# CHASM: Cross-frequency Harmonized Axis-Separable Mixing for Undersampled Imaging

**Pengcheng Fang**[1] **Hongli Chen**[2] **Yuxia Chen**[3]
**Tengjiao Sun**[1] **Jiaxin Liu**[4] **Xiaohao Cai**[1]

[1]University of Southampton, Southampton, U.K.
[2]The University of Queensland, Brisbane, Australia
[3]University of Electronic Science and Technology of China, Chengdu, China
[4]University of Illinois Urbana-Champaign, Urbana, IL, U.S.A.
p.fang@soton.ac.uk, hongli.chen@uq.edu.au
yuxia0313@std.uestc.edu.cn, Tengjiao.Sun@soton.ac.uk
jiaxin26@illinois.edu, x.cai@soton.ac.uk

## Abstract

Spectral token mixers based on Fourier transforms provide an efficient way to model global interactions in visual feature maps. Existing designs often either apply filter-wise spectral responses along fixed channel axes, or learn adaptive frequency-indexed channel mixing without explicitly aligning the channel directions used across frequencies. We propose CHASM, a Cross-frequency Harmonized Axis-Separable Mixer, as a structured middle ground. CHASM separates what should be shared from what should remain frequency-specific: all frequencies share a learned channel eigenbasis, while each frequency retains its own positive spectral gains. The shared basis makes channel directions comparable across the spectrum, whereas the positive gains preserve local spectral adaptivity. CHASM applies this structured operator separably along the height and width axes and is used as a drop-in replacement mixer inside existing backbones. We provide a structural characterization of the shared-basis operator family and evaluate CHASM through controlled same-backbone comparisons. Across accelerated MRI reconstruction, undersampled MRI segmentation, and natural-image reconstruction, CHASM consistently improves over same-backbone spectral-mixer baselines. Ablations show that removing the shared-basis constraint weakens performance, and randomizing coherent sampling geometry substantially reduces the gain, supporting cross-frequency harmonization as a useful inductive bias for undersampled imaging.

## 1 Introduction

Spectral token mixers have become a useful mechanism for modeling long-range interactions in visual feature maps. By transforming features into a frequency domain, applying learnable operations, and mapping the result back to the spatial domain, they expose each feature location to global context with a compact operator [Li et al., 2021, Guibas et al., 2022, Rao et al., 2021, Chi et al., 2020]. Such designs are especially attractive in inverse imaging and medical-image analysis, where undersampling, aliasing, and recovery often have explicit spectral structure [Lustig et al., 2007, Zbontar et al., 2018].

Most spectral mixers learn frequency-indexed responses: after a spatial Fourier transform, a frequency-dependent operation is applied at each retained Fourier bin. Existing designs often emphasize one of two forms of flexibility. Filter-wise methods give different frequencies different responses but largely operate along fixed channel axes. More adaptive spectral mixers allow frequency-indexed channel

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

transformations, but then each frequency may learn its own channel directions without an explicit mechanism for aligning those directions across the spectrum. What is missing is a structured middle ground: a mixer should learn useful channel directions, make them comparable across frequencies, and still allow each frequency to respond differently.

We propose CHASM, a Cross-frequency Harmonized Axis-Separable Mixer, to realize this middle ground. CHASM parameterizes each frequency-indexed channel operator with a shared learned eigenbasis and a frequency-specific positive gain vector. The shared basis defines a common channel coordinate system across the spectrum, whereas the gains preserve frequency-specific spectral adaptivity. In other words, CHASM harmonizes channel directions without forcing all frequencies to share identical responses.

CHASM is designed as a drop-in spectral mixer for controlled same-backbone evaluation. This modular setting isolates the effect of cross-frequency harmonization from changes to the surrounding architecture. We also provide a structural characterization of the direct shared-basis operator family, clarifying the local degree-of-freedom structure of the harmonized spectral operator.

We evaluate CHASM on accelerated MRI reconstruction, undersampled MRI segmentation, and natural-image reconstruction. Across CC359, fastMRI, BraTS, ADE20K, and ImageNet settings, CHASM improves over same-backbone spectral-mixer baselines when inserted into HiFi-Mamba-family, U-Net-, and ViT-based backbones [Chen et al., 2025, Fang et al., 2025, Ronneberger et al., 2015, Dosovitskiy et al., 2021]. Mechanistic ablations further support the proposed design: releasing the shared-basis constraint weakens performance, and disrupting coherent sampling geometry with random masks substantially reduces the gain.

**Contributions.** Our main contributions are threefold.

- First, we propose CHASM, an axis-separable spectral mixer that shares learned channel directions across frequencies while retaining frequency-specific positive gains.
- Second, we characterize the local degrees of freedom of the direct shared-basis operator family, connecting its parameterization to the common-principal-component (CPC) structure [Flury, 1984].
- Third, we validate CHASM through same-backbone comparisons across reconstruction and segmentation benchmarks, with ablations of basis sharing, orthogonality, positive gains, and sampling geometry.

## 2 Related Work

**FFT-based spectral token mixing.** Fourier-domain operations provide global mixing beyond local spatial neighborhoods. Fourier Neural Operators parameterize global convolutional operators in frequency space [Li et al., 2021], while Fast Fourier Convolution introduces spectral branches into convolutional networks [Chi et al., 2020]. In vision architectures, Global Filter Networks learn frequency-domain filters for long-range spatial dependencies [Rao et al., 2021], and Adaptive Fourier Neural Operators introduce adaptive spectral mixing over selected modes [Guibas et al., 2022]. Factorized Fourier neural operators (F-FNO) use separable spectral layers [Tran et al., 2023], while tensorized Fourier neural operators (TFNO) share factors in a spectral weight tensor [Kossaifi et al., 2024]. We compare both under the same reconstruction protocol in Appendix A.5. CHASM coordinates the frequency-indexed channel operators through a shared orthogonal eigenbasis and frequency-specific positive gains.

**Structured token mixers and replacement operators.** Recent architectures increasingly separate the token mixer from the surrounding network design. MLP-Mixer shows that token mixing can be implemented without convolution or attention [Tolstikhin et al., 2021], and MetaFormer abstracts a common block structure in which different token mixers can be exchanged [Yu et al., 2022]. Following this view, CHASM is evaluated as a drop-in mixer block inside HiFi-Mamba-family, U-Net, and ViT backbones [Chen et al., 2025, Fang et al., 2025, Ronneberger et al., 2015, Dosovitskiy et al., 2021].

**Spectral structure in MRI reconstruction.** Accelerated MRI reconstruction is naturally tied to frequency-domain structure: measurements are acquired in undersampled $k$-space, and reconstruction must recover missing spectral information. Classical compressed-sensing MRI exploits sparsity

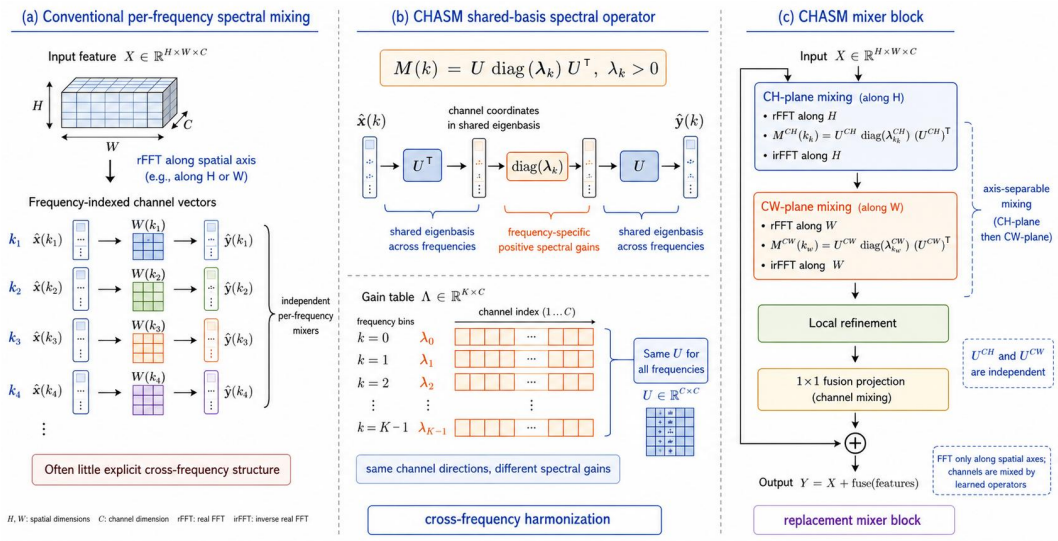


Figure 1: Overview of CHASM. (a) Standard spectral mixers often learn independent channel operators for different Fourier bins, with little explicit cross-frequency structure. (b) CHASM introduces cross-frequency harmonization by sharing a learned channel eigenbasis $U$ across frequencies while retaining frequency-specific positive gains $\lambda_k$. (c) The resulting operator is applied separably along the height–channel and width–channel planes, followed by local refinement and $1 \times 1$ fusion, yielding a drop-in replacement mixer block.

priors for rapid imaging [Lustig et al., 2007], and fastMRI provides large-scale benchmarks for learning-based reconstruction [Zbontar et al., 2018]. Recent MRI backbones further emphasize frequency-aware decomposition and high-fidelity detail recovery [Chen et al., 2025, Fang et al., 2025]. CHASM is orthogonal to these backbone-level designs, targeting the spectral mixer itself through shared-basis cross-frequency harmonization.

