DASH-LISTA: DISENTANGLING HISTORY FEEDBACK IN HYBRID-ISTA NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Choosing a feedback architecture and simplifying a trained solver require different evidence about history. DASH-LISTA exposes correction information and injection location within complete Hybrid-ISTA updates. Matched retraining, fixed-backbone replacement and equal-norm reflection distinguish architecture benefit, checkpoint dependence and directional effects. Completion and subsequent updates transform older-history directions. Under a shared gain cap, four-tap FIR improves on two-time and nonlinear current controls by 4.25 ± 0.66 and 9.25 ± 0.90 dB in three independently trained confirmation instances, each tested on 100,000 previously unused signals; all paired gains are positive. Original-gain effects remain tail-sensitive, with one short-history ordering reversal. In images, replacing history with six independent bounded current gains retains a dual-branch cost of about 0.026 dB after 6,400 adaptation steps in each of two cohorts, while slightly improving refinement-only reconstruction. On test50, reflection at stages four and five has a larger direct-path effect in all six checkpoints. Retrained image alternatives reach nearly coincident endpoint means, separating checkpoint dependence from architecture benefit.

1 INTRODUCTION

Finite-depth unfolded networks learn reconstruction within a limited iteration budget (Gregor & LeCun, 2010; Chen et al., 2018). Earlier increments provide correction directions whose effect depends on the receiving computation: a learned refiner can transform a direction, a threshold can suppress it, and a direct branch can transmit it to the next state. This interaction matters both for choosing a feedback architecture and for simplifying a trained solver.

Improvement over a parent solver can arise from learned current scaling as well as history. Architecture selection permits joint adaptation of the backbone and correction; simplifying an existing checkpoint holds that backbone fixed. Replacing history therefore tests a different computation from independently training its alternative. DASH-LISTA separates these effects within Hybrid-ISTA (Zheng et al., 2023). Its complete-state interface preserves proximal steps, learned refinement, mixture and persistent context while varying correction information and injection location. Matched retraining measures architecture benefit, fixed-backbone replacement measures model dependence, and interventions at a common state measure directional effects.

Shared-cap FIR gains over two-time and nonlinear current replicate in independent synthetic instances, each tested on 100,000 signals. Original-gain effects remain tail-sensitive. In images, dual-branch replacement costs persist after a 16-fold budget increase and relaxation to six independent stage gains; refinement-only replacement improves. Same-checkpoint interventions locate stronger direct-path effects at test50 stages four and five. Dependence varies across injection configurations despite nearly coincident retrained endpoint means.

Contributions. **A state-preserving feedback framework.** DASH controls information and injection location within complete Hybrid-ISTA updates, with exact zero-feedback parent references. **Recovery gains under shared gain constraints.** Matched retraining replicates FIR gains on independent instances. Separately, equal-norm interventions isolate useful older-history directions in frozen models. **Budget, function class and pathway effects.** Nested replacements retain dual-branch costs across two cohorts, and same-state interventions locate stronger direct-path effects at

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.

specific stages. Together with retraining, these findings distinguish checkpoint simplification from architecture selection.

2 Related work

Unfolding and memory. LISTA learns sparse-code approximations; weight coupling and support selection enable conditional recovery analyses (Gregor & LeCun, 2010; Chen et al., 2018). Hybrid-ISTA inserts a free-form network into a proximal mixture (Zheng et al., 2023). MADUN and MAUN introduce memory into image unfolding (Song et al., 2021; Hu et al., 2024); MAPUN and HUNet exploit features across stages (Song et al., 2023; Shen & Gan, 2025). DASH instead studies corrections formed from solver increments, with separate access to refinement and direct estimate inputs. Retaining the complete Hybrid-ISTA update makes the correction's information and injection location explicit experimental variables.

Learned deviations and safeguards. Safe-L2O combines learned candidates with a residual test and fallback (Heaton et al., 2023); forward–backward deviations admit history (Banert et al., 2024; Sadeghi et al., 2024). Learned-KM uses summable perturbations (Martin & Belgioioso, 2026), and residual-controlled DRS adapts solver parameters (Liu & Hu, 2026). These methods motivate studying both the size and the information content of learned deviations. We connect norm-controlled direction interventions to matched output fitting and independent retraining, which evaluate different roles of feedback. Our auxiliary fixed-reference bounds use inexact-iteration analysis (Combettes & Yamada, 2015; Schmidt et al., 2011; Liang et al., 2016).

Reconstruction alternatives. AMP/LAMP use Onsager corrections (Donoho et al., 2009; Borgerding et al., 2017); LDAMP adds learned denoisers (Metzler et al., 2017). ISTA-Net+ learns residual-domain image reconstruction (Zhang & Ghanem, 2018), while ALISTA uses analytic weights (Liu et al., 2019). Our image-CS and external-baseline comparisons retain their respective operators, preprocessing and budgets (Section 5.4; Appendices AE and Y).

3 DASH-LISTA: state-preserving trajectory feedback

3.1 A complete update with two injection locations

We recover $x^\star$ from $y = Ax^\star + n$, with $A \in \mathbb{R}^{M\times N}$. Here $x_k \in \mathbb{R}^N$ is the iterate and $\mathrm{soft}_\theta(z) = \mathrm{sign}(z)\max\{|z| - \theta, 0\}$ acts coordinatewise. For dictionary image CS, $A = \Phi D$ maps coefficients to measurements. Hybrid-ISTA combines learned refinement with a direct mixture path. DASH inserts a correction before refinement and optionally into the direct path. The CP-style instance makes both locations explicit:

$$\begin{aligned} v_k &= \mathrm{soft}_{\theta_{1,k}}\Big(x_k + \overline{W}_k^\top (y - Ax_k)\Big), \\ r_k &= v_k - x_k, \quad m_k = \mathcal{M}_k(r_0, \dots, r_k), \quad \widehat{v}_k = v_k + \gamma_k m_k, \\ u_k &= \widehat{v}_k + \mathrm{DNN}_k(\widehat{v}_k), \\ w_k &= \mathrm{soft}_{\theta_{2,k}}\Big(u_k + \widehat{W}_k^\top (y - Au_k)\Big), \\ x_{k+1} &= \alpha_k w_k + (1-\alpha_k)\widehat{v}_k. \end{aligned} \tag{1}$$

The increment r_k records the first proximal step's proposed change, using quantities available at inference. Refinement-only injection replaces the final $\widehat{v}_k$ by v_k, keeping the corrected refinement input. The CP mixture obeys $0 \le \alpha_k \le 1$ and $\alpha_k\theta_{2,k} \le (1-\alpha_k)\theta_{1,k}$; each other variant keeps its own completion rule. The image implementation also includes dictionary and pixel-domain operations (Appendix AH).

The image-CS completion. The image port performs refinement in pixel space and returns to coefficient space. With synthesis D_k, learned analysis E and the six-convolution thresholded residual R_k, it computes

$$p_k = D_k\widehat{v}_k, \qquad g_k = p_k - t_k\Phi^\top(\Phi p_k - y), \qquad u_k = E\big[g_k + R_k(g_k)\big]. \tag{2}$$

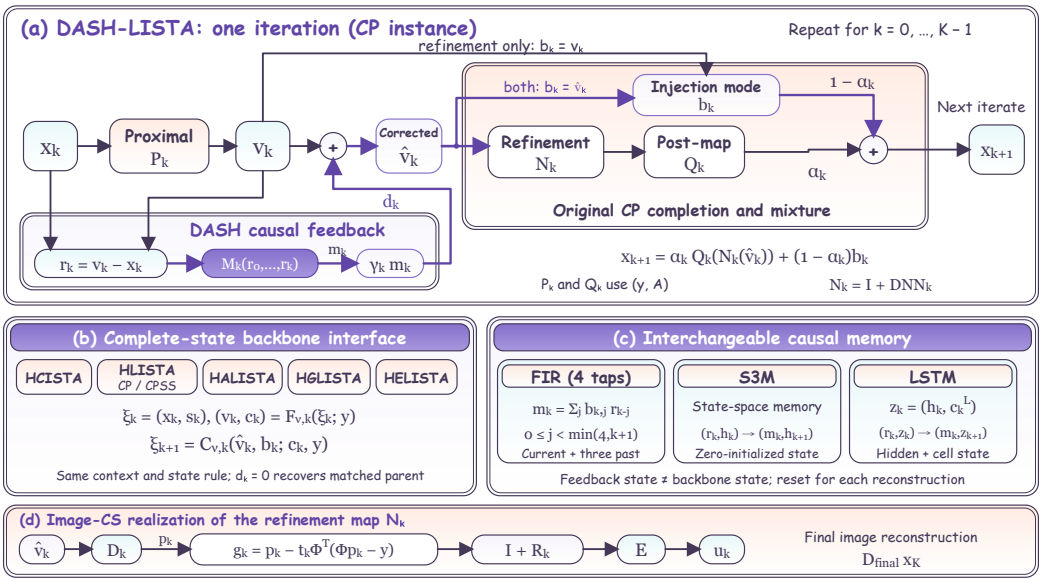


Figure 1: DASH-LISTA with explicit injection choices. (a) CP iteration: P_k and Q_k are the two proximal steps; $N_k = I + \mathrm{DNN}_k$. The fixed injection mode selects $b_k = v_k$ (refinement only) or $b_k = \widehat{v}_k$ (both). (b) Complete backbone state and context. (c) Causal memory; LSTM state z_k includes hidden and cell states. (d) Image-CS refinement with synthesis, pixel-domain consistency and learned analysis.

The second proximal step in Eq. (1) still uses the fixed $A = \Phi D$. Here $D \in \mathbb{R}^{256\times512}$, $E \in \mathbb{R}^{512\times256}$; $D_k = D$ until the last stage, where $D_k = D_{\mathrm{final}}$ is learned. The output is $D_{\mathrm{final}}x_K$. All matched image arms retain this routing; E is a learned analysis map, not an exact inverse (Appendix AH).

3.2 History and a matched current-increment control

Our primary memory is coordinate-wise FIR:

$$m_k = \sum_{j=0}^{J_k-1} b_{k,j} r_{k-j}, \qquad J_k = \min(J, k+1). \tag{3}$$

Four taps use the current increment and three preceding increments. Synthetic experiments use four tanh taps per layer; matched image controls share four taps with absolute-sum cap 0.5. The retrained current control replaces r_{k-j} by r_k in every active term, keeping the parameters, sharing and startup availability unchanged. This comparison isolates access to past increments from learned current scaling. The input restriction applies to the correction module; the solver retains its complete backbone state. For the six-stage image solver, let c_j denote the effective shared taps, $C = \sum_{j=0}^{3} c_j$, and η_k the current control's effective gain. Its complete gain schedule is

$$(\eta_0, \ldots, \eta_5) = (c_0, c_0 + c_1, c_0 + c_1 + c_2, C, C, C), \qquad \|c\|_1 \le 0.5. \tag{4}$$

In particular, $\eta_3 = \eta_4 = \eta_5$ and $|\eta_0| + \sum_{k=1}^{5} |\eta_k - \eta_{k-1}| = \|c\|_1$. This defines Current4, inheriting the FIR parameter budget and sharing. Independent retraining adapts the backbone as well. Fixed-backbone replacement fits Current4 and the wider TV6/Box6 classes, including six independent stage gains (Section 5.5).

FIR buffers reset between reconstructions. S3M and LSTM use the same interface with their own memory contracts (Appendix AK).

Table 1: Correction controls and their specific comparison targets. Current input restricts the correction, not the full solver state. The delayed term is zero at startup; a_k is the training-only RMS increment scale from the initialized parent.

Control	Definition and comparison target
Current gain	$g_k r_k$: can learned current scaling account for the gain?
Two-time	$b_{k,0} r_k + (\sum_{j=1}^{J_k-1} b_{k,j}) r_{k-1}$: retain four raw coefficients but only two increments.
Nonlinear current	$J_k \tanh(\mathrm{MLP}_k(r_k/a_k)) \odot r_k$: 25 parameters/layer versus four FIR taps; a coordinatewise current map, not equal capacity.
Equal-norm intervention	Vary direction through the same frozen completion; not a lower bound on retrained alternatives.

3.3 Preserving the complete backbone state

To extend the construction across variants, let $\xi_k = (x_k, s_k)$ denote the complete state of backbone ν and c_k its within-iteration context. Factor the parent update as

$$(v_k, c_k) = F_{\nu,k}(\xi_k; y), \qquad B_{\nu,k}(\xi_k; y) = C_{\nu,k}(v_k, v_k; c_k, y). \tag{5}$$

DASH changes only the two estimate inputs, with $d_k = \gamma_k m_k$:

$$\xi_{k+1} = C_{\nu,k}(v_k + d_k, v_k + d_k; c_k, y), \tag{6}$$

or $C_{\nu,k}(v_k + d_k, v_k; c_k, y)$ for refinement only. Context, gates, anchors and persistent backbone state remain under the original completion rule; feedback has separate state. This interface covers HCISTA, HLISTA-CP/CPSS, HALISTA, HGLISTA and HELISTA through the mappings in Appendix AG.

Proposition 1 (Zero-feedback identity)**.** *Under Eq.* (5)*, disabling feedback recovers the original full state at every finite iteration with the same parameters, initialization and coupled random draws. Corrected inputs must remain in the completion domain.*

Complete-state induction supplies the exact parent reference (Appendix AG.1). Separately trained arms preserve backbone structure, state semantics and update rules while learning weights independently (Section 5.1).

4 How historical corrections change recovery error

4.1 Completion-space alignment

Fix a trained model, a state, observations and a target $x^\star$. Let $T_k(d)$ denote the output of the complete iteration with correction d, including the second proximal step and mixture. For an alternative c, define $e_c = T_k(c) - x^\star$ and $\Delta = T_k(d) - T_k(c)$. Expanding squared error gives

$$\underbrace{\|e_c\|_2^2 - \|T_k(d) - x^\star\|_2^2}_{B_k(d,c)} = \underbrace{-2\langle e_c, \Delta\rangle}_{A_k(d,c)} - \underbrace{\|\Delta\|_2^2}_{S_k(d,c)}. \tag{7}$$

Improvement requires completion-space alignment A_k to exceed displacement cost S_k. These target-based quantities describe an executed contrast; reconstructing $A_k = B_k + S_k$ checks consistency. The receiving map below determines how a correction changes that contrast.

Why injection location changes the comparison. For fixed CP weights, a common state and correction d, both injection choices share w_k. Their outputs differ by $(1 - \alpha_k)d$. For $e_{\mathrm{ref}} = x_{k+1}^{\mathrm{ref}} - x^\star$, the squared-error reduction is

$$B_k^{\mathrm{both,ref}} = -2(1 - \alpha_k)\langle e_{\mathrm{ref}}, d\rangle - (1 - \alpha_k)^2 \|d\|^2. \tag{8}$$

The direct path bypasses the current refinement and second threshold. Its signed effect depends on the refinement-path error, motivating same-checkpoint path interventions alongside separately trained locations.

4.2 THE DIRECTIONAL CONTRIBUTION BEYOND RECENT INCREMENTS

For $k \geq 1$, let $U_k = \mathrm{span}\{r_k, r_{k-1}\}$, $p = P_{U_k} d$ and $h = d - p$. For effective FIR taps $c_{k,j}$, $h = (I - P_{U_k}) \sum_{j=2}^{J_k - 1} c_{k,j} r_{k-j}$ isolates a direction supplied by older increments. Reflection compares $d_+ = p + h$ with $d_- = p - h$, preserving both p and $\|d_\pm\|_2$. Write

$$a = \tfrac{1}{2}[T_k(d_+) + T_k(d_-)] - x^\star, \qquad u = \tfrac{1}{2}[T_k(d_+) - T_k(d_-)].$$

The two output errors are $a + u$ and $a - u$; subtracting their squared norms yields

$$\|T_k(d_-) - x^\star\|_2^2 - \|T_k(d_+) - x^\star\|_2^2 = -4\langle a, u \rangle. \tag{9}$$

For the synthetic ReLU/soft-threshold completion, $u = \overline{\mathcal{J}}_k h$, where $\overline{\mathcal{J}}_k$ averages affine-region Jacobians along the reflected segment. This relation includes activation and threshold crossings. On one region, with CNN Jacobian K_k, second-threshold mask $D_{2,k}$ and backprojection $W_k = \widehat{W}_k^\top$, differentiation gives

$$\mathcal{J}_k = \underbrace{\alpha_k D_{2,k}(I - W_k A)(I + K_k)}_{\mathcal{J}_{k,\mathrm{ref}}} + \underbrace{(1 - \alpha_k) I}_{\mathcal{J}_{k,\mathrm{direct}}}. \tag{10}$$

On a segment within one affine region, $\mathcal{J}_{k,\mathrm{ref}} h = 0$ makes h invisible through refinement, while the direct path transmits $(1 - \alpha_k)h$. The inner product in Eq. (9) determines its benefit. Thus equal input norms need not imply equal effects across paths; Figure 2 tests direction at fixed weights.

4.3 FROM A SINGLE INTERVENTION TO THE FINAL RECONSTRUCTION

After an intervention, let z contain the complete backbone state and FIR buffer, and let $\Psi_k(z)$ execute all remaining layers. For the two post-intervention states, set $\Delta_K = \Psi_k(z_d) - \Psi_k(z_c)$. Equation (7) holds at the endpoint with $e_c = \Psi_k(z_c) - x^\star$ and Δ_K. For the synthetic FIR model, the finite suffix is piecewise affine in z, so $\Delta_K = \overline{H}_k(z_d - z_c)$, with $\overline{H}_k$ the segment-averaged suffix Jacobian. The suffix recomputes increments and buffer entries. Transmission by a branch and alignment after the suffix jointly determine terminal effect; an immediate directional gain need not survive later updates. Appendix A gives the derivations.

4.4 THE IMPLEMENTED GAIN CONSTRAINT AND FINITE-DEPTH ERROR

In an ℓ_p norm, $1 \leq p \leq \infty$, effective FIR taps $c_{k,j} = \gamma_k b_{k,j}$ with $\sum_j |c_{k,j}| \leq G$ give

$$\|d_k\| \leq G \max_{0 \leq j < J_k} \|r_{k-j}\|. \tag{11}$$

The capped controls obey the same envelope. For complete-state error $e_k = \|\xi_k - \xi^\star\|$, assume domain-valid bounds on the parent ($qe_k + \epsilon_B$), injection sensitivity (L_d) and increment ($C_r e_k + \nu_r$). Then the triangle inequality gives

$$e_{k+1} \leq q e_k + a_G \max_{j < J_k} e_{k-j} + \eta. \tag{12}$$

Here $a_G = L_d G C_r$ and $\eta = \epsilon_B + L_d G \nu_r$ retain reference defects. For $M < N$, $\|I - WA\|_2 \geq 1$ gives $q_k \geq 1$ in the full-space bound, whereas geometric decay requires $q + a_G < 1$. Appendix A.6 derives a signed augmented-state Jacobian and a regional recovery certificate that retains threshold masks, buffer dynamics and target defects. Its matrix and domain conditions specify the checkpoint-level verification required for contraction. The experiments match permitted gain and measure signed recovery differences.

5 EXPERIMENTS

5.1 PROTOCOLS AND STATISTICS

Table 2 summarizes the two implementations. For method m and trained instance s, synthetic reconstruction error is

$$R_{m,s} = \frac{\sum_i \|\widehat{x}_{m,s,i} - x^\star_{s,i}\|_2^2}{\sum_i \|x^\star_{s,i}\|_2^2}, \qquad E_{m,s} = 10 \log_{10} R_{m,s}, \qquad d_s = E_{C,s} - E_{F,s}. \tag{13}$$

Positive d_s favors FIR. We report $\bar{d} \pm \sqrt{\sum_s (d_s - \bar{d})^2/(S-1)}$ across trained instances, applying this aggregation also to absolute scores. Image metrics are averaged within seed before pairing in the stated subtraction order. The replicate count S refers to trained instances, not test samples. CUDA FP32 inference disables TF32; synthetic energy sums accumulate in FP64 before the logarithm.

Table 2: Matched comparisons within each domain. Targets, completion, sharing and optimization vary across domains, precluding attribution to sparsity alone.

	Synthetic recovery	**Image controls**
Depth; training batch	16 layers; 64 signals	6 stages; 64 patches
FIR parameters	4 taps/layer (64 total)	4 globally shared taps
Gain constraint	Tanh taps; separate $G = 0.5$ arm	Absolute-sum cap 0.5
Training budget	48 phases; 480,000 updates	End-to-end; 26,000 updates
Checkpoint selection	Validation every 200 updates	Validation every 500 updates

5.2 Recovery benefit survives a shared gain constraint

Task and optimization. We use noiseless $y = Ax^\star$, $A \in \mathbb{R}^{250\times 500}$, and independent Bernoulli(0.1) masks times standard normal coefficients, giving random support with mean 50. Within instance, FIR, two-time and nonlinear current share the operator, initialization, training streams and CUDA platform. The backbone uses tied backprojection, three kernel-9 convolutions (width 16, ReLU), layer-specific thresholds and mixtures. Each depth has three 10,000-update Adam phases at $5\times10^{-4}, 10^{-4}, 10^{-5}$, using half summed squared error. New parameters train first, then all introduced parameters with older gradients scaled by 0.3 per layer of age. Phase boundaries reset Adam and restore the best checkpoint. A 1,000-signal validation set selects the final model (Appendices K and N).

Table 3: Same-CUDA recovery: absolute FIR NMSE (lower is better) and comparator-minus-FIR NMSE (positive favors FIR), in dB. Mean ± sample SD across problem instances; parentheses count positive paired differences. Original tests precede the larger evaluation; the three blocks distinguish original gain, shared-cap development, and shared-cap independent confirmation with previously unused tests.

Evaluation		**Absolute NMSE**	**Paired gain (dB)**	
Instances	**Signals**	**FIR (dB) ↓**	**Two-time − FIR**	**Nonlinear − FIR**
Original-gain recovery				
Original 0–2	10,000	-87.38 ± 4.09	9.97 ± 7.74 (3/3)	12.52 ± 3.53 (3/3)
Original 0–2	100,000	-79.18 ± 7.40	7.16 ± 8.60 (3/3)	7.08 ± 6.44 (3/3)
New 3–4	10,000	-85.89 ± 3.24	11.17 ± 1.55 (2/2)	17.77 ± 4.26 (2/2)
New 3–4	100,000	-67.36 ± 4.66	1.63 ± 2.85 (1/2)	4.97 ± 6.27 (2/2)
Shared-cap development ($G = 0.5$)				
Capped 0–2	10,000	-62.34 ± 1.45	4.35 ± 0.29 (3/3)	10.49 ± 0.40 (3/3)
Capped 0–2	100,000	-60.65 ± 0.49	3.80 ± 0.28 (3/3)	9.07 ± 0.29 (3/3)
Shared-cap independent confirmation ($G = 0.5$)				
Capped 5–7	100,000	-60.62 ± 0.99	4.25 ± 0.66 (3/3)	9.25 ± 0.90 (3/3)

Absolute error and paired gains. At original gain, all three initial instances retain positive paired gains on 100,000 signals. The two new instances retain positive mean gains, with a two-time ordering reversal for seed 3. Their FIR NMSE changes from -85.89 ± 3.24 to -67.36 ± 4.66 dB on the larger test set, using the same frozen checkpoints. The support decomposition below identifies the samples changing both absolute error and relative gains.

Controlling permitted feedback size. For $G = 0.5$ retraining, active tanh taps c_j are rescaled by $\min\{1, G/\sum_j |c_j|\}$ (scale one at zero). Two-time then collapses delayed coefficients; nonlinear

current uses $G\tanh(f_k(r_k/a_k))\odot r_k$. These rules enforce Eq. (11) from startup, matching permitted size while allowing different realized norms. The original capped cohort yields gains of 3.80 ± 0.28 and 9.07 ± 0.29 dB on arrays previously evaluated by original-gain models. A separate confirmation fixes seeds 5–7, the same protocol and 100,000 unused signals per instance before training. Its nine models yield gains of 4.25 ± 0.66 and 9.25 ± 0.90 dB, all six positive (Table 3). All 24 prespecified support-bin contributions are positive. Shared-cap gains thus replicate under new operators, initializations and training streams (absolute scores and prefix curves in Appendix F.1).

5.3 Older directions affect both completion and subsequent recovery

Norm-controlled interventions. We use the first 4,096 signals of the designated new block for the three original CUDA FIR checkpoints. For $p = P_{U_k} d$, the comparisons d versus $(\|d\|/\|p\|)p$, and $(\|p\|/\|d\|)d$ versus p, match norms at the FIR and projected scales, respectively. Each favors FIR's direction in 41/42 seed–layer cells over layers 3–16. Reflection $2p - d$ keeps the recent projection and full norm while reversing older-direction residual h; it also raises error in 41/42 cells. At the final layer, reflection raises NMSE by approximately 2.928, 0.010 and 4.596 dB for seeds 0–2. Their common sign conceals different magnitudes (Eq. (9)).

A single intervention at each of layers 4, 8, 12 and 16 is followed by ordinary FIR updates on the changed trajectory. Both projection and reflection increase terminal error in all 12 cases (Figure 2c). The suffix recomputes increments along each intervened trajectory. Table 3 instead compares retrained alternatives. On the three new capped checkpoints, a calibration-set autodiff score predicts all three replacement-loss signs but underestimates magnitude: MAE improves 4.6% over zero change. This directly tests transfer from local sensitivity to a finite replacement; accurate magnitude prediction remains unresolved (Appendix F.2).

