LoMACS-SVDNet

Orthogonality without Decompositions

Axial Gating + Low-Rank Projected Attention + One-step Ortho Refinement + Spectral Self-Calibration

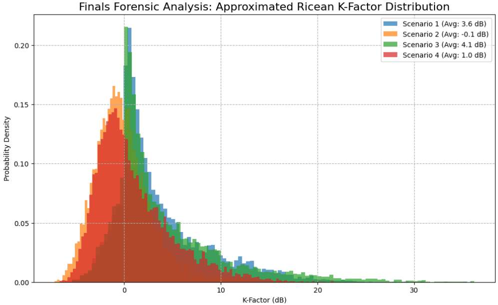
One-slide Overview (What we do)

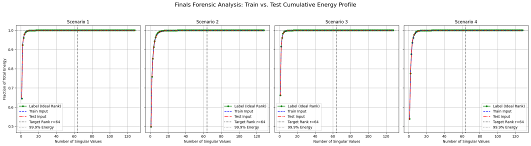
Core Claim

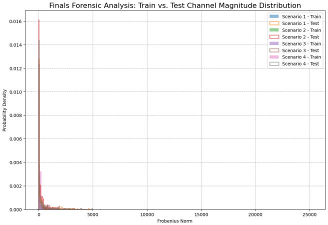
We learn the *diagonalizing transform* itself for large MIMO channels $\mathbf{H}$, producing $(\mathbf{U}, \mathbf{S}, \mathbf{V})$ with near-orthonormal $\mathbf{U}, \mathbf{V}$ and spectrally aligned $\mathbf{S}$, **without** SVD/QR/EVD/inversion inside the network.

- Physics priors: **angle/frequency sparsity**, few dominant paths, near-Hermitian structure; data show effective rank $\approx 2-4$, high column/row correlation.
- Architecture: **Axial Low-rank Freq Gate (ALF)** $\Rightarrow$ **Dual representation** $\Rightarrow$ **Grouped Projected Attention (GPA) + GatedConv** $\Rightarrow$ **Neural Ortho Refiner (NOR)** $\Rightarrow$ **Spectral Self-Calibration (SSC)**.
- Step-3: **Structured pruning** on k_{len} (projection length) and gate hidden size $\Rightarrow \mathrm{MACs} \downarrow$ measurably via PyTorch Profiler.

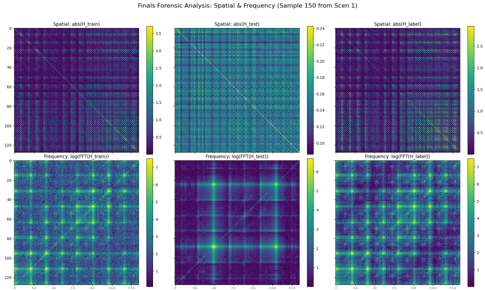
Qualitative Results

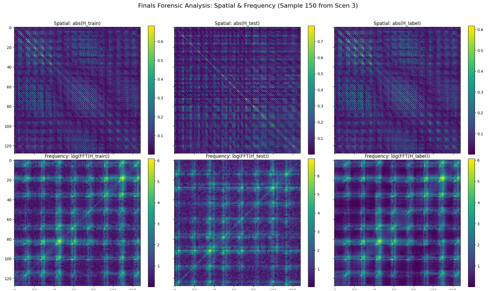


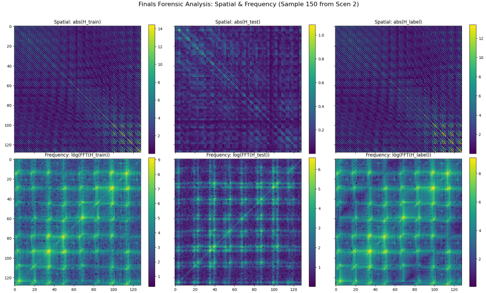


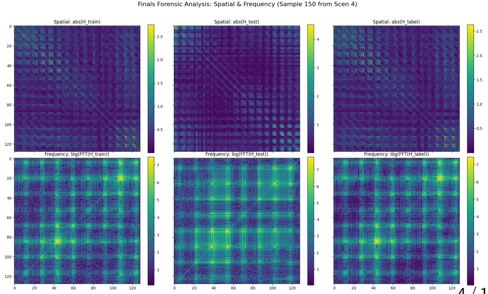


Qualitative Results









Data & Physics Priors ⇒ Architectural Bias

- **Low effective rank**: top-5 singular values capture ≈ all energy; few dominant paths/clusters.
- **Angle-domain sparsity & strong correlation**: FFT Top-1–5% energy high; column/row correlation $\mu \approx 0.98 \sim 0.997$.
- **Near-Hermitian & ill-conditioned**: $\left\|\mathbf{H}\mathbf{H}^{H} - \mathbf{H}^{H}\mathbf{H}\right\|_{\mathrm{F}} / \left\|\mathbf{H}\right\|_{\mathrm{F}}^{2} \approx 0$, condition number $\sim 10^{10}$.

