QuIR: Quotient-Informed Recovery Network for Jointly Sparse Blind Deconvolution

Anonymous authors
Paper under double-blind review

Abstract

Jointly sparse blind deconvolution (JSBD) seeks to recover multiple signals and a shared unknown convolution kernel from noisy observations, assuming that the signals have a common sparse support. Existing inverse-filter methods apply a candidate inverse filter to the observations and evaluate the joint sparsity of the resulting signals. This operation can amplify measurement noise at frequencies where the kernel response is weak, degrading recovery under severe ill-conditioning or very low signal-to-noise ratios. We propose Quotient-Informed Recovery Network (QuIR), which predicts a candidate support before inverse-filter estimation to guide signal and kernel recovery. To predict support from a limited number of noisy channels, we train a neural network to estimate normalized counts of cyclic coordinate differences between support locations. The network uses quotient features that do not depend on the unknown kernel in the noiseless model. On the evaluated noise-conditioning grid, QuIR achieves 73.6% exact support recovery, compared with 13.5% for inverse-filter JSBD.

1 Introduction

Blind deconvolution is underdetermined because each measurement mixes an unknown signal with an unknown convolution kernel. Multiple channels can constrain this factorization when they share structure. In jointly sparse blind deconvolution (JSBD), all channels use the same kernel and all signals share the same sparse support. These assumptions couple the channels through the kernel and restrict where the signals can be active, but convolution mixes the two unknown factors before observation. We study the circular-convolution model

$$y_\ell = g \circledast x_\ell + w_\ell, \quad \ell = 1, \dots, L, \tag{1}$$

where $g, x_\ell, y_\ell, w_\ell \in \mathbb{C}^N$ and $\circledast$ denotes circular convolution. Write $X = [x_1, \dots, x_L]$, $Y = [y_1, \dots, y_L]$, and $W = [w_1, \dots, w_L]$, so $Y = C(g)X + W$ with $C(g)$ the circular-convolution matrix. The signals share the row support

$$S = \{n \in \mathbb{Z}_N : \|X_{n,:}\|_2 > 0\}, \quad K = |S|, \quad 1 \le K < N, \tag{2}$$

where $\mathbb{Z}_N = \{0, \dots, N-1\}$ and indices are taken modulo N. All signals vanish outside S, while their coefficients within S can vary across channels. The goal is to recover X and g, allowing reciprocal complex scalings and opposite circular shifts of the two factors, which leave their convolutions unchanged.

The common support is not directly visible in these observations. The inverse-filter JSBD method of Wang et al. (2026) searches for an inverse whose outputs are jointly sparse. Each candidate inverse also filters the measurement noise, so weak kernel frequencies can amplify noise before sparsity is assessed. Let $\mathcal{F}$ denote the discrete Fourier transform (DFT), write $\widehat{g} = \mathcal{F}g$, and assume $\widehat{g}[k] \neq 0$ for all k. The ratio $\kappa(g) = \max_k |\widehat{g}[k]| / \min_k |\widehat{g}[k]|$ measures kernel conditioning. Severe ill-conditioning and low signal-to-noise ratio (SNR) make this inverse-filter criterion difficult to use. Estimating support structure first gives the inverse a direct target: it favors recovered signals whose energy is concentrated on the estimated active locations, before sparsity is judged after filtering the noise.

The shared kernel also suggests how to estimate support before inversion. In the Fourier domain, it multiplies every channel entry at a given frequency by the same complex scalar. In the noiseless

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.

model, squared magnitudes of normalized inner products between Fourier rows cancel this scalar. Because the signals share a support, comparisons between rows at different frequency separations aggregate information about pairwise circular offsets between active locations. We encode these normalized pair counts by the *support autocorrelation* a_S, which is unchanged by a common shift of S. Under independent zero-mean Gaussian active coefficients with equal variance, the comparisons converge to a_S after inverse Fourier transformation and zero-offset normalization as the channel count grows.

At finite channel count, direct application of the large-channel relation can produce a biased estimate of a_S. Noise at weak kernel frequencies also perturbs the normalized comparisons. *Quotient-Informed Recovery Network* (QuIR) learns this map from the comparisons and Fourier-row powers, using known training supports to form the targets. The powers provide information about the reliability of different frequency rows and remain kernel dependent. An approximate decoder converts the predicted autocorrelation into a candidate support $\widetilde{S}$. Since the counts are reflection invariant, the original observations are used to compare $\widetilde{S}$ and $-\widetilde{S}$.

The candidate determines the inverse objective. For $T \in \{\widetilde{S}, -\widetilde{S}\}$, let $u_\ell(h) = \mathcal{F}^{-1}(h \odot \widehat{y}_\ell)$ be the signal produced by inverse frequency response h, where $\odot$ denotes elementwise multiplication. QuIR minimizes the fraction of recovered energy outside T:

$$\widehat{h}_T \in \arg\min_{h \neq 0} \frac{\sum_{\ell=1}^{L} \sum_{n \notin T} |u_\ell(h)[n]|^2}{\sum_{\ell=1}^{L} \|u_\ell(h)\|_2^2}. \tag{3}$$

The ratio is defined for nonzero total output energy. Thus the learned representation specifies which output rows are penalized; for fixed T, the inverse is obtained from an eigenvalue problem. All recovered coordinates are retained, so the final support can differ from T. Support structure serves as an intermediate constraint for recovering the signals and kernel. Applying this strategy to another inverse problem requires both an observable structural target before inversion and a recovery criterion that can use its estimate.

Contributions. QuIR makes three contributions. First, it introduces a recovery architecture in which a learned estimate of support autocorrelation is decoded into a candidate support and used to define the relative off-support-energy criterion in eq. (3). Unlike inverse-filter methods that score each candidate by output sparsity, this criterion uses the decoded structure before fitting the inverse. Second, we analyze how representation and candidate errors propagate through orientation selection and inverse estimation to the final support, giving conditional recovery guarantees in terms of representation separation, inverse residuals, eigengaps, noise, and row-energy margins. Third, we establish the observation-to-representation link: under the stated coefficient model, a kernel-invariant Fourier-row statistic converges to normalized support autocorrelation as the channel count grows. We also show that the anchor-normalized statistic used here has total mass at most L and therefore cannot equal the target, whose mass is K, when $L < K$.

Related work. Blind-identification methods exploit cross-channel subspaces to identify a shared filter (Moulines et al., 1995), while theoretical studies of blind deconvolution establish recovery conditions under subspace, short-filter, compressive, and sparsity assumptions (Ahmed et al., 2014; Chi, 2016; Li et al., 2016; Paris and Jacques, 2019; Mulleti et al., 2020). Sparse multichannel methods incorporate the signal prior by optimizing sparsity over an inverse filter or by jointly optimizing the signal–kernel factors (Li and Bresler, 2019; Qu et al., 2020; Shi and Chi, 2021). The closest JSBD method promotes joint row sparsity within the inverse-filter optimization (Wang et al., 2026); hence, common-support information is used when candidate inverses are evaluated through their recovered outputs. QuIR changes the role and timing of this information. It first estimates a shift-invariant support autocorrelation from kernel-invariant Fourier-row statistics, decodes a realizable support candidate, and then uses that candidate to define the relative off-support-energy criterion for inverse recovery. The learned component corrects the finite-channel and noisy structural statistic, while the final inverse remains a model-based eigenvalue problem applied to the original observations. Related learning-based approaches include adaptive reconstruction across operators (Gossard and Weiss, 2024), image-specific variational priors (Huo et al., 2023), and unrolled sparse multichannel recovery (Tolooshams et al., 2023). Autocorrelation and moment representations are also used in sparse phase retrieval and multireference alignment (Ranieri et al., 2013; Baechler et al., 2019;

Bendory et al., 2018; Perry et al., 2019); in those settings, the representation is formed from the corresponding transformed observations or moments, whereas the unknown convolution in JSBD makes estimating the support autocorrelation itself a separate problem. Finally, quotient-manifold methods identify equivalent kernel–signal factorizations during optimization (Huang and Hand, 2018). The quotient in QuIR has a different role: it identifies Fourier rows that differ by a nonzero complex scalar, thereby removing the shared kernel scaling before support-structure estimation.

2 QuIR: Quotient-Informed Recovery Network

QuIR uses the decoded support candidate to specify which output rows are penalized during inverse fitting. Stage I constructs statistics that contain information about support differences. Stage II estimates their normalized counts and decodes a candidate. Stage III fits an inverse to the original observations using that candidate. Figure 1 summarizes these roles.

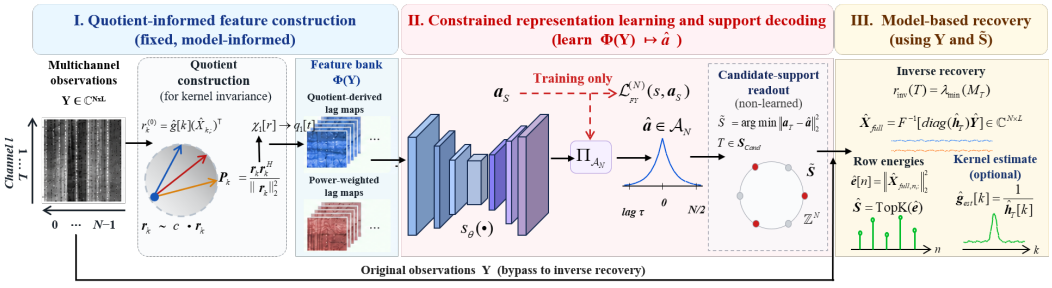

Figure 1: **Overview of QuIR. Stage I: Quotient-informed feature construction** forms statistics from normalized Fourier rows and row powers. **Stage II: Constrained representation learning and support decoding** predicts normalized support autocorrelation $\widehat{a}$ and decodes a candidate $\widetilde{S}$. **Stage III: Model-based recovery** uses $T \in \{\widetilde{S}, -\widetilde{S}\}$ to define the relative off-support energy objective on the original observations and recovers all signal coordinates.

2.1 Stage I: Quotient-informed feature construction

Support differences as the target. An absolute support origin cannot be identified because a signal shift can be absorbed by the opposite kernel shift. Relative locations can instead be recorded by their differences. A cyclic offset t is a displacement modulo N; for example, active positions i and j contribute one ordered pair at $t = i - j$. With $b_S[n] = \mathbf{1}\{n \in S\}$, define

$$c_S[t] = \sum_{n \in \mathbb{Z}_N} b_S[n] b_S[n-t], \qquad a_S[t] = \frac{c_S[t]}{|S|}. \tag{4}$$

The first quantity is the circular autocorrelation of the binary support indicator; the second normalizes its zero-offset count. They satisfy

$$a_{S+\tau} = a_{-S} = a_S, \qquad a_S[0] = 1, \qquad \mathbf{1}^{\mathsf{T}} a_S = |S|. \tag{5}$$

Reflection also preserves the counts. Distinct supports beyond shifts and reflections can share them; such supports are called homometric. Thus a_S encodes support geometry without necessarily identifying a unique support.

Removing the kernel from clean Fourier rows. Write $\widehat{Y} = \mathcal{F}Y$, applying the DFT columnwise. At frequency k, the Fourier row $r_k = (\widehat{Y}_{k,:})^{\mathsf{T}} \in \mathbb{C}^L$ collects one coefficient from each channel. In the noiseless model, $r_k^{(0)} = \widehat{g}[k](\widehat{X}_{k,:})^{\mathsf{T}}$, so the kernel only rescales this vector by a nonzero complex number. Its direction, including vectors that differ by a common complex phase, is represented by the rank-one orthogonal projector $P_k = r_k r_k^H / \|r_k\|_2^2$ for $r_k \neq 0$. A projector retains the complex line spanned by the row and is unchanged by $r_k \mapsto c r_k$, $c \neq 0$. This is the Fourier-row quotient, a point of $\mathbb{CP}^{L-1}$. It removes kernel scaling from clean rows; the shift and reflection invariances of a_S concern a separate ambiguity in the represented support.

Compare these row directions at frequency separation r through

$$\chi_1[r] = \frac{1}{N}\sum_k \operatorname{tr}(P_k P_{k-r}), \qquad q_1 = \mathcal{F}^{-1}\chi_1. \tag{6}$$

Here $\operatorname{tr}(P_k P_j) = |r_k^H r_j|^2/(\|r_k\|_2^2\|r_j\|_2^2)$ compares two normalized directions without their common phases. The inverse DFT converts frequency separation r into cyclic offset t, the coordinate of the support differences above.

The large-channel relation and its finite-channel limitation. Assume $\widehat{g}[k] \neq 0$ at every frequency and, for fixed N and S, independent $\mathcal{CN}(0,1)$ active coefficients across locations and channels. The normalized inner product between clean signal rows at separation r converges to $K^{-1}\sum_{n\in S} e^{-2\pi i r n/N}$. Its squared magnitude sums over pairs of active locations, with each term depending on their difference. Consequently, the noiseless statistic satisfies

$$q_1^{(0)} \longrightarrow q_S := \frac{c_S}{K^2} \quad (L\to\infty), \qquad \frac{q_S}{q_S[0]} = a_S. \tag{7}$$

This provides a target that can be estimated before fitting an inverse. At finite L, however, direct normalization $a_{\mathrm{ana}} := q_1/q_1[0]$ obeys a different mass constraint. For $R_L = \sum_k P_k$,

$$\mathbf{1}^{\mathsf{T}} a_{\mathrm{ana}} = \frac{(\operatorname{tr} R_L)^2}{\operatorname{tr}(R_L^2)} \le \operatorname{rank} R_L \le L. \tag{8}$$

Since $\mathbf{1}^{\mathsf{T}} a_S = K$, this particular estimator cannot equal a_S when $L < K$, even without noise. This restriction neither rules out other estimators nor establishes that learning is necessary. Finite-channel fluctuations also affect $L \ge K$, and noise perturbs the row directions and removes their kernel independence. QuIR therefore estimates a_S from these statistics together with row-power information. The feature tensor $\Phi(Y)$ includes unweighted and power-weighted comparisons and power summaries; the latter two remain kernel dependent. Derivations and the feature specification are in sections A.2, A.3 and D.