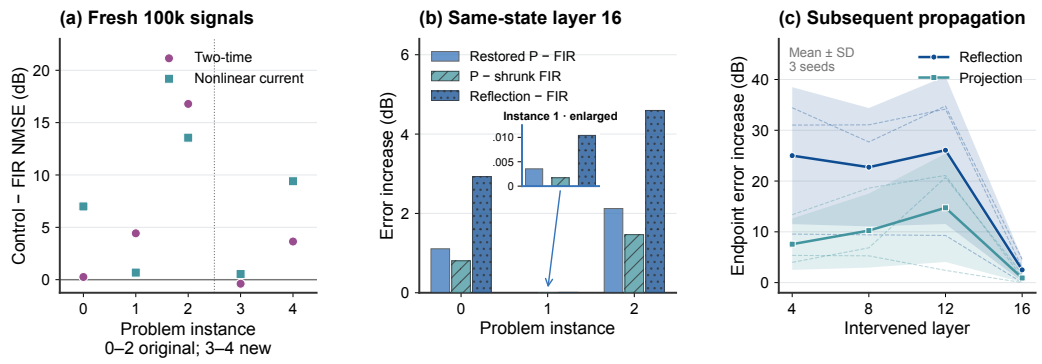


Figure 2: Magnitude and scope of history effects. (a) Five original-gain instances on 100,000 signals; cohorts as in Table 3. (b) Layer-16 norm-matched contrasts, with seed 1 enlarged; P is latest-two projection. (c) Propagated interventions: mean ± sample SD (thick/shaded) and individual seed traces (thin dashed). Panels (b,c) use three frozen checkpoints and 4,096 signals, separate from retraining in (a).

Support contributions to paired gains. For disjoint support bins B_b (≤ 40, 41–50, 51–60, ≥ 61), let $e_{m,i} = \|\widehat{x}_{m,i} - x_i^\star\|^2$ and $P = \sum_i \|x_i^\star\|^2$. Using the same sample groups and denominator for both methods gives

$$G_b = \frac{\sum_{i\in B_b}(e_{C,i} - e_{F,i})}{P}, \qquad R_C - R_F = \sum_b G_b, \qquad E_C - E_F = 10\log_{10}\left(1 + \frac{\sum_b G_b}{R_F}\right). \tag{14}$$

Bins with at most 60 nonzeros contribute positively in all five instances against both controls; the ≥ 61 bin is negative in three of ten comparisons. In new seed 3 versus two-time, lower-support gains of 1.058×10^{-8} are outweighed by -4.401×10^{-8} from 6,092 high-support signals, reversing the net sign. The original test instead gives a positive high-support contribution. Thus the decomposition locates the observed reversal without discarding samples. It describes these test sets, not unobserved tail risk (Appendix D).

5.4 IMAGE RECOVERY: PARENT IMPROVEMENT AND ARCHITECTURE CHOICE

Common image measurement and training. We use the released Hybrid-ISTA column-normalized Gaussian Φ and 256×512 dictionary. Grayscale images are split into nonoverlapping 16×16 patches. Patch means are removed and supplied as side information outside the nominal sampling rate. Reconstruction restores them, assembles and unpads the image, then clips to $[0, 255]$. We hash-partition BSD500 (Arbeláez et al., 2011) into 400 training, 50 validation and 50 test images. The 26,000-update study trains five MR25 arms with Adam (batch 64, learning rate 10^{-4}, symmetry weight 0.01, gradient-norm cap 1): parent, current and FIR at refinement/both locations. Within seed, all arms share backbone initialization and patch streams. Validation-patch PSNR selects checkpoints (Appendix R).

A separate four-rate, 13,000-update end-to-end comparison improves HLISTA-CP in all twelve rate–seed pairs, with both hybrid methods above ISTA-Net+ (Table 83); its schedule differs from the original authors' stagewise training.

Parent-relative gain and matched architecture comparison. Table 4 evaluates the 26,000-update models on validation50 and all 5,000 COCO val2017 images (Lin et al., 2014) without further training. Current-both has the highest mean PSNR on both sets. On COCO, FIR-both gains 0.0169 dB over the parent but changes PSNR by -0.0148 ± 0.0439 dB relative to current-both, with mixed seed signs. Fixed-size timing places current-both at 119.90 ± 0.18 ms versus FIR-both at 125.88 ± 0.43 ms. Timings use MTT S5000, 512×512 inputs and uniformly simplified inference, distinct from native-size COCO throughput (Appendix U).

Table 4: 26,000-update matched image controls: mean $\pm$ sample SD over three trained seeds. COCO evaluates the same frozen checkpoints on all 5,000 native-size images; latency uses a separate fixed 512×512 input with uniform inference optimization. Latency summarizes seed medians, not COCO throughput. Best/second means: bold red/regular blue.

Method	**PSNR (dB) ↑**		**Latency (ms) ↓**
	Validation50	**COCO5000**	**512 × 512 input**
HLISTA-CP	29.0552 ± 0.0179	28.1269 ± 0.0248	$\mathbf{104.20 \pm 0.10}$
Current, refinement	29.0784 ± 0.0138	28.1475 ± 0.0152	119.90 ± 0.50
Current, both	$\mathbf{29.0812 \pm 0.0116}$	$\mathbf{28.1585 \pm 0.0114}$	119.90 ± 0.18
FIR, refinement	29.0650 ± 0.0182	28.1444 ± 0.0096	125.86 ± 0.48
FIR, both	29.0751 ± 0.0212	28.1438 ± 0.0330	125.88 ± 0.43

5.5 REPLACEMENT COST ACROSS BUDGETS AND CURRENT-GAIN CLASSES

Current4 has the stage gains in Eq. (4), sharing its last three gains. We fit two wider classes $d_k = g_k r_k$:

$$\mathcal{C}_{\mathrm{TV6}} = \{g : |g_0| + \textstyle\sum_{k=1}^{5} |g_k - g_{k-1}| \le G\}, \quad \mathcal{C}_{\mathrm{Box6}} = \{g : \max_k |g_k| \le G\}, \quad G = 0.5. \tag{15}$$

Thus $\mathcal{C}_{\mathrm{Current4}} \subset \mathcal{C}_{\mathrm{TV6}} \subset \mathcal{C}_{\mathrm{Box6}}$: TV6 releases late-stage sharing; Box6 permits six independent bounded gains. Equal-budget FIR4 adaptation supplies the replacement reference, distinct from the original FIR used for retraining contrasts. Nine frozen backbones comprise both/refinement seeds 0–2 and both seeds 3–5. Original/zero starts optimize whole-image PSNR using 320 fitting and 80 selection images from train400. Adam budgets of 400, 1,600 and 6,400 updates select within each prefix, including feasible narrower-class candidates. Test50 and validation50 are reused (Appendix H).

Beyond four shared coefficients. Current4 costs change from 0.03024 to 0.03002 dB in the original both cohort and from 0.03049 to 0.03039 dB in the new cohort as budget increases 16-fold. Box6 lowers both costs to about 0.0262 dB, retaining positive differences in all six checkpoints (Table 5). TV6 and Box6 slightly improve all three refinement-only seeds. The earlier 400-update objective comparison remains in Appendix G.

Table 5: Lower replacement cost is better. Test50 replacement cost after 6,400 updates: adapted FIR minus current PSNR (dB), paired mean ± sample SD over three seeds. Positive favors FIR; all six dual-branch differences remain positive for every class.

Checkpoint cohort	Current4	TV6	Box6
Both, seeds 0–2	0.03002 ± 0.00511	0.02851 ± 0.00450	$\mathbf{0.02623 \pm 0.00463}$
Both, seeds 3–5	0.03039 ± 0.00201	0.02921 ± 0.00186	$\mathbf{0.02622 \pm 0.00322}$
Refinement, seeds 0–2	0.00019 ± 0.00062	-0.00167 ± 0.00092	$\mathbf{-0.00345 \pm 0.00206}$

All 18 Current4 trajectories pass the prespecified final stability check: three consecutive best selection-score increments below 0.001 dB. FIR4 and TV6 also pass 18/18; Box6 passes 17/18. For both seeds 0 and 3, a deterministic projected-gradient check leaves FIR costs of 0.03121 and 0.03240 dB. These are finite-search results. Independent retraining instead gives 0.0037 ± 0.0084 dB for FIR on seeds 3–5 (all stages in Appendix J).

5.6 Path-specific effects within the same trained model

For the six both checkpoints, we reflect the older component $d^- = p - h$ versus $d^+ = p + h$ at one stage and rerun the complete suffix. The routes $(++), (-+), (+-), (--)$ independently choose refinement/direct inputs. With final normalized image MSE ℓ, define

$$D_{\text{ref}} = \ell_{-+} - \ell_{++}, \quad D_{\text{direct}} = \ell_{+-} - \ell_{++}, \quad I = \ell_{--} - \ell_{-+} - \ell_{+-} + \ell_{++}. \tag{16}$$

Joint reflection gives $\ell_{--} - \ell_{++} = D_{\text{ref}} + D_{\text{direct}} + I$. The first two stages have zero effects. On test50 at stages four and five, $D_{\text{direct}} > D_{\text{ref}}$ in all six checkpoints; stage six has five such orderings, with one negative direct effect. Negative late-stage I means joint loss increases are smaller than the sum of separate effects (Figure 3). Even additive output changes can interact under squared error; Appendix H.4 separates this geometry from nonadditive output changes. On validation50, the new cohort has negative stage-six direct effects in all three checkpoints.

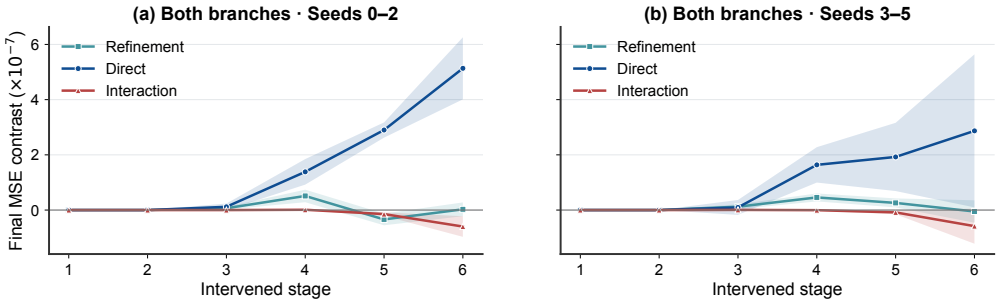


Figure 3: Single-stage reflections and final test50 error: mean ± sample SD (shading) over three checkpoint seeds per cohort. Corrections have equal norms and recent projections; all suffix states and history are recomputed. Positive contrast indicates increased error.

Synthetic replacement costs remain 17.06 ± 1.30 and 17.83 ± 0.98 dB at 6,400 updates. Both current fits still improve, leaving optimal replacement error unresolved (Appendix H.5).

6 Discussion and conclusion

Shared-cap recovery gains replicate on independent synthetic instances. Image dependence varies with injection despite nearly coincident retrained means. DASH-LISTA makes the relevant operation explicit: matched retraining for architecture selection, output-optimized replacement for checkpoint simplification, and direction interventions to locate pathway effects. Replacement conclusions depend on function class and fitting budget. Deployment requires a quality tolerance and paired precision. The common-reference calculation in Appendix I gives an empty joint-decision interval.

AI USE STATEMENT

Generative AI tools, including ChatGPT and Codex, assisted with literature retrieval, research and experiment design, mathematical derivations and proof drafting, implementation, result analysis, manuscript drafting and editing, and figure preparation. This included assistance with code for synthetic-signal generation; numerical signals were sampled from the specified Bernoulli–Gaussian distributions using seeded random-number generators. The authors independently reviewed the AI-assisted work and take full responsibility for the paper and accompanying materials.

REPRODUCIBILITY STATEMENT

The main text and supplement specify the method, mathematical assumptions and proofs, data processing, checkpoint selection and experimental protocols. The companion `DASH-LISTA-supplementary.zip` contains this complete PDF, a reading guide, machine-readable records and CPU recomputation scripts, indexed in Appendix AO. The package is self-contained for the listed saved-record checks, using Python and NumPy. Dataset collections and model weights are excluded, so these checks reproduce reported aggregations rather than training or checkpoint inference. The evidence map identifies the available records for each result family.

ETHICS STATEMENT

The image experiments use public datasets for reconstruction, without using object or segmentation labels to identify individuals. The companion archive excludes dataset image collections and third-party model weights; source licensing and redistribution conditions continue to apply. The reconstruction models have not been validated for clinical or other safety-critical use.

REFERENCES

P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011. URL https://github.com/BIDS/BSDS500.

Sebastian Banert, Jevgenija Rudzusika, Ozan Öktem, and Jonas Adler. Accelerated forward-backward optimization using deep learning. *SIAM Journal on Optimization*, 34(2):1236–1263, 2024. doi: 10.1137/22M1532548. URL https://arxiv.org/abs/2105.05210.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. doi: 10.1137/080716542. URL https://doi.org/10.1137/080716542.

Mark Borgerding, Philip Schniter, and Sundeep Rangan. AMP-inspired deep networks for sparse linear inverse problems. *IEEE Transactions on Signal Processing*, 65(16):4293–4308, 2017. doi: 10.1109/TSP.2017.2708040. URL https://arxiv.org/abs/1612.01183.

Stephen Boyd, Laurent El Ghaoui, Eric Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics, 1994. ISBN 0-89871-334-X. URL https://web.stanford.edu/~boyd/lmibook/.

Xiaohan Chen, Jialin Liu, Zhangyang Wang, and Wotao Yin. Theoretical linear convergence of unfolded ISTA and its practical weights and thresholds. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/cf8c9be2a4508a24ae92c9d3d379131d-Abstract.html.

Patrick L. Combettes and Isao Yamada. Compositions and convex combinations of averaged nonexpansive operators. *Journal of Mathematical Analysis and Applications*, 425(1):55–70, 2015. doi: 10.1016/j.jmaa.2014.11.044. URL https://pcombet.math.ncsu.edu/jmaa3.pdf.

David L. Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009. doi: 10.1073/pnas.0909892106. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC2767368/.

Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th International Conference on Machine Learning*, 2010. URL https://icml.cc/2010/papers/449.pdf.

Howard Heaton, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. Safeguarded learned convex optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 7848–7855, 2023. doi: 10.1609/aaai.v37i6.25950. URL https://ojs.aaai.org/index.php/AAAI/article/view/25950.

Qian Hu, Jiayi Ma, Yuan Gao, Junjun Jiang, and Yixuan Yuan. MAUN: Memory-augmented deep unfolding network for hyperspectral image reconstruction. *IEEE/CAA Journal of Automatica Sinica*, 11(5):1139–1150, 2024. doi: 10.1109/JAS.2024.124362. URL https://www.ieee-jas.net/article/doi/10.1109/JAS.2024.124362.

Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Convergence rates with inexact non-expansive operators. *Mathematical Programming*, 159:403–434, 2016. doi: 10.1007/s10107-015-0964-4. URL https://arxiv.org/abs/1404.4837.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *Computer Vision – ECCV 2014*, volume 8693 of *Lecture Notes in Computer Science*, pp. 740–755, 2014. doi: 10.1007/978-3-319-10602-1_48. URL https://link.springer.com/chapter/10.1007/978-3-319-10602-1_48.

Jialin Liu, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. ALISTA: Analytic weights are as good as learned weights in LISTA. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=B1lnzn0ctQ.

Kang Liu and Jianchen Hu. Residual-controlled Douglas–Rachford splitting for differentiable solver layers, 2026. URL https://arxiv.org/abs/2608.14470.

Andrea Martin and Giuseppe Belgioioso. Learning to accelerate Krasnosel'skii–Mann fixed-point iterations with guarantees, 2026. URL https://arxiv.org/abs/2601.07665.

Christopher A. Metzler, Ali Mousavi, and Richard G. Baraniuk. Learned D-AMP: Principled neural network based compressive image recovery. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/8597a6cfa74defcbde3047c891d78f90-Abstract.html.

Hamed Sadeghi, Sebastian Banert, and Pontus Giselsson. Incorporating history and deviations in forward–backward splitting. *Numerical Algorithms*, 96(4):1819–1867, 2024. doi: 10.1007/s11075-023-01686-8. URL https://arxiv.org/abs/2208.05498.

Mark Schmidt, Nicolas Le Roux, and Francis Bach. Convergence rates of inexact proximal-gradient methods for convex optimization. In *Advances in Neural Information Processing Systems*, volume 24, 2011. URL https://proceedings.neurips.cc/paper/2011/hash/8f7d807e1f53eff5f9efbe5cb81090fb-Abstract.html.

Feiyang Shen and Hongping Gan. HUNet: Homotopy unfolding network for image compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12799–12808, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Shen_HUNet_Homotopy_Unfolding_Network_for_Image_Compressive_Sensing_CVPR_2025_paper.html.

Jiechong Song, Bin Chen, and Jian Zhang. Memory-augmented deep unfolding network for compressive sensing. In *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 4249–4258, 2021. doi: 10.1145/3474085.3475562. URL https://arxiv.org/abs/2110.09766.

Jiechong Song, Bin Chen, and Jian Zhang. Deep memory-augmented proximal unrolling network for compressive sensing. *International Journal of Computer Vision*, 131(6):1477–1496, 2023. doi: 10.1007/s11263-023-01765-2.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. URL https://ece.uwaterloo.ca/~z70wang/research/ssim/.

Jian Zhang and Bernard Ghanem. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1828–1837, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_ISTA-Net_Interpretable_Optimization-Inspired_CVPR_2018_paper.html.

Ziyang Zheng, Wenrui Dai, Duoduo Xue, Chenglin Li, Junni Zou, and Hongkai Xiong. Hybrid ISTA: Unfolding ISTA with convergence guarantees using free-form deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3226–3244, 2023. doi: 10.1109/TPAMI.2022.3172214. URL https://arxiv.org/abs/2204.11640v2.

SCIENTIFIC SUPPLEMENT: EVIDENCE BY CLAIM

Table highlighting. Pale red with bold text marks the best displayed mean; pale blue with regular text marks the second-best distinct mean. Displayed ties share the same rank. Comparisons use matched alternatives within each operator, sampling rate, budget and stage; cross-backbone panels rank methods within each branch row. Higher PSNR/SSIM and lower NMSE, latency, memory or replacement cost are better. Colors describe numerical ranks, not significance. Protocol entries, diagnostic coefficients and contrasts with different references remain unranked.

The supplement begins with complete theory–experiment derivations and saved-record checks, followed by fresh-sample, direction, new-instance evidence and paired support contributions. Historical matched comparisons, image evidence, and full conditional proofs follow. Historical execution and duplicate presentations are indexed separately in Appendix AO. All protocols preserve their original splits, budgets, seeds and uncertainty units. Historical archive names in the protocol narratives refer to their original record collections. The companion package consolidates included records under the paths in Appendix AO; no separate archive download is required for the advertised checks.

Table 6: Core-claim navigation. Record families are consolidated in the companion ZIP; Appendix AO and its evidence map identify packaged paths and verification coverage.

Claim	Paper evidence	Data / execution entry
Completion geometry	Section 4; Appendix A	`records/stability/`; theory checks
Recovery range	Table 3; Appendices B, C	`records/stability/`; two separately reported cohorts
Directional value	Figure 2; Appendix B	`records/stability/results/direction/`
Paired sample contributions	Appendix D	`records/stability/`; paired-support check
Shared-cap replication	Appendices F, F.1	`research/gain_budget_20260911/`; `records/confirmation/`
Prospective replacement prediction	Appendix F.2	`records/confirmation/`; predictions and test errors
Fitting-objective control	Appendix G	`research/objective_control_20260913/`
Replacement budget and path effects	Section 5.5; Appendix H	`records/replacement/`; 90 trajectories and all evaluation rows
Dependence versus architecture	Section 5.4	`records/core/`; image-control and COCO records
Stage-resolved BSD test	Appendix J	`records/depth_curves/`; both image splits
Conditional stability	Section 4; Appendix AG.4	Full-state assumptions; finite-tap and exponential-memory rate proofs

A COMPLETION GEOMETRY, TRAJECTORY PROPAGATION AND FINITE GAIN

The analysis uses the same complete update, correction interventions and energy aggregation as the experiments. Its loss identities hold for a fixed deterministic completion; piecewise-affine structure supplies an explicit realization for the synthetic FIR implementation. The final stability result has the contraction-domain hypotheses stated below.

A.1 AN EXACT COMPLETION-SPACE IMPROVEMENT CRITERION

Proposition 2 (Alignment and displacement)**.** *Let T map correction vectors to reconstruction vectors and fix a target $x^\star$. For any c, d in its domain, define $e_c = T(c) - x^\star$ and $\Delta = T(d) - T(c)$. Then*

$$B = \|T(c) - x^\star\|_2^2 - \|T(d) - x^\star\|_2^2 = A - S, \quad A = -2\langle e_c, \Delta\rangle, \quad S = \|\Delta\|_2^2.$$

If $\Delta \neq 0$ and $e_c \neq 0$, improvement is equivalent to

$$-\frac{\langle e_c, \Delta \rangle}{\|e_c\|_2 \|\Delta\|_2} > \frac{\|\Delta\|_2}{2\|e_c\|_2}.$$

For a collection of common samples, the same identity holds after summing A, B, S over samples. Division by their common total target energy gives the difference of pooled linear NMSE values. If $E_d = \sum_i \|T_i(d_i) - x_i^\star\|^2 > 0$ and $E_c > 0$, their dB contrast is

$$10 \log_{10}(E_c/E_d) = 10 \log_{10}\left(1 + \frac{\sum_i (A_i - S_i)}{E_d}\right).$$

Thus effect size combines the signed completion-space improvement with the reconstruction error scale.

Proof. Since $T(d) - x^\star = e_c + \Delta$, expansion of $\|e_c + \Delta\|_2^2$ proves $B = A - S$. Dividing the strict inequality $-2\langle e_c, \Delta \rangle > \|\Delta\|_2^2$ by its positive normalization gives the angular criterion. Summation and division by a common positive denominator preserve the identity. Since $E_c = E_d + \sum_i B_i$, substitution into the logarithmic energy ratio proves the dB expression. If $\Delta = 0$, $B = 0$; if $e_c = 0$, $B = -S \leq 0$. □

In image CS, T includes domain conversions and the chosen reconstruction output. A correction fit in coefficient space targets its input; the improvement criterion acts on its reconstructed output. Ground truth is used for this offline diagnostic, while model inference uses measurements. The direct-branch specialization follows by substituting $\Delta = (1 - \alpha)d$ for a common state and fixed mixture in Eq. (48).

A.2 Orthogonal history and the reflection contrast

Proposition 3 (Norm-preserving directional contrast)**.** *Let U be a subspace of correction space, $p = P_U d$ and $h = d - p$. For $d_\pm = p \pm h$, $P_U d_+ = P_U d_- = p$ and $\|d_+\|_2 = \|d_-\|_2$. Define $a = (T(d_+) + T(d_-))/2 - x^\star$ and $u = (T(d_+) - T(d_-))/2$. The reflection loss contrast equals $-4\langle a, u \rangle$. If T is continuous with finitely many affine pieces on the segment, let $J(t)$ be an affine-piece linear part along $p + th$, $-1 \leq t \leq 1$, and set $\overline{J} = \frac{1}{2}\int_{-1}^{1} J(t)\, dt$. Then $u = \overline{J}h$.*

Proof. Orthogonality gives $\langle p, h \rangle = 0$, so both squared correction norms equal $\|p\|_2^2 + \|h\|_2^2$. The two reconstruction errors are $a + u$ and $a - u$; subtracting their squared norms gives $\|a - u\|_2^2 - \|a + u\|_2^2 = -4\langle a, u \rangle$. A finite piecewise-affine network restricted to a line segment is continuous and piecewise affine. Partition the segment at its breakpoints. On each interval the output derivative along the line is $J(t)h$. The integral of these interval derivatives equals $T(p + h) - T(p - h)$ by telescoping. If a segment lies on a region boundary, choose any adjacent affine piece agreeing on that segment; its tangential product $J(t)h$ is the same. Division by two proves the representation. □

Taking $U = \mathrm{span}\{r_k, r_{k-1}\}$ isolates the component outside the two most recent increments. At startup $k = 0$, use $U = \mathrm{span}\{r_0\}$; for $k = 0, 1$, FIR has $h = 0$. For later layers, four-tap FIR with effective taps $c_{k,j} = \gamma_k b_{k,j}$ satisfies

$$h = (I - P_U) \sum_{j=2}^{J_k - 1} c_{k,j} r_{k-j}.$$

The reflection flips this component while preserving the recent-increment projection and total norm. In a single affine region, $a = T(p) - x^\star$ and $u = Jh$, so the contrast is exactly $-4\langle T(p) - x^\star, Jh \rangle$. Across regions, the output midpoint a and the segment average $\overline{J}$ retain the exact formula.