Design implications

⇒ **Front-end gating** to keep informative angles/Delays; **low-rank global modeling** (projected attention) + **local smoothing** (gated conv); **decomposition-free** orthogonality control; explicit **diagonalization & spectral consistency**.

Objective & Train Signal (What we optimize)

Given $\mathbf{H} \in \mathbb{C}^{M \times N}$, predict $\mathbf{U} \in \mathbb{C}^{M \times r}$, $\mathbf{V} \in \mathbb{C}^{N \times r}$, $\mathbf{S} \in \mathbb{R}_{+}^{r}$ such that:

$$\hat{\mathbf{H}} = \mathbf{U}\,\mathrm{diag}(\mathbf{S})\mathbf{V}^{H}, \quad \mathbf{U}^{H}\mathbf{U} \approx \mathbf{I},\ \mathbf{V}^{H}\mathbf{V} \approx \mathbf{I}.$$

AEPlus loss:

$$L = \underbrace{\frac{\left\|\mathbf{H} - \mathbf{U}\,\mathrm{diag}(\mathbf{S})\mathbf{V}^{H}\right\|_{\mathrm{F}}}{\|\mathbf{H}\|_{\mathrm{F}}}}_{\text{Reconstruction } L_{\mathrm{rec}}} + \lambda \underbrace{\left(\left\|\mathbf{U}^{H}\mathbf{U} - \mathbf{I}\right\|_{\mathrm{F}} + \left\|\mathbf{V}^{H}\mathbf{V} - \mathbf{I}\right\|_{\mathrm{F}} \right)}_{\text{Orthogonality } L_{\mathrm{ortho}}}$$

$$+\, w_E \underbrace{\left(1 - \frac{\|\mathbf{M}\|_{\mathrm{F}}}{\|\mathbf{H}\|_{\mathrm{F}}}\right)}_{\text{Energy ratio } L_{\mathrm{energy}}} + w_D \underbrace{\frac{\|\mathbf{M} - \mathrm{diag}(\mathbf{M})\|_{\mathrm{F}}}{\|\mathbf{H}\|_{\mathrm{F}}}}_{\text{Off-diagonal penalty } L_{\mathrm{diag}}}$$

$$+\, w_S \underbrace{\overline{\left| \frac{\mathbf{S}}{\|\mathbf{H}\|_{\mathrm{F}}} - \frac{|\,\mathrm{diag}(\mathbf{M})|}{\|\mathbf{H}\|_{\mathrm{F}}} \right|}}_{\text{Spectral match } L_{\mathrm{smatch}}},$$

where $\mathbf{M} = \mathbf{U}^{H}\mathbf{H}\mathbf{V}$.

Role of Each Loss Term

Term	Purpose / Constraint enforced
L_{rec}	Accuracy anchor: force $(\mathbf{U}, \mathbf{S}, \mathbf{V})$ to reconstruct $\mathbf{H}$ faithfully.
L_{ortho}	Keep $\mathbf{U}, \mathbf{V}$ close to unitary (Stiefel manifold), stabilizes inverse-like operations and aligns with decomposition goal.
L_{energy}	Maximize total captured energy in $M = \mathbf{U}^H \mathbf{H} \mathbf{V}$, discouraging leakage outside the modeled subspace.
L_{diag}	Suppress off-diagonal energy in M, i.e. improve "near-diagonalization" given current $\mathbf{U}, \mathbf{V}$.
L_{smatch}	Align predicted singular spectrum $\mathbf{S}$ with measured diagonal magnitude $\lvert \text{diag}(M) \rvert$, combining geometric and spectral consistency.

Schedules: λ ramps up early (orthogonality first), w_E, w_D, w_S decay smoothly to avoid over-constraining late training.

Step-1: Axial Low-rank Frequency Gate (ALF)

Motivation (math/physics). In the 2D FFT domain, channel magnitude concentrates on few rows/cols (angles/delays). Use a separable gate to *select* informative bands:

$$\text{gate}_{m,n} = \sigma\left(\frac{r_m + c_n}{T}\right), \quad r = \text{MLP}_r\big(\mathrm{mean}_n|\mathbf{H}_f|\big), \;\; c = \text{MLP}_c\big(\mathrm{mean}_m|\mathbf{H}_f|\big).$$

Apply: $\mathbf{H}_f \mapsto \mathbf{H}_f \odot \text{gate}$, then $\mathbf{H}_d = \mathcal{F}^{-1}\{\cdot\}$.