# 3 Method

## 3.1 Real-valued representation and axis-wise spectral mixing

Let $\mathbf{X} \in \mathbb{R}^{H\times W\times C}$ denote a feature tensor with height $H$, width $W$, and channel dimension $C$. CHASM operates on real-valued network features. For complex-valued MRI inputs, real and imaginary components are stacked along the channel dimension before being processed by the network; thus $C$ denotes the resulting real feature-channel dimension.

CHASM uses one-dimensional real Fourier transforms along spatial axes. For the height–channel plane, abbreviated as the CH-plane, we write

$$\widehat{\mathbf{X}}^{\mathrm{CH}}[k_h, w, :] = \mathcal{F}_H\big(\mathbf{X}[:, w, :]\big)[k_h], \qquad k_h = 0, \dots, K_H - 1, \tag{1}$$

where $\mathcal{F}_H$ denotes the real FFT along the height axis and $K_H = \lfloor H/2 \rfloor + 1$ is the number of retained frequency bins. Analogously, the width–channel plane, or CW-plane, gives $\widehat{\mathbf{X}}^{\mathrm{CW}}[h, k_w, :]$ with $K_W = \lfloor W/2 \rfloor + 1$.

At each retained spatial frequency, the mixer applies a channel operator to the corresponding $C$-dimensional Fourier coefficient:

$$\widehat{\mathbf{Z}}^{\mathrm{CH}}[k_h, w, :] = \mathbf{M}^{\mathrm{CH}}(k_h)\widehat{\mathbf{X}}^{\mathrm{CH}}[k_h, w, :]. \tag{2}$$

Although the Fourier coefficient is complex-valued, $\mathbf{M}^{\mathrm{CH}}(k_h) \in \mathbb{R}^{C\times C}$ is a real channel-mixing matrix and is applied by complex linear extension. Equivalently, the same real matrix acts on the real and imaginary parts of each Fourier coefficient. This preserves the conjugate symmetry required by the inverse real Fourier transform. In particular, the DC bin and, when present, the Nyquist bin remain real-valued because they are multiplied by real matrices. No Fourier transform is taken along the channel dimension; channels are mixed by learned linear operators indexed by spatial frequency.

### 3.2 Cross-frequency harmonized spectral operator

The central design question is how the channel operators $\{M(k)\}$ should vary across frequencies. A frequency-indexed channel operator determines two aspects of the spectral response: the channel directions being mixed and the gain applied along those directions. Filter-wise designs keep the channel axes fixed, whereas fully independent operators allow both directions and gains to vary with frequency. CHASM separates these roles: channel directions are shared across frequencies, while spectral gains remain frequency-specific.

For the CH-plane, the harmonized spectral operator is

$$\mathbf{M}^{\mathrm{CH}}(k_h) = \mathbf{U}^{\mathrm{CH}} \operatorname{diag}\big(\boldsymbol{\lambda}^{\mathrm{CH}}_{k_h}\big)\big(\mathbf{U}^{\mathrm{CH}}\big)^\top, \tag{3}$$

where $\mathbf{U}^{\mathrm{CH}} \in \mathbb{R}^{C\times C}$ is a learned orthogonal basis shared across all height frequencies, and $\boldsymbol{\lambda}^{\mathrm{CH}}_{k_h} \in \mathbb{R}^C_{>0}$ is a frequency-specific positive gain vector. The CW-plane uses an independent basis and response table:

$$\mathbf{M}^{\mathrm{CW}}(k_w) = \mathbf{U}^{\mathrm{CW}} \operatorname{diag}\big(\boldsymbol{\lambda}^{\mathrm{CW}}_{k_w}\big)\big(\mathbf{U}^{\mathrm{CW}}\big)^\top. \tag{4}$$

Thus, $\mathbf{U}^{\mathrm{CH}}$ and $\mathbf{U}^{\mathrm{CW}}$ are separate parameters. The two axes share the same structural principle, but not the same learned channel basis.

For each axis $a \in \{\mathrm{CH}, \mathrm{CW}\}$, the orthogonal basis is parameterized with exactly $C(C-1)/2$ trainable scalars. Let $\boldsymbol{\theta}^a \in \mathbb{R}^{C(C-1)/2}$ denote the parameters assigned to the strictly lower triangular entries of a skew-symmetric matrix $\mathbf{A}^a$. We construct

$$A^a_{ij} = \begin{cases} \theta^a_{ij}, & i > j, \\ -\theta^a_{ji}, & i < j, \\ 0, & i = j, \end{cases} \qquad \mathbf{U}^a = \exp(\mathbf{A}^a). \tag{5}$$

The matrix exponential maps skew-symmetric matrices to $\mathrm{SO}(C)$, so $\mathrm{cond}_2(\mathbf{U}^a) = 1$ in exact arithmetic. The operator family is naturally described using orthogonal bases in $\mathrm{O}(C)$, but using an $\mathrm{SO}(C)$ representative does not reduce expressivity for the symmetric operators considered here: if an eigenbasis has determinant $-1$, flipping the sign of one eigenvector gives a determinant-1 basis and leaves $\mathbf{U}\operatorname{diag}(\boldsymbol{\lambda})\mathbf{U}^\top$ unchanged. The exponential coordinates are not globally one-to-one, but they provide the correct local degrees of freedom for the shared-basis operator family. We initialize the skew-symmetric parameters to zero, so the learned basis starts from the identity matrix.

The frequency-specific gains are represented by learnable tables. For axis $a$, we learn $\boldsymbol{\Gamma}^a \in \mathbb{R}^{B_a\times C}$, where $B_a$ is the number of learned frequency bins. At a given input resolution, the table is linearly interpolated along the normalized rFFT frequency coordinate to the retained frequency length $K_a$, and then mapped to positive gains:

$$\boldsymbol{\lambda}^a_k = \operatorname{softplus}\left(\operatorname{Interp}\big(\boldsymbol{\Gamma}^a\big)_{k,:}\right), \qquad k = 0, \ldots, K_a - 1. \tag{6}$$

When $B_a = K_a$, this reduces to a direct per-frequency gain table. The response tables are initialized so that the spectral core starts near an identity gain. We do not apply per-channel mean normalization.

For a complex Fourier coefficient $\mathbf{v} \in \mathbb{C}^C$, the operator is applied as

$$\mathbf{M}^a(k)\mathbf{v} = \mathbf{U}^a\left(\boldsymbol{\lambda}^a_k \odot \big((\mathbf{U}^a)^\top \mathbf{v}\big)\right), \tag{7}$$

where $\odot$ denotes elementwise multiplication. Since $\mathbf{U}^a$ and $\boldsymbol{\lambda}^a_k$ are real, the same operation is applied consistently to the real and imaginary parts. At the level of the spectral core, each retained-frequency channel operator is real symmetric with a positive spectrum.

### 3.3 CHASM mixer block

The proposed operator is used as the spectral core of a complete mixer block. CHASM applies the harmonized operator separably along the two spatial axes, factorizing spectral mixing into two one-dimensional frequency-indexed channel operations over the CH and CW planes. This reduces the parameterization required for frequency-indexed channel mixing while still enabling global interaction along both spatial axes through composition.