Jacobian of the executed synthetic completion. Use column vectors and the learned backprojection $W \in \mathbb{R}^{N \times M}$. With stored weights interpreted in real arithmetic, the frozen layer has $s = v + d$, $u_s = s + \mathrm{CNN}(s)$, $w = \mathrm{soft}_{\theta_2}((I - WA)u_s + Wy)$ and $T(d) = \alpha w + (1 - \alpha)s$. The scalar α

is determined by trained layer parameters and is constant across the paired interventions. On one affine region, differentiation gives

$$J = \alpha D_2(I - WA)(I + K) + (1 - \alpha)I, \qquad K = C_3 D_{\mathrm{ReLU},2} C_2 D_{\mathrm{ReLU},1} C_1.$$

Here C_i are the padded convolution operators, $D_{\mathrm{ReLU,i}}$ the activation masks, and D_2 the soft-threshold derivative mask. This is the three-convolution residual tail in the evaluated synthetic implementation. The learned tap values, thresholds and mixture are frozen. The projection and norm rescalings construct input vectors before applying this common completion. General backbones retain the loss identity through their own T; this explicit Jacobian specializes it to the synthetic CP model.

A.3 Exact finite-suffix propagation

Proposition 4 (Propagation with the feedback buffer)**.** *Let z_d, z_c be two states immediately after an intervention, containing both complete backbone state and the feedback buffer. Let Ψ_k execute the same frozen suffix and extract the terminal reconstruction. With $\Delta_K = \Psi_k(z_d) - \Psi_k(z_c)$ and $e_c = \Psi_k(z_c) - x^\star$, terminal improvement is $-2\langle e_c, \Delta_K\rangle - \|\Delta_K\|_2^2$. For a finite continuous piecewise-affine suffix,*

$$\Delta_K = \left[\int_0^1 D\Psi_k(z_c + t(z_d - z_c))\, dt\right](z_d - z_c),$$

where affine-piece linear parts specify the tangential derivative on region boundaries. If each subsequent augmented-state step has Lipschitz constant L_t and the terminal readout has constant L_o in compatible norms, then

$$\|\Delta_K\| \le L_o \prod_{t=k+1}^{K-1} L_t \, \|z_d - z_c\|.$$

Proof. Apply Proposition 2 to the common suffix. The integral formula follows by the same segment partition and telescoping as Proposition 3. Applying each Lipschitz inequality successively and finally the readout inequality proves the bound; an empty product is one. A synthetic FIR step consists of affine maps, ReLU, soft thresholding and buffer shifts, with frozen coefficients. Each step, and hence its finite composition, is continuous piecewise affine. □

The intervention changes the current reconstruction after computing its increment; existing buffer entries at that step are retained. Every later increment is recomputed from the altered state. The suffix therefore models the single-layer experiment's full trajectory. Keeping the original future increments would define a different map. Learned input-dependent gates or nonlinear memories use their own suffix regularity; the error identity continues to hold for their deterministic outputs.

A.4 Finite-tap gain: an envelope shared by the trained controls

Here tap coefficients include any external injection gain; the evaluated implementation absorbs that gain into its taps.

Proposition 5 (Active-tap envelope)**.** *For any norm, FIR coefficients with $\sum_{j=0}^{J_k-1} |b_{k,j}| \le G$ satisfy $\|d_k\| \le G \max_{j<J_k} \|r_{k-j}\|$. Collapsing the delayed coefficients into a two-time correction preserves this bound. For ℓ_p norms, $1 \le p \le \infty$, the nonlinear current correction $G\tanh(f_k(r_k)) \odot r_k$ satisfies $\|d_k\|_p \le G\|r_k\|_p$.*

Proof. The triangle inequality bounds FIR by $\sum_j |b_{k,j}|\|r_{k-j}\|$. In the two-time arm, $|b_{k,0}| + |\sum_{j\ge1} b_{k,j}| \le \sum_j |b_{k,j}|$. For nonlinear current correction, every coordinate multiplier has absolute value at most G, which gives the stated ℓ_p inequality. □

The gain experiment scales active raw tanh taps before collapsing delayed coefficients. It therefore enforces this envelope from the first layer. The original synthetic tanh parameterization has the layerwise bound $G_k \le J_k$; the capped condition uses $G = 0.5$. These are permitted-size bounds: realized norm and completion-space alignment are measured separately.

Constants for the trained synthetic CP update. At layer k, define $P_k(x) = \mathrm{soft}_{\theta_{1,k}}((I - WA)x + Wy)$ and $R_k(v) = \alpha_k \,\mathrm{soft}_{\theta_{2,k}}((I - WA)(v + \mathrm{CNN}(v)) + Wy) + (1 - \alpha_k)v$. Thus the zero-feedback update is $R_k(P_k(x))$ and injection gives $R_k(P_k(x) + d)$. In the Euclidean norm, set

$$s_W = \|I - WA\|_2, \quad L_C = \prod_{i=1}^{3} \|C_i\|_2, \quad L_{R,k} = \alpha_k s_W(1 + L_C) + (1 - \alpha_k).$$

Soft thresholding and ReLU are nonexpansive, yielding the explicit choices

$$L_{d,k} = L_{R,k}, \qquad q_k = L_{R,k} s_W, \qquad C_r = s_W + 1.$$

For a target $x^\star$, retain the reference defects

$$\epsilon_{B,k} = \|R_k(P_k(x^\star)) - x^\star\|_2, \qquad \nu_{r,k} = \|P_k(x^\star) - x^\star\|_2.$$

The triangle inequality now gives, for every x,

$$\|R_k(P_k(x)) - x^\star\|_2 \le q_k \|x - x^\star\|_2 + \epsilon_{B,k}, \qquad \|r_k(x)\|_2 \le C_r \|x - x^\star\|_2 + \nu_{r,k}.$$

These inequalities include the trained thresholds and CNN. For noiseless observations, $P_k(x^\star) = \mathrm{soft}_{\theta_{1,k}}(x^\star)$; if the target has s nonzeros, $\nu_{r,k} \le \sqrt{s}|\theta_{1,k}|$. With $y = Ax^\star + n$, nonexpansiveness gives the bound $\nu_{r,k} \le \|Wn\|_2 + \sqrt{s}|\theta_{1,k}|$. Taking maxima over the finite set of layers gives uniform constants. The constants are explicit functions of frozen weights and the target; the norm products provide bounds, while the completion identities quantify signed effects on the evaluated outputs.

Nullspace obstruction for these explicit constants. For every underdetermined $A \in \mathbb{R}^{M \times N}$, $M < N$, choose $0 \ne z \in \ker A$. For every W, $(I - WA)z = z$, and therefore $s_W = \|I - WA\|_2 \ge 1$. With $L_C \ge 0$ and $0 \le \alpha_k \le 1$,

$$L_{R,k} = \alpha_k s_W(1 + L_C) + (1 - \alpha_k) \ge 1, \qquad q_k = L_{R,k} s_W \ge 1.$$

Thus no nonnegative feedback budget makes these particular constants satisfy $q + a_G < 1$. This fact is independent of the older CP coherence condition. The finite-horizon recursion below remains valid with $q \ge 1$. The geometric conclusion concerns separately established constants on a specified domain. Actual completion Jacobians retain threshold masks and signed matrix products that these norm products bound separately; a tighter contraction certificate must account for that structure and the domain of all iterates. The inequality above concerns the displayed bounds, not the smallest Lipschitz constant of the trained network.

Theorem 6 (Finite-memory error envelope)**.** *Fix observations and a complete-state reference $\xi^\star$ whose iterate component is $x^\star$. In one norm and on a domain containing all perturbed states and their zero-feedback comparisons, suppose uniformly*

$$\begin{aligned} \|\mathcal{B}_k(\xi, 0) - \xi^\star\| &\le q\|\xi - \xi^\star\| + \epsilon_B, \\ \|\mathcal{B}_k(\xi, d) - \mathcal{B}_k(\xi, 0)\| &\le L_d\|d\|, \\ \|r_k\| &\le C_r\|\xi_k - \xi^\star\| + \nu_r. \end{aligned}$$

All constants are nonnegative. Assume a zero initial buffer, at most $J \ge 1$ taps, and the envelope in Proposition 5, with any explicit external gain absorbed into G. Set $a_G = L_d G C_r$ and $\eta = \epsilon_B + L_d G \nu_r$. Then $e_k = \|\xi_k - \xi^\star\|$ satisfies

$$e_{k+1} \le q e_k + a_G \max_{0 \le j < \min(J, k+1)} e_{k-j} + \eta.$$

Starting at $b_0 = e_0$, equality in this recursion defines a finite-horizon upper envelope $e_k \le b_k$. If $a_G > 0$ and $q + a_G < 1$, the unique root $\rho \in (q, 1)$ of $\rho^J - q\rho^{J-1} - a_G = 0$ gives

$$e_k \le e_0 \rho^k + \frac{\eta}{1 - q - a_G}.$$

For $a_G = 0$ and $q < 1$, $e_k \le q^k e_0 + \eta(1 - q^k)/(1 - q)$.

Proof. The triangle inequality and active-tap envelope give

$$e_{k+1} \le q e_k + \epsilon_B + L_d G \max_j (C_r e_{k-j} + \nu_r),$$

which is the stated recursion. Its right-hand side is monotone in each past error, so induction gives $e_k \le b_k$. For $a_G > 0$, $f(t) = t^{J-1}(t-q) - a_G$ is strictly increasing for $t > q$; $f(q) = -a_G < 0$ and $f(1) = 1 - q - a_G > 0$. Hence the root exists and is unique. Put $E = \eta/(1-q-a_G)$. If $e_t \le e_0\rho^t + E$ through time k, then $0 < \rho < 1$ implies

$$e_{k+1} \le e_0(q\rho^k + a_G\rho^{k-J+1}) + (q + a_G)E + \eta = e_0\rho^{k+1} + E.$$

At startup the smallest available time index exceeds $k - J + 1$, tightening the bound. The induction starts at $e_0 \le e_0 + E$. Summing a geometric series proves the $a_G = 0$ case. Taking $\epsilon_B = \nu_r = 0$ recovers pure geometric decay. Uniform hypotheses at every future step give $\limsup_k e_k \le E$; finite-horizon hypotheses give the finite prefix. □

The explicit CP constants supply finite-depth envelopes for frozen trained weights, including their reference defects. In the underdetermined setting they obey $q \ge 1$. Geometric decay in the conditional theorem requires a separate domain-based bound satisfying $q + a_G < 1$. The exponential-memory extension in Appendix AF treats other memory kernels. Its CP coherence calculation specifies the scope of the zero-defect sufficient-condition route.

A.5 Numerical correspondence with saved intervention records

Some saved alignment terms are reconstructed as $A = B + S$ from error and displacement energies. This is an accounting and execution-consistency check, not an independent prediction of a new model's history benefit. We reconcile all $3 \times 16 \times 4096 = 196{,}608$ saved sample–layer projection contrasts with Eq. (7). The stored projection-minus-FIR error equals B; the stored completion-difference energy equals S. Thus $A = B + S$ recovers the alignment term from saved scalar output energies. The original execution also saved 3,072 batch–layer checks: maximum reflection norm relative error was 3.44×10^{-8}, and the raw-vector loss-identity residual was at most 6.40×10^{-15}. The scalars support energy reconciliation; the piecewise Jacobian formula is established by the preceding derivation.

Table 7: Layer-16 completion-space terms recovered from the saved 4,096-signal projection diagnostic. Energy ratios use the corresponding total target energy. All three columns are scaled by 10^{10}. $A - S = B$ measures FIR benefit over the unscaled projection; the equal-norm reflection is a separate contrast.

Seed	A/P	S/P	B/P	A/S
0	3.7363	1.7127	2.0235	2.1815
1	7.0703	4.3354	2.7349	1.6308
2	8.0982	3.9028	4.1954	2.0750

The reflection error contrast favors FIR in 41 of 42 layer-3–16 cells. By Eq. (9), this corresponds to a negative pooled midpoint–direction inner product in those cells. The three seed and all layer records are retained. Separately, 100 piecewise-affine toy maps verify the segment-integral identity across activation crossings; 36 finite-gain settings, with 108 reference-defect cases, verify the stated rate envelope, including startup. Another 500 points on 50 small dense CP examples verify the derived sensitivity and noisy reference-defect bounds. These arithmetic tests and saved-record checks accompany the proofs. The source entry is `check_theory_alignment.py`, with all values in `record_checks.json`.

A.6 The signed Jacobian of the complete FIR state

The full-space product bound discards threshold masks and signed feedback couplings. We retain both to formulate a regional certificate for frozen weights. The matrix-inequality step uses quadratic Lyapunov analysis (Boyd et al., 1994); the state matrix below follows the executed DASH recursion. Observations are fixed, effective taps include the injection gain, and stored weights define real-arithmetic maps.

Write $P_k(x)$ for the first proximal step and $R_k(s)$ for the entire refinement, second proximal step and mixture with both-branch input s. For a positive integer J of taps, the pre-step state is $z_k =$

$(x_k, h_{k,1}, \ldots, h_{k,J-1})$, with $h_{k,j} = r_{k-j}$. Unavailable increments and their coefficients are zero. The augmented update is

$$s_k = (1 + c_{k,0})P_k(x_k) - c_{k,0}x_k + \sum_{j=1}^{J-1} c_{k,j}h_{k,j},$$
$$\mathcal{F}_k(z_k) = (R_k(s_k),\ P_k(x_k) - x_k,\ h_{k,1}, \ldots, h_{k,J-2}). \tag{17}$$

History shifts are part of the dynamics rather than external inputs. For $J = 1$, retain only the iterate component.

Proposition 7 (Active-region state matrix)**.** *Fix a joint activation cell of the first threshold and the synthetic ReLU/soft-threshold completion. Let $U_k = DP_k$, $V_k = DR_k$ and $T_k = (1 + c_{k,0})U_k - c_{k,0}I$ on that region. For $J \geq 2$, the Jacobian of Eq.* (17) *is*

$$Q_k = \begin{bmatrix} V_kT_k & c_{k,1}V_k & \cdots & c_{k,J-1}V_k \\ U_k - I & 0 & \cdots & 0 \\ 0 & I & \cdots & 0 \\ \vdots & & \ddots & \vdots \end{bmatrix}. \tag{18}$$

There are J block rows; the last $J - 2$ shift the history and are absent when $J = 2$. For $J = 1$, $Q_k = V_kT_k$. In the evaluated tied-backprojection CP model, with $L = I - WA$,

$$U_k = D_{1,k}L, \qquad V_k = \alpha_k D_{2,k}L(I + K_k) + (1 - \alpha_k)I,$$

where K_k is the refiner Jacobian and $D_{1,k}, D_{2,k}$ are threshold masks.

Proof. The derivatives of s_k with respect to its iterate and history blocks are T_k and $c_{k,j}I$. Composition with R_k gives the first block row. Differentiating the new increment gives $U_k - I$; the remaining derivatives are identity shifts. The formulas for U_k, V_k follow by differentiating the threshold, consistency and residual-refinement operations. □

For refinement-only injection, put $V_k^{\mathrm{ref}} = \alpha_k D_{2,k}L(I+K_k)$. The first block row becomes $[V_k^{\mathrm{ref}}T_k + (1-\alpha_k)U_k,\ c_{k,1}V_k^{\mathrm{ref}}, \ldots]$; the increment and shift rows stay unchanged. Thus the injection locations have different state matrices even when their refiners share weights.

A.7 Recovery references and a regional contraction theorem

An accurate reconstruction need not be a fixed point of each learned layer. Define $r_k^\star = P_k(x^\star) - x^\star$, with $r_t^\star = 0$ for $t < 0$, and the target-consistent buffer reference $z_k^\star = (x^\star, r_{k-1}^\star, \ldots, r_{k-J+1}^\star)$. Its exact defect is

$$b_k = \mathcal{F}_k(z_k^\star) - z_{k+1}^\star = \left(R_k\left(P_k(x^\star) + \sum_{j=0}^{J-1} c_{k,j}r_{k-j}^\star \right) - x^\star, 0, \ldots, 0 \right). \tag{19}$$

The buffer components cancel because the reference obeys their shift rule. All terms are computable from frozen weights, observations and the offline target. Setting the ideal buffer to zero would introduce extra defects whenever learned thresholds give $r_k^\star \neq 0$.

Use one positive-definite matrix H throughout the certified horizon and $\|v\|_H = (v^\top Hv)^{1/2}$. The following theorem applies to the full state or to a fixed, proved invariant subspace containing the reference states, expressed in orthonormal coordinates.

Theorem 8 (Regional finite-prefix recovery certificate)**.** *Fix a horizon K and nonnegative radii R_k^{ball}, $0 \leq k \leq K$. Let $\mathcal{D}_k = \{z : \|z - z_k^\star\|_H \leq R_k^{\mathrm{ball}}\}$. For $0 \leq k < K$, assume $\mathcal{F}_k$ is continuous with finitely many affine pieces on $\mathcal{D}_k$. For every affine region intersecting this ball, suppose its augmented matrix satisfies*

$$Q_{k,\sigma}^\top H Q_{k,\sigma} \preceq \rho_k^2 H, \qquad \rho_k \geq 0. \tag{20}$$

Set $\delta_k = \|b_k\|_H$ and require $\rho_k R_k^{\mathrm{ball}} + \delta_k \leq R_{k+1}^{\mathrm{ball}}$. If $z_0 \in \mathcal{D}_0$, all iterates remain in the stated balls and

$$\|z_K - z_K^\star\|_H \leq \left(\prod_{k=0}^{K-1} \rho_k \right) \|z_0 - z_0^\star\|_H + \sum_{t=0}^{K-1} \delta_t \prod_{k=t+1}^{K-1} \rho_k. \tag{21}$$

If $\rho_k \le \rho < 1$ and $\delta_k \le \delta$, this is at most $\rho^K \|z_0 - z_0^\star\|_H + \delta(1-\rho^K)/(1-\rho)$. For $H \succeq \underline{h} I$ with $\underline{h} > 0$, division by $\sqrt{\underline{h}}$ also bounds $\|x_K - x^\star\|_2$.

Proof. The segment joining z and $z_k^\star$ stays in the convex ball. Partition it into the finitely many affine pieces of $\mathcal{F}_k$. Equation (20) bounds the directional derivative on each piece in the H norm by $\rho_k \|z - z_k^\star\|_H$. Integrating along the segment gives $\|\mathcal{F}_k(z) - \mathcal{F}_k(z_k^\star)\|_H \le \rho_k \|z - z_k^\star\|_H$. Continuity supplies the same bound on region boundaries. Adding b_k gives $e_{k+1}^H \le \rho_k e_k^H + \delta_k$. The radius condition proves invariance by induction. Unrolling proves Eq. (21); the uniform bound is a geometric sum. Finally, the iterate error is at most the full-state Euclidean error, which is at most $e_K^H/\sqrt{\underline{h}}$. □

Zero defects give geometric recovery on the certified prefix. An infinite sequence additionally needs the domain and matrix conditions at every future step. With uniform $\rho < 1$ and $\delta_k \to 0$, recovery error tends to zero: split the defect sum at T, let its finite initial part decay, and bound its tail by $\sup_{t \ge T} \delta_t/(1-\rho)$. A finite untied network specifies no future layers automatically.

Checking the activation domain. A fixed activation pattern defines affine inequalities $\ell_i^\top z \le a_i$ for the first threshold, both ReLU layers and second threshold. Successive substitution through that pattern constructs these rows from the weights. A ball lies in this cell exactly when

$$\ell_i^\top z_k^\star + R_k^{\mathrm{ball}} \sqrt{\ell_i^\top H^{-1} \ell_i} \le a_i \quad \text{for every row } i. \tag{22}$$

Maximizing the linear functional over the ellipsoid proves the formula; the extremizing direction is $H^{-1}\ell_i$. Constant rows are tested directly. A ball crossing cells requires the matrix inequality for all intersected pieces, rather than one sampled Jacobian. Candidate metrics can be searched with linear matrix inequalities at fixed ρ and then checked against the domain, radius and initialization conditions. Formal numerical certification additionally needs validated rounding or rigorous error bounds for the inequalities.

A support-restricted instance. For a fixed coordinate set S, let E_S insert its coordinates and $\mathcal{E}_S = \mathrm{diag}(E_S, \dots, E_S)$ insert all J blocks. If both threshold masks are zero outside S throughout a cell, an augmented input with every block supported on S produces first estimates, increments, corrected estimates and final outputs in S. The refiner may mix coordinates; its second threshold and the direct path restore the output support. The restricted matrix is therefore $\widetilde{Q}_{k,\sigma} = \mathcal{E}_S^\top Q_{k,\sigma} \mathcal{E}_S$, and Eq. (20) is tested in dimension $J|S|$. This argument requires cell-wide support invariance, not merely the support observed in one output. Threshold masks and the restriction can remove the nullspace directions that cause the full-space norm lower bound; contraction then depends on the resulting signed state matrix.

Untied layers and block contraction. When each state and its reference share a verified joint activation cell, define the vector error $\mathbf{e}_k = z_k - z_k^\star$. Its exact recursion is $\mathbf{e}_{k+1} = Q_k \mathbf{e}_k + b_k$. Put $Q_{b:a} = Q_{b-1} \cdots Q_a$, with an empty product equal to the identity. Induction yields

$$\mathbf{e}_b = Q_{b:a} \mathbf{e}_a + \sum_{t=a}^{b-1} Q_{b:t+1} b_t.$$

Induced norms of verified products in one common metric give a block version of Eq. (21); intermediate cell containment remains required. Separate spectral radii below one do not control an untied product: $Q_1 = \left[\begin{smallmatrix} 1/2 & 2 \\ 0 & 1/2 \end{smallmatrix}\right]$ and $Q_2 = Q_1^\top$ each have spectral radius $1/2$, whereas $\rho(Q_2 Q_1) > 1$.

A.8 How signed history changes local recovery modes

In a stationary affine regime, suppose U, V share an invariant direction v with $Uv = uv$ and $Vv = mv$. Its J iterate/history coordinates form an invariant subspace of Eq. (18), with characteristic polynomial

$$p_{u,m}(\lambda) = \lambda^J - m[(1+c_0)u - c_0]\lambda^{J-1} - m(u-1)\sum_{j=1}^{J-1} c_j \lambda^{J-1-j}. \tag{23}$$

For $\lambda \neq 0$, the buffer block of $\lambda I - Q$ is triangular with determinant λ^{J-1}. Its Schur complement is $\lambda - m[(1+c_0)u - c_0] - m(u-1)\sum_{j=1}^{J-1} c_j \lambda^{-j}$. Multiplication gives Eq. (23); polynomial equality extends it to zero. The homogeneous modal recursion is asymptotically stable exactly when all roots lie inside the unit disk. Every rate strictly above the spectral radius admits a finite-prefactor exponential bound; repeated roots need not admit prefactor one at the radius itself.

The current correction with gain g has iterate multiplier $m[(1+g)u - g]$. History instead changes a polynomial. For an affine example, let $m = 1$ and let U have modes $u = 0.2, 0.8$. Under $|g| \leq 0.5$, the best common current gain has worst-mode factor 0.7: the second mode is $0.8 - 0.2g \geq 0.7$, and $g = 0.5$ attains this bound. Two taps $c_0 = c_1 = 0.25$ satisfy the same absolute-sum cap. Their polynomials are $\lambda^2 + 0.2$ and $\lambda^2 - 0.75\lambda + 0.05$, giving worst radius

$$\max\left\{\sqrt{0.2},\ \frac{0.75 + \sqrt{0.3625}}{2}\right\} = 0.6760399 < 0.7.$$

This establishes a signed-history benefit for the stated affine model and gain class. For the unrestricted nullspace mode $u = m = 1$, the polynomial becomes $\lambda^{J-1}(\lambda - 1)$ for all taps: mask or domain structure is essential for strict contraction of that mode. The actual untied CNN model uses the full state matrices without assuming stationary masks or simultaneous diagonalization.

A.9 Frozen-weight checks of the derived state matrix

We check the derived state matrix on the three original CUDA FIR checkpoints without fitting or retraining. For each seed, we use sample index zero from its existing fresh-signal stream, draw the original 10,000-signal block before indexing, and examine layer 16 at the target-consistent reference. Stored FP32 weights are evaluated in FP64 for this algebra check; these numbers do not replace the FP32 reconstruction benchmark.

Three full-state Jacobian–vector products per checkpoint agree with automatic differentiation to a maximum absolute discrepancy of 4.45×10^{-15}. At these references, both threshold masks vanish outside the target support, of size 46, 43 and 51. We form the restricted matrix $\widetilde{Q}$ and the candidate metric $H = \sum_{i=0}^{127} (\widetilde{Q}^i)^\top \widetilde{Q}^i$. The minimum eigenvalue of $H - \widetilde{Q}^\top H \widetilde{Q}$ is numerically 1 in all three cases. The factors in Table 8 are the directly evaluated induced H norms, not the spectral radii themselves.

Table 8: Local state-matrix diagnostics at layer 16 of three frozen FIR models. R_H is a conservative single-cell radius. †Entry distance uses the support-projected actual state; the actual state also has nonzero off-support entries and therefore lies outside the proposed restricted domain.

Seed	$\rho(\widetilde{Q})$	$\|\widetilde{Q}\|_H$	R_H	$\|b\|_H$	Entry distance†
0	0.6424	0.9880	6.73×10^{-5}	1.18×10^{-3}	1.18×10^{-2}
1	0.7231	0.9867	4.74×10^{-5}	7.58×10^{-4}	7.00×10^{-3}
2	0.6580	0.9901	7.65×10^{-5}	1.76×10^{-3}	1.36×10^{-2}

To compute R_H, we divide every threshold/ReLU activation margin by the Euclidean norm of its augmented-state gradient, take the minimum to obtain R_2, and set $R_H = \sqrt{\lambda_{\min}(H)} R_2$. Constant inequalities remain satisfied. This sufficient inner radius includes both thresholds and both ReLU layers; it is not an estimate of the largest admissible region. The reference defects exceed these radii, so the same-radius condition $\|\widetilde{Q}\|_H R_H + \|b\|_H \leq R_H$ fails in each case. The actual pre-layer states also have off-support maximum magnitudes 1.66×10^{-3}, 6.34×10^{-4} and 1.40×10^{-3}. Thus the sampled matrices have local contractive metrics, while this particular restricted, single-cell construction supplies neither entry nor a closed tube for the executed trajectory. A whole-network certificate requires these conditions across the layer sequence; the three local checks do not supply it.