2.2 Stage II: Constrained representation learning and support decoding

Stage II learns the map $\Phi(Y) \mapsto \widehat{a}$: the input contains finite-channel statistics, while the target a_S is defined by support-difference counts. The prediction is constrained to the following convex outer relaxation of valid autocorrelations:

$$\mathcal{A}_N := \{a \in \mathbb{R}^N : a[0] = 1,\ a[t] = a[-t],\ 0 \le a[t] \le 1 \text{ for all } t \in \mathbb{Z}_N\}. \tag{9}$$

Every normalized support autocorrelation belongs to $\mathcal{A}_N$, although not every point in $\mathcal{A}_N$ is realizable by a discrete support. The estimator uses hybrid circular local/spectral blocks and outputs a length-N score $s = s_\theta(\Phi(Y))$; widths, Fourier modes, and training details are given in section D.

Let $\Omega(a) = \frac{1}{2}\|a\|_2^2 + \iota_{\mathcal{A}_N}(a)$, where $\iota_{\mathcal{A}_N}$ is zero on $\mathcal{A}_N$ and $+\infty$ outside it, and let Ω^* denote its convex conjugate. The constrained prediction is $\widehat{a} = \nabla\Omega^*(s) = \Pi_{\mathcal{A}_N}(s)$. During supervised training, a_S is computed from the known support, and the Fenchel–Young loss (Blondel, 2019; Blondel et al., 2020) is

$$\mathcal{L}_{\mathrm{FY}}^{(N)}(s, a_S) = \frac{\Omega^*(s) + \frac{1}{2}\|a_S\|_2^2 - \langle s, a_S\rangle}{N}. \tag{10}$$

For this choice of Ω, $\mathcal{L}_{\mathrm{FY}}^{(N)}(s, a_S) \ge \|\widehat{a} - a_S\|_2^2/(2N)$, so the loss bounds the squared representation error for each supervised example. This standard loss supplies a score gradient compatible with the constrained prediction; its role is separate from the choice of representation and features. The projection enforces the same constraints during training and inference.

The network predicts a continuous structural quantity, but inverse recovery requires a discrete candidate. Since $\widehat{a}$ need not correspond to a discrete support, a non-learned decoder, reproducible under a fixed seed protocol, approximately solves $\widetilde{S} \approx \arg\min_{T\in\mathcal{S}_{\mathrm{cand}}} \|a_T - \widehat{a}\|_2^2$. The candidate family contains the support cardinalities considered by the decoder, and the implemented search is approximate. The decoder supplies a realizable candidate, but need not identify the true support. Since $a_T = a_{-T}$, Stage III compares the two candidate orientations using the original observations. This comparison is not guaranteed to distinguish all supports sharing the same autocorrelation; recovery with an inexact candidate is analyzed in Section 3.

2.3 Stage III: Model-based recovery

The candidate support specifies which output rows are penalized in eq. (3). We now express that energy ratio as a generalized Rayleigh quotient, which gives both an inverse estimate and a score for comparing the two orientations. For $T \in \{\widetilde{S}, -\widetilde{S}\}$, define $\mathcal{E}_T = \{u \in \mathbb{C}^N : u[n] = 0 \text{ for } n \notin T\}$ and let P_T be the orthogonal projector onto $\mathcal{F}\mathcal{E}_T$. With $D_\ell = \mathrm{diag}(\widehat{y}_\ell)$ and $p[k] = \sum_\ell |\widehat{y}_\ell[k]|^2$, set

$$D_p = \mathrm{diag}(p) = \sum_\ell D_\ell^H D_\ell, \quad B_T = \sum_\ell D_\ell^H (I - P_T) D_\ell. \tag{11}$$

For inverse candidate h, $h^H B_T h$ and $h^H D_p h$ equal N times the recovered signal energy outside T and the total recovered signal energy, respectively, under our DFT convention. When $D_p \succ 0$, QuIR solves

$$r_{\mathrm{inv}}(T) = \min_{h \neq 0} \frac{h^H B_T h}{h^H D_p h} = \lambda_{\min}(M_T), \quad M_T = D_p^{-1/2} B_T D_p^{-1/2}. \tag{12}$$

For a unit bottom eigenvector $\widehat{z}_T$ of M_T, the inverse estimate is $\widehat{h}_T = D_p^{-1/2}\widehat{z}_T$, so $\widehat{h}_T^H D_p \widehat{h}_T = 1$. The same Rayleigh residual is evaluated for $\widetilde{S}$ and $-\widetilde{S}$; the smaller value selects the orientation. Correct orientation selection requires the residual separation analyzed in section 3; comparing these two orientations does not search over all homometric alternatives.

The selected inverse is applied to the original observations, $\widehat{X}_{\mathrm{full}} = \mathcal{F}^{-1}[\mathrm{diag}(\widehat{h}_T)\widehat{Y}]$, with row energies $\widehat{e}[n] = \|\widehat{X}_{\mathrm{full},n:}\|_2^2$. All output coordinates are retained. In the support-recovery evaluation, $\widehat{S} = \mathrm{TopK}(\widehat{e})$ selects the K largest row energies and need not equal the intermediate candidate. The true scene cardinality is supplied only for this final readout; feature construction, representation estimation, and candidate search do not receive it. A forward-kernel estimate can be read out by taking the elementwise reciprocals of the nonzero entries of $\widehat{h}_T$; implementation details and numerical safeguards are deferred to sections D and D.5.

3 Recovery Analysis

The estimate $\widehat{a}$ affects recovery through the candidate that defines the inverse objective. We first quantify when representation and search errors preserve the true support class, and when the original observations distinguish its reflection. A candidate outside that class can still produce an accurate inverse because the solver retains all output coordinates. We therefore also bound inverse and waveform errors for an arbitrary candidate, then determine when the resulting row energies preserve the final support. These results give conditional guarantees for the recovery stages; they do not guarantee the prediction accuracy of a trained network.

Autocorrelation separation margin for support decoding. To quantify when representation error can change the decoded support class, write $[S] = \{S + \tau : \tau \in \mathbb{Z}_N\}$ for the shift class and $\mathcal{O}(S) = [S] \cup [-S]$ for the shift/reflection class. Define the distance from the true normalized support autocorrelation to the nearest non-equivalent candidate,

$$\Delta_{\mathrm{ac}}(S) = \min_{T \in \mathcal{S}_{\mathrm{cand}} \setminus \mathcal{O}(S)} \|a_T - a_S\|_2, \tag{13}$$

with an empty minimum interpreted as $+\infty$.

Proposition 1 (Autocorrelation-based support decoding)**.** *Suppose $S \in \mathcal{S}_{\mathrm{cand}}$ and the decoder satisfies*

$$\|a_{\widetilde{S}} - \widehat{a}\|_2 \leq \min_{T \in \mathcal{S}_{\mathrm{cand}}} \|a_T - \widehat{a}\|_2 + \varepsilon_{\mathrm{dec}}. \tag{14}$$

Then

$$2\sqrt{2N\mathcal{L}_{\mathrm{FY}}^{(N)}(s, a_S)} + \varepsilon_{\mathrm{dec}} < \Delta_{\mathrm{ac}}(S) \quad \Longrightarrow \quad \widetilde{S} \in \mathcal{O}(S). \tag{15}$$

The condition separates representation error, search error, and intrinsic ambiguity of the representation. In particular, a non-equivalent homometric support has the same autocorrelation and gives $\Delta_{\mathrm{ac}}(S) = 0$. The implemented search has no certified $\varepsilon_{\mathrm{dec}}$; section B.4 states the corresponding sufficient condition for an exact decoder.

Conditions for resolving reflection from the original measurements. Proposition 1 guarantees decoding only up to shift and reflection because $a_S = a_{-S}$. Stage III therefore uses the original observations to decide the remaining orientation. Assume $\widehat{g}[k] \neq 0$ for all k, let $h^\star[k] = 1/\widehat{g}[k]$, $\Lambda(S,T) = \{\tau : S + \tau \subseteq T\}$, and define the support stabilizer $\mathcal{H}_S := \{\tau \in \mathbb{Z}_N : S + \tau = S\}$. Superscript (0) denotes noiseless matrices with their own power normalization.

Proposition 2 (Clean inverse-filter nullspace)**.** *For noiseless data with $\widehat{g}[k] \neq 0$ for every k,*

$$h \in \ker B_T^{(0)} \iff d \circledast x_\ell \in \mathcal{E}_T \quad \forall \ell, \quad d = \mathcal{F}^{-1}(\widehat{g} \odot h). \tag{16}$$

If the channels span $\mathcal{E}_S$, then

$$\ker B_T^{(0)} = \{h : \mathcal{F}^{-1}(\widehat{g} \odot h) \in \mathcal{E}_{\Lambda(S,T)}\}. \tag{17}$$

Under this spanning condition, an aperiodic support satisfying $\mathcal{H}_S = \{0\}$ together with $T = S + \tau$ gives a one-dimensional shifted-inverse nullspace. A same-cardinality $T \notin [S]$ has a strictly positive clean inverse-filter residual when the clean power normalization is positive definite.

Thus a correct shifted orientation $T_+ \in [S]$ has zero residual, whereas a non-equivalent reflected $T_- \notin [S]$ has residual margin $\delta_{\mathrm{ref}} := \lambda_{\min}(M_{T_-}^{(0)}) > 0$. Its noisy ordering is preserved if $\|M_{T_+} - M_{T_+}^{(0)}\|_2 + \|M_{T_-} - M_{T_-}^{(0)}\|_2 < \delta_{\mathrm{ref}}$. Full span requires $L \geq K$; lower-channel sufficient rank conditions are given in section C.3.

Stability with an inexact candidate. The preceding results identify conditions under which the decoder recovers the true shift/reflection class. The stability analysis below also permits an arbitrary candidate T. Let $z_T^\star$ denote the scale/shift-aligned true inverse direction that minimizes the candidate Rayleigh residual; the explicit phase-ramp alignment is given in section C.4. Define $\eta_T = (z_T^\star)^H M_T z_T^\star$ and $\gamma_T = \lambda_2(M_T) - \lambda_1(M_T)$. Here η_T measures leakage of the aligned true inverse through the candidate constraint, while γ_T separates the best inverse direction from the next one.

Proposition 3 (Inverse stability with an inexact candidate)**.** *If $D_p \succ 0$ and $\gamma_T > 0$, a unit bottom eigenvector $\widehat{z}_T$ satisfies*

$$\sin^2 \angle(\widehat{z}_T, z_T^\star) \leq \frac{\eta_T - \lambda_1(M_T)}{\gamma_T} \leq \frac{\eta_T}{\gamma_T}, \tag{18}$$

where $\sin^2 \angle(u,v) = 1 - |u^H v|^2$ for unit vectors.

The result also applies when $T \notin \mathcal{O}(S)$: an inexact candidate support can still yield an accurate inverse direction if its true-inverse leakage is small relative to the eigengap. The leakage η_T uses the unknown true inverse and is an analysis quantity, not an inference-time test.

From inverse error to final support. Let $\tau_T^\star$, $X_T^\star$, and ν_T denote the candidate-dependent alignment shift, the corresponding scale/shift-aligned true waveform, and the filtered-noise term defined in sections C.4 and C.5. The key waveform error level is $\delta_X := \sqrt{2\eta_T/(N\gamma_T)} + \nu_T$.

Proposition 4 (Waveform and support stability)**.** *Under the assumptions of Proposition 3, for some $|\alpha| = 1$,*

$$\|\widehat{X}_{\mathrm{full}} - \alpha X_T^\star\|_F \leq \delta_X. \tag{19}$$

Write $e_T^\star[n] = \|X_{T,n:}^\star\|_2^2$ and $m_T := \max_n \|X_{T,n:}^\star\|_2$. Then

$$\|\widehat{e} - e_T^\star\|_\infty \leq \delta_X(2m_T + \delta_X). \tag{20}$$

If Δ_E is the gap between the smallest active and largest inactive row energies of $X_T^\star$, then

$$2\delta_X(2m_T + \delta_X) < \Delta_E \quad \Longrightarrow \quad \mathrm{TopK}(\widehat{e}) = S + \tau_T^\star. \tag{21}$$

An intermediate candidate can therefore differ from the final support. Leakage and eigengap control inverse-direction error; filtered noise and the row-energy margin then determine whether that error changes the Top-K support. Together, Propositions 1–4 trace the effect of representation and candidate errors through the recovery pipeline. The bound concerns the inverse-filtered signals, not direct error in the reciprocal forward-kernel readout. The derivation is in section C.5.

4 EVALUATION

We first compare analytic and learned autocorrelation estimates with the decoder and inverse solver fixed. This paired diagnostic measures the effect of replacing the representation estimate. The larger phase-grid experiments compare complete methods across noise, conditioning, and channel count. Further diagnostics examine inexact candidates and factor recovery, followed by measured-data evaluation. Exact support, latent-waveform error, and recovered-kernel correlation are reported separately.