Given $\mathbf{X} \in \mathbb{R}^{H\times W\times C}$, the CH-plane pass first computes an rFFT along the height axis, applies (3) at each height frequency, and returns to the spatial domain:

$$\widehat{\mathbf{X}}^{\mathrm{CH}}[k_h, w, :] = \mathcal{F}_H\big(\mathbf{X}[:, w, :]\big)[k_h], \tag{8}$$

$$\widehat{\mathbf{Z}}^{\mathrm{CH}}[k_h, w, :] = \mathbf{M}^{\mathrm{CH}}(k_h)\widehat{\mathbf{X}}^{\mathrm{CH}}[k_h, w, :], \tag{9}$$

$$\mathbf{Z}[:, w, :] = \mathcal{F}_H^{-1}\big(\widehat{\mathbf{Z}}^{\mathrm{CH}}[:, w, :]\big). \tag{10}$$

The CW-plane pass then applies the analogous operation to $\mathbf{Z}$:

$$\widehat{\mathbf{Z}}^{\mathrm{CW}}[h, k_w, :] = \mathcal{F}_W\big(\mathbf{Z}[h, :, :]\big)[k_w], \tag{11}$$

$$\widehat{\mathbf{T}}^{\mathrm{CW}}[h, k_w, :] = \mathbf{M}^{\mathrm{CW}}(k_w)\widehat{\mathbf{Z}}^{\mathrm{CW}}[h, k_w, :], \tag{12}$$

$$\mathbf{T}[h, :, :] = \mathcal{F}_W^{-1}\big(\widehat{\mathbf{T}}^{\mathrm{CW}}[h, :, :]\big). \tag{13}$$

Thus the default CHASM spectral core is

$$\Phi_{\mathrm{CHASM}} = \Phi_{\mathrm{CW}} \circ \Phi_{\mathrm{CH}}, \tag{14}$$

where the height-axis harmonized mixing is applied before the width-axis harmonized mixing.

The full CHASM mixer block wraps this spectral core with lightweight local refinement and channel fusion:

$$\mathbf{Y} = \mathbf{X} + \mathbf{P}_{\mathrm{fuse}}\left(\mathcal{L}\big(\Phi_{\mathrm{CHASM}}(\mathbf{X})\big)\right), \tag{15}$$

where $\mathcal{L}$ is a lightweight local convolutional refinement and $\mathbf{P}_{\mathrm{fuse}}$ is a $1 \times 1$ fusion projection. The proposed cross-frequency harmonization is contained in $\Phi_{\mathrm{CHASM}}$; the local refinement and fusion layers make the operator a practical drop-in mixer block for existing vision and reconstruction backbones.

The two axes use independent bases and frequency responses. The second pass acts on features transformed by the first, allowing the two spectral operations to interact through the channel space.

### 3.4 Structural characterization of the shared-basis operator

We characterize the direct-table spectral operator for one axis of CHASM, with a gain vector at each retained frequency. Its simultaneous diagonalization has the common-principal-component (CPC) structure of Flury [1984]. The dimension count below specializes this structure to frequency-indexed channel operators. For $K$ frequency bins and channel dimension $C$, define the positive-gain shared-basis family

$$\mathcal{S}_{K,+}^{C} = \left\{\big(\mathbf{M}(0), \ldots, \mathbf{M}(K-1)\big) : \mathbf{M}(k) = \mathbf{U}\,\mathrm{diag}(\boldsymbol{\lambda}_k)\mathbf{U}^{\top},\ \mathbf{U} \in \mathrm{O}(C),\ \boldsymbol{\lambda}_k \in \mathbb{R}_{>0}^{C}\right\}. \tag{16}$$

This is the family of real symmetric channel operators with positive spectra that are simultaneously diagonalizable by a shared orthogonal basis.

For eigenmode $j$, define its joint spectral signature as

$$\boldsymbol{s}_j = \big(\lambda_{0j}, \lambda_{1j}, \ldots, \lambda_{K-1,j}\big) \in \mathbb{R}_{>0}^{K}. \tag{17}$$

The generic stratum of $\mathcal{S}_{K,+}^{C}$ is the subset where $\boldsymbol{s}_i \neq \boldsymbol{s}_j$ for all $i \neq j$. On this stratum, the joint eigenspaces are one-dimensional.

**Proposition 1** (Local degrees of freedom of the direct shared-basis operator)**.** *On the generic stratum defined by pairwise distinct joint spectral signatures, the direct-table family $\mathcal{S}_{K,+}^{C}$ is a smooth manifold of intrinsic dimension*

$$\frac{C(C-1)}{2} + KC. \tag{18}$$

*The direct shared-basis spectral operator has the same number of local degrees of freedom: $C(C-1)/2$ for the basis and $KC$ for the frequency-specific gain table. Thus, the direct shared-basis parameterization matches the intrinsic local dimension of this generic structured family.*

*Proof.* The orthogonal group $\mathrm{O}(C)$ has dimension $C(C-1)/2$, and the positive gain table $\boldsymbol{\Lambda} \in \mathbb{R}_{>0}^{K\times C}$ contributes $KC$ additional coordinates. Since $\mathbb{R}_{>0}^{K\times C}$ is an open subset of $\mathbb{R}^{K\times C}$, the positivity constraint does not change the local dimension. On the generic stratum, the joint spectral signatures

are pairwise distinct, so the shared eigenspaces are one-dimensional. The remaining ambiguity is discrete: columns of $\mathbf{U}$ may be sign-flipped, and eigenmode columns may be permuted together with the corresponding columns of $\mathbf{\Lambda}$. Quotienting by this finite signed-permutation stabilizer does not change the manifold dimension. Thus the intrinsic dimension of the generic structured family is $C(C-1)/2 + KC$. This verifies that the direct shared-basis parameterization matches the intrinsic local dimension of the corresponding generic structured family. □

The dimension count separates coordination from adaptivity: the shared basis provides common channel directions across frequencies, while the gain table preserves frequency-specific responses along those directions.

The practical implementation uses a learned response table with $B$ frequency bins and interpolates it to the $K$ retained rFFT bins of the input. When $B < K$, this gives a smooth subparameterization of the direct-table family used in Proposition 1.

**Parameter-level frequency reindexing.** The shared-basis form has a simple parameter-level response to frequency-index permutations. Let $\sigma$ be a permutation of $\{0, \dots, K-1\}$ and define

$$(\sigma \cdot \mathbf{\Lambda})_{k,:} = \mathbf{\Lambda}_{\sigma^{-1}(k),:}. \tag{19}$$

Then

$$\mathbf{U}\,\mathrm{diag}\big((\sigma \cdot \mathbf{\Lambda})_{k,:}\big)\mathbf{U}^\top = \mathbf{M}(\sigma^{-1}(k)). \tag{20}$$

Thus, reindexing frequencies only permutes rows of the gain table and leaves the basis component unchanged.

### 3.5 Positive spectral gains

The positive shared-basis form is also compatible with the operator structure encountered in inverse problems. For a linear measurement operator $\mathcal{A}$, the normal operator $\mathcal{A}^*\mathcal{A}$ is Hermitian positive semidefinite. In single-coil Cartesian MRI, for example, $\mathcal{A} = \mathbf{P}_\Omega\mathbf{F}$ gives

$$\mathcal{A}^*\mathcal{A} = \mathbf{F}^*\mathbf{P}_\Omega^*\mathbf{P}_\Omega\mathbf{F}, \tag{21}$$

where $\mathbf{F}$ is the Fourier transform and $\mathbf{P}_\Omega$ is the sampling mask. In multi-coil MRI with coil sensitivities $\{\mathbf{S}_c\}$, the normal operator has the form

$$\mathcal{A}^*\mathcal{A} = \sum_c \mathbf{S}_c^*\mathbf{F}^*\mathbf{P}_\Omega^*\mathbf{P}_\Omega\mathbf{F}\mathbf{S}_c. \tag{22}$$

The positive-semidefinite structure of these normal operators motivates positive gains in CHASM's learned feature space. At each retained frequency, the channel operator $\mathbf{U}\,\mathrm{diag}(\boldsymbol{\lambda})\mathbf{U}^\top$ is Hermitian under complex linear extension, with a positive spectrum when $\boldsymbol{\lambda} > 0$. The shared eigenbasis supplies common channel directions across frequencies.

### 3.6 Parameters and computational cost

For one axis with $K$ retained frequency bins, the direct shared-basis spectral operator has

$$C(C-1)/2 + KC \tag{23}$$

trainable spectral-operator parameters. In the interpolated implementation, the learned gain table has $B$ frequency bins and is interpolated to $K$ retained bins, giving

$$C(C-1)/2 + BC \tag{24}$$

trainable spectral-operator parameters for that axis. For the two-axis CHASM spectral core, this becomes

$$C(C-1) + (B_H + B_W)C. \tag{25}$$

The full CHASM mixer block additionally includes the local refinement and fusion projection in (15).