The accompanying arithmetic checker independently verifies 100 modal characteristic polynomials, 50 ellipsoid support-function identities, complete-state recursions for one, two and four taps, non-unit external gains, and 54 forced-history cases including zero-gain/zero-decay limits. It also checks

the affine gain example and the untied-product counterexample. Checkpoint hashes and floating-point diagnostics are recorded separately. These are algebra and local numerical checks, not interval-arithmetic certification.

B FROZEN-SAMPLE STABILITY AND DIRECTION CONTROLS

Fixed samples and replay. This supplemental protocol was fixed after the original CUDA comparison. There are three existing problem/initialization/training instances, with no additional training in this block. Each uses ten independent blocks of 10,000 Bernoulli–Gaussian signals, density 0.1 and standard-normal amplitudes, generated by the unchanged NumPy routine with streams $9600000 + 100s + b$ ($b = 0, \ldots, 9$). All methods share each saved array. The aggregate NMSE sums all 100,000 error and target energies before the logarithm. Block dB values are not averaged into the primary metric. The original test and diagnostic arrays are retained separately. Original CUDA scores and prior CUDA replays of MUSA checkpoints match the new replay within 10^{-7} dB; historical MUSA-runtime values retain their separately recorded numerical differences. Repeated reconstruction is bitwise identical, and permutation/single-sample correspondence passes an absolute 2×10^{-5} output tolerance. The prior diagnostic uses exactly the reconstructed stream $760000 + s$ and 2,048 samples.

Table 9: Frozen new-sample confirmation. NMSE (dB), lower is better; mean ± sample SD across three paired problem instances, not 30 blocks.

Method	Seed 0	Seed 1	Seed 2	Mean ± SD
Two-time	-75.58	-69.60	-70.87	-72.02 ± 3.15
Nonlinear current	-68.85	-73.35	-74.10	-72.10 ± 2.84
FIR	$\mathbf{-75.84}$	$\mathbf{-74.03}$	$\mathbf{-87.66}$	$\mathbf{-79.18 \pm 7.40}$

The separately retained MUSA FIR checkpoints, executed on CUDA, score -79.53 ± 7.94 dB on these arrays; they are not pooled with CUDA FIR as new problem seeds. CUDA comparator-minus-FIR differences are $(0.2571, 4.4296, 16.7889)$ dB for two-time and $(6.9956, 0.6777, 13.5626)$ dB for nonlinear current correction. Their paired summaries are 7.16 ± 8.60 and 7.08 ± 6.44 dB. All original tables remain unchanged. Tail contributions use raw error energy, whereas per-signal quantiles use error/target ratios. In the prior MUSA seed-1 diagnostic, the three largest errors contribute 98.91% of total error, versus 24.29% from the largest ten errors in the original 10,000-signal set (the same top-0.1% rule with ceiling). The ≥ 61 support group contributes 99.07% of its diagnostic error. This locates the observed reversal in the error tail; it does not certify a general cause for every failure. No samples are removed.

Conditional support scan. For each fixed support size K, generate 2,048 signals at stream $9700000 + 100s + K$, with uniformly sampled supports and Gaussian amplitudes. This is a stress protocol, not the original Bernoulli test distribution.

Table 10: All fixed-support conditions, NMSE (dB), mean ± sample SD over three existing instances. Best mean per row is bold.

Support	FIR	Two-time	Nonlinear
30	$\mathbf{-98.63 \pm 1.52}$	-84.72 ± 1.49	-79.23 ± 0.90
40	$\mathbf{-96.75 \pm 0.93}$	-83.72 ± 2.23	-77.48 ± 1.16
50	$\mathbf{-93.57 \pm 0.82}$	-81.30 ± 2.83	-75.98 ± 0.65
60	$\mathbf{-88.31 \pm 0.89}$	-76.17 ± 3.42	-72.85 ± 1.43
70	$\mathbf{-59.09 \pm 6.74}$	-54.33 ± 7.90	-55.41 ± 1.02
80	$\mathbf{-36.46 \pm 0.89}$	-35.06 ± 1.14	-35.05 ± 1.43

Matched-norm controls in actual completion. Use the first 4,096 IDs of block $b = 0$ at each seed, reusing confirmation samples rather than claiming another independent set. In each unmodified CUDA FIR trajectory, project d_k by per-sample FP64 SVD onto $\mathrm{span}(r_k, r_{k-1})$ with

tolerance $500\epsilon_{64}\sigma_{\max}$ to obtain p_k. Evaluate original d, projection p, shrunk FIR $(\|p\|/\|d\|)d$, norm-restored projection $(\|d\|/\|p\|)p$, and reflection $2p - d$ through the actual completion. Targets and supports are used only for reporting. Normalization is undefined when $\|d\| \le 10^{-12}$ or $\|p\| \le \max(10^{-12}, 10^{-8}\|d\|)$; these cases would be separately flagged, with zero diagnostic correction and no equal-norm interpretation. There are zero such cases in this run. Normalized projection is an oracle diagnostic and can exceed the original coefficient budget. Native completion matches the formal implementation bitwise. First-two-layer projection residual fractions are below 10^{-10}; reflection norm errors are below 2×10^{-5}. Saved FP64 energies also verify

$$\|e^P\|^2 - \|e^P + \Delta x\|^2 = -2\langle e^P, \Delta x\rangle - \|\Delta x\|^2, \quad \Delta x = x^F - x^P.$$

This is an error-expansion identity, not a new convergence result. Both equal-norm pairings favor the FIR direction in 41/42 cells from layers 3–16; reflection is worse in 41/42. These are descriptive counts of dependent layers/instances. The complete all-layer records, including the exceptions, accompany the diagnostic arrays.

Table 11: Layer-16 same-state diagnostic NMSE (dB). These 4,096-signal values are not the 100,000-signal confirmation scores. Best per row bold.

Seed	Original	Projection	Shrunk FIR	Restored proj.	Reflection
0	**−91.100**	−90.094	−90.899	−89.992	−88.172
1	**−62.299**	−62.297	**−62.299**	−62.295	−62.288
2	**−90.650**	−88.926	−90.391	−88.528	−86.054

Single-layer propagation. Apply projection or reflection at exactly one of layers 4, 8, 12 or 16; all later steps use the original FIR rule with history rebuilt along the perturbed trajectory. Equal correction norm holds at the common intervention state, not across subsequent separated trajectories. Both interventions increase endpoint error in all 12 seed–layer cases below. This measures local intervention and propagation, not the quality of a retrained model.

Table 12: Final NMSE after a single intervention, dB. Unchanged original shown alongside each case; lower is better. All specified cases retained.

Seed	Layer	Original	Projection	Reflection
0	4	**−91.100**	-77.736	-60.082
0	8	**−91.100**	-72.487	-60.029
0	12	**−91.100**	-69.993	-56.937
0	16	**−91.100**	-90.094	-88.172
1	4	**−62.299**	-56.941	-52.743
1	8	**−62.299**	-57.027	-52.884
1	12	**−62.299**	-59.889	-53.013
1	16	**−62.299**	-62.297	-62.288
2	4	**−90.650**	-86.665	-56.209
2	8	**−90.650**	-83.802	-62.952
2	12	**−90.650**	-69.955	-55.875
2	16	**−90.650**	-88.926	-86.054

C Independent-instance retraining confirmation

Fixed design and training. We train FIR, two-time and nonlinear current-increment correction on two additional problem instances (seeds 3 and 4), yielding six models. The instances were fixed before training and reported separately from seeds 0–2. Each seed changes the sensing operator, backbone initialization and sample streams; this tests retraining on new instances, not zero-shot operator transfer. The task remains noiseless Bernoulli–Gaussian recovery with $N = 500$, $M = 250$, activation probability 0.1 and standard-normal amplitudes. The operator uses NumPy RandomState(seed), a 250×500 Gaussian draw, column normalization in FP64 and conversion to FP32. This rule reproduces the original three saved operators byte-for-byte.

All three arms share the seed-specific initial backbone and data streams, 16 layers and 480,000 updates: 10,000 steps in each of 48 phases. Adam uses learning rates $(5 \times 10^{-4}, 10^{-4}, 10^{-5})$ per depth, batch size 64, $\epsilon = 10^{-8}$ and fresh optimizer state per phase. The first phase trains newly introduced parameters; joint phases multiply gradients by 0.3^{k-a} for parameter age a at zero-based depth k. Validation is performed at step 1, every 200 updates and the phase endpoint. Phase-best weights are restored; the final checkpoint minimizes validation error among depth-16 candidates. Validation contains 1,000 signals (stream $310000+s$). Training streams are $410000+1000s+3k+p$ for phase p. No test result selects checkpoints or seeds.

The original gain condition is retained: FIR uses four layer-specific tanh taps with the active-history startup mask; two-time correction sums all active delayed coefficients onto the previous increment. Nonlinear current correction uses an eight-hidden-unit coordinatewise tanh MLP, a zero-initialized final layer and active-tap-count output scaling. Its per-layer input scales come from 1,024 training-distribution signals (stream $730000+s$) through the initialized parent. These runs have no additional absolute-sum gain cap. All six models complete their fixed budgets on CUDA with FP32 model computation and FP64 metric accumulation.

Two fixed evaluation sizes. The original endpoint uses 10,000 signals per instance at stream $910000 + s$. The fresh evaluation uses ten blocks of 10,000 signals at $9600000 + 100s + b$, $b = 0, \ldots, 9$, shared across methods. For each instance, we sum error and target energies across all signals before converting the ratio to dB. We report mean $\pm$ sample SD across the two instances; blocks and signals are not additional training replicates. All six models and both evaluation sets are retained.

Table 13: Original test: 10,000 signals per instance. NMSE (dB), lower is better; sample SD across two instances. Best values in each column are bold.

Method	Seed 3	Seed 4	Mean $\pm$ SD
FIR	$\mathbf{-83.6044}$	$\mathbf{-88.1833}$	$\mathbf{-85.8939 \pm 3.2378}$
Two-time	-73.5332	-75.9202	-74.7267 ± 1.6879
Nonlinear current	-68.8533	-67.4008	-68.1271 ± 1.0271

Table 14: Fresh test: 100,000 signals per instance. NMSE (dB), lower is better; sample SD across two instances. Best values in each column are bold.

Method	Seed 3	Seed 4	Mean $\pm$ SD
FIR	-64.0647	$\mathbf{-70.6576}$	$\mathbf{-67.3611 \pm 4.6619}$
Two-time	$\mathbf{-64.4516}$	-67.0128	-65.7322 ± 1.8110
Nonlinear current	-63.5255	-61.2533	-62.3894 ± 1.6066

Paired architectural gains. Comparator-minus-FIR NMSE is positive when FIR is better. On the original endpoint, two-time differences are $(10.0712, 12.2631)$ dB, or 11.1672 ± 1.5499 dB; nonlinear differences are $(14.7511, 20.7825)$ dB, or 17.7668 ± 4.2649 dB. On fresh samples, the corresponding differences are $(-0.3869, 3.6448)$ dB, or 1.6290 ± 2.8509 dB, and $(0.5392, 9.4043)$ dB, or 4.9718 ± 6.2686 dB. Thus the new instances retain a mean FIR advantage at both evaluation sizes, while the larger sample evaluation reduces its magnitude and reverses the two-time comparison in seed 3. This supports recovery benefits in additional instances without asserting uniform dominance or statistical significance from two repeats. The result complements the original-instance tail analysis rather than replacing its fixed endpoint or pooling protocols.

Original comparison and timing view. The original-instance comparison and its separately measured cost panel are retained here alongside the expanded four-group main-text table.

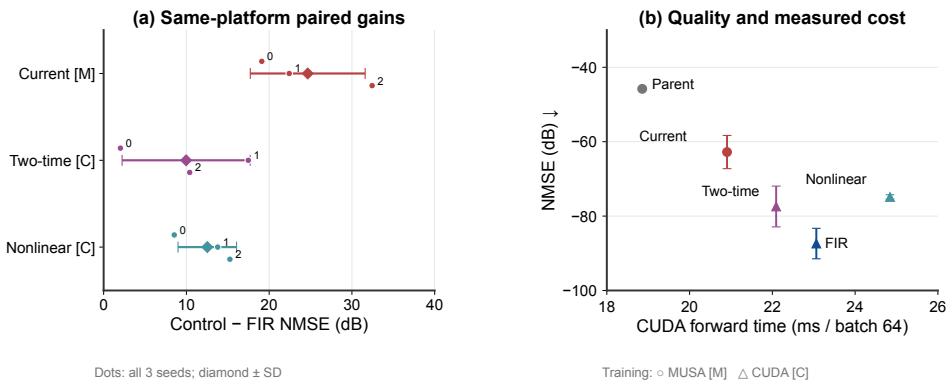


Figure 4: Targeted synthetic controls and cost. (a) All three paired seed reductions, with mean and sample SD. Each contrast uses a same-platform FIR: MUSA for current gain, CUDA for the two stronger controls. (b) Mean NMSE and seed SD against a single session of CUDA forward timing for device-resident batch 64; markers identify training platform. Neither panel implies statistical significance.

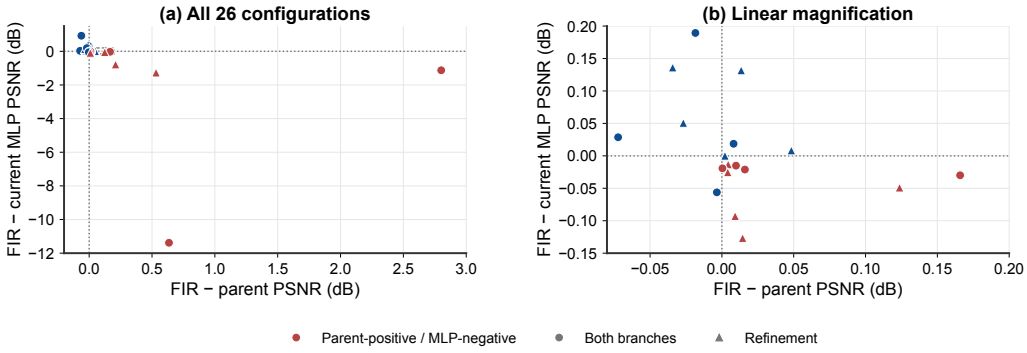


Figure 5: Changing the comparator changes the conclusion. All 26 frozen test200 configuration means are retained in (a); (b) magnifies the near-zero region on linear axes. Red marks positive FIR–parent but negative FIR–current-MLP means. Circles: both branches; triangles: refinement. The current MLP is LSTM-size-matched, not FIR-size-matched.

Historical absolute-score context. The original matched comparisons retain their platform-specific scores.

Table 15: Synthetic recovery: 24 trained models, three paired seeds per method/platform. Test NMSE (dB, lower is better), mean ± sample SD. All use 16 layers, 480,000 updates and validation selection. Bold marks the best mean within each training platform; FIR platforms are not pooled.

Method	**NMSE (dB) ↓**
MUSA training	
HLISTA-CP	-45.80 ± 0.42
Matched current-increment control	-62.78 ± 4.46
Two-tap learned correction	-59.50 ± 4.17
HLISTA-CPSS	-64.53 ± 1.66
DASH-LISTA FIR (four taps)	$\mathbf{-87.44 \pm 5.53}$
CUDA training	
Two-time control	-77.42 ± 5.48
Nonlinear current control	-74.86 ± 0.60
DASH-LISTA FIR (four taps)	$\mathbf{-87.38 \pm 4.09}$

D PAIRED SUPPORT-GROUP ERROR CONTRIBUTIONS

This post-hoc analysis reuses the five fixed fresh-test sets and their previously saved common support bins ($K \le 40$, $41 \le K \le 50$, $51 \le K \le 60$, $K \ge 61$). It uses no new inference, training or checkpoint selection. For FIR error energy $e_{F,i}$, comparator energy $e_{C,i}$ and target energy p_i, define

$$G_b = \frac{\sum_{i \in B_b} (e_{C,i} - e_{F,i})}{\sum_i p_i}, \qquad \widehat{R}_C - \widehat{R}_F = \sum_b G_b.$$

These are linear error ratios, not additive dB contributions. Every method uses the same input-defined bins and global denominator. We reconcile bin counts and energies against all ten saved blocks and verify the identity against pooled method errors. No method-specific error-ranked tails, sample removal or distribution-risk certificate is used.

Table 16: All paired bin contributions, in units of 10^{-8} linear error ratio. Positive supports FIR. Columns sum to Total before rounding; contrasts are not ranked or bolded. Seeds 0–2 and 3–4 remain separate cohorts.

Seed	**Control**	$\boldsymbol{K \le 40}$	**41–50**	**51–60**	$\boldsymbol{K \ge 61}$	**Total**
0	Two-time	0.0169	0.1889	0.3646	-0.4115	0.1589
0	Nonlinear	0.1148	1.2466	1.8679	7.2104	10.4398
1	Two-time	0.0382	0.4978	1.0724	5.4097	7.0181
1	Nonlinear	0.0777	0.9086	1.5853	-1.9030	0.6685
2	Two-time	0.0152	0.1631	0.3516	7.4825	8.0124
2	Nonlinear	0.0835	0.8584	1.1567	1.6234	3.7220
3	Two-time	0.0361	0.3456	0.6762	-4.4012	-3.3433
3	Nonlinear	0.0238	0.2514	0.3852	4.5246	5.1850
4	Two-time	0.0348	0.3448	0.7272	10.1924	11.2991
4	Nonlinear	0.1517	1.6661	2.7828	61.7365	66.3372

Every lower-support bin has positive aggregate contribution in each of the ten comparisons; this does not mean every signal improves. The highest support bin is adverse for seed-0 two-time, seed-1 nonlinear and seed-3 two-time. The last case reverses the total: 1.0579×10^{-8} from 93,908 lower-support signals is outweighed by -4.4012×10^{-8} from 6,092 higher-support signals. Its overall linear gap is -3.3433×10^{-8}, corresponding to the observed -0.3869 dB comparator-minus-FIR difference. A support group is not the same as a rare error-ranked tail; the analysis identifies contributions in the observed sample set, not the frequency or magnitude of unseen failures.

Weights and separate method contributions. The accompanying `paired_support_full_report.csv` and JSON, rebuilt by `compare_test_support.py`, report each group's count, target-energy share P_b/P, raw method error energies, normalized contributions $E_{F,b}/P$ and

$E_{C,b}/P$, and signed G_b. Percentages of the net advantage are not used: cancellation can make them arbitrarily large. Fresh-test within-bin positive/negative per-signal parts are not computed here; existing common-bin energy sums suffice for this group-level identity, not individual win rates.

Original versus fresh sets at fixed checkpoints. For seeds 3/4 we additionally read the saved original test arrays, count actual nonzeros and match indices and target energies against all three methods' original per-signal records. The same four bins cover all 10,000 signals. No checkpoint or training changes between the two test sizes.

Table 17: Fixed-checkpoint original/fresh comparison. Signed linear contributions and totals are multiplied by 10^8; the last column alone is dB. All cases are shown, without ranking contrasts.

Seed	Control / set	$K \leq 40$	41–50	51–60	$K \geq 61$	Total	dB
3	Two-time / 10k	0.038	0.327	0.633	3.000	3.997	10.071
3	Two-time / 100k	0.036	0.346	0.676	-4.401	-3.343	-0.387
3	Nonlinear / 10k	0.023	0.260	0.377	11.926	12.586	14.751
3	Nonlinear / 100k	0.024	0.251	0.385	4.525	5.185	0.539
4	Two-time / 10k	0.036	0.341	0.672	1.357	2.407	12.263
4	Two-time / 100k	0.035	0.345	0.727	10.192	11.299	3.645
4	Nonlinear / 10k	0.159	1.679	2.900	13.304	18.042	20.783
4	Nonlinear / 100k	0.152	1.666	2.783	61.737	66.337	9.404

Seed 3's high-support contribution against two-time changes from $+2.9996 \times 10^{-8}$ to -4.4012×10^{-8}, while the three lower-support contributions remain positive in both sets. Seed 4 presents a different accounting pattern: its total linear advantages increase, yet its dB advantages decrease. The conversion explicitly depends on FIR's own error level:

$$\Delta_{\mathrm{dB}} = 10 \log_{10} \frac{\widehat{R}_C}{\widehat{R}_F} = 10 \log_{10} \left(1 + \frac{\sum_b G_b}{\widehat{R}_F} \right).$$

E HISTORY UNDER A SHARED FEEDBACK-GAIN BUDGET

Question and matched training. This experiment tests whether FIR retains its advantage when feedback gain is bounded consistently across FIR, two-time and nonlinear current controls. We train all nine combinations of the three methods and original problem seeds 0, 1 and 2 with $G = 0.5$. Each uses 16 layers, tied backbone weights, 48 training phases and 10,000 updates per phase (480,000 per model; 4,320,000 additional updates in total), with validation every 200 updates. Within seed, operators, initialization, calibration, training streams, validation data and 10,000 test signals match the original CUDA comparison. We reuse all nine original-gain checkpoints. These retrainings are distinct from new-instance and fresh-sample tests.

Exact constraint. For the J active raw taps $c_j = \tanh(\omega_j)$, FIR and two-time apply

$$c'_j = c_j \min\left\{1, \frac{G}{\sum_{\ell=0}^{J-1} |c_\ell|}\right\}, \qquad G = 0.5,$$

using scale one when the denominator is zero. The two-time control then collapses delayed coefficients onto the preceding increment, so its effective absolute coefficient sum is also at most G; cancellation may make it smaller. Startup includes active taps only. For the nonlinear current control, $J\tanh(\mathrm{MLP}(r_k/s_k))\,r_k$ is replaced by $G\tanh(\mathrm{MLP}(r_k/s_k))\,r_k$, coordinatewise. The original condition uses its original per-tap parameterization without this extra cap. The constraint matches allowed feedback upper bounds, not realized vector norms, model capacity or training convergence. Its numerical value matches the image cap, while depth, sharing and optimization protocols remain distinct.

Table 18: NMSE on the original 10,000-signal tests under original and capped gain. Mean ± sample SD across three instances; lower is better. Bold denotes the best method within each gain condition.

Method	Original gain	$G = 0.5$
FIR (four taps)	$\mathbf{-87.3838 \pm 4.0939}$	$\mathbf{-62.3382 \pm 1.4472}$
Two-time	-77.4164 ± 5.4759	-57.9861 ± 1.3298
Nonlinear current	-74.8593 ± 0.6007	-51.8433 ± 1.6946

Paired effects and interpretation. For comparator C, define $\Delta_C(G) = \mathrm{NMSE}_C(G) - \mathrm{NMSE}_{\mathrm{FIR}}(G)$, so positive values favor FIR. Define the interaction $I_C = \Delta_C(\mathrm{original}) - \Delta_C(0.5)$; positive values indicate a smaller FIR advantage under the cap. Statistics are computed from the three within-seed differences, with sample SD ($\mathrm{ddof} = 1$), rather than from independent method SDs. NMSE pools signal errors and target energies before taking the logarithm.

Table 19: Seed-paired FIR advantages and gain interactions (dB). All capped advantages are positive; interactions have mixed signs. Contrasts against different comparators are not ranked.

Contrast	Versus two-time	Versus nonlinear current
Original advantage	9.9674 ± 7.7354	12.5245 ± 3.5349
Capped advantage	4.3521 ± 0.2870	10.4949 ± 0.4028
Interaction	5.6153 ± 7.5724	2.0296 ± 3.2683
Capped advantage, seed 0	4.0576	10.1093
Capped advantage, seed 1	4.3679	10.9129
Capped advantage, seed 2	4.6309	10.4625
Interaction, seed 0	−2.0319	−1.5765
Interaction, seed 1	13.1107	2.8687
Interaction, seed 2	5.7670	4.7965

FIR wins every capped comparison; its average advantage contracts, with negative interactions for seed 0. This does not explain the image-domain contrast. Records: `research/gain_budget_20260911/PLAN.md` and `research/gain_budget_20260911/summary.json`.

F SHARED-CAP CONFIRMATION ON THE COMMON 100,000 SIGNALS

All nine completed $G = 0.5$ checkpoints are frozen before this extension. We reuse the ten saved 10,000-signal blocks per original instance, without changing checkpoint selection or generating a favorable replacement test. These arrays were previously evaluated by the original-gain models; this is a paired extension on an opened set, not a new blind confirmation. FP32 inference uses the same CUDA implementation with TF32 disabled; error and target energies are accumulated in FP64 before taking logarithms. Per-block target energies and support counts match the existing records.

Table 20: Absolute NMSE on the common 100,000 signals. Mean ± sample SD across three instances; lower is better, best mean bold within each condition.

Method	Original gain	Cap 0.5
fir	$\mathbf{-79.1751 \pm 7.4034}$	$\mathbf{-60.6462 \pm 0.4898}$
two-time	-72.0166 ± 3.1545	-56.8417 ± 0.2096
nonlinear	-72.0965 ± 2.8403	-51.5713 ± 0.4646

Table 21: Capped 100,000-signal NMSE and comparator-minus-FIR gaps (dB). All original instances are retained. Bold marks the lowest absolute NMSE within each row; diagnostic gaps are not ranked.