Protocol. All learnable components and recovery rules were fixed before the final evaluation. In the synthetic evaluation, a scene is one generated problem instance, and a cell fixes one configuration of the generator parameters. QuIR is evaluated on 114,200 scenes across 1,142 cells. NeuralBD-JSBD and VDIP-JSBD are task-specific JSBD ports of components from Schirninger et al. (2026) and Huo et al. (2023), respectively, and are evaluated on the same scenes. Traditional JSBD uses a preserved archive on matched grids but separate scenes, so that pooled comparison is unpaired. The cardinality-informed baseline uses the same QuIR scenes and receives the true cardinality during structural estimation. Other components also differ, so this comparison does not isolate the effect of cardinality access. The learned representation estimator in QuIR does not receive scene-specific K, SNR, or κ; K enters only at the shared final Top-K readout. Protocol details are in sections D and E.1 to E.3. The primary metric is exact support recovery up to circular shift, without reflection alignment.

What the learned representation contributes. Over 2,900 diagnostic scenes, learning reduces mean autocorrelation MSE from 1.438×10^{-3} to 7.461×10^{-4}. On a paired 290-scene subset (10 scenes per cell), with the decoder and inverse solver fixed, replacing the analytic estimate by the learned one raises exact recovery from $245/290$ to $270/290$. A separate ablation uses 2,100 scenes per variant, for 6,300 scene-runs, with the same downstream stages. With FY fixed, restoring the full row-power-augmented features raises final exact recovery from 65.71% to 71.76%. With full features fixed, replacing projected MSE by FY changes it from 71.52% to 71.76%. Thus the larger observed final-recovery gain comes from the features; details are in section E.4.

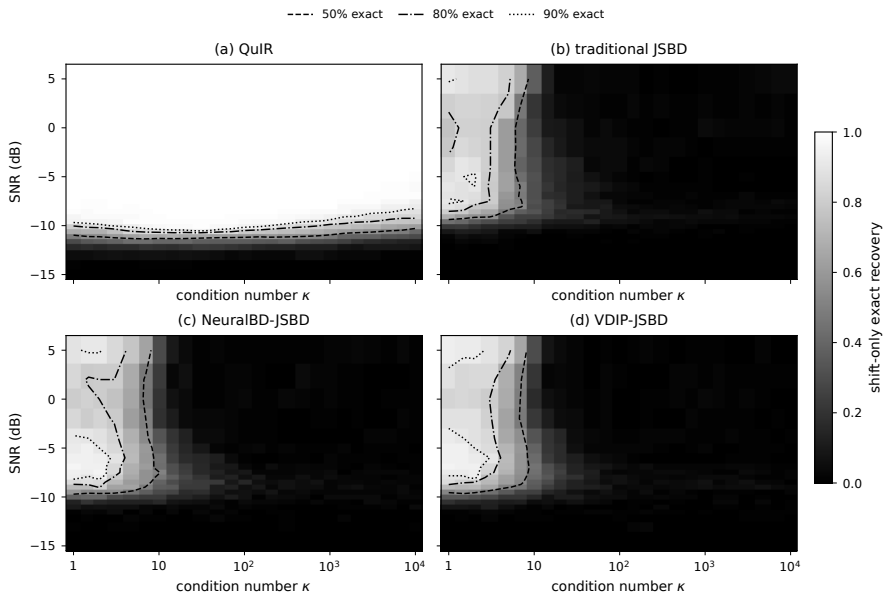

Figure 2: **Noise–conditioning phase.** Exact-support recovery on the same 525-cell SNR–κ grid for (a) QuIR, (b) traditional JSBD, (c) NeuralBD-JSBD, and (d) VDIP-JSBD; contours mark 50%/80%/90% recovery. Panels (a), (c), and (d) use the same scenes; panel (b) uses independent scenes at the same grid settings.

Robustness to noise and kernel conditioning. Across the SNR–κ grid, exact recovery is 73.6% for QuIR, compared with 13.5% for traditional JSBD on the matched grid and 15.5%/14.8% for

Table 1: **Same-scene full-pipeline comparison with a cardinality-informed baseline.** The baseline receives true cardinality during autocorrelation estimation.

slice	scenes	QuIR	baseline	Δ (pp)
all	114,200	87.32	84.04	+3.28
SNR–κ	52,500	73.61	69.03	+4.57
low-L K–L	1,800	97.22	84.89	+12.33
low-L N–L	4,200	87.05	71.81	+15.24

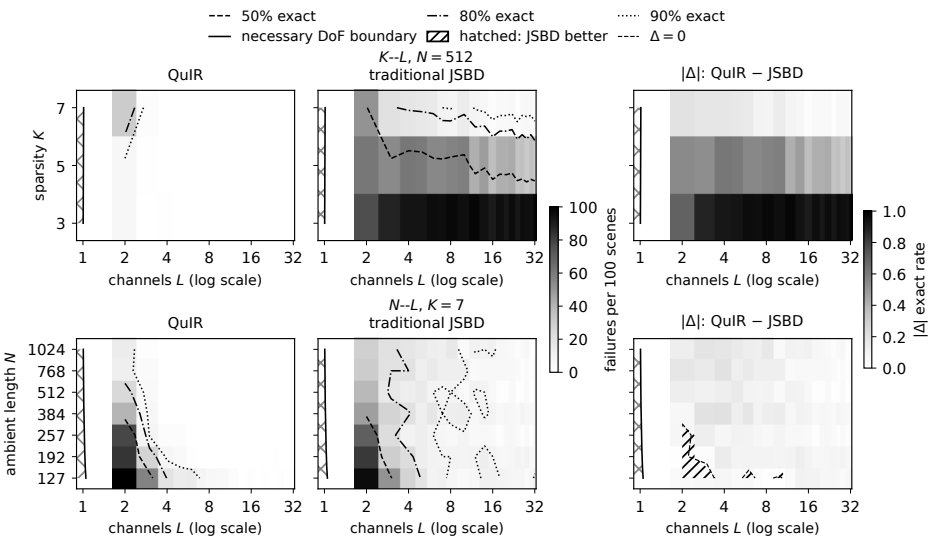

Figure 3: **Finite-channel and finite-N boundaries.** QuIR, traditional JSBD, and their exact-rate difference after merging core and low-L scans; the solid curve is the necessary dimension boundary.

same-scene NeuralBD-JSBD/VDIP-JSBD. The phase map shows the largest separation in the noisy and ill-conditioned region; high-condition-number extensions are reported in fig. 6 and section E.6.

Complete-pipeline comparison with the cardinality-informed baseline. On shared scenes, the baseline succeeds in 95,969/114,200 cases versus 99,715/114,200 for QuIR (5,063 QuIR-only; 1,317 baseline-only). Each axis reuses 100 seeds across cells, so counts are descriptive, not independent trials. Gains are largest at low L and modest on saturated core grids.

Finite-channel recovery boundary. The low-L scans show a finite-channel boundary: recovery degrades sharply at the smallest channel counts even among parameter cells satisfying the necessary dimension count in eq. (43). Figure 3 also shows where traditional JSBD is locally stronger.

Recovery from inexact candidate supports. In a separate 2,000-scene diagnostic, 607 of the 895 candidates outside the true shift/reflection class still yield exact final support. The full-coordinate readout can therefore succeed with an inexact intermediate candidate. Across the final evaluation, larger observed inverse eigengaps are associated with successful recovery (fig. 4). The archive does not contain the true-inverse leakage η_T, so this association alone does not test the bound in Proposition 3. In the 2,000-scene diagnostic, median aligned waveform normalized mean squared error (NMSE) is 0.402 overall and 0.375 conditional on exact support. Exact support therefore does not imply high-fidelity waveform recovery.

Measured-data transfer. On the public RENEW LensFD raw-IQ data described by Chen et al. (2023), the fixed 1-D model is evaluated without retraining or true path cardinality. Across 14 indoor locations, equal-weight mean common-waveform correlation is 0.797 for QuIR versus 0.444 for traditional JSBD, and channel row-norm correlation is 0.897 versus 0.782. The pilot is used only to form a band-limited reference after recovery; these correlations do not measure full-band physical

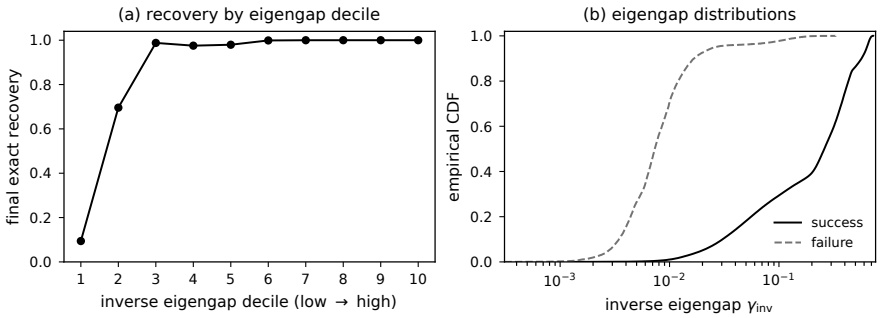

Figure 4: **Inverse-direction separation and final recovery.** Final recovery versus observed inverse eigengap across the 114,200-scene final evaluation.

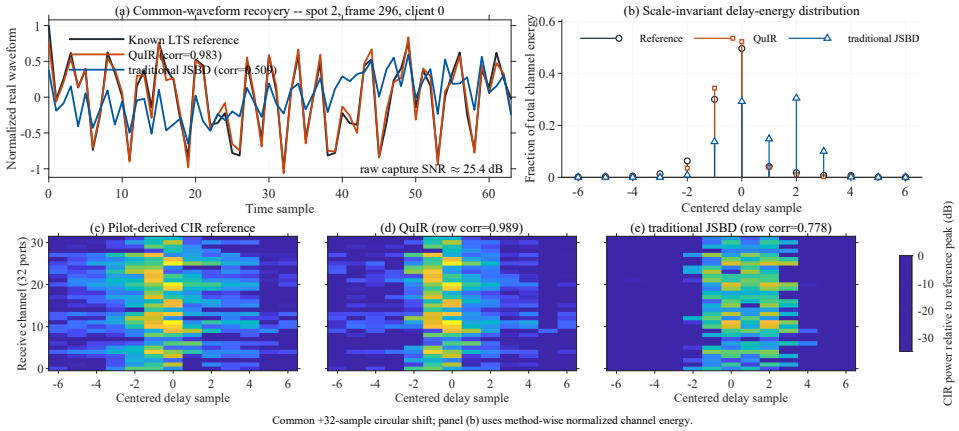

Figure 5: **Measured 1-D signal-and-kernel recovery on LensFD.** Illustrative capture selected for high QuIR–reference correlations. (a) Common-waveform recovery against the long-training-sequence reference; (b) scale-invariant delay-energy distribution; (c)–(e) reference, QuIR, and traditional JSBD channel impulse responses across 32 receive ports. References are used only post hoc; circular alignment resolves the intrinsic shift ambiguity.

ground-truth error. The 2-D microscopy extension uses a separate synthetically trained model and provides qualitative results without latent-image or measured-kernel ground truth. Acquisition, alignment, and 2-D details are in section E.5.

5 Discussion and Conclusion

QuIR estimates normalized support autocorrelation before inverse filtering and uses a decoded candidate to minimize the relative energy outside that candidate. Shared kernel scaling and a common sparse support make this intermediate representation accessible through Fourier-row statistics under the stated coefficient model. Learning improves on direct normalization in the paired diagnostic, while the complete-method experiments show gains across the evaluated noise and conditioning grids and on measured 1-D data.

The approach suggests estimating structure before an unknown operator when two links are available: observations contain information about the structure without operator inversion, and the estimated structure defines a subsequent recovery criterion. JSBD provides these links here; extending the idea requires establishing them for the new model. The autocorrelation representation retains reflection and possible homometric ambiguities, and the recovery guarantees remain conditional on representation, candidate, and noise errors. The waveform diagnostic and the near-noiseless distribution-shift test also show limits: exact support need not yield accurate amplitudes, and learning need not outperform the analytic estimate in every regime.

AI USE STATEMENT

Generative-AI tools assisted with literature retrieval and discovery, as well as the organization, drafting, and language editing of this paper. They also assisted with mathematical checking, figure planning, code and experiment workflows, and preparation of supplementary materials. Numerical claims were computed from preserved fixed-protocol result files. The authors reviewed all AI-assisted material and take responsibility for the cited work, final text, mathematical claims, code, and experimental results.

REFERENCES

A. Ahmed, B. Recht, and J. Romberg. Blind deconvolution using convex programming. *IEEE Transactions on Information Theory*, 60(3):1711–1732, 2014.

G. Baechler, M. Krekovic, J. Ranieri, A. Chebira, Y. M. Lu, and M. Vetterli. Super resolution phase retrieval for sparse signals. *IEEE Transactions on Signal Processing*, 67(18):4839–4854, 2019.

T. Bendory, N. Boumal, C. Ma, Z. Zhao, and A. Singer. Bispectrum inversion with application to multireference alignment. *IEEE Transactions on Signal Processing*, 66(4):1037–1050, 2018.

J. J. Benedetto and M. Fickus. Finite normalized tight frames. *Advances in Computational Mathematics*, 18:357–385, 2003.

M. Blondel. Structured prediction with projection oracles. *Advances in Neural Information Processing Systems*, 32:12145–12156, 2019.

M. Blondel, A. F. T. Martins, and V. Niculae. Learning with Fenchel–Young losses. *Journal of Machine Learning Research*, 21(35):1–69, 2020.

Z. Chen, C. N. Barati, J. Veihl, C. Shepard, and A. Sabharwal. LensFD: Using lenses for improved sub-6 GHz massive MIMO full-duplex. *IEEE Transactions on Vehicular Technology*, 72(6):7517–7529, 2023. doi:10.1109/TVT.2023.3240558.

