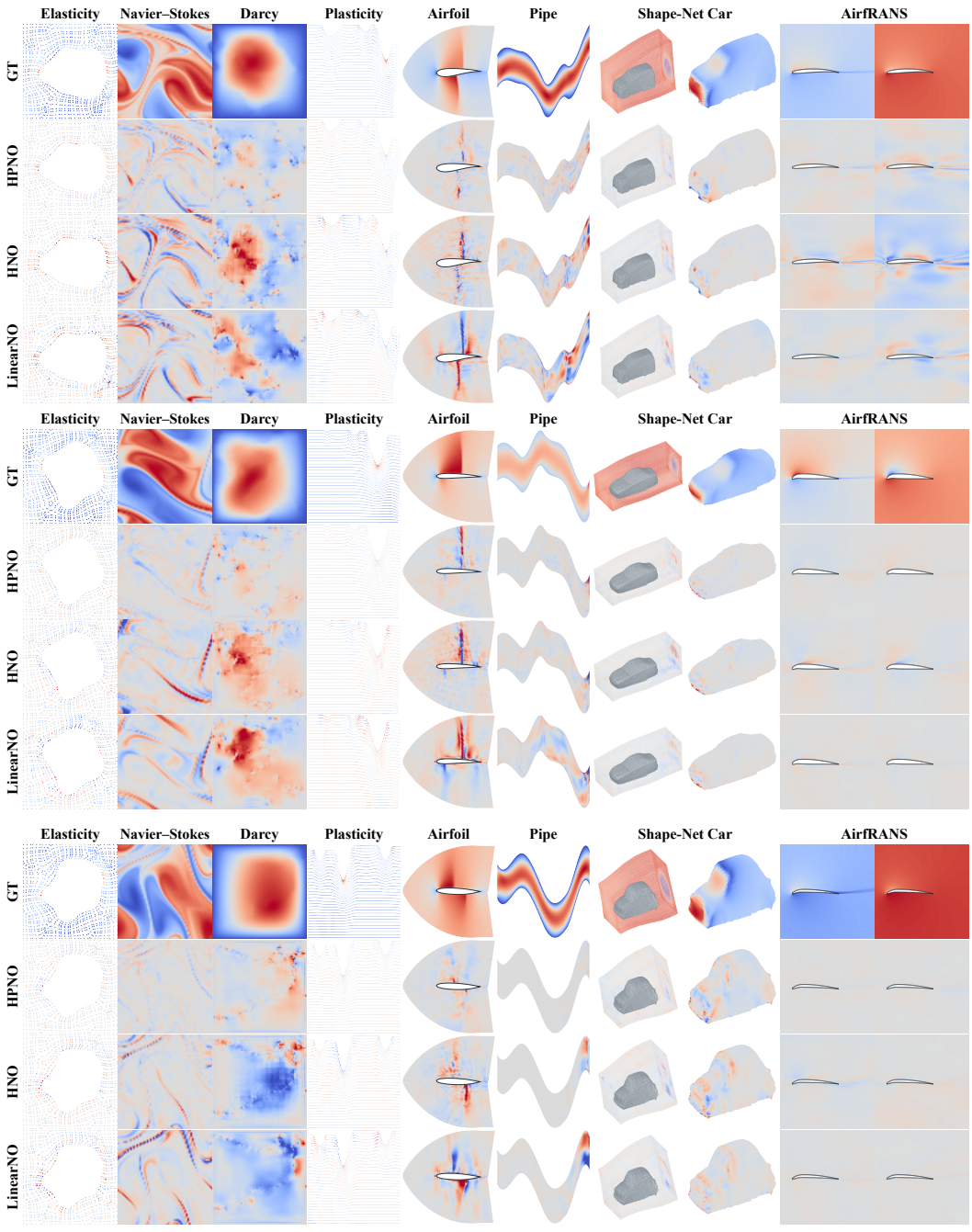

Figure F.1: Ground-truth fields and signed prediction errors (continued).