Instance	FIR	Two-time	Nonlinear	Δ_{two}	$\Delta_{\text{nonlinear}}$
0	$\mathbf{-60.7339}$	-56.9043	-51.9479	3.8297	8.7860
1	$\mathbf{-60.1185}$	-56.6080	-51.0521	3.5105	9.0663
2	$\mathbf{-61.0862}$	-57.0130	-51.7137	4.0733	9.3725

For the fixed support groups $\leq$ 40, 41–50, 51–60 and $\geq$ 61, the signed contribution is $G_b = \sum_{i \in b}(e_{C,i} - e_{F,i}) / \sum_i p_i$. All groups use common inputs and denominators; their contributions sum back to the pooled linear gap. These are descriptive sample contributions, not bounds on unobserved tails. The cap retains positive paired gains in all six comparisons; it does not improve absolute quality relative to the original-gain models or equate realized correction norms.

Evidence: `research/reviewer_followup_20260913/evaluate_gain.py`, `gain_summary.json`, and per-block `gain_results/` arrays. The nine checkpoint identities and training protocol remain those in `research/gain_budget_20260911/`. This adds evaluation, not training repeats.

F.1 Independent confirmation under the shared gain cap

Fixed protocol and independent instances. We fix consecutive instance seeds 5–7 before training and train FIR, two-time and nonlinear-current models for each instance. The operator, initialization and streams change together across instances; within an instance all three arms share them. The architecture, $G = 0.5$ constraint, 16-layer schedule, 480,000-update budget and validation selection follow Section 5.2. Each operator is a 250×500 standard Gaussian draw from NumPy RandomState(s), column-normalized in FP64 and stored in FP32. Signals follow the same Bernoulli–Gaussian law. Validation uses 1,000 signals from stream $310000 + s$; training streams use $410000 + 1000s + 3(k-1) + p$ for depth $k = 1, \ldots, 16$ and phase $p = 0, 1, 2$. The new 100,000-signal test stream is $2026091400 + s$. Test arrays are prepared before training; evaluation starts only after all nine models and the predictions below are fixed. The original and confirmation cohorts remain separately reported in Table 3.

Table 22: Independent capped cohort on 100,000 previously unused signals per instance. NMSE (dB), lower is better; mean ± sample SD across three trained instances.

Method	Seed 5	Seed 6	Seed 7	Mean ± SD
FIR	**−59.9361**	**−61.7515**	**−60.1577**	**−60.6151 ± 0.9904**
Two-time	−55.3869	−57.0465	−56.6633	−56.3656 ± 0.8690
Nonlinear current	−50.9991	−51.4914	−51.6072	−51.3659 ± 0.3229

Paired gains and complete trajectories. FIR improves on two-time by 4.5493, 4.7050 and 3.4945 dB in instances 5, 6 and 7, respectively (4.2496 ± 0.6586 dB). Its gains over nonlinear current are 8.9370, 10.2601 and 8.5505 dB (9.2492 ± 0.8966 dB). Both contrasts are positive in every instance. Figure 6 retains all prefixes: the final FIR advantage is not a uniform ordering at intermediate layers. Each curve evaluates prefixes of the selected 16-layer checkpoint, with no shallow-model retraining or prefix selection.

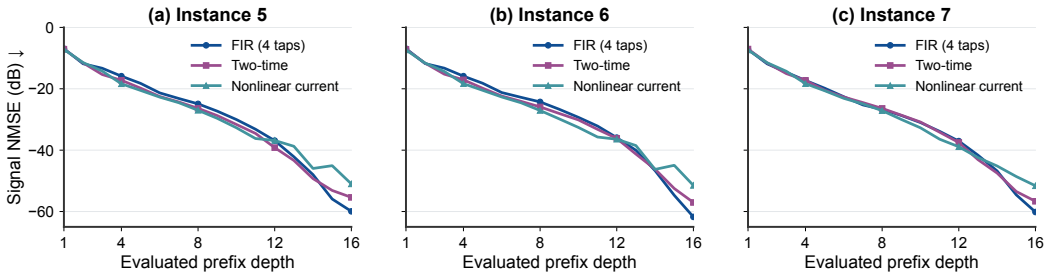


Figure 6: Independent capped cohort: all 16 prefixes on 100,000 unused signals per instance. Lower pooled NMSE is better. Each panel has its own method labels and represents one trained problem instance. These are prefixes of depth-16-trained networks, not separately optimized depths.

Common support groups. Table 25 uses the four groups fixed before test and the shared denominator in Eq. (14). All 24 signed contributions are positive: three instances, four groups and two controls. High-support signals contribute the largest share of FIR's error in each instance while also contributing positively to its advantage over either control. Thus error concentration and the sign of method differences are separate properties even in this positive cohort. The result replicates recovery benefit within the fixed operator family, signal law, depth and optimization protocol; it measures observed sample contributions rather than a bound on unobserved tails.

Instance-level endpoint comparison. Figure 7 places the original capped cohort and independent confirmation side by side on a common gain scale. All twelve endpoint contrasts favor FIR under the shared cap, while the two cohorts retain their distinct test histories. Absolute NMSE values remain in Table 3 and the per-instance tables above.

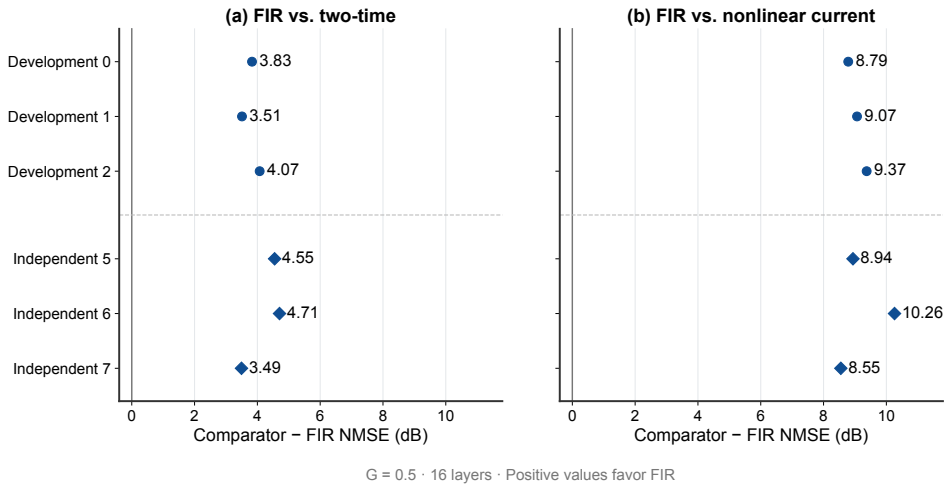


Figure 7: Paired endpoint gains under the common $G = 0.5$ constraint, with 16 layers and 480,000 training updates. Each point is comparator NMSE minus FIR NMSE for one problem/training instance on the same 100,000 signals; positive values favor FIR. NMSE is computed from pooled error and target energies before conversion to dB. Circles identify development instances 0–2 evaluated on the previously opened common test arrays; diamonds identify independent instances 5–7 evaluated on unused test arrays. Values are calculated from full-precision records and displayed to two decimals. Each point is one paired instance, so no within-instance training uncertainty bar is assigned.

F.2 Prospective prediction of frozen replacement loss

Fitting and selection without test access. For each new FIR checkpoint, only its 64 layerwise tap parameters are adjustable; every backbone tensor remains fixed. We fit current replacement and FIR adaptation under the same gain cap, original/zero coefficient starts, and 400 Adam steps per start (batch 64; 10^{-3} for 300 steps, 2×10^{-4} for 100). Each minibatch minimizes squared reconstruction error divided by target energy. The fitting set contains 4,096 signals from stream $2026091500 + s$; the selection set contains 1,024 signals from $2026091600 + s$. Pooled selection NMSE at steps 0, 100, 200, 300 and 400 selects one checkpoint across both starts. All current replacements select the zero-start step 400. FIR selects original-start step 0 for seeds 5/6 and original-start step 400 for seed 7. Selection error is still decreasing at the chosen current endpoints: these are finite-budget replacement costs, not optimum errors over all current replacements.

Autodiff sensitivity versus correction magnitude. Let d_k^F be the original FIR correction and d_k^C the fitted current correction. For a scalar λ, execute the entire solver with

$$d_k^\lambda = (1-\lambda) d_k^F(\mathcal{H}_k^\lambda) + \lambda d_k^C(r_k^\lambda), \qquad P = \mathrm{AD}_{\lambda=0}\, R_{\mathrm{cal}}(\lambda),$$
$$H = \frac{\sum_{i,k} \| d_k^C(r_{k,i}^0) - d_k^F(\mathcal{H}_{k,i}^0) \|_2^2}{\sum_i \|x_i^\star\|_2^2}. \tag{24}$$

Here $\mathcal{H}_k^\lambda$ and r_k^λ are regenerated along the interpolated trajectory, and $R_{\mathrm{cal}}(\lambda) = \sum_i \|\widehat{x}_i(\lambda) - x_i^\star\|_2^2 / \sum_i \|x_i^\star\|_2^2$ is pooled calibration risk. The operator AD denotes automatic differentiation of the executed graph, including its conventions at ReLU and threshold knots. It equals the classical derivative when the ordinary chain rule applies along that graph; at nonsmooth points it specifies an algorithmic score, without asserting that a classical derivative exists. A third calibration set of 1,024 signals from $2026091700 + s$ determines P by automatic differentiation and H on the original trajectory. The two predictions and a zero-change baseline are locked before test. The target is the full test-set change $\Delta R = R_{\mathrm{current}} - R_{\mathrm{original\ FIR}}$ in linear NMSE, not dB. P is used as a unit-displacement first-order prediction; H is an unscaled, nonnegative magnitude baseline. Neither is a finite-loss bound. Implementation checks verify that $\lambda = 0$ and $\lambda = 1$ exactly reproduce the original FIR and fitted-current endpoints.

Table 23: Frozen-backbone fitting, NMSE (dB). Bold compares equally budgeted adapted FIR with current replacement; original FIR is the unadapted reference. All selection uses calibration signals.

Condition	Seed 5	Seed 6	Seed 7	Mean ± SD
Original FIR	−59.9361	−61.7515	−60.1577	-60.6151 ± 0.9904
FIR (adapted)	**−59.9361**	**−61.7515**	**−60.1460**	$\mathbf{-60.6112 \pm 0.9931}$
Current replacement	−42.7920	−42.9338	−43.6911	-43.1390 ± 0.4834

Table 24: Prospective predictions versus observed replacement loss, in pooled linear NMSE units. Predictions were locked before test. Positive observed change means current replacement raises error.

Seed	Observed loss	AD score P	Magnitude H	Zero
5	5.1562×10^{-5}	1.4332×10^{-7}	1.1131×10^{-1}	0
6	5.0221×10^{-5}	3.0353×10^{-6}	1.2565×10^{-1}	0
7	4.1781×10^{-5}	3.4052×10^{-6}	8.6521×10^{-2}	0

Observed loss and prediction accuracy. Current replacement raises NMSE by 17.1441, 18.8178 and 16.4666 dB relative to the original FIR, whereas matched FIR adaptation retains nearly the same quality. The autodiff score has the correct sign in all three tests but substantially underestimates each full replacement loss. Its mean absolute error is 4.5660×10^{-5}, compared with 4.7855×10^{-5} for zero change (a 4.586% reduction) and 0.10778 for H. Magnitude H also has three correct signs because all observed changes are positive; sign agreement alone does not separate the predictors. The prospective test establishes a measurable local sensitivity, while the finite replacement effect is much larger than the first-order prediction. It does not establish a transferable predictor of replacement magnitude or retrained architecture benefit.

F.3 PER-GROUP EVIDENCE AND RECORD VERIFICATION

Table 25: Common support groups for the new capped cohort. Energy is target-energy percentage; the three error contributions and two signed gains are in 10^{-6} linear NMSE units, using the same total energy denominator per instance. Gains sum exactly to the complete linear gap before display rounding.

Seed	K	n	Energy (%)	FIR	Two-time	Nonlinear	G_{two}	G_{nonlin}
5	≤ 40	7410	5.59	**0.0051**	0.0113	0.0860	0.0062	0.0809
5	41–50	46531	43.08	**0.0529**	0.1527	1.0054	0.0998	0.9525
5	51–60	39838	43.39	**0.1329**	0.4777	2.2760	0.3448	2.1431
5	≥ 61	6221	7.94	**0.8240**	2.2511	4.5775	1.4271	3.7536
6	≤ 40	7447	5.64	**0.0064**	0.0152	0.0960	0.0088	0.0896
6	41–50	46248	42.68	**0.0664**	0.2325	1.1625	0.1662	1.0961
6	51–60	39965	43.60	**0.1700**	0.6853	2.6338	0.5154	2.4638
6	≥ 61	6340	8.08	**0.4253**	1.0409	3.2012	0.6156	2.7759
7	≤ 40	7551	5.68	**0.0059**	0.0138	0.0830	0.0079	0.0771
7	41–50	46383	42.87	**0.0740**	0.1890	0.9347	0.1151	0.8607
7	51–60	39870	43.53	**0.2328**	0.6365	2.1947	0.4037	1.9619
7	≥ 61	6196	7.91	**0.6517**	1.3168	3.6944	0.6651	3.0427

Record-level checks. The completed record set retains all nine trained models and all six selected fitting endpoints. Recomputing the 1,500,000 model–signal rows checks finite errors, common target energies and supports, pooled scores, paired means/sample standard deviations, test-block sums, and support-bin additivity. The three trained instances, rather than signals, blocks or layers, are the uncertainty units. Training histories retain all 48 phases per model; selection records and queue timestamps preserve the order of training, locked prediction and test. A 64-signal replay of all 15 saved endpoints reproduces the recorded errors within FP32 tolerance and checks final/prefix agreement and frozen backbone tensors. This is a checkpoint replay check, not a fresh training replication.

Evidence locations. `research/reviewer_confirm_20260914/frozen_plan.json` identifies the fixed protocol; `train.py`, `predict.py` and `evaluate.py` implement its stages. The local `completed_evidence_20260915.zip` contains model/selection records, operators, initializations, per-signal results and source files. Full test arrays remain with the experiment; the package includes their stream definitions and generation code. `verify_completed.py`, `verified_results.json` and `checkpoint_replay_check.json` record the separate recomputation and replay checks. The nine new models are additional to the original capped cohort and are not implicitly contained in earlier evidence bundles.

G FITTING-OBJECTIVE CONTROL FOR FROZEN IMAGE REPLACEMENT

Earlier comparison summary. Table 26 retains the original calibration, objective-fitting and retraining contrasts. The budget and function-class extension appears in Appendix H.

Table 26: Common-CUDA validation50 contrasts, paired mean ± sample SD (PSNR dB). Only independent retraining updates the backbone; all fits freeze it. Negative favors the stated FIR reference.

Current-input alternative minus reference	Refinement	Both branches
Vector fit − original FIR	-0.00017 ± 0.00045	-0.02783 ± 0.00310
Image-PSNR fit − original FIR	$+0.00094 \pm 0.00045$	-0.02696 ± 0.00300
Independent retraining − original FIR	$+0.01376 \pm 0.01059$	$+0.00595 \pm 0.03243$
Patch-MSE fit − adapted FIR	$+0.00081 \pm 0.00008$	-0.03019 ± 0.00377
Image-PSNR fit − adapted FIR	$+0.00012 \pm 0.00039$	-0.02995 ± 0.00360

Matched design. Six existing FIR checkpoints cross both/refinement injection with seeds 0–2. For each we adapt four effective shared coefficients either as FIR or as current-input correction. Both alternatives start from the original FIR coefficients and from zero, and share the same 400 full-image draws, Adam schedule (0.001 for 300 updates, 0.0002 for 100), and projection onto $\|c\|_1 \leq 0.5$. Backbone weights, dictionary, thresholds and mixing parameters are frozen and checked for exact equality after each trajectory. All six stages are recomputed on the adapted trajectory at every update.

The only changed treatment is the fitting objective: normalized demeaned patch MSE, including padding, versus negative whole-image PSNR after restoring patch means, assembling the image, removing padding and clipping to $[0, 255]$. Both consume all patches of the same full image. This comparison includes objective-specific weighting and boundary treatment, not just the logarithm. Identical Adam settings control the optimizer, without asserting that both objectives have identical optimization difficulty.

The original hash-ordered 400-image training set is partitioned into 320 fitting and 80 selection images. Both objectives select by mean full-image PSNR on those 80 images, checking steps 0, 100, 200, 300 and 400 and retaining the best candidate across both starts. Validation50 never selects coefficients. This gives 48 trajectories and 24 selected fits, all evaluated on the same previously opened validation50. This is not a new independent dataset. The full-image objective agrees with the existing evaluator within 1.8×10^{-6} dB in the startup check; both coefficient gradients are finite and nonzero. Total concurrent runtime on RTX 5060 Ti was 89.6 minutes, which is experiment runtime rather than an inference-cost comparison.

Table 27: Current-minus-FIR PSNR after equal-budget adaptation (dB). All three dual-branch differences favor FIR under both fitting objectives.

Location	Objective	Seed 0	Seed 1	Seed 2
both	MSE	-0.02998	-0.03407	-0.02653
both	PSNR	-0.02998	-0.03353	-0.02633
refinement	MSE	+0.00082	+0.00072	+0.00088
refinement	PSNR	-0.00003	-0.00018	+0.00056

Objective effect versus replacement effect. For both-branch models, changing MSE to image PSNR improves current replacement by 0.00281 ± 0.00049 dB and FIR adaptation by 0.00256 ± 0.00030 dB. Their gap therefore remains near 0.030 dB. This does not support the fitting objective as the principal explanation for this observed gap under the fixed optimization protocol. Refinement contrasts remain small and do not establish equivalence. The result concerns these frozen backbones and four-coefficient replacements, not the best performance achievable by retraining a current-input architecture.

Evidence: `research/objective_control_20260913/PLAN.md`, `run.py`, the `mse/` and `psnr/` histories and per-image `results.json`, and `verified_summary.json`. All

48 frozen-tensor checks, 24 train-only selections and 1,200 image records were checked. The intervening exploratory adaptation protocols are retained separately; this comparison uses the final matched objective design. No new COCO replacement evaluation or independent human author audit is asserted.

H REPLACEMENT BUDGET, STAGE FREEDOM AND PATH INTERVENTIONS

H.1 FROZEN-BACKBONE DESIGN AND NESTED CURRENT CLASSES

This study reuses nine six-stage MR25 FIR checkpoints: both/refinement seeds 0–2 and both seeds 3–5. It fits four feedback classes while freezing the dictionary, thresholds, mixtures and all remaining backbone tensors. The original cohorts remain separate. For correction class a with parameters η, the fitting objective is

$$\min_{\eta \in \mathcal{C}_a} \frac{1}{320} \sum_{I \in \mathcal{D}_{\mathrm{fit}}} -\mathrm{PSNR}\big(f^a_{\theta,\eta}(y_I), I\big)\,. \tag{25}$$

The full six-stage trajectory is recomputed at every update; each image uses mean restoration, patch assembly, unpadding and clipping to $[0, 255]$. The existing hash-ordered train400 list provides the first 320 fitting images and the remaining 80 for candidate selection. These images were available during backbone training. Test50 is the primary evaluation and validation50 the earlier-protocol counterpart; both had been evaluated in preceding project experiments. This study does not introduce a new blind dataset or independent backbone training.

FIR4 uses four globally shared effective taps with $\|c\|_1 \le 0.5$. Current4 collapses active taps onto r_k, giving $g = (c_0, c_0 + c_1, c_0 + c_1 + c_2, C, C, C)$, $C = \sum_{j=0}^{3} c_j$. TV6 optimizes $a = (g_0, g_1 - g_0, \ldots, g_5 - g_4)$ with $\|a\|_1 \le 0.5$ and uses $g = \mathrm{cumsum}(a)$. Box6 directly optimizes six gains in $[-0.5, 0.5]$. The Current4 total variation equals $\|c\|_1$, while $\max_k |g_k| \le \|a\|_1$, proving the nested classes in Eq. (15). TV6 projects in a coordinates, rather than claiming Euclidean projection in g coordinates. These classes deliberately expand replacement freedom and are not an equal-capacity architecture comparison.

Budgets, starts and selection. Each of the four classes starts from the original effective FIR coefficients (exactly mapped for TV6/Box6) and from zero. Each trajectory runs 6,400 Adam updates, drawing one complete fitting image at a time. Learning rates are 10^{-3} through update 300, 2×10^{-4} through 1,600, then 5×10^{-5}. FIR4/Current4 use an ℓ_1 projection, TV6 an a-coordinate ℓ_1 projection and Box6 coordinate clipping. Selection PSNR is evaluated on all 80 images at steps 0, 100, 200, 300, 400 and every 400 steps thereafter. Budgets 400, 1,600 and 6,400 select the best score across both starts within that prefix. At each budget, TV6 also evaluates the selected Current4 candidate and Box6 the selected TV6 candidate. These inherited candidates cannot cross budgets. Their additional fitting and selection cost means that nested pools do not represent equal total search cost. Pool nesting guarantees a nondecreasing best selection score, not a test-set ordering.

There are 72 image adaptation trajectories, each with 6,400 updates. Backbone tensors are checked for exact equality to their initial copies. Complete histories record selection80, a fixed fitting subset of 32 images, and fitting320 at step zero and budget endpoints. Every reported mean first averages images within checkpoint, and every paired difference is computed within checkpoint before the three-seed mean and sample SD.

H.2 COMPLETE IMAGE SUMMARIES

Tables 28–31 give both metrics at all budgets; Tables 32 and 33 distinguish the original FIR from the adapted reference. The six both checkpoints retain positive test PSNR costs in every class and budget. The refinement-only TV6 and Box6 comparisons are negative at the final budget in all three seeds. The seed-level costs appear in Table 34.

Table 28: Lower replacement cost is better. Test50: adapted FIR minus current PSNR (dB). Each budget selects within its trajectory prefix; mean ± sample SD across three paired seeds.

Updates	Current4	TV6	Box6
Both, seeds 0–2			
400	0.03024 ± 0.00520	0.02942 ± 0.00501	$\mathbf{0.02813 \pm 0.00506}$
1600	0.03019 ± 0.00514	0.02907 ± 0.00500	$\mathbf{0.02669 \pm 0.00485}$
6400	0.03002 ± 0.00511	0.02851 ± 0.00450	$\mathbf{0.02623 \pm 0.00463}$
Both, seeds 3–5			
400	0.03049 ± 0.00191	0.03033 ± 0.00109	$\mathbf{0.02800 \pm 0.00257}$
1600	0.03049 ± 0.00192	0.02947 ± 0.00172	$\mathbf{0.02661 \pm 0.00336}$
6400	0.03039 ± 0.00201	0.02921 ± 0.00186	$\mathbf{0.02622 \pm 0.00322}$
Refinement, seeds 0–2			
400	0.00020 ± 0.00054	-0.00124 ± 0.00105	$\mathbf{-0.00215 \pm 0.00161}$
1600	0.00018 ± 0.00052	-0.00148 ± 0.00111	$\mathbf{-0.00312 \pm 0.00213}$
6400	0.00019 ± 0.00062	-0.00167 ± 0.00092	$\mathbf{-0.00345 \pm 0.00206}$

Table 29: Lower replacement cost is better. Test50: adapted FIR minus current SSIM. Each budget selects within its trajectory prefix; mean ± sample SD across three paired seeds.

Updates	Current4	TV6	Box6
Both, seeds 0–2			
400	0.001297 ± 0.000274	0.001207 ± 0.000189	$\mathbf{0.000965 \pm 0.000266}$
1600	0.001279 ± 0.000289	0.001183 ± 0.000190	$\mathbf{0.000871 \pm 0.000337}$
6400	0.001308 ± 0.000281	0.001180 ± 0.000183	$\mathbf{0.000899 \pm 0.000370}$
Both, seeds 3–5			
400	0.000537 ± 0.000211	0.000570 ± 0.000206	$\mathbf{0.000280 \pm 0.000310}$
1600	0.000534 ± 0.000185	0.000611 ± 0.000163	$\mathbf{0.000198 \pm 0.000372}$
6400	0.000559 ± 0.000165	0.000527 ± 0.000245	$\mathbf{0.000193 \pm 0.000415}$
Refinement, seeds 0–2			
400	0.000009 ± 0.000016	-0.000086 ± 0.000037	$\mathbf{-0.000131 \pm 0.000052}$
1600	0.000009 ± 0.000013	-0.000104 ± 0.000033	$\mathbf{-0.000185 \pm 0.000072}$
6400	0.000013 ± 0.000013	-0.000103 ± 0.000024	$\mathbf{-0.000205 \pm 0.000063}$

Table 30: Lower replacement cost is better. Validation50: adapted FIR minus current PSNR (dB). Each budget selects within its trajectory prefix; mean ± sample SD across three paired seeds.