Y. Chi. Guaranteed blind sparse spikes deconvolution via lifting and convex optimization. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):782–794, 2016.

C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

A. Gossard and P. Weiss. Training adaptive reconstruction networks for blind inverse problems. *SIAM Journal on Imaging Sciences*, 17(2):1314–1346, 2024. doi:10.1137/23M1545628.

D. Huo, A. Masoumzadeh, R. Kushol, and Y.-H. Yang. Blind image deconvolution using variational deep image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):11472–11483, 2023. doi:10.1109/TPAMI.2023.3283979.

W. Huang and P. Hand. Blind deconvolution by a steepest descent algorithm on a quotient manifold. *SIAM Journal on Imaging Sciences*, 11(4):2757–2785, 2018.

Y. Lau, Q. Qu, H.-W. Kuo, P. Zhou, Y. Zhang, and J. Wright. Short-and-sparse deconvolution—a geometric approach. In *International Conference on Learning Representations*, 2020.

Y. Li and Y. Bresler. Multichannel sparse blind deconvolution on the sphere. *IEEE Transactions on Information Theory*, 65(11):7415–7436, 2019.

Y. Li, K. Lee, and Y. Bresler. Identifiability in blind deconvolution with subspace or sparsity constraints. *IEEE Transactions on Information Theory*, 62(7):4266–4275, 2016.

E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue. Subspace methods for the blind identification of multichannel FIR filters. *IEEE Transactions on Signal Processing*, 43(2):516–525, 1995.

S. Mulleti, K. Lee, and Y. C. Eldar. Identifiability conditions for compressive multichannel blind deconvolution. *IEEE Transactions on Signal Processing*, 68:4627–4642, 2020.

V. Niculae, A. F. T. Martins, M. Blondel, and C. Cardie. SparseMAP: Differentiable sparse structured inference. In *ICML*, volume 80 of PMLR, pp. 3799–3808, 2018.

A. Paris and L. Jacques. Identifiability conditions for multi-channel blind deconvolution with short filters. arXiv:1902.09151, 2019.

A. Perry, J. Weed, A. S. Bandeira, P. Rigollet, and A. Singer. The sample complexity of multi-reference alignment. *SIAM Journal on Mathematics of Data Science*, 1(3):497–517, 2019.

Q. Qu, X. Li, and Z. Zhu. Exact recovery of multichannel sparse blind deconvolution via gradient descent. *SIAM Journal on Imaging Sciences*, 13(3):1630–1652, 2020.

J. Ranieri, A. Chebira, Y. M. Lu, and M. Vetterli. Phase retrieval for sparse signals: Uniqueness conditions. arXiv:1308.3058, 2013.

C. Schirninger, R. Jarolim, A. M. Veronig, M. Rempel, and F. Wöger. Neural blind deconvolution to reconstruct high-resolution ground-based solar observations. *Astronomy & Astrophysics*, 708:A170, 2026. doi:10.1051/0004-6361/202658932.

L. Shi and Y. Chi. Manifold gradient descent solves multi-channel sparse blind deconvolution provably and efficiently. *IEEE Transactions on Information Theory*, 67(7):4784–4811, 2021.

G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, 1990.

B. Tolooshams, S. Mulleti, D. Ba, and Y. C. Eldar. Unrolled compressed blind-deconvolution. *IEEE Transactions on Signal Processing*, 71:2118–2129, 2023.

W. Wang, B. Guo, Z. Yang, S. Chen, and J. Shi. Jointly sparse blind deconvolution via Riemannian optimization. arXiv:2608.04465, 2026.

J. C. Ye, Y. Han, and E. Cha. Deep convolutional framelets: A general deep learning framework for inverse problems. *SIAM Journal on Imaging Sciences*, 11(2):991–1048, 2018.

Y. Zhang, H.-W. Kuo, and J. Wright. Structured local minima in sparse blind deconvolution. *Advances in Neural Information Processing Systems*, 31, 2018.

A Representation and Normalized Fourier-Row Statistics

We derive the normalized support-autocorrelation identities and the normalized Fourier-row statistics used in section 2, followed by finite-channel and fixed-weight noise bounds. Section B treats autocorrelation estimation and support decoding, section C analyzes inverse recovery from the observations, and section D gives implementation details. In the appendix, clean denotes noiseless observations.

A.1 Notation and normalization

Circular indices lie in $\mathbb{Z}_N$. The operator $C(g)$ is circular convolution by g, and $\mathrm{rsupp}(X) = \{n : \|X_{n,:}\|_2 > 0\}$. We use

$$\widehat{u}[k] = \sum_{n=0}^{N-1} u[n] e^{-2\pi i k n/N}, \quad u[n] = \frac{1}{N} \sum_{k=0}^{N-1} \widehat{u}[k] e^{2\pi i k n/N};$$

$\widehat{Y}$ applies this transform columnwise. Define the noiseless observations and their Fourier transform by $Y^{(0)} = C(g)X$ and $\widehat{Y}^{(0)} = \mathcal{F}Y^{(0)}$. With $r_k = \widehat{Y}^{\mathsf{T}}_{k,:}$, $r_k^{(0)} = (\widehat{Y}^{(0)}_{k,:})^{\mathsf{T}}$, and $w_k = \widehat{W}^{\mathsf{T}}_{k,:}$, we have $r_k = r_k^{(0)} + w_k$. For $r_k^{(0)} \neq 0$, let $P_k^{(0)} = r_k^{(0)}(r_k^{(0)})^H / \|r_k^{(0)}\|_2^2$. Given the same fixed realized nonnegative weights in both expressions and a positive denominator, define

$$\chi_\omega[r] = \frac{\sum_k \omega_k \omega_{k-r} \operatorname{tr}(P_k P_{k-r})}{\sum_k \omega_k \omega_{k-r}}, \quad \chi_\omega^{(0)}[r] = \frac{\sum_k \omega_k \omega_{k-r} \operatorname{tr}(P_k^{(0)} P_{k-r}^{(0)})}{\sum_k \omega_k \omega_{k-r}}.$$

For a support T, write $[T] = \{T+\tau : \tau \in \mathbb{Z}_N\}$ and $[-T] = \{-T+\tau : \tau \in \mathbb{Z}_N\}$. Let $b_T[n] = \mathbf{1}\{n \in T\}$, $c_T[t] = \sum_n b_T[n]b_T[n-t]$, and $a_T = c_T/|T|$. Let $\mathcal{E}_T = \{x : \mathrm{supp}(x) \subseteq T\}$, $\mathcal{V}_T = \mathcal{F}\mathcal{E}_T$, and let P_T be the orthogonal projector onto $\mathcal{V}_T$. Define $r_{\mathrm{inv}}(T) = \min_{h \neq 0} h^H B_T h/(h^H D_p h)$. Finally, $h^\star[k] = 1/\widehat{g}[k]$, $\widehat{z}_T = D_p^{1/2}\widehat{h}_T/\|D_p^{1/2}\widehat{h}_T\|_2$, and, for $m_\tau[k] = e^{-2\pi i k\tau/N}$,

$$h^{\star,\tau} = m_\tau \odot h^\star, \qquad z^{\star,\tau} = \frac{D_p^{1/2}h^{\star,\tau}}{\|D_p^{1/2}h^\star\|_2},$$
$$\tau_T^\star \in \arg\min_\tau (z^{\star,\tau})^H M_T z^{\star,\tau}, \quad h_T^\star = h^{\star,\tau_T^\star}, \quad z_T^\star = z^{\star,\tau_T^\star}.$$

A.2 Normalized support autocorrelation and projector identities

For $\widehat{b}_S[k] = \sum_{n\in S} e^{-2\pi i k n/N}$, circular autocorrelation gives

$$(\mathcal{F}a_S)[k] = |\widehat{b}_S[k]|^2/|S|. \tag{22}$$

Summing pair counts gives $\sum_t c_S[t] = |S|^2$, while $c_S[0] = |S|$, establishing the normalized support-autocorrelation properties in the main text. Define $\mathcal{Q}_N = \{q \in \mathbb{R}^N : q \geq 0,\ q[t] = q[-t],\ \mathbf{1}^\mathsf{T} q = 1,\ q[t] \leq q[0]\}$. The projector-statistic properties and asymptotic relation follow below. For a nonzero row y, $P(y) = yy^H/\|y\|^2$ is unchanged by $y \mapsto cy$ for $c \in \mathbb{C}^\times$. Conversely, equality of two rank-one orthogonal projectors implies equality of their one-dimensional ranges, hence the underlying rows differ by a nonzero scalar. Thus the collection (P_k) completely identifies the nonzero Fourier rows up to independent nonzero complex scalings.

For q_1, expand

$$\begin{aligned} q_1[t] &= \frac{1}{N^2}\sum_{r,k} \mathrm{tr}(P_k P_{k-r}) e^{2\pi i r t/N} \\ &= \frac{1}{N^2}\sum_{k,j} \mathrm{tr}(P_k P_j) e^{2\pi i (k-j)t/N} = \frac{1}{N^2}\mathrm{tr}(H_t H_t^H), \end{aligned}$$

where $H_t = \sum_k P_k e^{2\pi i k t/N}$. Hence

$$q_1[t] = N^{-2}\|H_t\|_F^2. \tag{23}$$

Nonnegativity and evenness follow immediately. Also $\sum_t q_1[t] = \chi_1[0] = 1$. Since $\mathrm{tr}(P_k P_j) \geq 0$,

$$q_1[t] \leq \frac{1}{N^2}\sum_{k,j} \mathrm{tr}(P_k P_j) = q_1[0],$$

so $q_1 \in \mathcal{Q}_N$.

For noiseless data with $\widehat{g}[k] \neq 0$ for all k, assume that the active coefficients are independent $\mathcal{CN}(0,1)$ variables, and let $\widehat{b}_S[k] = \sum_{n\in S} e^{-2\pi i k n/N}$. For one channel, $\xi_{k,\ell}$ and $\xi_{j,\ell}$ are proper complex Gaussian with variances $K = |S|$ and covariance $\widehat{b}_S[k-j]$. Hence, as $L \to \infty$,

$$\frac{\xi_j^H \xi_k}{\|\xi_k\|\|\xi_j\|} \to \frac{\widehat{b}_S[k-j]}{|S|} \quad \text{a.s.}$$

Taking the squared magnitude removes the phase of the limiting correlation, giving $|\widehat{b}_S[r]|^2/|S|^2$. The kernel-induced scalar was already removed by the noiseless row projectors. The inverse DFT maps $|\widehat{b}_S|^2$ to c_S, proving $q_1^{(0)}[t] \to q_S[t]$.

A.3 Anchor normalization and finite-channel effective rank

The anchor transformation and its inverse are

$$\mathcal{N}(q) := q/q[0], \quad \mathcal{N}^{-1}(a) = a/(\mathbf{1}^\mathsf{T} a). \tag{24}$$

Let $r = \operatorname{rank} R_L$ and write $R_L = (N/r)P_{\mathcal{U}} + \Delta_L$, where $P_{\mathcal{U}}$ projects onto its range. Then

$$\mathbf{1}^{\mathsf{T}} a_{\mathrm{ana}} = \frac{r}{1 + (r/N^2)\|\Delta_L\|_F^2}, \tag{25}$$

and, in the clean model where $r \le K$,

$$K - \mathbf{1}^{\mathsf{T}} a_{\mathrm{ana}} = (K - r) + \frac{r^2 \|\Delta_L\|_F^2}{N^2 + r\|\Delta_L\|_F^2}. \tag{26}$$

If $q \in \mathcal{Q}_N$, then $a = \mathcal{N}(q) = q/q[0]$ satisfies eq. (9). Conversely, for $a \in \mathcal{A}_N$, $q = \mathcal{N}^{-1}(a) = a/(\mathbf{1}^{\mathsf{T}} a)$ has unit mass, is nonnegative and even, and satisfies $q[t] \le q[0]$ because $a[t] \le 1 = a[0]$. These constructions are mutual continuous inverses, proving eq. (24).

For $R_L = \sum_k P_k$, $\operatorname{tr} R_L = N$ and

$$q_1[0] = \frac{1}{N^2} \operatorname{tr}(R_L^2).$$

Therefore

$$\mathbf{1}^{\mathsf{T}} a_{\mathrm{ana}} = \frac{1}{q_1[0]} = \frac{(\operatorname{tr} R_L)^2}{\operatorname{tr}(R_L^2)},$$

the participation-ratio effective rank. For a positive-semidefinite matrix this is at most $\operatorname{rank} R_L$, which is at most L. In the clean JSBD model, every ξ_k lies in the span of the active row vectors of X, while the restricted Fourier matrix on S has column rank K; hence $\operatorname{span}\{\xi_k\}_k = \operatorname{span}\{X_{n,:} : n \in S\}$ and $\operatorname{rank} R_L = \operatorname{rank} X_S$.

Let $r = \operatorname{rank} R_L$ and $P_{\mathcal{U}}$ project onto its range. Writing $R_L = (N/r)P_{\mathcal{U}} + \Delta_L$ with $\operatorname{tr} \Delta_L = 0$ gives

$$\operatorname{tr}(R_L^2) = \frac{N^2}{r} + \|\Delta_L\|_F^2,$$

because the cross term vanishes. Equations 25–26 follow. For a $K \times L$ matrix with absolutely continuous entries, $\operatorname{rank} X_S = \min(K, L)$ almost surely.