For the CH-plane pass, the FFT and inverse FFT cost $\mathcal{O}(HWC\log H)$, and the channel mixing costs $\mathcal{O}(WK_HC^2)$ when implemented as $\mathbf{U}\,\mathrm{diag}(\boldsymbol{\lambda}_k)\mathbf{U}^\top$. For the CW-plane pass, the corresponding costs are $\mathcal{O}(HWC\log W)$ and $\mathcal{O}(HK_WC^2)$. The total spectral core cost is therefore

$$\mathcal{O}\big(HWC(\log H + \log W) + (WK_H + HK_W)C^2\big), \tag{26}$$

plus the cost of constructing the two orthogonal bases. The matrix exponentials add an $\mathcal{O}(C^3)$ basis-construction cost per forward pass. This basis-construction cost is independent of the number of spatial positions and is amortized over all frequency bins and locations. At inference, the learned bases are input-independent and can be computed once and cached. Appendix A.10 reports isolated-core latencies with and without this cache.

# 4 Experiments

We evaluate CHASM under a controlled mixer-replacement protocol across cross-backbone MRI reconstruction, cross-domain transfer, and mechanistic ablations. All methods replace the same mixer location inside a fixed backbone, and each baseline is evaluated as its native complete mixer block with intrinsic components retained. The ablations isolate the effects of the shared-basis constraint and coherent sampling geometry by keeping the CHASM wrapper and training protocol fixed while changing only the spectral-core constraint or the sampling mask. Additional stability results, ablation definitions, computational cost, reproducibility details, cross-domain protocols, and baseline taxonomy are provided in Appendix A.

**Protocol and metrics.** We evaluate HiFi-Mamba-family, U-Net-based, and ViT-based backbones. For reconstruction, we report PSNR and SSIM; for segmentation, we report mIoU and Dice. Main reconstruction tables report case-level variation; the segmentation standard deviations in Table 2 are seed-level. Further seed stability and per-case segmentation statistics are reported in Appendices A.1 and A.2. CHASM uses $B_H = B_W = 96$ gain bins throughout; the gain-resolution sweep is reported in Appendix A.7.

Table 1: CC359 accelerated MRI reconstruction under fourfold and eightfold undersampling. All methods replace the same mixer block. Params counts the complete replacement block at channel dimension $C = 96$, including its intrinsic refinement and projection layers; model-level FLOPs are provided in Appendix A.9. Values are reported as mean with case-level standard deviation shown in smaller text.

| AR | Method | Params | HiFi-Mamba | | U-Net | | ViT | |
|---|---|---|---|---|---|---|---|---|
| | | | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
| ×8 | Baseline | 0.00M | $26.82_{\pm 0.72}$ | $0.791_{\pm 0.015}$ | $24.79_{\pm 0.57}$ | $0.737_{\pm 0.014}$ | $24.16_{\pm 0.55}$ | $0.722_{\pm 0.014}$ |
| | GFNet | 0.80M | $28.13_{\pm 0.75}$ | $0.826_{\pm 0.015}$ | $25.38_{\pm 0.60}$ | $0.751_{\pm 0.014}$ | $24.70_{\pm 0.57}$ | $0.733_{\pm 0.014}$ |
| | AFNO | 0.01M | $27.30_{\pm 0.73}$ | $0.805_{\pm 0.015}$ | $25.28_{\pm 0.59}$ | $0.748_{\pm 0.014}$ | $24.66_{\pm 0.57}$ | $0.732_{\pm 0.014}$ |
| | FFC | 1.19M | $27.32_{\pm 0.73}$ | $0.805_{\pm 0.015}$ | $25.13_{\pm 0.59}$ | $0.744_{\pm 0.014}$ | $24.50_{\pm 0.56}$ | $0.729_{\pm 0.014}$ |
| | FNO-D | 4.74M | $26.98_{\pm 0.72}$ | $0.796_{\pm 0.015}$ | $25.05_{\pm 0.59}$ | $0.742_{\pm 0.014}$ | $24.62_{\pm 0.57}$ | $0.731_{\pm 0.014}$ |
| | DiffFNO | 4.77M | $27.92_{\pm 0.74}$ | $0.819_{\pm 0.015}$ | $25.31_{\pm 0.60}$ | $0.749_{\pm 0.014}$ | $24.58_{\pm 0.57}$ | $0.730_{\pm 0.014}$ |
| | FSEL | 0.06M | $28.16_{\pm 0.75}$ | $0.827_{\pm 0.015}$ | $25.42_{\pm 0.60}$ | $0.752_{\pm 0.014}$ | $24.55_{\pm 0.57}$ | $0.730_{\pm 0.014}$ |
| | SpectFormer | 0.02M | $28.10_{\pm 0.75}$ | $0.825_{\pm 0.015}$ | $25.36_{\pm 0.60}$ | $0.751_{\pm 0.014}$ | $24.78_{\pm 0.58}$ | $0.735_{\pm 0.014}$ |
| | **CHASM** | 0.05M | $\mathbf{28.68}_{\pm 0.74}$ | $\mathbf{0.840}_{\pm 0.016}$ | $\mathbf{25.72}_{\pm 0.61}$ | $\mathbf{0.761}_{\pm 0.014}$ | $\mathbf{25.19}_{\pm 0.59}$ | $\mathbf{0.745}_{\pm 0.014}$ |
| ×4 | Baseline | 0.00M | $34.53_{\pm 0.75}$ | $0.936_{\pm 0.009}$ | $28.81_{\pm 0.60}$ | $0.852_{\pm 0.010}$ | $27.86_{\pm 0.58}$ | $0.830_{\pm 0.010}$ |
| | GFNet | 0.80M | $35.70_{\pm 0.75}$ | $0.947_{\pm 0.009}$ | $29.55_{\pm 0.62}$ | $0.866_{\pm 0.010}$ | $28.62_{\pm 0.59}$ | $0.844_{\pm 0.010}$ |
| | AFNO | 0.01M | $35.61_{\pm 0.75}$ | $0.941_{\pm 0.009}$ | $29.31_{\pm 0.61}$ | $0.860_{\pm 0.010}$ | $28.36_{\pm 0.59}$ | $0.838_{\pm 0.010}$ |
| | FFC | 1.19M | $35.62_{\pm 0.75}$ | $0.946_{\pm 0.009}$ | $29.24_{\pm 0.61}$ | $0.858_{\pm 0.010}$ | $28.45_{\pm 0.59}$ | $0.840_{\pm 0.010}$ |
| | FNO-D | 4.74M | $34.77_{\pm 0.75}$ | $0.939_{\pm 0.009}$ | $29.20_{\pm 0.61}$ | $0.858_{\pm 0.010}$ | $28.34_{\pm 0.59}$ | $0.838_{\pm 0.010}$ |
| | DiffFNO | 4.77M | $34.13_{\pm 0.74}$ | $0.933_{\pm 0.009}$ | $29.28_{\pm 0.61}$ | $0.859_{\pm 0.010}$ | $28.28_{\pm 0.59}$ | $0.836_{\pm 0.010}$ |
| | FSEL | 0.06M | $36.22_{\pm 0.76}$ | $0.951_{\pm 0.008}$ | $29.61_{\pm 0.62}$ | $0.867_{\pm 0.010}$ | $28.56_{\pm 0.59}$ | $0.843_{\pm 0.010}$ |
| | SpectFormer | 0.02M | $35.91_{\pm 0.75}$ | $0.948_{\pm 0.008}$ | $29.50_{\pm 0.62}$ | $0.865_{\pm 0.010}$ | $28.65_{\pm 0.59}$ | $0.845_{\pm 0.010}$ |
| | **CHASM** | 0.05M | $\mathbf{37.05}_{\pm 0.76}$ | $\mathbf{0.956}_{\pm 0.007}$ | $\mathbf{30.03}_{\pm 0.63}$ | $\mathbf{0.875}_{\pm 0.009}$ | $\mathbf{29.10}_{\pm 0.60}$ | $\mathbf{0.856}_{\pm 0.009}$ |

## 4.1 Accelerated MRI reconstruction

We first evaluate accelerated MRI reconstruction on CC359 [Souza et al., 2018] and fastMRI [Zbontar et al., 2018]. Table 1 reports CC359 results across three backbone families. CHASM improves over the same-backbone spectral-mixer baselines under both $4\times$ and $8\times$ acceleration in HiFi-Mamba, U-Net, and ViT.