Updates	Current4	TV6	Box6
Both, seeds 0–2			
400	0.02995 ± 0.00360	0.02944 ± 0.00370	$\mathbf{0.02788 \pm 0.00356}$
1600	0.02987 ± 0.00357	0.02928 ± 0.00370	$\mathbf{0.02708 \pm 0.00307}$
6400	0.02966 ± 0.00358	0.02869 ± 0.00313	$\mathbf{0.02671 \pm 0.00282}$
Both, seeds 3–5			
400	0.03204 ± 0.00317	0.03088 ± 0.00312	$\mathbf{0.02845 \pm 0.00430}$
1600	0.03199 ± 0.00318	0.03060 ± 0.00328	$\mathbf{0.02719 \pm 0.00478}$
6400	0.03179 ± 0.00332	0.03028 ± 0.00337	$\mathbf{0.02661 \pm 0.00440}$
Refinement, seeds 0–2			
400	-0.00012 ± 0.00039	-0.00160 ± 0.00095	$\mathbf{-0.00229 \pm 0.00165}$
1600	-0.00011 ± 0.00050	-0.00191 ± 0.00106	$\mathbf{-0.00352 \pm 0.00249}$
6400	-0.00027 ± 0.00064	-0.00195 ± 0.00134	$\mathbf{-0.00392 \pm 0.00257}$

Table 31: Lower replacement cost is better. Validation50: adapted FIR minus current SSIM. Each budget selects within its trajectory prefix; mean ± sample SD across three paired seeds.

Updates	Current4	TV6	Box6
Both, seeds 0–2			
400	0.001187 ± 0.000179	0.001129 ± 0.000138	$\mathbf{0.000964 \pm 0.000189}$
1600	0.001173 ± 0.000189	0.001118 ± 0.000141	$\mathbf{0.000907 \pm 0.000218}$
6400	0.001191 ± 0.000187	0.001113 ± 0.000134	$\mathbf{0.000921 \pm 0.000234}$
Both, seeds 3–5			
400	0.000658 ± 0.000170	0.000687 ± 0.000180	$\mathbf{0.000468 \pm 0.000261}$
1600	0.000655 ± 0.000152	0.000727 ± 0.000143	$\mathbf{0.000408 \pm 0.000299}$
6400	0.000672 ± 0.000136	0.000662 ± 0.000203	$\mathbf{0.000394 \pm 0.000320}$
Refinement, seeds 0–2			
400	-0.000005 ± 0.000024	-0.000091 ± 0.000047	$\mathbf{-0.000129 \pm 0.000071}$
1600	-0.000004 ± 0.000025	-0.000108 ± 0.000040	$\mathbf{-0.000191 \pm 0.000101}$
6400	-0.000007 ± 0.000029	-0.000106 ± 0.000046	$\mathbf{-0.000213 \pm 0.000094}$

Table 32: Test50 absolute quality after 6,400-step adaptation. Original FIR is the unadapted reference. Bold compares adapted alternatives within each cohort; mean ± sample SD across three seeds.

Method	PSNR (dB) ↑	SSIM ↑
Both, seeds 0–2		
Original FIR	27.98672 ± 0.02208	0.819615 ± 0.000708
Adapted FIR	$\mathbf{27.98951 \pm 0.02156}$	$\mathbf{0.819902 \pm 0.000680}$
Current4	27.95949 ± 0.02578	0.818594 ± 0.000876
TV6	27.96100 ± 0.02590	0.818722 ± 0.000863
Box6	27.96329 ± 0.02598	0.819003 ± 0.000839
Both, seeds 3–5		
Original FIR	27.98707 ± 0.02266	0.818962 ± 0.000794
Adapted FIR	$\mathbf{27.99286 \pm 0.02404}$	$\mathbf{0.819546 \pm 0.001061}$
Current4	27.96246 ± 0.02540	0.818987 ± 0.001126
TV6	27.96364 ± 0.02476	0.819019 ± 0.000948
Box6	27.96664 ± 0.02554	0.819352 ± 0.000788
Refinement, seeds 0–2		
Original FIR	27.98185 ± 0.01435	0.818807 ± 0.000776
Adapted FIR	27.98218 ± 0.01444	0.818795 ± 0.000778
Current4	27.98200 ± 0.01436	0.818782 ± 0.000768
TV6	27.98385 ± 0.01428	0.818897 ± 0.000757
Box6	$\mathbf{27.98564 \pm 0.01400}$	$\mathbf{0.819000 \pm 0.000730}$

Table 33: Validation50 absolute quality after 6,400-step adaptation. Original FIR is the unadapted reference. Bold compares adapted alternatives within each cohort; mean ± sample SD across three seeds.

Method	PSNR (dB) ↑	SSIM ↑
Both, seeds 0–2		
Original FIR	29.07516 ± 0.02112	0.845284 ± 0.000689
Adapted FIR	$\mathbf{29.07815 \pm 0.02093}$	$\mathbf{0.845530 \pm 0.000675}$
Current4	29.04849 ± 0.02325	0.844339 ± 0.000761
TV6	29.04946 ± 0.02345	0.844417 ± 0.000797
Box6	29.05143 ± 0.02369	0.844609 ± 0.000832
Both, seeds 3–5		
Original FIR	29.07336 ± 0.01641	0.844843 ± 0.000724
Adapted FIR	$\mathbf{29.07993 \pm 0.01790}$	$\mathbf{0.845333 \pm 0.000928}$
Current4	29.04813 ± 0.01935	0.844662 ± 0.000937
TV6	29.04965 ± 0.01886	0.844672 ± 0.000824
Box6	29.05331 ± 0.01921	0.844939 ± 0.000720
Refinement, seeds 0–2		
Original FIR	29.06477 ± 0.01851	0.844493 ± 0.000731
Adapted FIR	29.06580 ± 0.01835	0.844514 ± 0.000728
Current4	29.06606 ± 0.01789	0.844521 ± 0.000700
TV6	29.06775 ± 0.01825	0.844620 ± 0.000700
Box6	$\mathbf{29.06972 \pm 0.01800}$	$\mathbf{0.844727 \pm 0.000672}$

Table 34: Lower replacement cost is better. Test50 paired PSNR differences at 6,400 updates. Seed columns follow the ascending IDs of each cohort; positive favors adapted FIR.

Current class	First seed	Second seed	Third seed
Both, seeds 0–2			
Current4	$+0.031553$	$+0.034190$	$\mathbf{+0.024321}$
TV6	$+0.027947$	$+0.033268$	$\mathbf{+0.024321}$
Box6	$\mathbf{+0.022656}$	$\mathbf{+0.031452}$	$+0.024567$
Both, seeds 3–5			
Current4	$+0.032402$	$+0.028391$	$+0.030390$
TV6	$+0.031335$	$+0.027837$	$+0.028464$
Box6	$\mathbf{+0.029837}$	$\mathbf{+0.023666}$	$\mathbf{+0.025152}$
Refinement, seeds 0–2			
Current4	$+0.000874$	$+0.000028$	-0.000346
TV6	-0.000674	-0.001847	-0.002489
Box6	$\mathbf{-0.001299}$	$\mathbf{-0.003647}$	$\mathbf{-0.005413}$

H.3 Optimization stability and a second numerical route

The prespecified engineering stability criterion requires the last three increments of the best full-selection PSNR to be below 0.001 dB. At 6,400 steps, FIR4, Current4 and TV6 each pass in 18/18 trajectories; Box6 passes in 17/18. The exception is the zero start of both seed 5: its last increments are 0.001233, 0.000847 and 0.000803 dB. This is a criterion for observed late improvement, not global optimality.

For the two prespecified both checkpoints, seeds 0 and 3, Current4 additionally starts from its selected Adam candidate and takes at most 50 deterministic projected-gradient steps on fitting320. A maximum of 12 backtracking trials starts at step size 0.1, halves it each time, and tests an Armijo coefficient of 10^{-4}. Selection80 chooses among the initial point and the trajectory. Each run records 36,080 image forward evaluations and 16,000 image backward evaluations. These costs are additional to Adam, not an equal-compute optimizer comparison.

Table 35: Current4 projected-gradient check. Gains are relative to the selected Adam result; test cost is adapted FIR minus the selected projected-gradient candidate, in PSNR dB.

Seed	Selected step	Selection gain	Test gain	Test cost
0	50	0.0000328	0.0003445	0.0312088
3	0	0	0	0.0324022

The second route yields a small additional improvement for seed 0 and retains the Adam solution for seed 3. The dual-branch replacement gap remains at these two endpoints. Figure 8 shows both starts for these prespecified checkpoints; all nine sets of histories are retained in the study records.

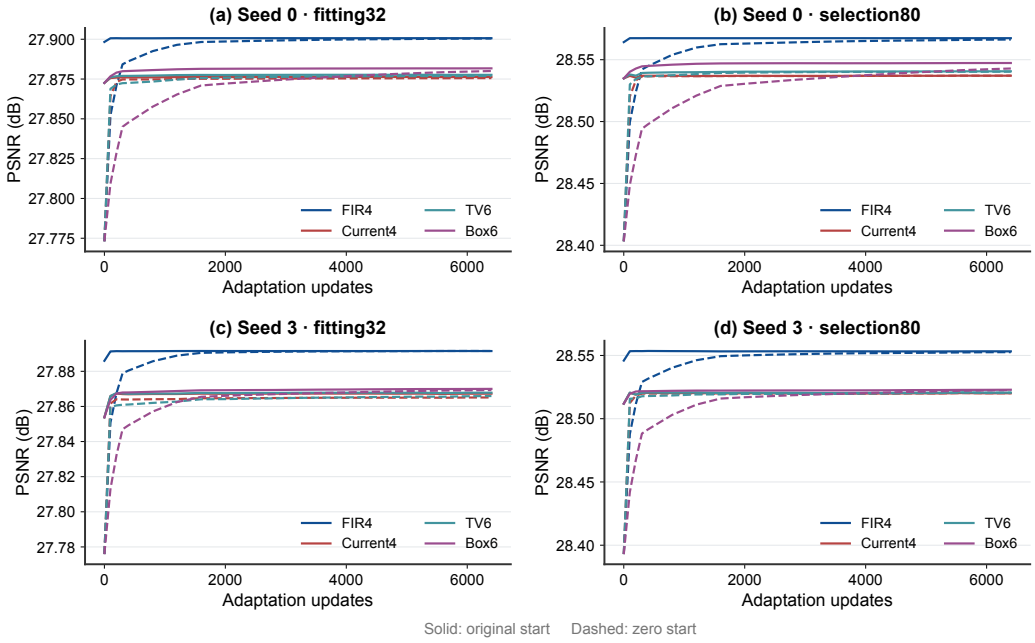


Figure 8: Adaptation curves for the two checkpoints designated for the second-optimizer check. Top: both seed 0; bottom: both seed 3. Left panels evaluate the fixed fitting32 subset; right panels use selection80. Solid/dashed lines denote original/zero starts, not independent backbone replicates. The complete image evaluation uses the selected candidates.

H.4 Path interventions and terminal image error

For each of six both checkpoints, each of six stages and each image, the entering state, observation, weights and previous increments are common to the four routes. A float64 SVD projects the original correction onto the span of the latest two increments, giving p and $h = d - p$. Reflection uses $p - h$ in refinement, direct or both branches. The unmodified $(++)$ result is the native model. All downstream states, increments and buffer entries are recomputed after the single intervention. Stages one and two have no older component and yield zero contrasts. The recorded maximum relative norm error is below 7.01×10^{-15} and orthogonality error below 1.73×10^{-15}.

Loss is per-image MSE divided by 255^2, averaged over the same 50 images before forming Eq. (16). Table 36 retains both evaluation splits and all nonzero-stage contrasts. At stages four and five, direct-path reflection increases test error more than refinement reflection in every checkpoint. At stage six, this ordering holds in five; new seed 5 has a negative direct contrast. Stage three has no uniform path ordering. In the new cohort on validation50, stage-six direct reflection reduces MSE in all three checkpoints, with mean contrast $(-3.0197 \pm 0.4333) \times 10^{-7}$. Thus the test50 path ordering does not extend uniformly to validation50. Negative late interactions indicate subadditivity of the two terminal-loss increases. The branch intervention identifies within-model path sensitivity; it is distinct from optimizing a replacement or retraining a new architecture.

Output geometry of the interaction. For one image with n pixels, write the four terminal reconstructions in $[0,1]$ units as $f_{++}, f_{-+}, f_{+-}, f_{--}$ and the target as x. Set $e = f_{++} - x$, $u = f_{-+} - f_{++}$, $v = f_{+-} - f_{++}$ and $w = f_{--} - f_{-+} - f_{+-} + f_{++}$. Expanding the four squared losses yields

$$I = \frac{2\langle u, v\rangle + 2\langle e + u + v, w\rangle + \|w\|^2}{n}. \tag{26}$$

Indeed, the joint error is $e + u + v + w$; subtracting the two single-route errors and adding $\|e\|^2$ leaves the numerator above. Averaging this identity over images with their own n gives the reported mean-MSE interaction. When the output changes are additive, $w = 0$ and $I = 2\langle u, v\rangle/n$: negative interaction then reflects opposing output changes. Nonzero w adds the contribution of nonadditive changes through the suffix and image reconstruction. Thus the observed negative I establishes subadditivity of the loss effects, rather than identifying nonlinearity by itself. The recorded losses determine I; separating its terms numerically would require the corresponding output vectors.

Table 36: Single-stage reflection effects on final image MSE, in units of 10^{-7} for pixels normalized to $[0,1]$. Mean $\pm$ sample SD across three checkpoint seeds. Stages 1–2 are exactly zero for all routes.

Stage	D_{ref}	D_{direct}	I
Test50; Both, seeds 0–2			
3	0.0661 ± 0.0479	0.1176 ± 0.1186	-0.0003 ± 0.0008
4	0.5110 ± 0.2272	1.3840 ± 0.4633	0.0103 ± 0.0120
5	-0.3400 ± 0.2123	2.8994 ± 0.2759	-0.1459 ± 0.0645
6	0.0267 ± 0.2502	5.1348 ± 1.1245	-0.5974 ± 0.3736
Test50; Both, seeds 3–5			
3	0.1213 ± 0.0888	0.0952 ± 0.2705	0.0060 ± 0.0085
4	0.4574 ± 0.1448	1.6371 ± 0.6411	-0.0041 ± 0.0217
5	0.2590 ± 0.1625	1.9246 ± 1.2360	-0.0852 ± 0.0676
6	-0.0520 ± 0.4063	2.8688 ± 2.7727	-0.5822 ± 0.6373
Validation50; Both, seeds 0–2			
3	0.0386 ± 0.0106	0.0436 ± 0.0654	-0.0005 ± 0.0004
4	0.2030 ± 0.1023	0.7554 ± 0.2399	-0.0250 ± 0.0244
5	-0.2353 ± 0.1590	1.1311 ± 0.4864	-0.1232 ± 0.0591
6	0.1777 ± 0.1681	1.2172 ± 1.2004	-0.5229 ± 0.3326
Validation50; Both, seeds 3–5			
3	0.1016 ± 0.0630	-0.2269 ± 0.3748	0.0038 ± 0.0062
4	-0.0121 ± 0.3333	0.7090 ± 0.6169	0.0077 ± 0.0170
5	0.0410 ± 0.0917	-0.2274 ± 0.2508	-0.0874 ± 0.0518
6	-0.1592 ± 0.3466	-3.0197 ± 0.4333	-0.5083 ± 0.5611

H.5 Synthetic replacement and remaining optimization dependence

The three existing capped FIR checkpoints (seeds 5–7) retain their 16-layer backbone and fit FIR taps, sum-current taps, or direct current gains. Sum-current collapses effective taps onto the current increment; direct-current optimizes 16 gains in $[-0.5, 0.5]$. Both represent the same effective current-gain function class under the active-tap startup rule. FIR and sum-current use the existing tanh/rescaling parameterization. Original and zero starts each run 6,400 Adam updates, batch 64, with the image study's learning-rate schedule and budget/checkpoint intervals. The objective is batch pooled normalized squared error. Separate streams provide 4,096 fitting and 1,024 selection signals; the existing 100,000 test signals per instance are reused after selecting the candidates. The test score pools error and target energy before taking the logarithm.

Table 37: Fixed-backbone synthetic adaptation: absolute NMSE (dB, lower is better) on the existing 100,000-signal arrays, mean ± sample SD across seeds 5–7.

Updates	Adapted FIR	Sum-current	Direct-current
400	$\mathbf{-60.6116 \pm 0.9929}$	-43.1381 ± 0.4819	-39.8151 ± 1.6179
1600	$\mathbf{-60.6116 \pm 0.9929}$	-43.4908 ± 0.5433	-41.7506 ± 0.8907
6400	$\mathbf{-60.6116 \pm 0.9929}$	-43.5513 ± 0.5713	-42.7849 ± 0.5236

Table 38: Lower replacement cost is better. Synthetic replacement loss: current NMSE minus adapted FIR NMSE (dB), paired mean ± sample SD. Both current parameterizations continue improving at the budget limit.

Updates	Sum-current − FIR	Direct-current − FIR
400	$\mathbf{17.4735 \pm 1.2136}$	20.7965 ± 2.2592
1600	$\mathbf{17.1207 \pm 1.2825}$	18.8609 ± 1.3486
6400	$\mathbf{17.0603 \pm 1.3012}$	17.8267 ± 0.9766

The final per-seed sum-current losses are 16.7564, 18.4865 and 15.9379 dB; direct-current losses are 17.7230, 18.8509 and 16.9061 dB. All twelve current-start trajectories still fail the 0.001 dB late improvement criterion. Five of the six selected current fits occur at update 6,400, with sum-current seed 6 selected at 6,000. FIR seeds 5 and 6 select their original step-zero coefficients and seed 7 selects step 400. Thus the finite-budget replacement losses remain large while the current fits continue improving. They measure dependence of these trained FIR backbones, not a lower bound on replacement error or the independent-retraining gains in Table 3.

Record-level verification. The study contains 90 adaptation trajectories, two projected-gradient checks, 11,900 image replacement records, 14,400 path-intervention records and 30 arrays of 100,000 synthetic model–signal outcomes. Recomputation checks finite metrics, unique keys, common inputs, pooled NMSE, PSNR and paired sample SD against the summaries. Selection records restrict every candidate to its declared budget. Source code, histories, selected parameters and per-image/per-signal records are indexed under `research/replacement_path_20260915/`; `verify_results.py` recomputes the saved-record checks and `build_paper_assets.py` generates these tables. These checks are distinct from full training reproduction and independent human author verification.

I ONE QUALITY TOLERANCE FOR TWO DECISIONS

Common reference and estimands. For each fixed image-training seed, let Q_F be the original FIR model's validation50 mean PSNR, Q_C the independently retrained current-input model's score, and Q_R the frozen-backbone, whole-image-PSNR replacement's score. All are evaluated on the same CUDA implementation and images. Define $D_{\text{train}} = Q_C - Q_F$ and $D_{\text{replace}} = Q_R - Q_F$. The equal-budget adapted FIR control remains a separate comparison in Table 26; it is not substituted for Q_F when comparing these two decisions.

For a task-specified tolerable PSNR loss $\delta > 0$, a joint decision to accept retraining but reject this replacement requires

$$L_{\text{train}} > -\delta, \qquad U_{\text{replace}} < -\delta, \quad \text{hence} \quad \max\{0, -L_{\text{train}}\} < \delta < -U_{\text{replace}}.$$

We examine the full interval rather than choose a favorable δ from the observed scores. No application-specific tolerance was fixed in this study, and this analysis does not establish one.

Exploratory uncertainty calculation. To expose the precision requirement, we use a stated normal-seed model: the three training-seed differences for each estimand are independent normal observations on this fixed image set. With $n = 3$, set

$$L_{\text{train}} = \bar{D}_{\text{train}} - t_{0.975,2} s_{\text{train}}/\sqrt{3},$$
$$U_{\text{replace}} = \bar{D}_{\text{replace}} + t_{0.975,2} s_{\text{replace}}/\sqrt{3}, \qquad t_{0.975,2} = 4.30265.$$

Each one-sided bound has 97.5% coverage under this model, giving at least 95% simultaneous coverage for the two bounds at one location by the union bound. Dependence between the two contrasts through Q_F does not require independence for this union bound. Coverage is not joint across locations. Three seeds cannot substantiate the normality assumption; images are not extra training repeats. The calculation is post-hoc sensitivity analysis, not a preregistered noninferiority test, distribution-free certificate or external confirmation. It conditions on the recorded training, selection and adaptation protocols.

Table 39: Common-reference paired PSNR differences and illustrative bounds (dB). The sample SD is computed from within-seed differences.

Quantity	Refinement	Both branches
D_{train}, mean $\pm$ SD	0.01376 ± 0.01059	0.00595 ± 0.03243
D_{replace}, mean $\pm$ SD	0.00094 ± 0.00045	-0.02696 ± 0.00300
L_{train}	-0.01254	-0.07461
U_{replace}	$+0.00207$	-0.01950
Required lower threshold for δ	0.01254	0.07461
Required upper threshold for δ	-0.00207	0.01950
Nonempty joint-decision interval	No	No

Dual-branch replacement is below its original FIR reference in all three seeds. Nevertheless, the retraining contrast's uncertainty prevents the two opposite decisions from being established at a common tolerance by this calculation. Refinement does not show the required replacement loss. Thus observed differences between tests do not yet certify different deployment decisions. The uncertainty concerns training repeats; more images alone do not supply additional training-seed information.

A future confirmation would first specify an externally justified quality tolerance, endpoint, multiplicity rule and fixed number of paired training seeds, using this batch for planning only. It would retain the same FIR reference and report the confirmation cohort separately. Three extra seed pairs are a possible resource allocation, not a guaranteed adequate sample size. No such confirmation was run here. The exact records, reconstructed means and bounds are provided by `reproducibility/reproduce_decisions.py` and `generated/decision_audit_20260913/summary.json`.

J STAGE-RESOLVED BSD EVALUATION ON ADDITIONAL TRAINING SEEDS

Training cohort and four comparison conditions. Seeds 3, 4 and 5 were fixed consecutively before training, separately from the earlier seeds 0–2. At MR25, each seed independently trains both-branch FIR and its current-increment control for 26,000 Adam updates, using matched backbone initialization, patch streams and the original fixed sensing matrix and dictionary. Patch batch size is 64, learning rate 10^{-4}, symmetry weight 0.01 and gradient clipping norm 1. Validation checks occur at step 1 and every 500 updates; the best validation checkpoint is retained. The BSD500 hash partition remains 400 training, 50 validation and 50 test images.

Each frozen FIR checkpoint also supplies two calibration conditions: current-input replacement and equal-budget FIR adaptation. Both optimize four shared coefficients with ℓ_1 cap 0.5 through all six stages, starting from original coefficients and zero. The 400-step image-PSNR objective, Adam schedule and 320/80 train-only fitting/selection protocol are those of Appendix G. Thus six trained models and their calibrated variants define four conditions per seed; the calibrations do not create additional independent backbone-training runs.

Frozen-prefix evaluation. After reviewing the validation endpoints, we added the stage curves and test50 evaluation, holding every checkpoint and calibration coefficient fixed. Test50 belongs to the existing project partition; this is a supplementary within-corpus evaluation. No test image selects weights, coefficients, seeds or an exit. Each curve executes stages 1–6 of the same six-stage-trained model. Internal dictionary routing is unchanged, and each intermediate code is decoded with the trained final dictionary D_{final}, matching the earlier prefix convention. This measures intermediate reconstruction quality, not independently trained depths or matched-latency performance.

Evaluation uses grayscale 16×16 patches, CUDA FP32 with TF32 disabled, patch batch 64, mean restoration, image assembly, unpadding and clipping to $[0, 255]$. Each trace starts with a fresh solver/history state. Image PSNR uses full-image pixel MSE and range 255; SSIM reuses the original grayscale evaluator. We first average 50 image scores within each seed, then report the mean and sample SD across three seeds. Paired SD is computed from the three within-seed differences, not from marginal method SDs.

Table 40: New seeds 3–5 on BSD test50: trained FIR minus independently trained current, mean $\pm$ sample SD of paired differences. Every stage is retained; all-stage validation and test curves are in Appendix J.

Stage	ΔPSNR (dB)	ΔSSIM	Positive PSNR seeds
1	$+0.2237 \pm 0.0083$	$+0.010125 \pm 0.000699$	3/3
2	$+0.1951 \pm 0.0837$	$+0.008247 \pm 0.003259$	3/3
3	$+0.1708 \pm 0.1651$	$+0.005946 \pm 0.009621$	3/3
4	$+0.4135 \pm 0.0362$	$+0.024929 \pm 0.002663$	3/3
5	$+0.0930 \pm 0.0502$	$+0.003077 \pm 0.002996$	3/3
6	$+0.0037 \pm 0.0084$	$+0.000453 \pm 0.000785$	2/3

Table 41: Six-stage endpoint quality for the additional seeds 3–5, mean $\pm$ sample SD. Bold identifies the higher displayed mean within each trained pair or equal-budget calibration pair on the same split; it does not denote significance.

Split	Condition	PSNR (dB)	SSIM
Validation50	FIR (trained)	$\mathbf{29.0734 \pm 0.0164}$	$\mathbf{0.844843 \pm 0.000724}$
	Current (trained)	29.0680 ± 0.0270	0.844401 ± 0.000733
	FIR (adapted)	$\mathbf{29.0799 \pm 0.0179}$	$\mathbf{0.845328 \pm 0.000929}$
	Current (replacement)	29.0478 ± 0.0194	0.844669 ± 0.000952
Test50	FIR (trained)	$\mathbf{27.9871 \pm 0.0227}$	$\mathbf{0.818962 \pm 0.000794}$
	Current (trained)	27.9834 ± 0.0304	0.818509 ± 0.000868
	FIR (adapted)	$\mathbf{27.9928 \pm 0.0240}$	$\mathbf{0.819538 \pm 0.001062}$
	Current (replacement)	27.9623 ± 0.0253	0.819001 ± 0.001142

Table 42: Within-seed endpoint contrasts, mean ± sample SD across seeds 3–5. Positive favors the named FIR reference. Different-reference contrasts are not ranked.