Pain point solved

Front-end **denoise + dimensionality bias**: spend compute only on high-energy angular slices; hidden size is **structurally prunable**.

Dual Representation: Raw & Coarse

- Build two views from $\mathbf{H}_d$: a raw map and a lightly down/up-sampled coarse map; concatenate Re / Im to form features before projection.
- **Why**: statistics show strong alignment ($\rho \approx 0.99$) and stable amplitude/phase bias; combining fine details + global contour increases robustness to timing/estimation errors.

Global–Local Backbone: GPA + GatedConv

Grouped Projected Attention (GPA). Split channels into groups (dim = gdim). For each group,

$$Q = XW_q,\ K = XW_k,\ V = XW_v; \quad K_{\text{red}} = K^\top P_k,\ V_{\text{red}} = V^\top P_v,$$

with $P_\bullet \in \mathbb{R}^{T \times k_{\text{len}}}$. Complexity reduces $O(T^2) \to O(Tk_{\text{len}})$.
Why it fits data. High column/row correlation $\Rightarrow$ few *global* bases suffice ($k_{\text{len}} \ll T$).

GatedConv1D (DW $\to$ sigmoid gate $\to$ PW). Provides *local* smoothing and selective nonlinearity to fix rough regions (weaker sparsity scenes).

Pain points solved

Controllable compute; interpretable *column-dictionary* that is **prunable** by column energy; local robustness via low-MACs convolution.

Step-2: Neural Ortho Refiner (NOR) — no QR/SVD

Goal. Reduce orthogonality error without banned operators.
One-step refinement (complex case):

$$\mathbf{U}' = \mathbf{U} - a\,\mathbf{U}\ \mathrm{sym}(\mathbf{U}^H\mathbf{U} - \mathbf{I}), \qquad \mathbf{V}' = \mathbf{V} - b\,\mathbf{V}\ \mathrm{sym}(\mathbf{V}^H\mathbf{V} - \mathbf{I}),$$

where $\mathrm{sym}(\mathbf{G}) = \frac{1}{2}(\mathbf{G} + \mathbf{G}^H)$, $a, b \in (0, 0.5]$ learned through a sigmoid; then *column normalization* (allowed).

Why it works

$f(\mathbf{U}) = \left\|\mathbf{U}^H\mathbf{U} - \mathbf{I}\right\|_{\mathrm{F}}^2$ decreases for small steps (first-order descent on the Stiefel manifold with simple retraction via column norm). **No** SVD/QR/GS/Householder/Givens/inversion.

Step-2: Spectral Self-Calibration (SSC)

Align singular spectrum to current geometry. Let $\mathbf{M} = \mathbf{U}^H \mathbf{H} \mathbf{V}$, $S_{\text{best}} = |\operatorname{diag}(\mathbf{M})|$.

$$\mathbf{S} = (1-\tau)\,\mathbf{S}_{\text{pred}} + \tau\, S_{\text{best}}, \qquad \tau:\ 0.90 \downarrow 0.60.$$

Pain point solved

When labels/noise mismatch exists, $\mathbf{S}$ is pulled toward the *measured* diagonal energy under current $\mathbf{U}, \mathbf{V}$. Early training trusts the head ($\mathbf{S}_{\text{pred}}$), later favors physical consistency.

Normalization & Robust Augmentation

- **Per-sample Frobenius normalization**: train/test consistent; during inference, **S** is rescaled by sample norm (orthogonality preserved).
- **Noise injection & antenna dropout**: power-aware complex noise, random row/column drop; matches lower SNR and larger row/col CV in harder scenes.

Step-3: Structured Pruning & Finetune (Real MACs $\downarrow$)

Two structural knobs.

1. $\boldsymbol{k}_{\text{len}}$: keep top columns by energy score $\|P_k\|_2^2 + \alpha\|P_v\|_2^2$.
2. **Gate hidden size**: keep channels by coupled in/out norm score across row/col MLPs.

Implementation detail

We *rebuild a smaller isomorphic network* and copy compatible weights (*no* sparse masks). Graph shrinks $\Rightarrow$ PyTorch Profiler MACs drop *for real*. Short finetuning recovers AE.

Module $\leftrightarrow$ Data Alignment

Pain Point	**Module Hook**
Low effective rank	ALF gating; GPA with small k_{len}
Angle sparsity & high correlation	Axial gate; column-dictionary attention
Ill-conditioning, near-Hermitian	Decomposition-free NOR; SSC
Non-diagonal residual energy	$L_{\text{energy}}, L_{\text{diag}}$
Uneven SNR / row/col CV	GatedConv; noise & antenna dropout
Compute budget (Score $= 100 \cdot \text{AE} + \text{MACs}$)	Structured pruning on k_{len}, gate hidden