Table 2: Unified evaluation across fastMRI reconstruction, natural-image reconstruction, and undersampled MRI segmentation. The last row reports the maximum standard deviation across methods: case-level for reconstruction and seed-level for segmentation.

| Mixer block | fastMRI ×8 | | fastMRI ×4 | | ADE20K | | ImageNet | | BraTS ×4 | | BraTS ×8 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | mIoU↑ | Dice↑ | mIoU↑ | Dice↑ |
| Baseline | 31.24 | 0.741 | 34.31 | 0.849 | 23.67 | 0.750 | 27.89 | 0.835 | 0.727 | 0.841 | 0.612 | 0.757 |
| GFNet | 31.56 | 0.761 | 34.50 | 0.853 | 23.97 | 0.762 | 28.20 | 0.846 | 0.737 | 0.848 | 0.632 | 0.773 |
| AFNO | 31.32 | 0.752 | 34.47 | 0.850 | 23.74 | 0.753 | 27.97 | 0.838 | 0.731 | 0.844 | 0.621 | 0.765 |
| FFC | 31.40 | 0.757 | 34.51 | 0.852 | 23.73 | 0.752 | 27.95 | 0.837 | 0.730 | 0.843 | 0.619 | 0.763 |
| FNO-D | 31.32 | 0.756 | 34.49 | 0.852 | 23.76 | 0.755 | 28.00 | 0.840 | 0.733 | 0.845 | 0.625 | 0.768 |
| DiffFNO | 31.28 | 0.755 | 34.59 | 0.854 | 23.79 | 0.756 | 28.04 | 0.841 | 0.735 | 0.846 | 0.628 | 0.770 |
| FSEL | 31.61 | 0.761 | 34.57 | 0.854 | 23.98 | 0.765 | 28.22 | 0.849 | 0.739 | 0.850 | 0.639 | 0.778 |
| SpectFormer | 31.76 | 0.764 | 34.65 | 0.855 | 24.01 | 0.766 | 28.25 | 0.850 | 0.742 | 0.853 | 0.643 | 0.781 |
| **CHASM** | **32.31** | **0.779** | **35.22** | **0.872** | **24.52** | **0.783** | **28.58** | **0.864** | **0.792** | **0.884** | **0.699** | **0.822** |
| *Max std.* | 1.85 | 0.076 | 2.24 | 0.052 | 1.80 | 0.045 | 1.65 | 0.040 | 0.035 | 0.030 | 0.050 | 0.045 |

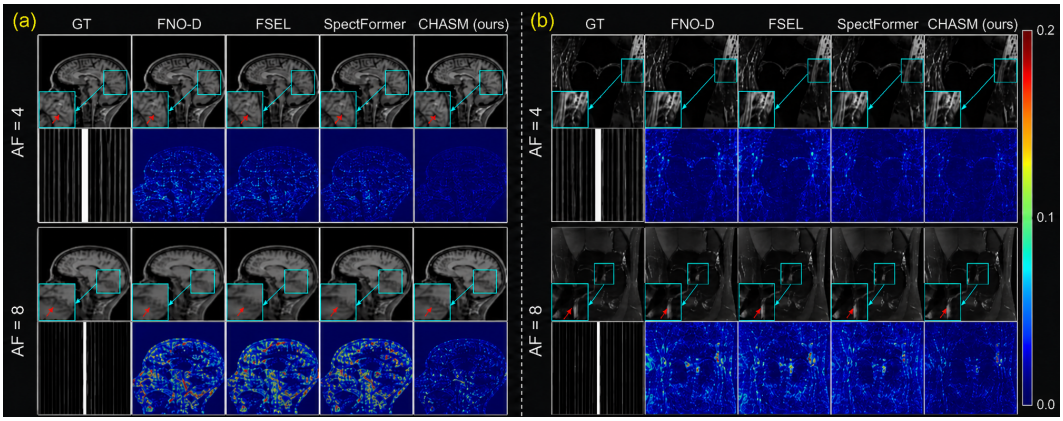


Figure 2: Qualitative comparison on the fastMRI and CC359 datasets under single-coil settings. (a) Reconstruction results on the fastMRI knee dataset with acceleration factors AF=4 and AF=8. (b) Reconstruction results on the CC359 brain dataset under the same acceleration factors. The second row of each subplot shows the corresponding error maps. The blue boxes and red arrow highlight the details in the reconstruction results.

### 4.2 Additional reconstruction and segmentation benchmarks

Table 2 evaluates CHASM across datasets and tasks. Each comparison uses the same backbone, input operator, objective, and training protocol across mixer blocks (Appendix A.11).

We evaluate fastMRI Knee reconstruction with the HiFi-Mamba-family backbone. We also evaluate natural-image reconstruction on ADE20K [Zhou et al., 2017] and ImageNet [Deng et al., 2009], and undersampled brain tumor segmentation on BraTS [Menze et al., 2015]. For natural images, all methods use the same downsampling degradation and reconstruction objective. For BraTS, inputs are retrospectively undersampled and models predict tumor segmentation masks.

Table 2 shows that CHASM improves fastMRI reconstruction, natural-image reconstruction, and undersampled MRI segmentation, suggesting that cross-frequency harmonization is useful beyond a single dataset or task. Figure 2 further provides qualitative reconstruction examples on fastMRI and CC359. The error maps show that CHASM reduces residual aliasing and better preserves fine anatomical structures under both fourfold and eightfold acceleration.

Table 3: Random-mask falsification and ablation studies with the HiFi-Mamba backbone. (a) Random-mask falsification reports ΔPSNR of CHASM over the strongest spectral-mixer baseline in the main reconstruction tables. Structured-mask gains are computed from the corresponding main reconstruction results. Random-mask gains are measured by replacing structured masks with random masks under the same model and training setting. Drop is computed as $1-\Delta_{\mathrm{random}}/\Delta_{\mathrm{structured}}$. (b) Spectral-core basis constraints are evaluated on CC359 $\times 8$. (c) CH/CW axis compositions are evaluated on CC359 $\times 8$.

**(a) Random-mask falsification**

| Setting | Struct. | Random | Drop |
|---|---|---|---|
| CC359 $\times 4$ | +0.83 | +0.10 | 88.0% |
| CC359 $\times 8$ | +0.52 | +0.08 | 84.6% |
| fastMRI $\times 4$ | +0.57 | +0.08 | 86.0% |
| fastMRI $\times 8$ | +0.55 | +0.06 | 89.1% |
| **Mean** | **+0.62** | **+0.08** | **86.9%** |

**(b) Shared-basis constraint**

| Variant | PSNR↑ | SSIM↑ |
|---|---|---|
| Identity basis | 27.86 | 0.820 |
| Untied basis | 28.43 | 0.833 |
| Shared, signed gain | 28.51 | 0.836 |
| Shared, complex | 28.56 | 0.838 |
| **CHASM** | **28.68** | **0.840** |

**(c) Axis factorization**

| Variant | PSNR↑ | SSIM↑ |
|---|---|---|
| CH only | 28.31 | 0.831 |
| CW only | 28.27 | 0.829 |
| CW → CH | 28.54 | 0.837 |
| CH + CW | 28.48 | 0.835 |
| **CH → CW** | **28.68** | **0.840** |

### 4.3 Mechanistic analysis

We next test whether the gain comes from the proposed cross-frequency harmonization. Unless otherwise stated, ablations are conducted on CC359 $\times 8$ reconstruction with the HiFi-Mamba-family backbone. The shared-basis ablation keeps the CHASM wrapper, local refinement, fusion projection, and training protocol fixed while changing only the spectral core; the random-mask study changes the sampling geometry and measures how the CHASM gain varies with spectral structure. Definitions of the ablation variants and three-seed stability for the shared-basis ablation are provided in Appendix A.8 and Appendix A.3.

Table 3(a) shows that randomizing the sampling geometry reduces CHASM's mean gain from $0.62$ to $0.08$ dB. Cross-frequency harmonization is most effective under the coherent spectral relationships of structured masks. In Table 3(b), untying the basis lowers PSNR from $28.68$ to $28.43$ dB with the wrapper unchanged, demonstrating the benefit of shared channel directions in this setting. Table 3(c) further shows that the default CH→CW axis-separable composition gives the strongest performance among the tested variants.

**Constraint interactions and fixed-wrapper controls.** Additional controls separate basis sharing, orthogonality, and the surrounding block (Appendices A.4 and A.6). CHASM reaches $28.68$ dB, compared with $28.39$ for an unconstrained free-sandwich core and $28.04$ for a shared, tied non-orthogonal core. Removing orthogonality with tying retained costs $0.64$ dB; relaxing both constraints costs $0.29$ dB, showing that their effects interact. Removing positive gains at fixed orthogonal tying costs $0.17$ dB. With the CHASM wrapper held fixed, CHASM also exceeds the SpectFormer and FSEL cores by $0.32$ and $0.46$ dB, respectively. Under the same CC359 $\times 8$ protocol, F-FNO reaches $28.46$ dB and TFNO $28.06$ dB (Appendix A.5).