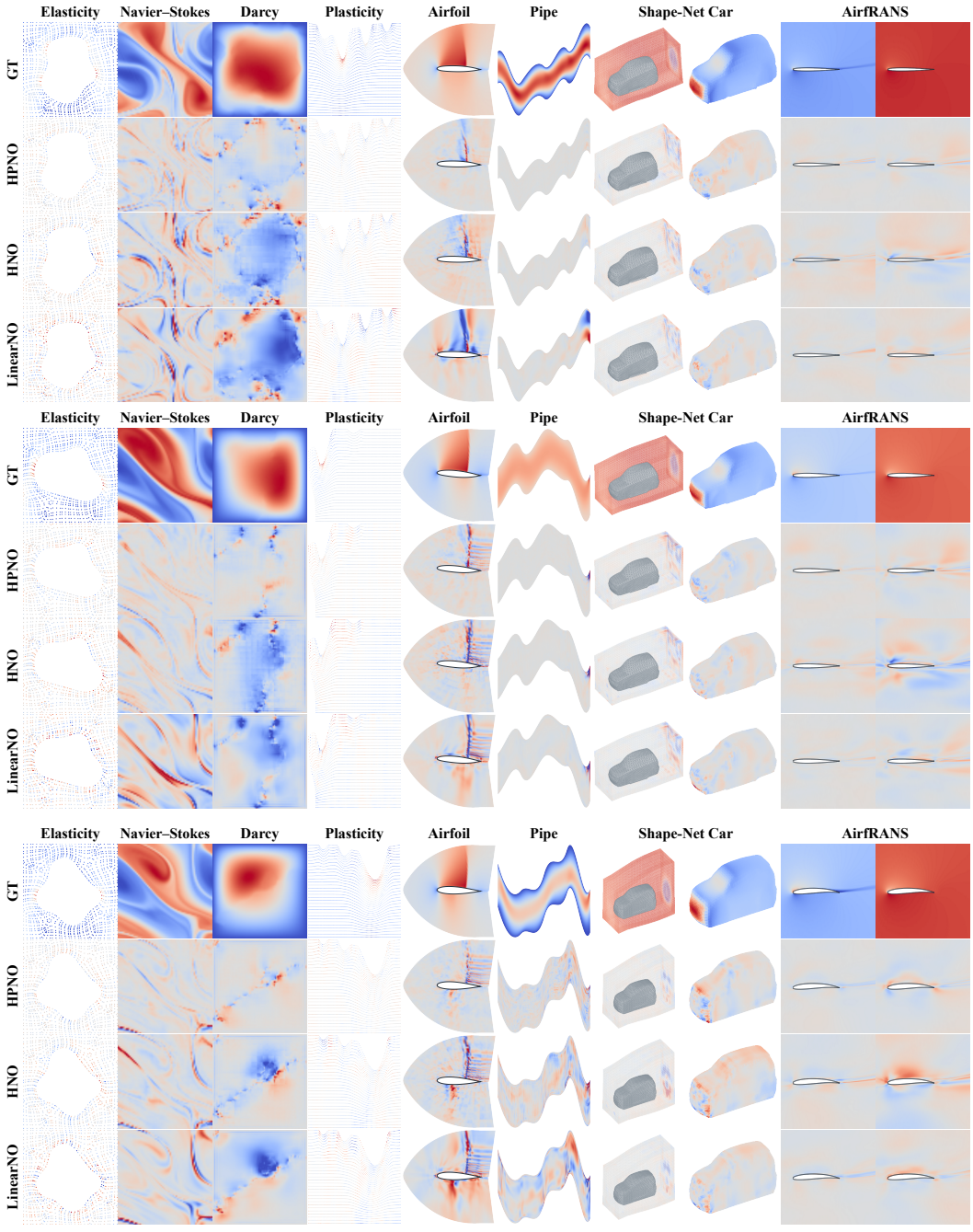

Figure F.1: Ground-truth fields and signed prediction errors (continued).