Split	Contrast	ΔPSNR (dB)	ΔSSIM
Validation50	Trained FIR − trained current	$+0.00534 \pm 0.01133$	$+0.000442 \pm 0.000643$
	Original FIR − replacement	$+0.02552 \pm 0.00350$	$+0.000173 \pm 0.000301$
	Adapted FIR − replacement	$+0.03204 \pm 0.00317$	$+0.000658 \pm 0.000170$
Test50	Trained FIR − trained current	$+0.00369 \pm 0.00839$	$+0.000453 \pm 0.000785$
	Original FIR − replacement	$+0.02475 \pm 0.00263$	-0.000039 ± 0.000406
	Adapted FIR − replacement	$+0.03049 \pm 0.00191$	$+0.000537 \pm 0.000211$

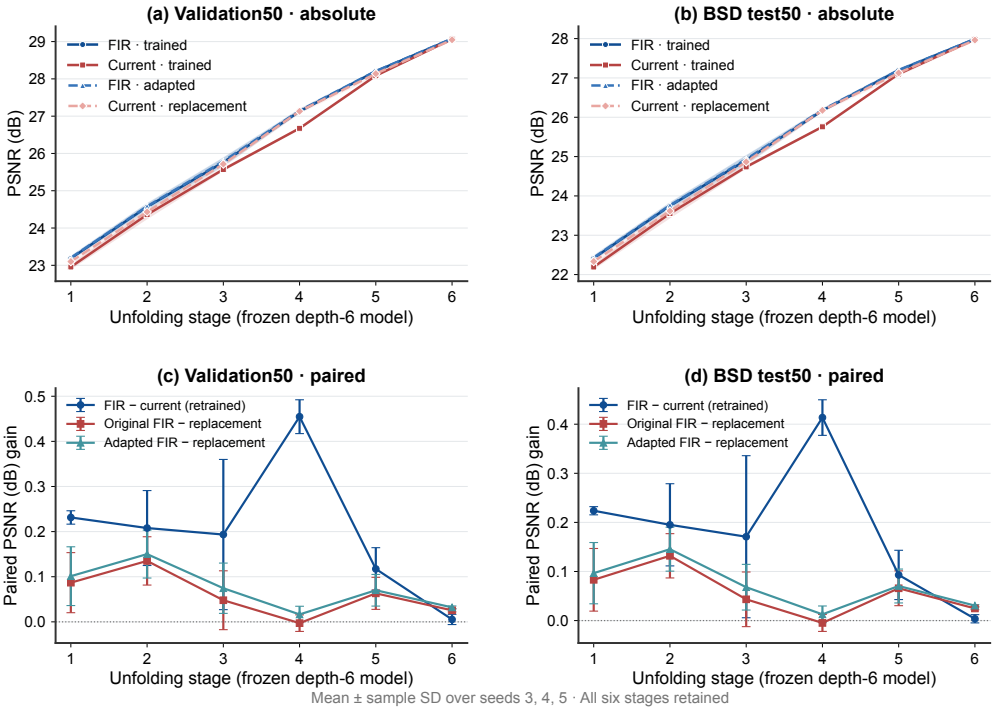


Figure 9: Frozen-prefix PSNR on the additional CUDA cohort, seeds 3–5. Panels (a,b) show all four conditions on validation50 and test50; panels (c,d) show three within-seed contrasts. Bands and error bars are mean ± sample SD over seeds. Each panel has its own legend; positive paired differences favor the named FIR reference. All six stages are retained.

Intermediate quality and endpoint comparison. The stage-four FIR-minus-retrained-current PSNR gap is $+0.4547 \pm 0.0375$ dB on validation50 and $+0.4135 \pm 0.0362$ dB on test50, positive in all three seeds on both sets. At stage six these gaps are $+0.0053 \pm 0.0113$ and $+0.0037 \pm 0.0084$ dB, respectively, with mixed seed signs. The complete trajectory therefore reveals a larger intermediate separation than the endpoint alone. The two architectures are independently trained, so this profile reflects their full learned trajectories rather than an isolated effect of inserting history at stage four.

Frozen replacement is a separate comparison. At the test50 endpoint, original FIR exceeds its calibrated current-input replacement by 0.02475 ± 0.00263 dB in PSNR; equal-budget adapted FIR exceeds that replacement by 0.03049 ± 0.00191 dB. Both PSNR gaps are positive for all three seeds. Original-FIR-minus-replacement SSIM is -0.000039 ± 0.000406, whereas adapted-FIR-minus-replacement SSIM is $+0.000537 \pm 0.000211$. These metric-specific results distinguish checkpoint replacement from independent retraining, whose terminal PSNR ordering has mixed signs. They do not define a statistical noninferiority or equivalence margin.

Table 43: All-stage trained-FIR-minus-trained-current contrasts. Mean ± sample SD of three within-seed differences; each prefix is decoded with D_{final}.

Split	Stage	ΔPSNR (dB)	ΔSSIM
Validation50	1	$+0.2314 \pm 0.0148$	$+0.009116 \pm 0.000619$
	2	$+0.2079 \pm 0.0830$	$+0.007633 \pm 0.002765$
	3	$+0.1935 \pm 0.1664$	$+0.005812 \pm 0.008477$
	4	$+0.4547 \pm 0.0375$	$+0.022017 \pm 0.002208$
	5	$+0.1173 \pm 0.0470$	$+0.003948 \pm 0.002359$
	6	$+0.0053 \pm 0.0113$	$+0.000442 \pm 0.000643$
Test50	1	$+0.2237 \pm 0.0083$	$+0.010125 \pm 0.000699$
	2	$+0.1951 \pm 0.0837$	$+0.008247 \pm 0.003259$
	3	$+0.1708 \pm 0.1651$	$+0.005946 \pm 0.009621$
	4	$+0.4135 \pm 0.0362$	$+0.024929 \pm 0.002663$
	5	$+0.0930 \pm 0.0502$	$+0.003077 \pm 0.002996$
	6	$+0.0037 \pm 0.0084$	$+0.000453 \pm 0.000785$

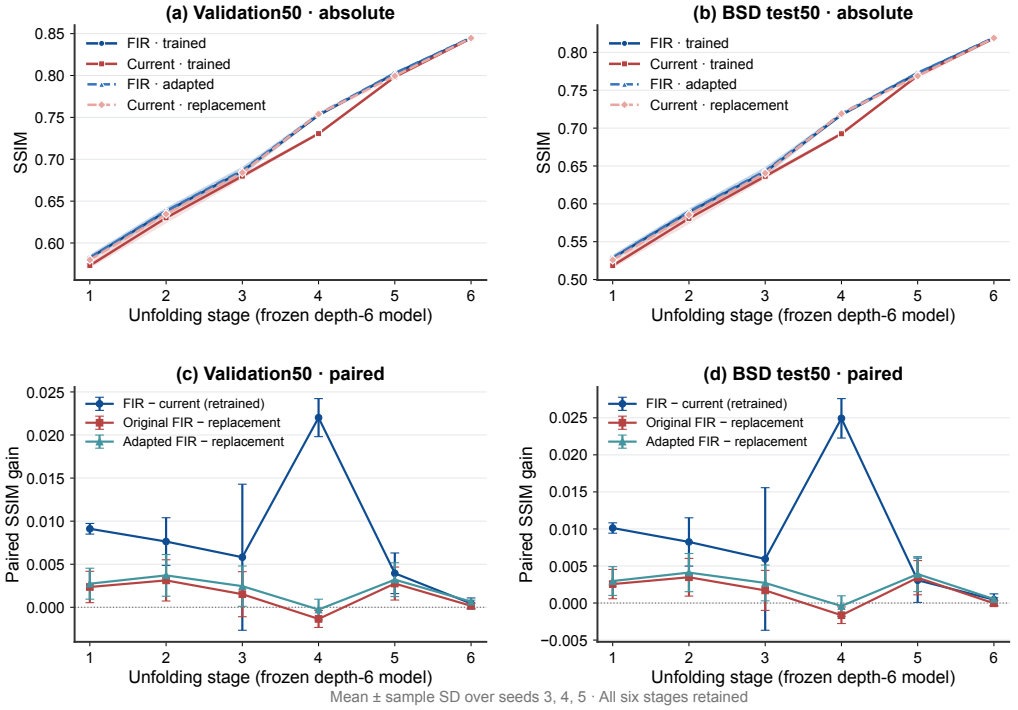


Figure 10: Frozen-prefix SSIM with the same conditions, seeds, images and prefixes as Figure 9. Each panel includes its own legend. Shading and error bars denote sample SD across three seed means; they are not confidence intervals. Every method uses the same image processing and metric implementation.

Record verification and reproducibility. The 144 unique condition–stage–split summaries contain 7,200 finite per-image records, with disjoint validation50/test50 IDs and unchanged frozen tensors. Every prefix passes the reference startup-patch check; stage six matches the native forward exactly and reproduces the prior batch-64 validation metrics with zero discrepancy. These are automated record and inference checks. Raw seed files, `absolute.csv`, `paired.csv`, `per_image.csv`, evaluation and figure/table sources are in `research/decision_confirmation_20260913/depth_curves/`; its parent directory retains the original training and calibration records.

K CORRECTED MATCHED SYNTHETIC SPARSE RECOVERY

Evidence identity and implementation scope. This is a new nine-run experiment (three methods, three prespecified paired seeds), distinct from the 18 historical records in Appendix AO. The backbone is the tied-weight, scalar-threshold HLISTA-CP configuration from the original synthetic implementation. We use an independently implemented PyTorch/MUSA training path with common supplied initial states, not the original TensorFlow optimization trajectory.

The inspected original parent converts its convolution output from $[B, N, 1]$ directly to $[N, B]$ by reshape. For $B > 1$, this does not preserve sample correspondence. The inspected control instead squeezes and transposes. All three new arms use the same sample-preserving tail; all also use the same 10^{-8} stabilization in the mixing denominator. The original parent source hash matches the historical provenance, but the available control source hash differs from its historical record. We therefore do not attribute the old numerical ranking conclusively to this discrepancy. The corrections are implemented in a separate synthetic experiment directory; no image training or evaluation source was changed for this rerun. The image evidence remains under its separately documented implementations and is not relabeled as this corrected protocol.

Task, model and matching. Measurements are $y = Ax^\star$ with $A \in \mathbb{R}^{250\times 500}$ loaded from the original seed-specific problem files. Coordinates of $x^\star$ are independent Bernoulli(0.1) masks times standard normal variables; no observation noise is added. The number of active coordinates is random, not fixed at 50. The 16-layer backbone shares its backprojection and three convolution kernels across layers; the convolutions have kernel size 9, hidden width 16, no biases, and ReLU in the first two layers. Thresholds and mixture parameters are scalar and layer-specific. Each seed supplies a common backbone initial state to all three arms. The parent has 127,640 parameters; each feedback arm has 127,704.

At layer k, FIR uses four layer-specific coefficients $b_{k,j} = \tanh(a_{k,j})$, initially zero, with only $J_k = \min(4, k+1)$ active lags. There is no additional absolute-sum cap; $\sum_{j<J_k} |b_{k,j}| < J_k$. FIR injects $\sum_{j<J_k} b_{k,j} r_{k-j}$, whereas the matched current control injects $(\sum_{j<J_k} b_{k,j}) r_k$. Both correct the CNN input and direct mixture branch. This matches coefficient count, startup and available aggregate gain range, not merely the name of the control. It differs from the image protocol's four globally shared taps with absolute-sum cap 0.5. No claim of equality between these two training protocols is made.

Training, selection and fresh evaluation. Every model receives 480,000 updates with batch size 64: 10,000 updates in each of three phases at each of 16 depths. Learning rates are 5×10^{-4}, 10^{-4} and 10^{-5}. The first phase trains newly introduced parameters; the next two train all introduced parameters, scaling older-parameter gradients by 0.3 per layer of age. Adam is reset each phase; the loss is half the summed squared error. The minimum-validation-error state is restored after each phase. This fixed finite budget replaces the legacy adaptive stopping schedule.

Validation uses 1,000 fresh signals per seed and occurs every 200 updates, as well as at the first and last update of each phase. The final checkpoint minimizes validation global NMSE over all depth-16 evaluations. Only then are 10,000 separate fresh test signals evaluated for that seed. Seed indices 0, 1, and 2 jointly identify operator, initialization, training stream, validation and test streams; all arms within each index share these inputs. Test generators use seeds $910000 + s$, validation generators $310000 + s$, and training streams use $410000 + 1000s + 3(k-1) + p$ for depth k and phase $p \in \{0, 1, 2\}$. No test scores were used to change the nine-run roster, training budget or checkpoint rule. A scheduling amendment moved the single unstarted parent run from GPU7 to idle GPU0 without changing the scientific configuration.

Metric and numerical precision. For each seed, the primary statistic is

$$E_s = 10 \log_{10} \left(\frac{\sum_{i=1}^{10000} \|\widehat{x}_{s,i} - x^\star_{s,i}\|_2^2}{\sum_{i=1}^{10000} \|x^\star_{s,i}\|_2^2} \right). \tag{27}$$

We report the mean and sample SD of the three E_s values. This energy-weighted ratio is neither the mean of per-signal log NMSE nor the log of an unweighted mean of per-signal ratios. Denominators are positive and checked; no clipping, epsilon floor or log truncation is applied. Models use FP32

with TF32 convolution disabled and no mixed precision; error and target energies accumulate in FP64. These values must not be combined with image metrics using a numerical floor.

Table 44: All prespecified synthetic test outcomes. NMSE in dB, lower is better. SD describes three paired problem/initialization/training/test seeds, not pure training-seed variance. Differences are computed within seed before aggregation.

Method or contrast	Seed 0	Seed 1	Seed 2	Mean ± SD
HLISTA-CP	-45.8809	-45.3426	-46.1669	-45.7968 ± 0.4185
Current increment	-61.9892	-67.5901	-58.7737	-62.7843 ± 4.4617
FIR	$\mathbf{-81.0982}$	$\mathbf{-90.0149}$	$\mathbf{-91.2161}$	$\mathbf{-87.4431 \pm 5.5275}$
FIR minus parent	-35.2173	-44.6723	-45.0492	-41.6463 ± 5.5708
Current minus parent	-16.1083	-22.2475	-12.6069	-16.9876 ± 4.8801
FIR minus current	-19.1091	-22.4247	-32.4424	-24.6587 ± 6.9417

Checks, provenance and interpretation. Zero-feedback outputs agree across all 16 prefixes and three CPU initializations; GPU zero-feedback and batch-permutation tests also pass. A corrected TensorFlow equation graph and PyTorch implementation agree on forward outputs and selected input/weight gradients with supplied common weights. After disabling TF32, maximum tested MUSA forward discrepancy is approximately 3.58×10^{-7}. These checks do not assert identical Adam trajectories or certify trained backbone contraction. All nine archives were checked, global test NMSE was independently recomputed from 10,000 per-signal error/target energy records, and validation-selected checkpoint indices and all 48 completed phases were verified. Frozen configuration, source/input hashes, phase checkpoints and validation histories are retained with the result artifacts.

The paired FIR advantage is present in all three seeds. It establishes a strong positive instance against the specified retrained current control, not superiority over every memory-free alternative. The final selected model's prefix metrics are not independently trained shallow-model comparisons, nor a convergence-speed measurement. We do not infer synthetic latency from image inference costs. The image controls and external test provide a different empirical answer; differences in depth, target distribution, sharing and training prevent attributing that contrast to sparsity alone. Neither outcome verifies the conditional theory's backbone assumptions.

L Bounded synthetic competitive comparisons

Roster, budget and relationship to the existing test. This block adds exactly six trained models: two-tap learned correction and HLISTA-CPSS, each at seeds 0, 1 and 2. The original nine-run study is unchanged. All six use the same seed-specific operators, initialization, training and validation streams, depth 16, batch 64 and 480,000-update schedule in Appendix K. Validation receives the same checkpoint selection opportunities. The two configurations were fixed before training; there was no seed search, test-directed retuning or selection among extra runs. The 10,000 test signals per seed had already been evaluated for the original three methods. Their reuse enables paired method comparison, not a new independent external confirmation. No exhaustive competitor hyperparameter search or proof of each architecture's optimized ceiling is claimed.

Two-tap learned correction. The correction is $d_k = \tanh(a_{k,0})r_k + \tanh(a_{k,1})r_{k-1}$ after startup; at $k = 0$ only the current tap is active. Both taps are layer-specific, initialized to zero, and inserted at both branches of the common corrected backbone. There are 32 additional parameters (127,672 total), versus 64 for four-tap FIR. There is no extra absolute-sum cap; the available aggregate absolute gain is below 2, versus 4 for four active FIR taps. This natural short-history competitor is not FISTA and is not a capacity- or gain-matched history-length ablation. Its lower accuracy cannot alone identify whether length, parameter count, gain range or their interaction causes the difference.

HLISTA-CPSS adaptation. The model retains the common sample-preserving CNN tail and stabilized mixture of the corrected HLISTA-CP implementation, with 127,640 parameters and no feedback taps. Both threshold operations use support selection. At layer k (zero-indexed),

$p_k = \min(0.8(k+1), 10)$ percent. Coordinates bypass shrinkage only when their magnitudes strictly exceed both the learned threshold and the nearest empirical $(100 - p_k)$th percentile; the selection mask is detached. The increment 0.8 is the inspected source's default increment; the cap 10 is a fixed benchmark choice motivated by the expected 10% activity, not that source's default cap 0 (which disables support selection). The cap was not optimized after inspecting test results. This is an adapted HLISTA-CPSS competitor, not numerical reproduction of a legacy TensorFlow checkpoint.

Table 45: All added competitors and the existing four-tap comparator: test NMSE in dB (lower is better), mean ± sample SD across paired problem/training seeds. Positive comparator-minus-FIR differences favor four-tap FIR.

Method/contrast	Seed 0	Seed 1	Seed 2	Mean ± SD
Two-tap correction	-64.2373	-56.3990	-57.8632	-59.4998 ± 4.1676
HLISTA-CPSS	-63.9495	-63.2330	-66.3954	-64.5259 ± 1.6582
Four-tap FIR	$\mathbf{-81.0982}$	$\mathbf{-90.0149}$	$\mathbf{-91.2161}$	$\mathbf{-87.4431 \pm 5.5275}$
Two-tap minus FIR	16.8609	33.6159	33.3529	27.9432 ± 9.5985
CPSS minus FIR	17.1488	26.7819	24.8207	22.9171 ± 5.0909

Verification and scope. At zero feedback the two-tap implementation matches the common parent, and it matches four-tap FIR with the last two taps zeroed on the checked inputs. CPSS shrinkage matches an independent NumPy nearest-percentile reference; CPU/GPU and sample-permutation checks pass. All six archives pass integrity checks. All 48 training phases completed, the validation-selected checkpoint indices were checked, and global test NMSE was independently recomputed from per-signal energies. Target energies match the original same-seed test records. All six outcomes are retained. Four-tap FIR outperforms both specified competitors in each seed, strengthening competitive evidence for this configuration. The comparison supplies neither a synthetic latency result nor an experimentally certified convergence guarantee, and it does not change the separate image-domain history/cost conclusions.

M Stronger synthetic controls, frozen intervention and inference cost

This section preserves the earlier cross-platform comparison and the original frozen diagnostics. The completed same-platform training and new independent timing session are reported separately in Appendix N; no historical result is overwritten.

Controlled alternatives. This fixed extension trains two alternatives at seeds 0, 1 and 2 without outcome-dependent seed selection. It reuses the primary synthetic study's sensing matrices, backbone initialization and sample streams, with 16 layers, three phases per layer and 10,000 updates per phase (480,000 updates/model). Batch size is 64; validation occurs every 200 updates. Phase-best restoration and depth-16 validation selection precede evaluation of 10,000 test signals per seed. All six runs completed. The original two-tap competitor remains a separate model: it has fewer raw coefficients and a smaller aggregate gain range, whereas the new two-time control retains four raw coefficients.

Let $q_k = \min(k+1, 4)$ denote the active tap count and $c_{k,j} = \tanh(\theta_{k,j})$. The new two-time correction is

$$d_k = c_{k,0} r_k + \left(\sum_{j=1}^{q_k - 1} c_{k,j} \right) r_{k-1},$$

with the second term absent at initialization. It preserves the four per-layer parameters and startup mask while tying all past inputs to the preceding increment. This matches raw parameterization and coefficient ranges, not independent effective dimensions. The nonlinear current-only alternative uses

$$d_k = q_k \tanh(f_k(r_k / s_k)) \odot r_k,$$

where f_k is a coordinatewise $1 \to 8 \to 1$ tanh MLP shared across signal coordinates, with a zero-initialized final layer. Its 25 parameters per layer exceed FIR's four; it is not capacity-matched. The RMS scale s_k is calibrated from initialized-parent increments on 1,024 training-distribution signals (stream $730000 + s$ for seed s). No test targets enter calibration or model selection.

Training platform and numerical verification. The new controls use PyTorch 2.7.1+cu128 on an RTX 5060 Ti, whereas the parent, current and FIR references were trained on MUSA. Inputs and streams are paired, but training hardware is not. Frozen reevaluation of the nine reference checkpoints on CUDA differs from their original NMSE by at most 0.00022 dB. This checks inference portability, not training-trajectory equivalence. Table 46 therefore retains training-platform labels; historical reference scores are not overwritten.

Table 46: Synthetic test NMSE (dB, lower is better). Means and sample SDs use three training seeds; each seed first aggregates error and target energies across 10,000 signals. Ranks compare means within each training platform.

Method	Training platform	NMSE (dB)
HLISTA-CP	MUSA	-45.80 ± 0.42
Current gain	MUSA	-62.78 ± 4.46
Four-tap FIR	MUSA	$\mathbf{-87.44 \pm 5.53}$
Two-time control	CUDA	$\mathbf{-77.42 \pm 5.48}$
Nonlinear current control	CUDA	-74.86 ± 0.60

Table 47: Paired NMSE reductions relative to FIR: control minus FIR, so positive favors FIR. All fixed seeds are retained. These are cross-platform training comparisons; sample SD is not a confidence interval.

Control	Seed 0	Seed 1	Seed 2	Mean $\pm$ SD
Two-time control	0.30	18.92	10.87	10.03 ± 9.34
Nonlinear current control	6.81	15.22	15.73	12.58 ± 5.01

FIR's lower NMSE in every seed supports recovery gains beyond the tested linear-current, collapsed-history and nonlinear-current alternatives. The two-time comparison has substantial seed dispersion; the platform boundary remains part of its interpretation.

Frozen trajectory intervention. For each existing FIR checkpoint, stagewise scalar gains are fitted on 4,096 training-distribution signals (stream $740000 + s$):

$$a_k = \frac{\sum_i \langle r_{k,i}, d_{k,i} \rangle}{\sum_i \|r_{k,i}\|_2^2}.$$

All denominators are positive. Unlike the image calibration, this scalar fit has no additional gain constraint. We freeze these gains and evaluate $d_k^{(\lambda)} = a_k r_k + \lambda(d_k - a_k r_k)$ for the fixed values $\lambda \in \{0, 0.5, 1\}$. Each intervention reruns the complete trajectory, rebuilding history from intervened states; independent batches reset state. The full-history endpoint returns the original correction directly. No test signal enters the gain fit.

Table 48: Frozen synthetic FIR intervention on CUDA. Test NMSE (dB), mean $\pm$ sample SD across the same three checkpoints; lower is better.

Correction	NMSE (dB)
Fitted current gain ($\lambda = 0$)	-18.03 ± 0.79
Half history residual ($\lambda = 0.5$)	-48.23 ± 0.17
Original FIR ($\lambda = 1$)	$\mathbf{-87.44 \pm 5.53}$

The ordered response demonstrates dependence of these trained FIR trajectories on their history corrections. It does not prove that a current-only model cannot recover after retraining, nor that the fitted scalar is an optimal replacement for final reconstruction quality.

Actual synthetic inference cost. All five methods and three seeds are measured on the same RTX 5060 Ti, separately from training, in FP32 with TF32 and automatic mixed precision disabled. Inputs are device-resident measurements; timing encloses model forward execution with CUDA synchronization. For each batch size and checkpoint, 30 warmups precede 100 timed repetitions. The reported latency is mean $\pm$ sample SD across the three per-seed medians, not the SD of 300 independent training runs. Desktop WDDM graphics activity remains in the environment. Peak allocated memory includes resident model/input allocations; it is neither incremental workspace nor reserved memory.

Table 49: Synthetic forward latency (ms, lower is better). Batch 64 is time per batch, not per signal. All frozen checkpoints execute on CUDA.

Method	Batch 1	Batch 64
HLISTA-CP	$\mathbf{5.80 \pm 0.25}$	$\mathbf{6.28 \pm 0.06}$
Current gain	6.48 ± 0.07	7.11 ± 0.04
Four-tap FIR	7.14 ± 0.00	7.67 ± 0.05
Two-time control	6.77 ± 0.08	7.45 ± 0.10
Nonlinear current control	7.92 ± 0.07	8.51 ± 0.07

FIR takes more time than the parent, current gain and two-time control, but less than the nonlinear current control in this implementation. Its quality advantage over two-time correction is a quality–cost tradeoff, not a uniform speedup. These synthetic timings do not replace the separate image reconstruction measurements.