A.4 Clean finite-L concentration

Assume noiseless observations with $\widehat{g}[k] \neq 0$ for all k and i.i.d. proper complex Gaussian active coefficients with unit variance. Let $\xi_k = (\widehat{X}_{k,:})^{\mathsf{T}}$ and $\varrho_S[r] = \widehat{b}_S[r]/|S|$. For universal constants $c, C > 0$ and $0 < t \le 1/2$,

$$\Pr\{\|q_1^{(0)} - q_S\|_\infty > 8t\} \le CN^2 e^{-cLt^2}. \tag{27}$$

Normalize $z_{k\ell} = \xi_{k\ell}/\sqrt{K}$ and let

$$\widehat{\Sigma}_{kj} = \frac{1}{L} \sum_{\ell=1}^{L} z_{k\ell} \overline{z_{j\ell}}, \quad \rho_{kj} = \mathbb{E}\widehat{\Sigma}_{kj} = \varrho_S[k - j].$$

Products of jointly proper unit-variance complex Gaussian variables are uniformly sub-exponential. Apply real-valued Bernstein bounds to the real and imaginary parts of each centered product, followed by a union bound over both parts and the N^2 covariance entries. This gives universal $c, C > 0$ such that, for $0 < t \le 1$,

$$\Pr\left(\max_{k,j} |\widehat{\Sigma}_{kj} - \rho_{kj}| > t\right) \le CN^2 e^{-cLt^2}.$$

On the complementary concentration event with $t \le 1/2$, all empirical diagonal entries lie in $[1 - t, 1 + t]$, and the normalized empirical correlations satisfy

$$\left| \frac{\widehat{\Sigma}_{kj}}{\sqrt{\widehat{\Sigma}_{kk}\widehat{\Sigma}_{jj}}} - \rho_{kj} \right| \le \frac{2t}{1 - t} \le 4t.$$

Both correlations have magnitude at most one, so their squared magnitudes differ by at most $8t$. Averaging over k and applying the normalized inverse DFT cannot increase the sup norm, proving eq. (27).

If $a^{(0)}_{\mathrm{ana}} = \mathcal{N}(q^{(0)}_1)$ and $e = q^{(0)}_1 - q_S$ with $\|e\|_\infty \le \delta < 1/K$, then exactly

$$a^{(0)}_{\mathrm{ana}} - a_S = \frac{e - e[0]a_S}{1/K + e[0]}, \tag{28}$$

and hence

$$\|a^{(0)}_{\mathrm{ana}} - a_S\|_2 \le \frac{2K\sqrt{N}\,\delta}{1 - K\delta}. \tag{29}$$

This concentration statement applies to the ideal exact-normalization projector statistic. The implementation uses machine-scale denominator protection and clipping in its numerical q-map, so the exact identities above apply to the ideal statistic rather than implying bitwise equality for the deployed floating-point map.

A.5 Fixed-weight projector-statistic noise bound

Use the same fixed realized weights in both expressions in section A.1, and assume the clean and noisy rows are nonzero. Set $\epsilon_k = \min\{1, \|w_k\|_2/\|r^{(0)}_k\|_2\}$. Then

$$|\chi_\omega[r] - \chi^{(0)}_\omega[r]| \le 4\frac{\sum_k \omega_k\omega_{k-r}(\epsilon_k + \epsilon_{k-r})}{\sum_k \omega_k\omega_{k-r}}. \tag{30}$$

For nonzero $r^{(0)}$ and $r = r^{(0)} + w$,

$$\left\| \frac{r}{\|r\|} - \frac{r^{(0)}}{\|r^{(0)}\|} \right\|_2 \le 2\min\{1, \|w\|/\|r^{(0)}\|\}.$$

For unit u, v, u_0, v_0, the map $|\langle u, v\rangle|^2$ is 2-Lipschitz in the inner product and the inner product is 1-Lipschitz in each argument, giving $4(\epsilon_k + \epsilon_{k-r})$ per row pair. Weighted averaging proves eq. (30). This bound holds conditional on fixed weights; it does not account for changes in observation-dependent weights when they are recomputed from noisy rather than clean data. The preceding subsection separately quantifies clean coefficient-sampling error.

B Constrained Autocorrelation Estimation and Support Decoding

B.1 Projection onto the relaxed autocorrelation domain

The constraints separate over pairs $\{t, -t\}$. Minimizing the squared distance within each pair gives its mean, then clipping enforces the interval constraint; coordinate zero is fixed:

$$\widehat{a}[0] = 1, \quad \widehat{a}[t] = \min\{1, \max\{0, (s[t] + s[-t])/2\}\}, \quad t \ne 0. \tag{31}$$

The formula includes indices satisfying $t = -t \pmod N$, such as $t = N/2$ for even N. The conjugate used in eq. (10) is

$$\Omega^*(s) = \sup_{a \in \mathcal{A}_N} \{\langle s, a\rangle - \tfrac{1}{2}\|a\|_2^2\}. \tag{32}$$

B.2 Fenchel–Young score gradient and autocorrelation-error bound

Let $\widehat{a} = \Pi_{\mathcal{A}_N}(s)$. By the definition of the conjugate,

$$\Omega^*(s) = \langle s, \widehat{a}\rangle - \tfrac{1}{2}\|\widehat{a}\|_2^2.$$

Therefore, for any target $a \in \mathcal{A}_N$,

$$\begin{aligned} N\mathcal{L}^{(N)}_{\mathrm{FY}}(s, a) &= \langle s, \widehat{a} - a\rangle + \tfrac{1}{2}(\|a\|_2^2 - \|\widehat{a}\|_2^2) \\ &= \langle s - \widehat{a}, \widehat{a} - a\rangle + \tfrac{1}{2}\|\widehat{a} - a\|_2^2. \end{aligned}$$

Euclidean projection optimality gives $\langle s - \widehat{a}, a - \widehat{a}\rangle \le 0$, so the first term is nonnegative and the lower bound in eq. (10) follows. Differentiating the conjugate gives $\nabla_s \mathcal{L}^{(N)}_{\mathrm{FY}} = (\widehat{a} - a)/N$.

B.3 Population Fenchel–Young target

For a fixed N and a distribution of observation–support pairs, let $\Phi^{(0)}$ be an ideal noiseless feature representation containing $q_1^{(0)}$ and Φ^{dep} the deployed feature representation. Define $\mu_\star(\Phi^\star) = \mathbb{E}[a_S \mid \Phi^\star]$ and $V_\star = \mathbb{E}\|a_S - \mu_\star(\Phi^\star)\|_2^2$, where $\star \in \{0, \mathrm{dep}\}$. The following are population-risk identities, not finite-training-sample generalization bounds:

$$V_0 \le \mathbb{E}\|a_S - a_{\mathrm{ana}}^{(0)}\|_2^2, \tag{33}$$

$$\mathbb{E}\|\widehat{a} - a_S\|_2^2 \le V_{\mathrm{dep}} + 2N\mathcal{E}_{\mathrm{FY}}^{\mathrm{Bayes}}, \tag{34}$$

where $\mathcal{E}_{\mathrm{FY}}^{\mathrm{Bayes}}$ is the excess population FY risk over all measurable score maps, defined below. For mixed dimensions these statements apply conditional on N.

For either feature representation $\Phi^\star$, condition on $\Phi^\star = \phi$ and write $\mu_\star = \mathbb{E}[a_S \mid \Phi^\star = \phi]$. Convexity of $\mathcal{A}_N$ gives $\mu_\star \in \mathcal{A}_N$. The conditional FY risk is

$$R_{\phi,\star}(s) = \frac{1}{N}\left[\Omega^*(s) - \langle s, \mu_\star\rangle + \mathbb{E}[\Omega(a_S) \mid \Phi^\star = \phi]\right].$$

Fenchel–Young duality implies that its minimum is attained whenever $\Pi_{\mathcal{A}_N}(s) = \mu_\star$, and

$$R_{\phi,\star}(s) - R_{\phi,\star}^* = \mathcal{L}_{\mathrm{FY}}^{(N)}(s, \mu_\star) \ge \frac{1}{2N}\|\Pi_{\mathcal{A}_N}(s) - \mu_\star\|_2^2.$$

For the deployed score map define

$$\mathcal{R}_{\mathrm{FY}}(s) := \mathbb{E}\,\mathcal{L}_{\mathrm{FY}}^{(N)}(s(\Phi^{\mathrm{dep}}), a_S), \quad \mathcal{E}_{\mathrm{FY}}^{\mathrm{Bayes}} := \mathcal{R}_{\mathrm{FY}}(s_\theta) - \inf_{s \text{ measurable}} \mathcal{R}_{\mathrm{FY}}(s).$$

Averaging the conditional inequality yields $\mathbb{E}\|\widehat{a} - \mu_{\mathrm{dep}}(\Phi^{\mathrm{dep}})\|_2^2 \le 2N\mathcal{E}_{\mathrm{FY}}^{\mathrm{Bayes}}$. Since $\mathbf{1}^\mathsf{T} a_S = K$,

$$\mathbf{1}^\mathsf{T}\mu_\star(\Phi^\star) = \mathbb{E}[K \mid \Phi^\star].$$

Moreover, for every $\Phi^\star$-measurable predictor $\widetilde{a}$,

$$\mathbb{E}\|\widetilde{a} - a_S\|_2^2 = \mathbb{E}\|\widetilde{a} - \mu_\star(\Phi^\star)\|_2^2 + \mathbb{E}\|a_S - \mu_\star(\Phi^\star)\|_2^2. \tag{35}$$

For the ideal bank, $a_{\mathrm{ana}}^{(0)} = \mathcal{N}(q_1^{(0)})$ is $\Phi^{(0)}$-measurable, proving eq. (33); for fixed N and S, clean Gaussian consistency of $q_1^{(0)}$ as $L \to \infty$ then gives $V_0 \to 0$. For the deployed bank, let $q_{\mathrm{dep}} := (\Phi^{\mathrm{dep}})_{1,:}$ be its actual first lag map. Define the measurable comparator $a_{\mathrm{dep}} := \mathcal{N}(q_{\mathrm{dep}})$ when $q_{\mathrm{dep}}[0] > 0$, and $a_{\mathrm{dep}} := \mathbf{0} \in \mathbb{R}^N$ otherwise. The comparator need not lie in $\mathcal{A}_N$ for the orthogonal decomposition above. Then

$$V_{\mathrm{dep}} \le \mathbb{E}\|a_S - a_{\mathrm{dep}}\|_2^2, \tag{36}$$

an exact comparator statement for the deployed feature tensor. Taking $\widetilde{a} = \widehat{a}$ and combining with the deployed Bayes-excess bound proves eq. (34).

B.4 Autocorrelation-margin proof

Write

$$e_{\mathrm{ac}} = \|\widehat{a} - a_S\|_2 \le \sqrt{2N\mathcal{L}_{\mathrm{FY}}^{(N)}(s, a_S)}. \tag{37}$$

Since S is feasible in the candidate family, eq. (14) gives $\|a_{\widetilde{S}} - \widehat{a}\|_2 \le e_{\mathrm{ac}} + \varepsilon_{\mathrm{dec}}$. Consequently $\|a_{\widetilde{S}} - a_S\|_2 \le 2e_{\mathrm{ac}} + \varepsilon_{\mathrm{dec}}$ by the triangle inequality. If this is below $\Delta_{\mathrm{ac}}(S)$, the candidate must lie in $\mathcal{O}(S)$, proving Proposition 1. For exact decoding, the sufficient condition is

$$\mathcal{L}_{\mathrm{FY}}^{(N)}(s, a_S) < \Delta_{\mathrm{ac}}(S)^2/(8N). \tag{38}$$

A zero margin prevents this argument from distinguishing a homometric competitor.

C Inverse Identifiability and Recovery Stability

C.1 Inverse-filter nullspace and translational embedding

Because $B_T = \sum_\ell D_\ell^H P_T^\perp D_\ell \succeq 0$,

$$h \in \ker B_T \iff P_T^\perp D_\ell h = 0 \quad \forall \ell.$$

In the clean model $D_\ell h = \mathrm{diag}(\widehat{x}_\ell)(\widehat{g} \odot h)$. Setting $z = \widehat{g} \odot h$ and $d = \mathcal{F}^{-1} z$, the convolution theorem gives

$$P_T^\perp D_\ell h = 0 \iff d \circledast x_\ell \in \mathcal{E}_T,$$

proving eq. (16). If the channels span $\mathcal{E}_S$, this is equivalent to $d \circledast \mathcal{E}_S \subseteq \mathcal{E}_T$. Testing the canonical impulses δ_s, $s \in S$, shows

$$d \circledast \mathcal{E}_S \subseteq \mathcal{E}_T \iff \mathrm{supp}(d) \subseteq \bigcap_{s \in S} (T - s) = \Lambda(S, T).$$

Because multiplication by the nonzero $\widehat{g}$ and the DFT are invertible, this proves eq. (17).

If $|T| = |S|$, the inclusion $S + \tau \subseteq T$ is necessarily equality. Hence for $T = S + \tau_0$, $\Lambda(S, T) = \tau_0 + \mathcal{H}_S$, while for any same-cardinality candidate outside the shift orbit $\Lambda(S, T) = \varnothing$. If $\mathcal{H}_S = \{0\}$, the correct nullspace is one-dimensional and its unique direction is the shifted inverse $h^{\star,\tau_0}$. If in addition $D_p \succ 0$, an empty nullspace gives a strictly positive generalized Rayleigh minimum. Applying this to a reflected candidate proves clean orientation separation whenever $[-S] \neq [S]$ in the full-span regime.