## 5 Conclusion and Limitations

We presented CHASM, a structured spectral mixer that harmonizes channel directions across frequencies while preserving frequency-specific gains. By using a shared learned eigenbasis and frequency-dependent positive gains, CHASM provides a simple way to coordinate spectral channel mixing without removing local frequency adaptivity. Across accelerated MRI reconstruction, natural-image reconstruction, and undersampled MRI segmentation, CHASM improves over same-backbone spectral-mixer baselines under controlled replacement settings. The ablations support the shared-basis design and show that its advantage depends on coherent sampling geometry, supporting cross-frequency harmonization for undersampled imaging. This work is limited to image-domain reconstruction and segmentation. Future work can study broader dense prediction, real sensor-noise restoration, video reconstruction, and a deeper characterization of when cross-frequency mixing is most effective across sampling patterns, feature resolutions, and backbones.

## Acknowledgments and Disclosures

The authors received no third-party funding or support for this work and declare no competing interests.

## References

Hongli Chen, Pengcheng Fang, Yuxia Chen, Yingxuan Ren, Jing Hao, Fangfang Tang, Xiaohao Cai, Shanshan Shan, and Feng Liu. HiFi-Mamba: Dual-stream W-laplacian enhanced Mamba for high-fidelity MRI reconstruction. *arXiv preprint arXiv:2508.09179*, 2025. doi: 10.48550/arXiv.2508.09179. URL https://arxiv.org/abs/2508.09179.

Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. In *Advances in Neural Information Processing Systems*, volume 33, pages 4479–4488, 2020.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

Pengcheng Fang, Hongli Chen, Guangzhen Yao, Jian Shi, Fangfang Tang, Xiaohao Cai, Shanshan Shan, and Feng Liu. HiFi-MambaV2: Hierarchical shared-routed MoE for high-fidelity MRI reconstruction. *arXiv preprint arXiv:2511.18534*, 2025. doi: 10.48550/arXiv.2511.18534. URL https://arxiv.org/abs/2511.18534.

Bernhard N. Flury. Common principal components in $k$ groups. *Journal of the American Statistical Association*, 79(388):892–898, 1984. doi: 10.1080/01621459.1984.10477108.

John Guibas, Morteza Mardani, Zongyi Li, Andrew Tao, Anima Anandkumar, and Bryan Catanzaro. Adaptive fourier neural operators: Efficient token mixers for transformers. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=EXHG-A3jlM.

Jean Kossaifi, Nikola Kovachki, Kamyar Azizzadenesheli, and Anima Anandkumar. Multi-grid tensorized fourier neural operator for high-resolution PDEs. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=AWiDlO63bH.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007. doi: 10.1002/mrm.21391.

Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10): 1993–2024, 2015. doi: 10.1109/TMI.2014.237769.

Yongming Rao, Wenliang Zhao, Zheng Zhu, Jiwen Lu, and Jie Zhou. Global filter networks for image classification. In *Advances in Neural Information Processing Systems*, volume 34, pages 980–993, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28.

Roberto Souza, Oeslle Lucena, Julia Garrafa, David Gobbi, Marina Saluzzi, Simone Appenzeller, Letícia Rittner, Richard Frayne, and Roberto Lotufo. An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement. *NeuroImage*, 170:482–494, 2018. doi: 10.1016/j.neuroimage.2017.08.021.

Ilya O. Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, and Alexey Dosovitskiy. MLP-Mixer: An all-MLP architecture for vision. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. In *International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=tmIiMPl4IPa`.

Weihao Yu, Mi Luo, Pan Zhou, Chenyang Si, Yichen Zhou, Xinchao Wang, Jiashi Feng, and Shuicheng Yan. MetaFormer is actually what you need for vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10819–10829, 2022. doi: 10.1109/CVPR52688.2022.01055.

Jure Zbontar, Florian Knoll, Anuroop Sriram, Matthew J. Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, et al. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*, 2018.

Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641, 2017.

# A Additional Experimental Details

## A.1 Seed Stability

We evaluate run-to-run stability on the main MRI reconstruction settings. We report three-seed stability for CHASM and the strongest spectral-mixer baseline in the corresponding main reconstruction table. This appendix reports seed-level standard deviations, separate from the case-level standard deviations used in the main test-set tables.

Table 4: Three-seed stability on the main MRI reconstruction settings. CHASM is compared against the strongest spectral-mixer baseline in the corresponding main table. Values are reported as mean $\pm$ standard deviation.

| Setting | Comparator | Base PSNR↑ | CHASM PSNR↑ | ΔPSNR | Base SSIM↑ | CHASM SSIM↑ |
|---|---|---|---|---|---|---|
| CC359 $\times 8$ | FSEL | 28.16±0.03 | **28.68±0.02** | **+0.52** | 0.827±0.002 | **0.840±0.001** |
| CC359 $\times 4$ | FSEL | 36.22±0.05 | **37.05±0.03** | **+0.83** | 0.951±0.001 | **0.956±0.001** |
| fastMRI $\times 8$ | SpectFormer | 31.76±0.04 | **32.31±0.05** | **+0.55** | 0.764±0.002 | **0.779±0.003** |
| fastMRI $\times 4$ | SpectFormer | 34.65±0.02 | **35.22±0.03** | **+0.57** | 0.855±0.001 | **0.872±0.002** |

## A.2 Segmentation Statistics on BraTS

Table 5 reports variability across training seeds. Per-case paired $t$-tests against SpectFormer on $125$ patients give $t = 3.73$, $p < 0.001$ at $\times 4$ and $t = 2.61$, $p = 0.010$ at $\times 8$ for mIoU. These paired tests support the mIoU improvements at both acceleration factors. The corresponding standardized paired effects are $d_z = t/\sqrt{125}$, approximately $0.334$ at $\times 4$ and $0.233$ at $\times 8$, computed from the reported $t$ statistics.

Table 5: BraTS means and run-to-run standard deviations over three seeds.

| Mixer | $\times 4$ mIoU | $\times 4$ Dice | $\times 8$ mIoU | $\times 8$ Dice |
|---|---|---|---|---|
| Seed-level variation | | | | |
| FSEL | $0.739 \pm 0.016$ | $0.850 \pm 0.009$ | $0.639 \pm 0.012$ | $0.778 \pm 0.011$ |
| SpectFormer | $0.742 \pm 0.013$ | $0.853 \pm 0.010$ | $0.643 \pm 0.019$ | $0.781 \pm 0.013$ |
| CHASM | $0.792 \pm 0.012$ | $0.884 \pm 0.006$ | $0.699 \pm 0.011$ | $0.822 \pm 0.009$ |

Table 6: Case-level mIoU variability on BraTS. Standard deviations here describe variation across patients, separately from the seed variation in Table 5.

| Mixer | $\times 4$ mIoU | $\times 8$ mIoU |
|---|---|---|
| FSEL | $0.739 \pm 0.23$ | $0.639 \pm 0.20$ |
| SpectFormer | $0.742 \pm 0.24$ | $0.643 \pm 0.24$ |
| CHASM | $0.792 \pm 0.19$ | $0.699 \pm 0.22$ |

## A.3 Shared-Basis Ablation Stability

We additionally evaluate the stability of the shared-basis ablation on the main CC359 $\times 8$ reconstruction setting. The CHASM wrapper, local refinement, fusion projection, optimizer, data split, mask, and training schedule are kept fixed; only the spectral-core basis constraint is changed.

## A.4 Constraint Interactions

Table 8 compares constraint relaxations under the CC359 $\times 8$, HiFi-Mamba protocol. The wrapper and training protocol are fixed. The non-monotonic ordering of the relaxations shows that the orthogonality and tying effects interact. The $0.64$ dB comparison conditions on tying, while the $0.29$ dB comparison relaxes both constraints.

Table 7: Three-seed stability for the shared-basis ablation on CC359 $\times 8$ with the HiFi-Mamba-family reconstruction backbone. Values are reported as mean $\pm$ standard deviation.

| Variant | PSNR↑ | SSIM↑ |
|---|---|---|
| Untied basis | 28.43±0.04 | 0.833±0.002 |
| CHASM | **28.68±0.02** | **0.840±0.001** |

Table 8: Constraint relaxations on CC359 $\times 8$. Reported PSNR deviations are standard deviations over three seeds; SSIM means are shown.

| Variant | PSNR↑ | SSIM↑ |
|---|---|---|
| Shared, tied non-orthogonal | $28.04 \pm 0.03$ | 0.825 |
| Free sandwich | $28.39 \pm 0.02$ | 0.833 |
| Shared orthogonal, signed gains | 28.51 | 0.836 |
| **CHASM** | $\mathbf{28.68 \pm 0.02}$ | **0.840** |

### A.5 Additional Spectral-Operator Baselines

We additionally compare against two operator families that share spectral structure in different ways: the factorized Fourier neural operator (F-FNO; Tran et al., 2023), which is axis-separable but keeps independent per-frequency weights, and the tensorized Fourier neural operator (TFNO; Kossaifi et al., 2024), which factorizes the spectral weight tensor through shared, unconstrained factor matrices. Both are evaluated under the protocol of Section 4 on CC359 $\times 8$ with the HiFi-Mamba-family backbone.