C.2 Noisy reflection stability

Let $M_T^{(0)}$ denote the clean version of M_T, formed with its own clean power normalization. Assume $\lambda_{\min}(M_{T_+}^{(0)}) = 0$ and both clean and noisy power normalizations are positive definite. If the correct orientation is T_+ and the reflected candidate is T_-, then the residual ordering is preserved whenever

$$\delta_{\mathrm{ref}} = \lambda_{\min}(M_{T_-}^{(0)}) > 0, \quad \|M_{T_+} - M_{T_+}^{(0)}\|_2 + \|M_{T_-} - M_{T_-}^{(0)}\|_2 < \delta_{\mathrm{ref}}. \tag{39}$$

This follows by perturbing the two minimum Rayleigh quotients; the matrix perturbations include the observation-dependent normalization.

C.3 A sufficient low-channel rank condition

For arbitrary channel span define

$$\mathcal{K}_X(T) = \{d : d \circledast x_\ell \in \mathcal{E}_T \ \forall \ell\}.$$

Always $\mathcal{E}_{\Lambda(S,T)} \subseteq \mathcal{K}_X(T)$. Define the minimum anchor-set size

$$w_{\mathrm{anc}}(S, T) = \min \left\{ |R| : R \subseteq S, \ \bigcap_{r \in R} (T - r) = \Lambda(S, T) \right\}. \tag{40}$$

If $L \geq w_{\mathrm{anc}}(S, T)$, choose at one coefficient realization the first w_{anc} channels as impulses at the corresponding anchor locations and set the remaining channels to zero. At that realization $\mathcal{K}_X(T) = \mathcal{E}_{\Lambda(S,T)}$, so the stacked linear constraint operator has rank $N - |\Lambda(S, T)|$. This witness need not have positive probability under the coefficient law: it shows that a corresponding maximal minor, polynomial in the active channel coefficients, is not identically zero. Therefore, under any absolutely continuous coefficient law,

$$L \geq w_{\mathrm{anc}}(S, T) \quad \Longrightarrow \quad \dim \mathcal{K}_X(T) = |\Lambda(S, T)| \quad \text{a.s.} \tag{41}$$

This is a sufficient generic-rank condition, not a necessary channel threshold.

A useful corollary follows from a unique ordered support difference. If $a, b \in S$ are the unique ordered pair realizing $b - a$, then $(S - a) \cap (S - b) = \{0\}$, so the correct aperiodic candidate has $w_{\mathrm{anc}}(S, S) \leq 2$. For a reflected candidate $T = -S + \tau_0$, every two anchors intersect because $\tau_0 - a - b \in (T - a) \cap (T - b)$; hence two anchors cannot resolve reflection. If $[-S] \neq [S]$, some $c \in S$ satisfies $a + b - c \notin S$, and then $(T - a) \cap (T - b) \cap (T - c) = \varnothing$. Thus under this sufficient support condition, two generic channels certify the correct one-dimensional nullspace while three certify a positive wrong-reflection residual. These anchor counts are not claimed to be necessary; generic dense two-channel data can be more informative than two impulse anchors.

C.4 Inverse stability with an inexact candidate

Because the task accepts circular shifts, the true inverse has a corresponding phase-ramp gauge. For $m_\tau[k] = e^{-2\pi i k\tau/N}$, let

$$h^{\star,\tau} = m_\tau \odot h^\star, \quad z^{\star,\tau} = \frac{D_p^{1/2} h^{\star,\tau}}{\beta}, \quad \beta = \|D_p^{1/2} h^\star\|_2.$$

For an arbitrary candidate T, choose

$$\tau_T^\star \in \arg\min_{\tau\in\mathbb{Z}_N} (z^{\star,\tau})^H M_T z^{\star,\tau}, \quad h_T^\star = h^{\star,\tau_T^\star}, \quad \bar{h}_T^\star = h_T^\star/\beta, \quad z_T^\star = z^{\star,\tau_T^\star} = D_p^{1/2}\bar{h}_T^\star.$$

If $T \in [S]$, this is the accepted support-shift gauge and gives zero clean leakage. Let $v_1, v_2, \ldots$ be an orthonormal eigenbasis of M_T with eigenvalues $\lambda_1 \le \lambda_2 \le \cdots$, and expand $z_T^\star = \sum_i c_i v_i$. Then

$$\eta_T = \sum_i \lambda_i |c_i|^2 \ge \lambda_1 + (\lambda_2 - \lambda_1) \sum_{i\ge 2} |c_i|^2.$$

Since $\sum_{i\ge2} |c_i|^2 = \sin^2 \angle(z_T^\star, v_1)$, eq. (18) follows; $M_T \succeq 0$ gives the second inequality. The proof uses a direct variational decomposition. Davis–Kahan and related results provide classical context for eigenspace stability (Davis and Kahan, 1970; Stewart and Sun, 1990).

C.5 Waveform and row-energy stability

Using the candidate-aligned quantities defined in section C.4 and the same observation-defined D_p as for $\widehat{h}_T$, both $\bar{h}_T^\star$ and $\widehat{h}_T$ satisfy $h^H D_p h = 1$, and $z_T^\star = D_p^{1/2}\bar{h}_T^\star$. Choosing a phase $|\alpha| = 1$ that aligns $\widehat{z}_T$ and $z_T^\star$,

$$\|\widehat{z}_T - \alpha z_T^\star\|_2^2 \le 2\sin^2 \angle(\widehat{z}_T, z_T^\star) \le 2\eta_T/\gamma_T.$$

Moreover,

$$\|D_p^{1/2}(\widehat{h}_T - \alpha\bar{h}_T^\star)\|_2^2 = \sum_\ell \|D_\ell(\widehat{h}_T - \alpha\bar{h}_T^\star)\|_2^2.$$

Under the forward/inverse DFT convention in section A.1, the Parseval identity makes N^{-1} times this quantity the squared Frobenius difference between the two inverse-filtered outputs. The noiseless inverse-filter expression agrees with the main-text definition:

$$\mathcal{F}^{-1}[\mathrm{diag}(\bar{h}_T^\star)\widehat{Y}^{(0)}] = T_{\tau_T^\star} X/\beta = X_T^\star, \quad \nu_T = \|\mathcal{F}^{-1}[\mathrm{diag}(\bar{h}_T^\star)\widehat{W}]\|_F.$$

Triangle inequality gives the scale/shift-aligned waveform bound

$$\|\widehat{X}_{\mathrm{full}} - \alpha X_T^\star\|_F \le \delta_X, \quad \delta_X := \sqrt{2\eta_T/(N\gamma_T)} + \nu_T. \tag{42}$$

Let $m_T = \max_n \|X_{T,n:}^\star\|_2$ and let Δ_E denote the gap between the smallest true-support row energy and the largest off-support row energy. Since Frobenius error bounds every row error,

$$\|\widehat{e} - e_T^\star\|_\infty \le \delta_X(2m_T + \delta_X).$$

Thus the condition in eq. (21) is sufficient for the final row-energy Top-K readout to preserve the support ordering. This proves Proposition 4 for any selected candidate satisfying the leakage and gap assumptions. Exact orbit decoding gives an additional sufficient route, but is not required by the waveform bound.

D Implementation and Training

D.1 Feature construction

The feature extractor Fourier transforms Y and forms row powers $p_k = \|\widehat{Y}_{k,:}\|_2^2$. It uses one unweighted feature group and three power-weighted groups. For each power-weighted group, the threshold τ is the empirical 0.08, 0.20, or 0.40 quantile of (p_k), and its row weight is $\rho_k = \mathrm{clip}((p_k - \tau)/(p_k + \tau + 10^{-30}), 0, 1)$. The unweighted group uses $\rho_k = 1$. Each group computes

a stabilized weighted coherence curve G as in χ_ω of section A.1, with the normalized rows and denominator protected against numerical zero.

For one such curve, let $q(G) = \mathrm{even}(\Re\mathcal{F}^{-1}G)$, $f = 1/L$, and let $R(v)$ subtract the median of $v[t]$ over $t \neq 0$, divide by 1.4826 times its median absolute deviation plus 10^{-9}, and clip the result to $[-20, 20]$. The four feature maps are $\mathrm{clip}(q(G), -1, 1)$, $R(q(G))$, $R(q((G-f)_+))$, and $R(q(((G-f)_+/(1-f+10^{-12}))^2))$, for 16 lag-indexed maps across the four groups. The unweighted group is a numerically stabilized counterpart of the ideal row-scaling-invariant statistic q_1; exact invariance applies to the ideal statistic, while the three weighted groups remain kernel dependent. The 12 global descriptors are $\log_{10} N$, $\log_{10} L$, the log of $\sqrt{(\max_k p_k + 10^{-30})/(\min_k p_k + 10^{-30})}$, and logs of nine quantiles of $p_k/\operatorname{median}_j p_j$ at levels 0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, and 0.99; their implementation uses small numerical floors. Repeating these descriptors across lag produces a $28 \times N$ real tensor. No true K, SNR, or κ enters the feature extractor.

D.2 Neural architecture and training

The 215,244-parameter operator lifts 28 channels to width 36 and uses three hybrid circular local/spectral blocks with 24 Fourier modes. A controller that reads only the 12 observable global descriptors weights the feature groups and local/spectral gates. The final score is cyclically even; projection onto $\mathcal{A}_N$ symmetrizes, clips to $[0, 1]$, and fixes coordinate zero to one.

The final autocorrelation-estimation fine-tuning stage uses 800 training and 160 validation scenes, four epochs, and Adam with learning rate 7×10^{-4} and weight decay zero. Its optimization loss is exclusively the N-normalized quadratic Fenchel–Young loss in eq. (10). Checkpoint selection uses held-out MSE between the projected prediction $\widehat{a}$ and a_S; in this run it selects the same epoch as validation FY.

The initialization transfers 38 compatible tensors from a separately trained model using a 240/60 train/validation split. Direct K-dependent global inputs are not transferred: the first controller layer keeps only cardinality-independent global columns, and the lift remaps only normalized-row lag maps and cardinality-independent descriptors. Downstream compatible block/head weights are used only as an initialization and are then optimized with the current autocorrelation loss. This warm start is part of the method; the results do not establish identical behavior from random initialization. Training configurations span $N \in \{127, 192, 256, 257, 384, 512\}$, $L \in \{4, 8, 12, 16, 24, 32\}$, $K \in \{2, 3, 4, 5, 6, 7\}$, SNR in $\{-10, -8, -5, 0, 10, 20\}$ dB, and $\kappa \in \{1, 10, 30, 100, 10^3, 10^4\}$ subject to eq. (43). The random-seed sets for training, validation, preliminary checks, and final evaluation are mutually disjoint.

D.3 Candidate-support search and orientation selection

The candidate-support search considers cardinalities 2 through 7. For each candidate size, the non-learned approximate decoder uses six autocorrelation-ranked greedy starts, one explicitly seeded random start, and three local-swap passes. The seed is an explicit protocol input, so the output is reproducible under the protocol but is not a deterministic function of Y alone. Candidates are ranked by the full autocorrelation-space residual. Conditional on the selected candidate, the scale-invariant inverse-filter solver uses the Gram form of eq. (12), with a matrix of dimension $|T|L$; this reduces the eigensolve dimension when $|T|L < N$. The two reflected orientations are compared by the observable inverse-filter residual, with eigengap used only as a deterministic tie-breaker. The only numerical floor is machine-scale division protection and is not a tunable regularizer.

D.4 Low-rank form of the inverse-filter solve

Let V_T be the Fourier synthesis matrix whose columns are the transforms of the canonical impulses indexed by T. With the Fourier convention used in the implementation, $V_T^H V_T = NI$ and $P_T = V_T(V_T^H V_T)^{-1} V_T^H = V_T V_T^H/N$. Hence

$$B_T = D_p - \frac{1}{N}\sum_\ell D_\ell^H V_T V_T^H D_\ell.$$

Define

$$U_T = \left[D_p^{-1/2} D_1^H V_T/\sqrt{N} \;\; \cdots \;\; D_p^{-1/2} D_L^H V_T/\sqrt{N}\right] \in \mathbb{C}^{N \times |T|L}.$$

Then $M_T = I - U_T U_T^H$. Therefore the smallest eigenvector of M_T is the top eigenvector of $U_T U_T^H$, which can be recovered from the $|T|L \times |T|L$ Gram matrix $U_T^H U_T$. This Gram-form eigensolve reduces the matrix dimension only when $|T|L < N$; the identity itself holds without that restriction. The machine-scale floor used when forming $D_p^{-1/2}$ only prevents division by numerical zero and is not a statistical regularizer.

D.5 Forward-kernel readout

The inverse response is normalized by $\widehat{h}_T^H D_p \widehat{h}_T = 1$. When its entries are nonzero, a forward-kernel estimate is obtained by $\widehat{g}_{\mathrm{est}}[k] = 1/\widehat{h}_T[k]$ followed by the inverse DFT. This reciprocal readout retains the inherent scale/shift ambiguity and can be sensitive to small estimated inverse entries; the main stability bounds concern inverse-filtered signals and support.

E Experimental Protocol and Supplementary Results

E.1 Synthetic generator

For a fixed support of size K, the complex unknowns have $N + KL - 1$ degrees of freedom after removing scale. A necessary dimension count for generic local identification from NL observations is therefore

$$(N - K)L \geq N - 1. \tag{43}$$

This count is not a sufficient identifiability condition.

The final synthetic evaluation uses a controlled circular-complex JSBD generator. Only cells satisfying the necessary dimension count in eq. (43) are sampled. Each support is drawn uniformly without replacement subject to aperiodicity; the configuration `min_sep=1` imposes no separation beyond distinct support rows. Active coefficients are i.i.d. proper complex Gaussian, $\mathcal{CN}(0, 1)$. In the Fourier domain, kernel magnitudes are log-uniform over the requested dynamic range, with exact extrema enforcing $\kappa = \max_k |\widehat{g}[k]| / \min_k |\widehat{g}[k]|$, and phases are independent uniform on $[-\pi, \pi)$. For $Y^{(0)} = C(g)X$ and requested $\mathrm{SNR}_{\mathrm{dB}}$, the noise entries are independent $\mathcal{CN}(0, \sigma^2)$ with $\sigma^2 = \|Y^{(0)}\|_F^2 \, 10^{-\mathrm{SNR}_{\mathrm{dB}}/10}/(NL)$. Thus $\mathrm{SNR}_{\mathrm{dB}} = 10 \log_{10}(\|Y^{(0)}\|_F^2/(NL\sigma^2))$ specifies an expected noise power; the realized ratio $\|Y^{(0)}\|_F^2/\|W\|_F^2$ fluctuates across scenes. The final evaluation changes only the declared grid variables within this generator family, and all sampled settings are retained. The resulting phase maps characterize recovery within this controlled model family.

E.2 Phase protocol and baseline comparisons

Traditional JSBD provenance. The traditional-JSBD benchmark uses a fixed single-start 2,000-step protocol with `ls_max=8` and 100 scenes per cell. Its random-seed set is independent of the QuIR final-evaluation seeds, so phase-grid differences are distribution-matched rather than paired. All archived baseline results are reported without post hoc rerunning, retuning, or cell filtering.

Neural baseline configurations. NeuralBD-JSBD and VDIP-JSBD are task-specific ports, not native implementations of complex multichannel JSBD in the cited papers. They use, respectively, the official NeuralBD SIREN and positional-encoding components and the official VDIP fully connected network. Each network parameterizes a one-dimensional complex inverse-kernel Fourier response q and is optimized independently for each scene using the same scale-normalized joint-row-sparsity proxy and the barrier $10^{-5} \, \mathrm{mean}_k(|q[k]| + 10^{-3})^{-1}$. NeuralBD-JSBD uses 240 Adam steps (learning rate 0.005); VDIP-JSBD uses 320 (learning rate 0.002). Neither is pretrained across scenes. Both passed a predefined 100-scene easy-regime qualification (100/100 shift-only exact) before the complete grid evaluation; their configurations were then fixed. Neither recovery function receives the true scene-specific K, SNR, or κ; true K enters only at the common final Top-K row-energy readout. These results evaluate the specified JSBD ports, rather than a native reproduction of the original imaging tasks.

The dense SNR–κ core fixes $(N, L, K) = (256, 16, 3)$ and uses 21 SNR values from -15 to 5 dB and 25 condition numbers from 1 to 10^4 (525 cells). The K–L core fixes $(N, \mathrm{SNR}, \kappa) = (512, 10\,\mathrm{dB}, 30)$

with $K \in \{3, 5, 7\}$ and 13 values $L = 8, 10, \ldots, 32$ (39 cells). The N–L core fixes $(K, \mathrm{SNR}, \kappa) = (7, 10\,\mathrm{dB}, 30)$ with $N \in \{127, 192, 257, 384, 512, 768, 1024\}$ and the same L values (91 cells). The high-SNR κ–L grid fixes $(N, K, \mathrm{SNR}) = (512, 5, 20\,\mathrm{dB})$, uses 31 condition numbers from 1 to 10^5, and the same 13 L values (403 cells). Boundary scans add $L = 2, \ldots, 7$ for $K \in \{3, 5, 7\}$ and for the seven N values, plus $K \in \{4, 6\}$ intermediate-K stress. Every cell contains 100 scenes.

Table 2: **Fixed-protocol exact-support recovery by phase axis.** Traditional JSBD uses an independent random-seed set on the same numerical cells, so its comparison is distribution-matched rather than paired. NeuralBD-JSBD and VDIP-JSBD are evaluated on the same QuIR scenes. The $K \in \{4, 6\}$ row is an intermediate-K stress inside the declared training scope.

Phase axis	scenes	QUIR	trad. JSBD	NeuralBD-JSBD	VDIP-JSBD
SNR–κ core	52,500	.736	.135	.155	.148
K–L core	3,900	1.000	.524	.462	.514
N–L core	9,100	.997	.931	.898	.928
κ–L core	40,300	1.000	.414	.385	.401
low-L K–L	1,800	.972	.433	.340	.400
low-L N–L	4,200	.870	.765	.637	.715
$K = 4, 6$ intermediate-K	2,400	1.000	–	.432	.497

Same-scene cardinality-informed baseline. The baseline evaluation uses the same QuIR instances, indexed by $(\mathrm{axis}, \mathrm{cell}, \mathrm{rep})$, and a predefined decoder seed convention. Across all 114,200 pairs, QuIR/baseline exact counts are 99,715/95,969, with 5,063/1,317 one-sided discordances. The complete axis breakdown is:

axis	scenes	QUIR	baseline	Δ (pp)	QuIR-only / baseline-only
SNR–κ	52,500	73.61	69.03	+4.57	3624 / 1224
K–L	3,900	100.00	99.67	+0.33	13 / 0
N–L	9,100	99.75	98.99	+0.76	85 / 16
κ–L	40,300	99.97	98.98	+0.99	408 / 8
low-L K–L	1,800	97.22	84.89	+12.33	226 / 4
low-L N–L	4,200	87.05	71.81	+15.24	705 / 65
intermediate K	2,400	100.00	99.92	+0.08	2 / 0

The baseline uses true scene cardinality during autocorrelation estimation; QuIR does not. Other pipeline components also differ, so these paired results compare complete pipelines and do not isolate cardinality access or the representation estimate.

Inverse and factor-recovery diagnostic. The 2,000-scene diagnostic uses 20 predeclared cells with 100 new scenes each. Candidate relations before orientation selection and final-support outcomes are summarized below:

candidate relation	scenes	final exact	exact rate
shift orbit	522	522	1.000
reflection orbit	583	582	0.998
other	895	607	0.678

Overall support exactness is 1,711/2,000. Median shift/scale-aligned full-waveform NMSE is 0.402 overall and 0.375 conditional on exact support; median row-energy relative error is 0.168 overall and 0.148 conditional on exact support. The saved `inverse_shift_aligned_sin2` diagnostic is a squared-sine angular discrepancy in the raw h coordinates, whereas eq. (18) is stated in the whitened $z = D_p^{1/2} h$ coordinates, so the former is not used as a numerical test of the theorem.

E.3 CARDINALITY ACCESS, RUNTIME, AND REPRODUCIBILITY

The high-level inference procedure calls

`rec = recover_full(Y, seed)` then `extract_tdoa(rec["X_full"], K)`.

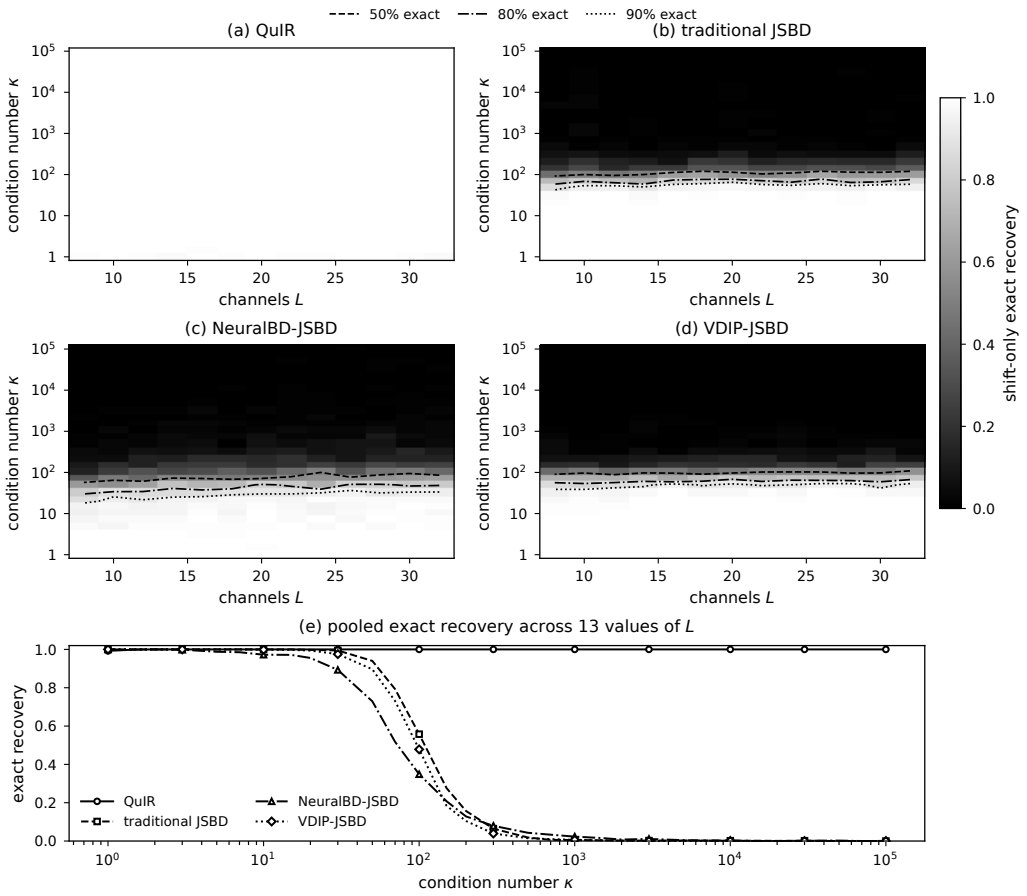

Figure 6: **High-SNR conditioning phase across channel counts.** Exact-support recovery at 20 dB for (a) QuIR, (b) traditional JSBD, (c) NeuralBD-JSBD, and (d) VDIP-JSBD over the common κ–L grid. All four heatmaps use the same exact-recovery scale and the same 50%/80%/90% contour convention. Panel (e) averages exact recovery across the 13 sampled values of L at each condition number.

An input-dependency check verifies that the normalized-row lag maps, neural operator, Fenchel–Young projection, candidate-support search, and scale-invariant inverse-filter solver do not receive the scene-specific value K. The first explicit scene-cardinality argument is the final Top-K support extraction. Before the final evaluation, the implementation was verified on two predefined held-out sanity sets. The implementation and checkpoints remained unchanged throughout the final evaluation.

Across all final-evaluation scenes, the median runtime is 4.13 s, with p90 7.14 s, p95 7.87 s, and p99 9.63 s. The evaluation machine has an Intel Core i9-14900HX CPU (32 logical processors) and an NVIDIA GeForce RTX 4070 Laptop GPU (8,188 MiB), with driver 560.94 and PyTorch 2.6.0+cu124 (CUDA runtime 12.4). These times depend on the implementation and hardware.

Reproducibility. The supplementary experimental package contains QuIR inference code and checkpoints, training and validation protocols, the seeded non-learned decoder, and the scale-invariant inverse-filter solver. It also contains the 114,200 QuIR final-evaluation records, same-scene cardinality-informed baseline results, the available autocorrelation-estimation ablation records, plotting data and scripts, and NeuralBD-JSBD/VDIP-JSBD adapter code and result archives. Scene-level records and audit code for the 290-scene paired downstream subset are included, together with scene-level records for the full 2,900-scene structural diagnostic and the separate 2,000-scene inverse diagnostic. Final-evaluation seeds are disjoint from training, validation, and development seeds, and no inference rule or evaluation criterion was modified after the evaluation protocol was fixed.

E.4 AUTOCORRELATION-ESTIMATION DIAGNOSTICS AND ABLATIONS

A separate diagnostic compares the analytic estimate $a_{\mathrm{ana}} = \mathcal{N}(q_1)$ with the network output while leaving the decoder, inverse-filter solver, and final metric unchanged. This experiment is explanatory and is not part of the primary 114,200-scene final evaluation.

Across 2,900 fresh diagnostic scenes, mean normalized support-autocorrelation MSE decreases from 1.438×10^{-3} for a_{ana} to 7.461×10^{-4} for $\widehat{a}$; the network has lower MSE on 71.7% of scenes. The median cosine between the network correction $\widehat{a} - a_{\mathrm{ana}}$ and the target correction $a_S - a_{\mathrm{ana}}$ is 0.732, and is positive on 82.7% of scenes. On a paired subset of 290 of these scenes (10 per cell), using the same decoder and inverse-filter solver for both estimates, final exact support recovery up to circular shift changes from $245/290$ for a_{ana} to $270/290$ for $\widehat{a}$.

On 500 out-of-distribution scenes at 60 dB, the analytic estimate has lower mean MSE than the network (1.815×10^{-4} versus 8.969×10^{-4}), with median correction cosine -0.054.

Controlled autocorrelation-estimation ablation. To separate the roles of the feature tensor and the training objective, we evaluate three autocorrelation-estimation variants under the same 2,100-scene protocol per variant, for 6,300 scene-runs in total, while keeping the downstream decoder, inverse-filter solver, and final support readout unchanged. The unweighted-projector variant retains only the numerically stabilized unweighted projector-based features and the Fenchel–Young objective; the projected-MSE variant uses the full QuIR feature tensor but replaces the Fenchel–Young objective with MSE between the projected prediction and a_S; the third variant is the complete QuIR autocorrelation estimator.