Table 9: Additional spectral-operator baselines on CC359 $\times 8$ (HiFi-Mamba-family backbone). F-FNO reports mean and standard deviation over three seeds; TFNO is a single run.

| Operator | PSNR↑ | SSIM↑ |
|---|---|---|
| TFNO (tensorized, shared factors) | 28.06 | 0.825 |
| F-FNO (axis-separable, per-frequency weights) | $28.46 \pm 0.03$ | $0.833 \pm 0.003$ |
| **CHASM** (explicit shared orthogonal basis) | **28.68** | **0.840** |

CHASM exceeds F-FNO by $0.22$ dB and TFNO by $0.62$ dB in Table 9.

### A.6 Baseline Spectral Cores in the CHASM Wrapper

To separate the spectral core from the surrounding block, we hold the CHASM wrapper fixed — its depthwise local refinement and $1\times1$ fusion projection — and replace only the spectral core with that of a strong baseline. For FSEL we take its frequency branch (real/imaginary $1\times1$ convolutions with a magnitude gate), dropping its own spatial branch and fusion; for SpectFormer we take its magnitude-gating branch, dropping its own fusion.

In Table 10, with the wrapper held fixed, the CHASM operator is $0.32$ dB above the SpectFormer core and $0.46$ dB above the FSEL core, and the same wrapper around an identity-basis core reaches $27.86$, $0.82$ dB below CHASM.

### A.7 Gain-Table Resolution

The gain tables are learned on a canonical grid of $B$ bins per axis and interpolated to the retained frequency bins of the input resolution. We use $B = 96$ except in this sweep. Table 11 sweeps $B$ on CC359 $\times 8$ with the HiFi-Mamba-family backbone over three seeds.

Performance varies by $0.16$ dB as $B$ increases from 32 to 128. At $B = 128$, CHASM reaches $28.63$ dB, within $0.04$ dB of the default, showing that performance is stable across this fourfold range of gain-table resolutions.

The default-configuration mean in this sweep is $28.67$ dB, within $0.01$ dB of the main-table mean of $28.68$ dB and within the reported seed variation.

Table 10: Spectral core swapped inside a fixed CHASM wrapper, CC359 $\times 8$ with the HiFi-Mamba-family backbone. Native rows are reproduced from Table 1.

| Spectral core | Wrapper | PSNR↑ | SSIM↑ |
|---|---|---|---|
| **CHASM** (shared orthogonal basis) | CHASM | **28.68** | **0.840** |
| SpectFormer core | CHASM | 28.36 | 0.831 |
| FSEL core | CHASM | 28.22 | 0.828 |
| FSEL core | FSEL native | 28.16 | 0.827 |
| SpectFormer core | SpectFormer native | 28.10 | 0.825 |
| Identity-basis core ($\mathbf{U} = \mathbf{I}$) | CHASM | 27.86 | 0.820 |

Table 11: Sensitivity to the gain-table resolution $B$ on CC359 $\times 8$ (three seeds). $B = 96$ is the default configuration used throughout the paper, with the same input resolution and interpolation rule across all settings.

| $B$ | 32 | 64 | 96 (default) | 128 |
|---|---|---|---|---|
| PSNR↑ | 28.51$\pm$0.03 | 28.59$\pm$0.03 | **28.67**$\pm$0.02 | 28.63$\pm$0.03 |
| SSIM↑ | 0.836$\pm$0.004 | 0.839$\pm$0.004 | **0.840**$\pm$0.003 | 0.838$\pm$0.003 |

### A.8 Ablation Variant Definitions

**Basis constraints.** The identity-basis variant fixes the shared basis to $U = I$ and learns only the frequency-specific gains. The untied-basis variant learns an independent orthogonal basis at each frequency, $\mathbf{M}(k) = \mathbf{U}_k \operatorname{diag}(\boldsymbol{\lambda}_k)\mathbf{U}_k^\top$. The tied non-orthogonal variant replaces the orthogonal basis with a general invertible matrix, giving $\mathbf{W} \operatorname{diag}(\boldsymbol{\lambda}_k)\mathbf{W}^{-1}$. The free-sandwich variant instead learns two independent unconstrained matrices, $\mathbf{P} \operatorname{diag}(\boldsymbol{\lambda}_k)\mathbf{Q}$. All variants keep the CHASM wrapper, local refinement, fusion projection, optimizer, mask, split, and training schedule fixed. The signed-gain variant keeps the shared basis but removes the positivity constraint on the gains, replacing the softplus mapping with unconstrained real-valued gains. The complex variant replaces the real-valued spectral-core gains with complex-valued frequency-domain gains while preserving the real-output constraint of the inverse real Fourier transform.

**Axis compositions.** The CH-only and CW-only variants apply the harmonized operator only along the height–channel or width–channel plane, respectively. The CW→CH and CH→CW variants apply the two axis-wise operators sequentially in the indicated order. The CH+CW variant applies the CH-plane and CW-plane operators in parallel and combines their outputs before the local refinement and fusion projection.

### A.9 Computational Cost

Table 12: Model-level FLOPs for CC359 reconstruction backbones after replacing the mixer block, under the same input setting as Table 1.

| Mixer block | HiFi-Mamba | U-Net | ViT |
|---|---|---|---|
| Baseline | 24.64G | 12.56G | 11.54G |
| GFNet | 24.62G | 12.56G | 11.01G |
| AFNO | 24.62G | 12.56G | 11.01G |
| FFC | 25.57G | 12.74G | 11.13G |
| FNO-D | 29.06G | 13.36G | 10.56G |
| DiffFNO | 21.26G | 11.88G | 11.01G |
| FSEL | 34.96G | 14.71G | 12.44G |
| SpectFormer | 26.43G | 12.96G | 11.27G |
| CHASM | 28.22G | 13.27G | 11.48G |

### A.10 Measured Runtime and Basis Caching

We time the isolated spectral core on an NVIDIA H100 80GB with input $2 \times C \times 64 \times 64$, fp32, PyTorch 2.5.1, 10 warm-up iterations, and 50 timed iterations. In the uncached path, each axis

basis is constructed once per forward and its transpose is reused. In the cached path, both bases are constructed once before timing and reused across inputs. Two independent measurements agreed within $4\%$.

Table 13: Isolated CHASM spectral-core latency in milliseconds.

| $C$ | 96 | 128 | 192 | 256 |
|---|---|---|---|---|
| Uncached | 1.18 | 1.81 | 1.81 | 1.97 |
| Cached | 0.47 | 0.47 | 0.56 | 0.75 |
| Speedup | $2.5\times$ | $3.9\times$ | $3.2\times$ | $2.6\times$ |

Caching removes basis construction from the inference path and reduces the measured latency by $2.5$–$3.9\times$. The basis-construction measurements in Table 14 stay between $0.41$ and $0.46$ ms for $C = 128$–$512$ on this hardware.

Table 14: Time to construct one orthogonal basis. The last row gives the theoretical cubic arithmetic scaling relative to $C = 128$.

| $C$ | 96 | 128 | 192 | 256 | 384 | 512 |
|---|---|---|---|---|---|---|
| Construction (ms) | 0.19 | 0.41 | 0.41 | 0.45 | 0.44 | 0.46 |
| Relative cubic cost | 0.42 | 1 | 3.4 | 8 | 27 | 64 |

### A.11 Reproducibility Details

All comparisons use the same backbone, dataset split, undersampling mask, training objective, optimizer, and training schedule within each experimental setting. Each method replaces the same mixer location as a complete non-cumulative mixer block. Components intrinsic to each baseline mixer, such as projection layers, local convolutional branches, or frequency-selection modules, are retained in the main comparisons. Baseline mixer configurations follow their original implementations, with channel- and resolution-dependent dimensions adjusted to match the host backbone. The fixed-wrapper experiment is specified separately in Appendix A.6. The *Baseline* row uses the unmodified host backbone, with an identity at the replacement location, and therefore has zero replacement-block parameters.