Table 3: **Controlled autocorrelation-estimation ablation.** Each variant is evaluated on a 2,100-scene protocol. Candidate exactness is measured up to circular shift and reflection; final exactness allows circular shifts only. Exact-recovery entries are percentages; the noise column reports final exact-support recovery on the 1,200-scene noise subset.

variant	mean autocorr. MSE ↓	candidate exact ↑	final exact ↑	noise final ↑
unweighted projector features + FY	1.837×10^{-3}	**32.95**	65.71	43.58
full features + projected MSE	1.651×10^{-3}	29.05	71.52	**55.17**
full features + FY (QUIR)	$\mathbf{1.587 \times 10^{-3}}$	32.52	**71.76**	54.75

Holding the Fenchel–Young objective fixed, restoring the full row-power-augmented feature tensor lowers mean autocorrelation MSE and raises final exact recovery from 65.71% to 71.76%; on the noise subset, the increase is from 43.58% to 54.75%. Among the reported subsets, the largest observed final-recovery difference occurs on the noise subset. Holding the full feature tensor fixed, replacing projected MSE with the Fenchel–Young objective changes final exact recovery from 71.52% to 71.76%, lowers autocorrelation MSE from 1.651×10^{-3} to 1.587×10^{-3}, and increases candidate exact recovery up to shift and reflection from 29.05% to 32.52%. In this ablation, the objective choice has a larger observed effect on autocorrelation error and candidate accuracy than on final exact recovery. The near-equal final rates despite different candidate accuracy are also consistent with the observed tolerance of the inverse-filter solver to inexact candidates.

E.5 MEASURED-DATA EVALUATION

The following experiments are distribution-shift evaluations separate from the controlled final synthetic evaluation. The 1-D and 2-D evaluations use separate checkpoints, each fixed throughout its measured-data evaluation. No measured example is used for training, fine-tuning, or checkpoint selection.

LensFD raw-IQ protocol. The 1-D transfer uses the public RENEW LensFD no-lens indoor measurements (Chen et al., 2023). The raw captures use a 3.6-GHz carrier and 5-MHz sampling with a 64-point OFDM DFT, a 16-sample cyclic prefix, an LTS pilot, and 52 active subcarriers. The predefined preprocessing selects active frames before recovery using a pilot-slot-to-noise-slot power-ratio threshold and fixes 32 receive ports from the available base-station array. The OFDM sample

offset supplied to QuIR is estimated from cyclic-prefix repetition; known LTS samples are withheld during blind recovery. The 1-D checkpoint is applied without real-data retraining or fine-tuning, and true path cardinality is not supplied. Traditional JSBD uses one deterministic start with up to 2,000 iterations and `ls_max=8`.

The known LTS is used only after recovery to form evaluation references. It has 52 active subcarriers; the remaining guard and DC bins do not provide pilot observations. The resulting CIR reference is band-limited and is not exact physical path ground truth. The corresponding full 64-bin blind factorization is therefore not claimed to be globally identifiable on every carrier. For the inspected raw file `no-indoor-spot-1`, the source HDF5 array contains 401 frames, two client streams, 80 base-station antenna ports, and 1,120 complex samples per capture; the same preprocessing rules are used before the multi-location comparison.

Measured 1-D recovery. For each capture, let $\widehat{g}$ be the recovered common waveform and g_{ref} the band-limited pilot-derived reference. The reported common-waveform correlation is $\max_\tau |\langle \mathrm{roll}_\tau \widehat{g}, g_{\mathrm{ref}} \rangle| / (\|\widehat{g}\|_2 \|g_{\mathrm{ref}}\|_2)$; the maximizing shift also minimizes waveform error after one optimal complex scale. For the channel comparison, the recovered latent stack is similarly aligned by a common circular shift and complex scale. We then compute the Pearson correlation between the two real row-norm profiles $r_X[n] = \sqrt{\sum_\ell |X_{n,\ell}|^2}$ and $r_{\widehat{X}}[n] = \sqrt{\sum_\ell |\widehat{X}_{n,\ell}|^2}$; this is the reported channel row-norm correlation.

The aggregate comparison contains 5,328 measured captures across 14 physical indoor locations. We first average valid-capture correlations within each location and then average the 14 location means with equal weight. The mean common-waveform correlation is 0.797 for QuIR and 0.444 for traditional JSBD, while the mean channel row-norm correlation is 0.897 and 0.782, respectively; QuIR is higher at all 14 locations for both summaries. The illustrative high-fidelity example in fig. 5 was selected by maximizing the geometric mean of the two QuIR correlations against the post-hoc references, without using the traditional JSBD outcome. It is spot 2, frame 296, client 0, with raw capture SNR approximately 25.4 dB. In that example, the common-waveform correlations are 0.983 for QuIR and 0.509 for traditional JSBD, and the CIR row-norm correlations are 0.989 and 0.778. All delay-domain panels use the same reference-defined circular shift of $+32$ samples; the delay-energy panel normalizes each method by its own total recovered channel energy to remove global scale ambiguity.

Measured two-dimensional microscopy transfer. The algebraic operators in the 1-D method have 2-D analogues: the cyclic index set $\mathbb{Z}_N$ becomes $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2}$, pair differences become two-dimensional displacements, and the DFT and circular convolution act in two dimensions. The normalized support autocorrelation, normalized-row projector construction, constrained Fenchel–Young estimator, non-learned support readout, and inverse-filter recovery retain their roles. For microscopy, the QuIR-2D architecture and inference rules are fixed throughout measured-data evaluation; the synthetic training distribution and real-data interface are adapted to nonnegative microscopy data. The QuIR-2D checkpoint is trained from random initialization on synthetic microscopy-matched scenes only, with checkpoint selection performed on synthetic validation loss and no measured frame used for training or model selection.

For evaluation we use $L = 32$ consecutive measured frames and a 31×31 region of interest (ROI) selected from the observations alone by the temporal-standard-deviation map. Preprocessing applies per-frame 10th-percentile background subtraction, nonnegative projection, and a single RMS normalization computed over the entire block. The displayed 320-mW block at frames 17001–17032 was fixed before the comparative recovery and was not selected using outputs from any method. The temporal standard deviation is an observation-derived structural proxy, not latent-image ground truth. The measured sequence provides neither latent-image nor measured-PSF ground truth, so this experiment supplies qualitative evidence for applying the method in two dimensions, rather than an exact-recovery benchmark. Recovered image structure is displayed from the full recovered ROI; the lower-row surfaces in fig. 7 show inferred forward PSFs rather than measured PSFs.

Fixed 2-D baselines. Huber+ℓ_1 follows the public two-dimensional multi-channel sparse blind deconvolution (MCS-BD) formulation of Qu et al. (2020). We use the Huber Riemannian stage from the public implementation with $\mu = 10^{-2}$ and line search, followed by the ℓ_1 rounding stage

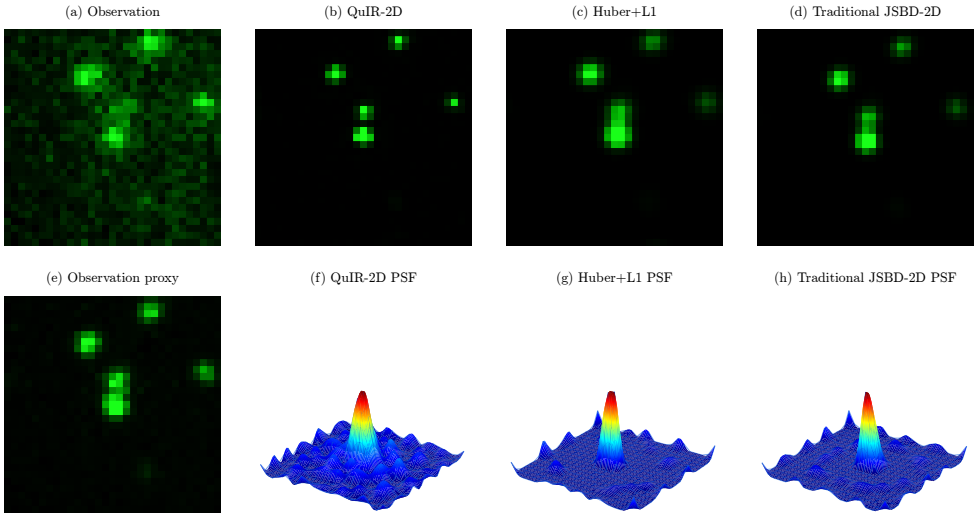

Figure 7: **Measured 2-D microscopy transfer without additional injected noise.** Top: mean observation and image-level recoveries from QuIR-2D, Huber+ℓ_1, and traditional JSBD-2D. Bottom: an observation-derived proxy (not ground truth) and the inferred forward PSFs associated with the three recovered latent stacks. The displayed block is 320 mW, frames 17001–17032; no latent-image or measured-PSF ground truth is used in this comparison.

described in that paper. Before measured-data evaluation, we set the random seed to 123 and the phase-I/phase-II iteration budgets to 120/120; these settings are reused for every block and added-noise level.

Traditional JSBD-2D implements the joint row-sparsity objective of Wang et al. (2026). Our 2-D implementation uses fast Fourier transforms and preserves the preconditioner and smoothed row-ℓ_2/ℓ_1 objective of the supplied dense 2-D implementation. To avoid forming its dense convolution/Hessian tensor on the 31×31 ROI, we replace negative-curvature escape with three deterministic starts and select by the final observable objective. We fix seed 123 and 120 first-order iterations, and verify the implementation gradients by finite differences before evaluation. Neither baseline is tuned per measured block or noise level.

Controlled-noise robustness. We stress the same measured block with signed additive noise at signal-to-added-noise ratios of 20, 10, 5, and 0 dB, defined as $10\log_{10}(\|Y_{\text{native}}\|_F^2/\|W_{\text{add}}\|_F^2)$. For each native block, one zero-mean Gaussian noise direction is standardized to unit RMS and reused across all methods and specified noise levels, with only its scalar amplitude changed. These values are not physical camera-SNR estimates; negative noisy samples are not clipped, so the experiment is a controlled additive-noise stress rather than a camera simulator. The methods are not retrained or retuned across noise levels. For display, only the intrinsic circular-translation ambiguity of blind deconvolution is resolved; rotation, reflection, warping, rescaling, and content-dependent cropping are not used. Because latent ground truth is unavailable, proxy correlation and discrete total variation (DTV) are not used as recovery-accuracy claims. Figure 8 shows all evaluated levels down to 0 dB.

E.6 Scope and boundary cases

The model and candidate-support search are evaluated for $2 \le K \le 7$. Low L remains a finite-sample boundary even among parameter cells satisfying the necessary dimension count in eq. (43). Traditional JSBD is locally stronger in several very-low-L/small-N cells and by one or two scenes in a few well-conditioned high-SNR cells; these cells remain visible in the figures. The normalized support autocorrelation is reflection-invariant and can be shared by homometric supports, so the original observations provide the orientation score subject to the conditions in section C.2. The final-evaluation records include the inverse residual and eigengap but not the oracle leakage η_T; the 2,000-scene diagnostic adds factor-recovery diagnostics without changing the inference procedure.

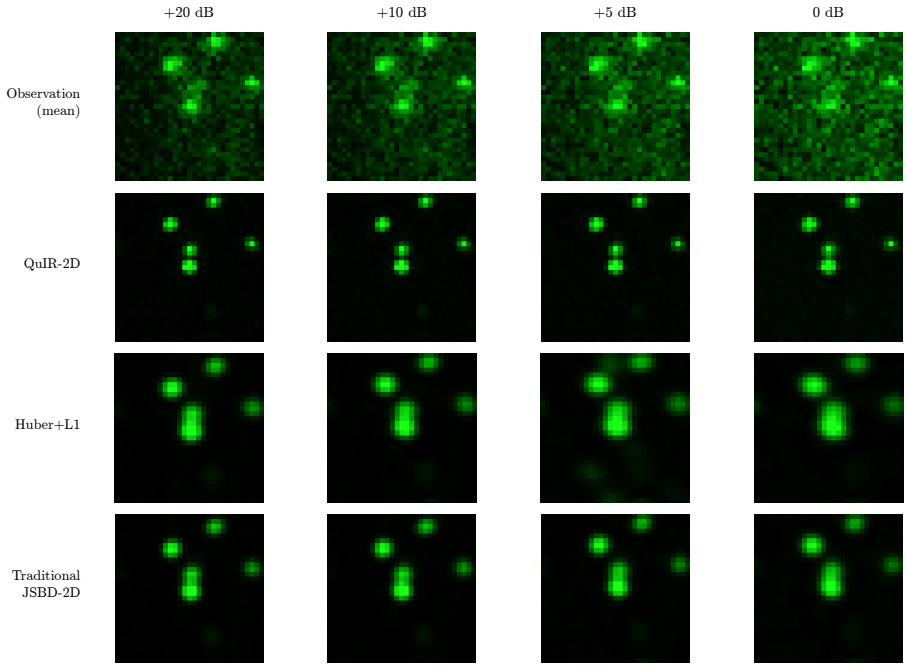

Figure 8: **Controlled-noise recovery on the same measured 2-D block.** Columns show 20, 10, 5, and 0 dB signal-to-added-noise ratio; rows show the mean observation, QuIR-2D, Huber+ℓ_1, and traditional JSBD-2D. All panels use the same measured scene and recovery protocols, with no method-specific geometric registration beyond circular translation.

A separate conditioning stress test, not part of the primary 114,200-scene final evaluation, fixes $(N, K) = (512, 5)$, uses $L \in \{4, 8, 16, 32\}$ and SNR in $\{0, 20\}$ dB, and samples nine condition numbers from 10^4 to 10^8 with 100 scenes per cell (7,200 scenes total). At $\kappa = 10^8$, exact recovery is 100% for $L \in \{8, 16, 32\}$ at 20 dB and for $L \in \{16, 32\}$ at 0 dB; the corresponding low-channel settings degrade, including 84% for $L = 4$ at 20 dB, and 69% for $L = 8$ and 3% for $L = 4$ at 0 dB. These values summarize the sampled test range rather than continuous critical-κ estimates. Distribution-shift evaluation is separate from the controlled generator study reported here.