**Datasets.** We evaluate accelerated MRI reconstruction on CC359 and fastMRI, natural-image reconstruction on ADE20K and ImageNet, and undersampled MRI segmentation on BraTS. For MRI reconstruction, the same acceleration factors and undersampling protocols are used for all compared methods. For natural-image reconstruction, all methods use the same degradation operator and reconstruction loss. For BraTS segmentation, inputs are retrospectively undersampled and models predict tumor segmentation masks.

**Cross-domain settings.** For Table 2, fastMRI reconstruction uses the HiFi-Mamba-family reconstruction backbone. ADE20K and ImageNet reconstruction use the same U-Net-style reconstruction backbone with bicubic downsampling at $4\times$, trained with an $\ell_1$ reconstruction loss. BraTS segmentation uses a U-Net segmentation backbone with retrospective Cartesian undersampling at $4\times$ and $8\times$, trained with a Dice–cross-entropy loss. Its undersampled inputs use the same construction as the reconstruction experiments.

**Metrics.** For reconstruction, we report PSNR and SSIM. For segmentation, we report mIoU and Dice. Higher is better for PSNR, SSIM, mIoU, and Dice.

**Training protocol.** All methods are trained under matched settings within each task. Unless otherwise specified, we use AdamW with a learning rate of $5 \times 10^{-4}$, batch size $8$, and train for $100$ epochs. The same training and validation splits, undersampling masks, loss function, optimizer, learning-rate setting, batch size, training length, and data augmentation are used for all mixer blocks within each experimental setting. Hyperparameters are fixed across methods and test results are reported using the corresponding validation-selected checkpoint.

**Compute.** Experiments are run on NVIDIA A100 and H100 GPU workers. Model-level FLOPs are reported in Appendix A.9, and H100 spectral-core latencies with their measurement settings in Appendix A.10.

**Implementation.** CHASM uses real FFT and inverse real FFT along spatial axes, implemented with `torch.fft.rfft` and `torch.fft.irfft`. The channel operators are real-valued and applied to complex Fourier coefficients by complex linear extension, equivalently applying the same real matrix to the real and imaginary parts. The orthogonal bases are constructed from skew-symmetric matrices through the matrix exponential, and frequency-specific gains are mapped to positive values using softplus. We learn $B_H = B_W = 96$ gain bins and interpolate them to the retained rFFT bins. The gain resolution and CH→CW order are selected once on the CC359 development set and kept fixed across datasets. CC359 slices have spatial size $256 \times 256$.

**Block wrapper details.** The local refinement module $\mathcal{L}$ is implemented as a depthwise $3 \times 3$ convolution with padding 1, a GELU activation, and a $1 \times 1$ pointwise convolution. The fusion projection $\mathbf{P}_{\text{fuse}}$ is a $1 \times 1$ convolution that maps the refined features back to $C$ channels. Both pointwise convolutions preserve the channel width. At $C = B_H = B_W = 96$, the spectral core has $27{,}552$ parameters. The depthwise and two pointwise convolutions, including biases, add $19{,}584$, giving $47{,}136$ parameters ($0.05$M rounded) for the complete block. In the mechanistic ablations, $\mathcal{L}$, $\mathbf{P}_{\text{fuse}}$, optimizer, mask, split, and training schedule are kept fixed; only the spectral-core constraint is changed.

**Cross-domain protocol details.** Table 15 summarizes the protocol used for the cross-domain transfer experiments in Table 2, listing each task's backbone, input operator, and objective.

Table 15: Protocol details for the cross-domain transfer evaluation.

| Setting | Backbone | Input operator | Objective | Purpose |
|---|---|---|---|---|
| fastMRI | HiFi-Mamba-family | Cartesian undersampling | reconstruction loss | MRI transfer |
| ADE20K | U-Net-style | $4\times$ bicubic downsampling | $\ell_1$ | natural-image reconstruction |
| ImageNet | U-Net-style | $4\times$ bicubic downsampling | $\ell_1$ | natural-image reconstruction |
| BraTS | U-Net segmentation | Cartesian undersampling | Dice–CE | undersampled segmentation |

### A.12 Taxonomy of Spectral Mixer Baselines

Table 16 summarizes how CHASM differs from existing spectral mixers: it explicitly imposes a shared channel eigenbasis across frequencies while preserving frequency-specific gains.

Table 16: Taxonomy of spectral mixer designs. "Shared basis" means that the method directly ties channel directions across frequency bins through a learned shared eigenbasis.

| Mixer | Spectral response | Channel mixing | Shared basis | Positive gains |
|---|---|---|---|---|
| GFNet | learned filters | diagonal/filter-wise | no | no |
| AFNO | adaptive Fourier modes | shared block/channel | no explicit | no |
| FFC | Fourier branch | conv./spectral | no explicit | no |
| FNO-D | learned FNO response | mode-wise | no explicit | no |
| DiffFNO | dynamic Fourier response | frequency-conditioned | no explicit | no |
| FSEL | frequency selection | learned enhancement | no explicit | no |
| SpectFormer | spectral attention | token/channel | no explicit | no |
| **CHASM** | shared-basis gains | symmetric channel | **yes** | **yes** |

GFNet learns frequency-specific filters along fixed channel axes. AFNO shares blockwise channel transformations across retained modes, providing implicit cross-frequency tying. FNO variants learn mode-specific spectral operators, with deformation or dynamic responses providing additional adaptivity. Spectral-attention methods couple features through attention. CHASM coordinates frequencies through a shared orthogonal eigenbasis and frequency-specific positive gains.

## NeurIPS Paper Checklist

1. **Claims** Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction describe CHASM's shared-basis spectral mixer, its structural characterization, and its evaluation through same-backbone comparisons and mechanistic ablations.

2. **Limitations** Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper includes a Limitations section discussing the task scope, image-domain focus, and the need for further theoretical characterization.

3. **Theory assumptions and proofs** Question: For each theoretical result, does the paper provide the full set of assumptions and a complete proof?

   Answer: [Yes]

   Justification: Proposition 1 states the generic-stratum assumption for the direct shared-basis spectral-operator family. A proof is provided in Section 3.4.

4. **Experimental result reproducibility** Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions?

   Answer: [Yes]

   Justification: The paper specifies the evaluated mixer blocks, backbones, datasets, metrics, and controlled comparison protocol. Appendix A.11 provides additional training and implementation details.

5. **Open access to data and code** Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results?

   Answer: [No]

   Justification: The paper cites the original sources of its public datasets and describes the method and experimental protocol. Source code is not provided with the paper.

6. **Experimental setting/details** Question: Does the paper specify all the training and test details necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies datasets, acceleration factors, backbone families, metrics, and the replacement-block protocol. Appendix A.11 reports the shared optimizer, learning rate, batch size, training length, GPU workers, implementation details, and cross-domain settings; Appendix A.8 defines the ablation variants.

7. **Experiment statistical significance** Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Main test-set tables report case-level standard deviations for reconstruction metrics, and Table 2 reports the maximum case-level standard deviation for reconstruction and seed-level standard deviation for segmentation. Appendix A.2 separates both sources of variability and reports paired significance tests; further seed stability is reported in Appendix A.1.

8. **Experiments compute resources** Question: For each experiment, does the paper provide sufficient information on the computer resources needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix A.11 reports the use of NVIDIA A100 and H100 GPU workers and the shared training schedule. The main paper reports complete replacement-block parameter counts. Appendix A.10 provides spectral-core latency measurements and their hardware settings.

9. **Code of ethics** Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics?

   Answer: [Yes]

   Justification: The work uses standard public research datasets and proposes a general-purpose spectral mixer for reconstruction and segmentation. We are not aware of any deviation from the NeurIPS Code of Ethics.

10. **Broader impacts** Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: CHASM may benefit computational imaging and medical-image analysis. It is not a deployed clinical system, and any clinical use would require task-specific validation, regulatory review, and expert oversight.

11. **Safeguards** Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse?

    Answer: [N/A]

    Justification: The paper does not release high-risk models, scraped datasets, generative systems, or other assets with obvious dual-use misuse risks.

12. **Licenses for existing assets** Question: Are the creators or original owners of assets used in the paper properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper cites the datasets and methods used in the experiments, including CC359, fastMRI, ADE20K, ImageNet, and BraTS. These assets are used for research evaluation.

13. **New assets** Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not introduce a new dataset or benchmark asset.

14. **Crowdsourcing and research with human subjects** Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or new human-subject experiments.

15. **Institutional review board approvals or equivalent for research with human subjects** Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether IRB approvals or equivalent approvals were obtained?

    Answer: [N/A]

    Justification: The paper does not involve new human-subject research.

16. **Declaration of LLM usage** Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research?

    Answer: [N/A]

    Justification: LLMs are not used as part of the proposed method, experiments, or scientific contribution.