N Completed same-platform CUDA confirmation

Fixed training and validation. Three additional FIR models are trained from the common initialization at seeds 0, 1 and 2 on the same CUDA environment as the two-time and nonlinear controls. There are no warm starts from MUSA checkpoints and no outcome-dependent retries. Each run completes all 48 phases and 480,000 updates with the original validation-only selection rule. The model source is byte-identical to the earlier CUDA implementation; the training source diff only admits FIR in the argument parser. Sensing, initialization, validation and test files are byte-identical; training streams use the same NumPy generator schedule. All raw test error/target energies were recomputed, and target energies agree across paired methods.

Table 50: Same-platform test NMSE (dB, lower is better). Three fixed seeds, 10,000 test signals per seed; bold marks the best values in each column.

Method	Seed 0	Seed 1	Seed 2	Mean $\pm$ SD
Two-time (CUDA)	-80.80	-71.10	-80.35	-77.42 ± 5.48
Nonlinear (CUDA)	-74.29	-74.80	-75.49	-74.86 ± 0.60
FIR (CUDA)	$\mathbf{-82.83}$	$\mathbf{-88.58}$	$\mathbf{-90.75}$	$\mathbf{-87.38 \pm 4.09}$

Table 51: Paired differences, comparator minus CUDA FIR (dB). Positive means CUDA FIR has lower NMSE. The final row compares training platforms for FIR and is not an additional architectural control.

Comparator	Seed 0	Seed 1	Seed 2	Mean $\pm$ SD
Two-time (CUDA)	+2.03	+17.48	+10.40	9.97 ± 7.74
Nonlinear (CUDA)	+8.53	+13.78	+15.26	12.52 ± 3.53
FIR (MUSA)	+1.73	-1.44	-0.47	-0.06 ± 1.62

The stronger-control advantage persists when training platform is held fixed. This does not make training-platform effects identically zero: CUDA-minus-MUSA FIR differences have mixed signs, and their paired mean is $+0.06 \pm 1.62$ dB. The six FIR runs are not pooled into six independent problem seeds, and no better platform is selected post hoc. The earlier fresh-signal frozen diagnostics

remain associated with the MUSA FIR; their mixed ordering against nonlinear current correction is unchanged.

New isolated timing session. After training completes, all six model/platform groups are measured on the RTX 5060 Ti in FP32, with AMP and TF32 disabled. Each seed cyclically rotates the fixed method order. Device-resident observations are evaluated at batches 1 and 64, with 30 warmups and 100 CUDA-synchronized wall-clock forward timings per checkpoint/batch. All 3,600 values are archived. Each cell reports the mean and sample SD of three seed-level medians. Desktop WDDM activity is not eliminated. Absolute latencies of all reference methods differ from the earlier timing session; comparisons below and in Figure 4 are confined to this session, not attributed to a change in model weights across sessions.

Table 52: New-session forward latency (ms per batch). Lower is better; training platforms label distinct checkpoints, all executed on CUDA.

Checkpoint group	Batch 1	Batch 64
Parent (MUSA)	$\mathbf{17.86 \pm 0.33}$	$\mathbf{18.86 \pm 0.32}$
Current (MUSA)	19.98 ± 0.14	20.90 ± 0.17
FIR (MUSA)	21.98 ± 0.36	23.32 ± 0.31
FIR (CUDA)	22.11 ± 0.09	23.06 ± 0.34
Two-time (CUDA)	20.82 ± 0.33	22.09 ± 0.22
Nonlinear (CUDA)	23.84 ± 0.38	24.84 ± 0.07

At batch 64, CUDA FIR takes 23.06 ± 0.34 ms versus 22.09 ± 0.22 ms for two-time and 24.84 ± 0.07 ms for nonlinear current correction. Thus its higher recovery accuracy trades against time for two-time correction, while both observed metrics favor FIR over the tested nonlinear control. This is a comparison of these complete implementations, not a hardware-independent speed guarantee or a formal Pareto-optimality claim.

O Comparator choice across frozen cross-backbone configurations

This posthoc analysis reuses the fixed selected13 test200 records. Each of 13 operator–backbone–rate groups has two injection locations and three seeds. For each seed, we subtract the matching parent or current-input MLP before computing means and sample SDs. No standard deviation is inferred from marginal method statistics. There are 273 distinct method–group–seed records, each evaluated on 200 images. All 26 configurations are retained; the MLP is matched to LSTM parameter count, not FIR. The configurations share images, seeds and some parents: quadrant counts are descriptive, not independent-replicate tests. Positive PSNR means occur in 19 FIR–parent and 12 FIR–MLP contrasts. Six are positive under both comparators, 13 only under the parent, six only under MLP, and one under neither.

Table 53: All 26 paired PSNR (dB) contrasts. Mean ± sample SD over seeds 0/1/2; signs are in seed order. Different reference columns are not ranked.

Group	Location	FIR–parent	Signs	FIR–MLP	Signs
QR/HAL/10	Both	$+0.0001 \pm 0.0253$	+ + −	$+0.2227 \pm 0.1739$	+ + +
QR/HAL/10	Ref.	$+0.0484 \pm 0.0149$	+ + +	$+0.0083 \pm 0.0891$	− + −
QR/HAL/25	Both	-0.0722 ± 0.0385	− − −	$+0.0286 \pm 0.1171$	− − +
QR/HAL/25	Ref.	-0.0269 ± 0.0224	− − −	$+0.0508 \pm 0.0468$	+ + −
QR/HCI/25	Both	$+2.8010 \pm 0.0189$	+ + +	-1.1278 ± 0.1875	− − −
QR/HCI/25	Ref.	$+0.5321 \pm 0.2234$	+ + +	-1.2647 ± 0.2111	− − −
QR/HG-gain/25	Both	-0.0621 ± 0.0593	− − −	$+0.9256 \pm 0.8759$	+ + +
QR/HG-gain/25	Ref.	-0.0343 ± 0.0695	− − +	$+0.1361 \pm 0.1706$	+ − +
QR/HG-overshoot/25	Both	-0.0007 ± 0.0373	+ − +	$+0.3084 \pm 0.1831$	+ + +
QR/HG-overshoot/25	Ref.	$+0.0121 \pm 0.0476$	+ − +	$+0.3421 \pm 0.2150$	+ + +
QR/CPSS/10	Both	$+0.0160 \pm 0.0040$	+ + +	-0.0211 ± 0.0208	− − −
QR/CPSS/10	Ref.	$+0.0044 \pm 0.0077$	− + +	-0.0132 ± 0.0214	− − +
QR/CPSS/25	Both	$+0.0081 \pm 0.0030$	+ + +	$+0.0186 \pm 0.0072$	+ + +
QR/CPSS/25	Ref.	$+0.0144 \pm 0.0043$	+ + +	-0.1267 ± 0.0063	− − −
IID/HAL/10	Both	$+0.1659 \pm 0.0494$	+ + +	-0.0300 ± 0.0252	− − −
IID/HAL/10	Ref.	$+0.1237 \pm 0.1368$	+ + +	-0.0490 ± 0.0811	− − +
IID/HAL/25	Both	-0.0184 ± 0.0066	− − −	$+0.1894 \pm 0.3487$	+ − +
IID/HAL/25	Ref.	$+0.0134 \pm 0.0237$	+ + −	$+0.1317 \pm 0.0780$	+ + +
IID/HCI/25	Both	$+0.6350 \pm 0.0044$	+ + +	-11.3820 ± 0.0946	− − −
IID/HCI/25	Ref.	$+0.2100 \pm 0.0223$	+ + +	-0.7827 ± 0.0404	− − −
IID/HG-gain/25	Both	$+0.0004 \pm 0.0026$	− + −	-0.0192 ± 0.0056	− − −
IID/HG-gain/25	Ref.	$+0.0021 \pm 0.0041$	+ − +	$+0.0001 \pm 0.0184$	− + −
IID/CPSS/10	Both	$+0.0098 \pm 0.0072$	+ + +	-0.0151 ± 0.0204	− − +
IID/CPSS/10	Ref.	$+0.0041 \pm 0.0085$	+ − +	-0.0253 ± 0.0131	− − −
IID/CPSS/25	Both	-0.0035 ± 0.0187	+ − +	-0.0562 ± 0.0381	− − −
IID/CPSS/25	Ref.	$+0.0092 \pm 0.0163$	− + +	-0.0928 ± 0.0280	− − −

Table 54: All 26 paired SSIM contrasts. Mean ± sample SD over seeds 0/1/2; signs are in seed order. Different reference columns are not ranked.

Group	Location	FIR–parent	Signs	FIR–MLP	Signs
QR/HAL/10	Both	-0.0024 ± 0.0040	+ − −	$+0.0107 \pm 0.0092$	+ + +
QR/HAL/10	Ref.	$+0.0012 \pm 0.0055$	+ − −	$+0.0021 \pm 0.0069$	+ + −
QR/HAL/25	Both	-0.0047 ± 0.0019	− − −	$+0.0009 \pm 0.0059$	− − +
QR/HAL/25	Ref.	-0.0017 ± 0.0009	− − −	$+0.0010 \pm 0.0031$	+ + −
QR/HCI/25	Both	$+0.1084 \pm 0.0010$	+ + +	-0.0398 ± 0.0078	− − −
QR/HCI/25	Ref.	$+0.0125 \pm 0.0085$	+ + +	-0.0547 ± 0.0072	− − −
QR/HG-gain/25	Both	-0.0022 ± 0.0028	− − −	$+0.0492 \pm 0.0453$	+ + +
QR/HG-gain/25	Ref.	-0.0015 ± 0.0029	− − +	$+0.0062 \pm 0.0079$	+ − +
QR/HG-overshoot/25	Both	-0.0000 ± 0.0015	+ − +	$+0.0152 \pm 0.0088$	+ + +
QR/HG-overshoot/25	Ref.	$+0.0005 \pm 0.0020$	+ − +	$+0.0160 \pm 0.0108$	+ + +
QR/CPSS/10	Both	$+0.0009 \pm 0.0003$	+ + +	-0.0006 ± 0.0013	− − +
QR/CPSS/10	Ref.	$+0.0002 \pm 0.0005$	− + +	-0.0005 ± 0.0013	− − +
QR/CPSS/25	Both	$+0.0001 \pm 0.0002$	− + +	$+0.0007 \pm 0.0003$	+ + +
QR/CPSS/25	Ref.	$+0.0003 \pm 0.0003$	+ + +	-0.0029 ± 0.0001	− − −
IID/HAL/10	Both	$+0.0143 \pm 0.0018$	+ + +	$+0.0054 \pm 0.0016$	+ + +
IID/HAL/10	Ref.	$+0.0062 \pm 0.0050$	+ + +	-0.0055 ± 0.0041	− − −
IID/HAL/25	Both	-0.0014 ± 0.0007	− − −	$+0.0091 \pm 0.0164$	− − +
IID/HAL/25	Ref.	$+0.0006 \pm 0.0017$	+ + −	$+0.0067 \pm 0.0041$	+ + +
IID/HCI/25	Both	$+0.0265 \pm 0.0002$	+ + +	-0.4149 ± 0.0056	− − −
IID/HCI/25	Ref.	$+0.0009 \pm 0.0007$	+ + +	-0.0283 ± 0.0025	− − −
IID/HG-gain/25	Both	$+0.0001 \pm 0.0001$	+ + −	-0.0005 ± 0.0003	− − −
IID/HG-gain/25	Ref.	$+0.0001 \pm 0.0002$	+ − +	$+0.0000 \pm 0.0006$	− + −
IID/CPSS/10	Both	$+0.0007 \pm 0.0002$	+ + +	-0.0004 ± 0.0012	− − +
IID/CPSS/10	Ref.	$+0.0001 \pm 0.0003$	− − +	-0.0010 ± 0.0005	− − −
IID/CPSS/25	Both	-0.0002 ± 0.0006	+ − +	-0.0015 ± 0.0012	− − −
IID/CPSS/25	Ref.	$+0.0002 \pm 0.0005$	− − +	-0.0021 ± 0.0008	− − −

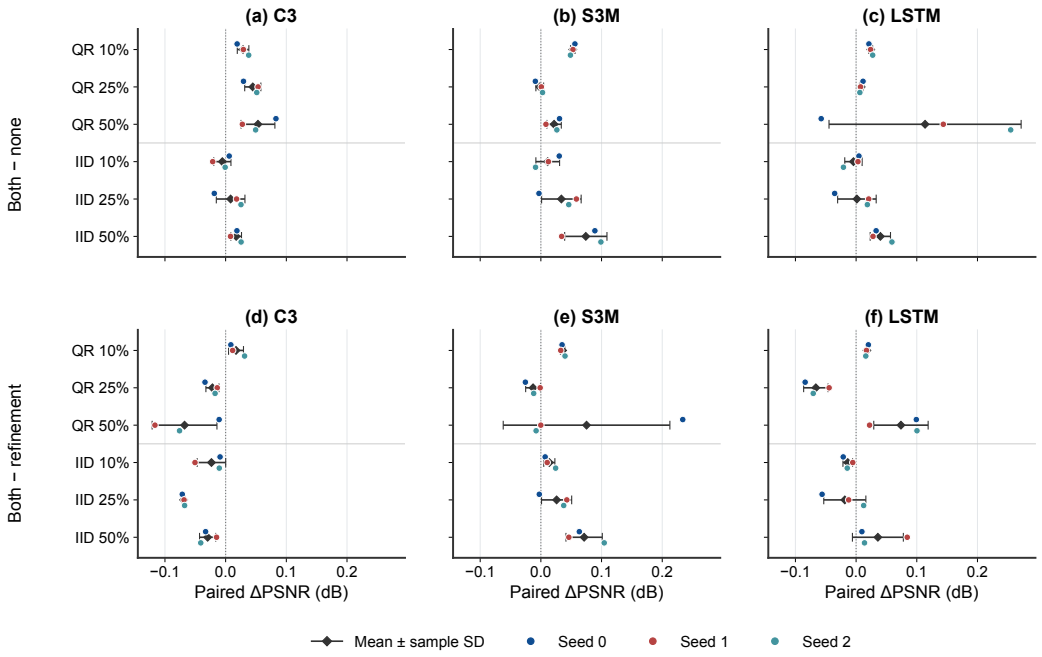


Figure 11: All 36 paired module/location PSNR contrasts at 13,000 updates on HCP. Top: both branches minus no feedback. Bottom: both branches minus refinement only. Positive values favor both-branch feedback. Each panel retains all six operator/rate settings and all three training seeds; diamonds and bars give mean ± sample SD of the same-seed differences. SD is not a confidence interval. These are selected validation results, not held-out significance tests.

P Frozen short-history expressivity and completion diagnostics

Fixed diagnostic. The protocol was fixed before reading diagnostic outputs, after the original model results were known. This is posthoc mechanism analysis, not a new training experiment. Each seed uses 2,048 newly generated Bernoulli–Gaussian signals from stream $760000 + s$, all 16 layers and the same five frozen models. Parent, current and FIR are the original MUSA-trained models; two-time and nonlinear controls are CUDA-trained. These diagnostics were not rerun on the new CUDA FIR models. True support is used only for offline error accounting. At an unmodified FIR state, form $R_k = [r_k, r_{k-1}]$, retaining available columns at startup, and compute

$$d_k^{\parallel} = R_k R_k^{\dagger} d_k, \qquad d_k^{\perp} = d_k - d_k^{\parallel}.$$

Projection is per signal, using FP64 SVD with singular-value threshold $500\epsilon_{64}\sigma_{\max}$. The reported residual fraction aggregates $\|d_k^{\perp}\|^2/\|d_k\|^2$ by summing numerator and denominator over signals before division. It tests unrestricted sample-specific scalar combinations of two increments, not arbitrary coordinatewise nonlinear current maps. It still requires the full FIR correction and cannot serve as a deployable short-history comparator.

We execute the original and projected corrections from the same backbone state and observations through the actual completion. The instrumented original completion matches the native implementation bitwise on all checked batches; the first two layers have projection residual fractions below 10^{-10}. From layer three onward, 39/42 seed–layer cells have higher target error after same-state projection. Three cells do not, so nonzero projection residual is not universally beneficial information. Full projected rollouts rebuild history from intervened states at every step. They are distinct from the same-state comparisons.

Error accounting. For known support S, we report the exact split $\|\widehat{x} - x^{\star}\|^2 = \|(\widehat{x} - x^{\star})_S\|^2 + \|\widehat{x}_{S^c}\|^2$. The first includes missed recovery and amplitude error; it is not pure amplitude error. At

the final FIR output, its fraction of total error is 0.8493, 0.7258 and 0.7942 across the three seeds. The machine-readable analysis retains every method, layer and seed. The FIR/nonlinear ordering

Table 55: Fresh diagnostic batch: final NMSE (dB), all three frozen seeds. This batch is separate from the 10,000-signal primary test. Bold marks the lowest value in each metric column; mixed seed ordering is retained.

Method	Seed 0	Seed 1	Seed 2	Mean ± SD
Parent	-46.14	-45.40	-45.93	-45.82 ± 0.38
Current	-63.17	-61.26	-57.75	-60.73 ± 2.75
FIR	**−90.68**	-71.23	**−90.53**	$\mathbf{-84.15 \pm 11.19}$
Two-time	-80.88	-62.63	-80.04	-74.52 ± 10.31
Nonlinear	-74.86	**−74.07**	-75.63	-74.86 ± 0.78

reverses for seed 1 in this new batch. Original test results are not replaced; the fresh-signal result limits an all-batch interpretation of the original all-seed ordering.

Finite-depth replacement propagation. Let z_k include the complete backbone and all feedback-buffer state, so both the original and replaced rollouts are deterministic one-step maps. Let T_k be a fixed replacement map and define $\epsilon_k = \|z_{k+1}^F - T_k(z_k^F)\|$ on the original trajectory. If $\|T_k(z_k^F) - T_k(z_k^C)\| \le L_k \|z_k^F - z_k^C\|$ for the compared states, the triangle inequality gives

$$\|z_K^F - z_K^C\| \le \sum_{k=0}^{K-1} \epsilon_k \prod_{t=k+1}^{K-1} L_t$$

under identical initialization. Induction proves the statement; empty products equal one. Contraction is not required. This standard finite-depth bound separates local completion differences from their later propagation. No L_k certificate is measured here, and projection near a rank change need not admit a uniform Lipschitz bound. The formula is therefore an interpretation framework, not a numerical certificate for these checkpoints or a comparison of separately retrained architectures. Backbone-output comparisons above observe the iterate component, not this whole augmented state norm.

Q Synthetic signal distributions and training histories

These are descriptive analyses of the existing nine trained checkpoints and saved records, not additional training or test observations. For each seed we compute $q_i = \|\hat{x}_i - x_i^\star\|^2 / \|x_i^\star\|^2$. Quantiles use linear interpolation in the linear ratios; no log floor is imposed. The primary statistic remains Eq. (27). FIR has lower per-signal error than current-input correction in 100.00%, 99.99%, 100.00% of the test signals for seeds 0, 1 and 2, respectively. These are conditional sample fractions, not independent training repetitions. The advantage therefore extends beyond a few repaired high-error samples in these evaluated models.

Table 56: Per-signal linear NMSE quantiles, separately for every trained seed.

Method	Seed	Median	90th	99th	99.9th
parent	0	8.592e-06	2.841e-05	2.070e-04	1.960e-03
current	0	1.415e-07	5.368e-07	5.355e-06	4.369e-05
fir	0	**4.111e-10**	**1.281e-09**	**6.205e-09**	**2.744e-08**
parent	1	8.573e-06	3.073e-05	2.326e-04	1.662e-03
current	1	3.102e-08	1.247e-07	8.471e-07	4.288e-06
fir	1	**2.896e-10**	**1.347e-09**	**9.408e-09**	**3.996e-08**
parent	2	8.960e-06	3.098e-05	2.203e-04	1.287e-03
current	2	4.822e-07	1.349e-06	9.038e-06	1.088e-04
fir	2	**2.964e-10**	**1.202e-09**	**7.897e-09**	**3.229e-08**

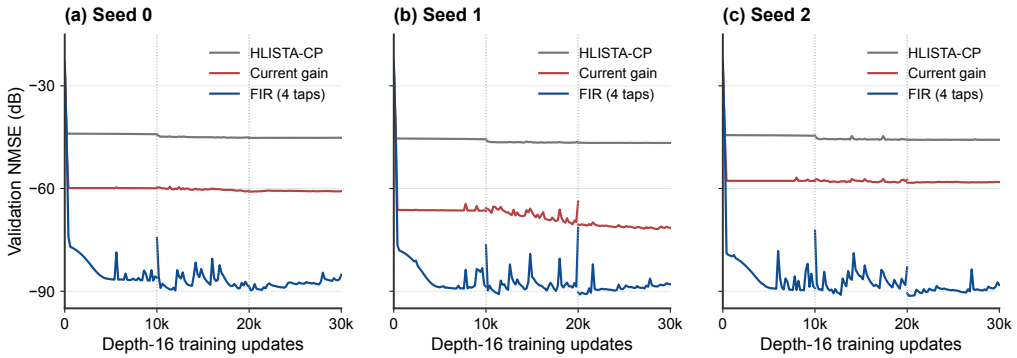


Figure 12: Saved depth-16 validation NMSE, split into optimizer phases at 10,000 and 20,000 updates. Phase-best restoration and optimizer resets occur at boundaries. These are training histories at fixed depth, not inference convergence or separately optimized shallow models. All-seed test CDFs follow below.

At the first logged depth-16 update, FIR has 0.72–0.94 dB lower validation NMSE than the current control. At the final logged update the gap is 16.49–30.12 dB across the three seeds. This is separation during depth-16 training, not inference acceleration.

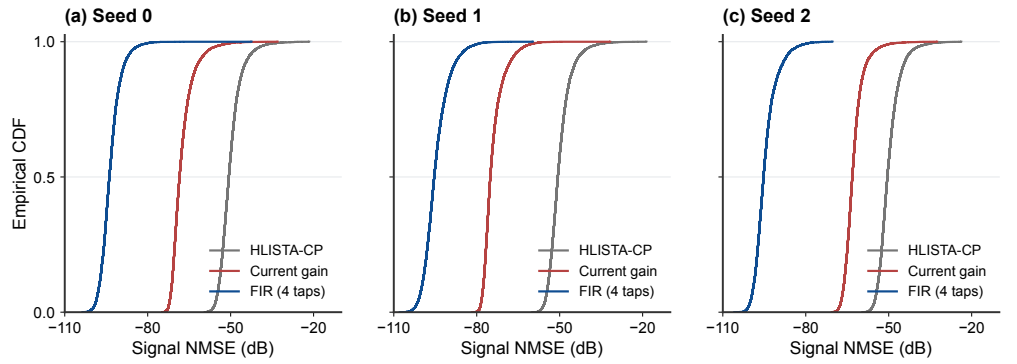


Figure 13: Original all-seed empirical CDFs on 10,000 test signals per seed. The 99.99% statement concerns FIR versus the original linear current control, not the added controls.

R RETRAINED CURRENT-INPUT AND HISTORY CONTROLS

Table 57: Retrained MR25 HLISTA-CP controls: validation50 image metrics, mean ± sample SD over three seeds, 26,000 updates. Bold marks the highest mean per column. All feedback arms use four shared parameters.

Correction	PSNR (dB)	SSIM
Parent	29.0552 ± 0.0179	0.844393 ± 0.000771
Current, refinement	29.0784 ± 0.0138	0.845295 ± 0.000401
Current, both	$\mathbf{29.0812 \pm 0.0116}$	0.845186 ± 0.000552
FIR, refinement	29.0650 ± 0.0182	0.844518 ± 0.000723
FIR, both	29.0751 ± 0.0212	$\mathbf{0.845303 \pm 0.000692}$

The frozen plan uses one six-phase HLISTA-CP backbone, the released MR25 Gaussian sensing matrix and dictionary, and the same BSD500 400/50/50 within-corpus partition. Five arms are trained from scratch with seeds 0/1/2 for 26,000 Adam updates (batch64, learning rate 10^{-4}, symmetry weight 0.01, gradient clip1, validation every 500 steps). All arms share backbone initialization and sampled patches for a seed. Feedback arms have four zero-initialized trainable taps, shared across phases, with absolute-sum cap 0.5. The current control applies the sum of available taps to the current increment; FIR applies them to the corresponding history. This matches gain range, startup availability and parameterization, not an arbitrary nonlinear function class. Parent has 880,030 parameters; each correction arm has 880,034.

Validation patches select checkpoints; image metrics below use validation50. Neither test50 nor locked COCO was accessed during training and checkpoint selection. The subsequent frozen COCO evaluation is reported in Appendix T. This is development evidence, not independent confirmation. All 15 runs and checkpoints were downloaded; per-image means, image IDs and shared split/matrix identities were verified.

Table 58: Same-seed contrasts for retrained controls. SD is across three training seeds, not a confidence interval.

Contrast	ΔPSNR	ΔSSIM
FIR, refinement − Current, refinement	-0.0134 ± 0.0100	-0.000777 ± 0.000817
FIR, both − Current, both	-0.0061 ± 0.0327	0.000118 ± 0.000635
FIR, both − FIR, refinement	0.0101 ± 0.0365	0.000786 ± 0.001357
Current, both − Current, refinement	0.0028 ± 0.0046	-0.000109 ± 0.000524

Table 59: Additional absolute validation metrics, mean ± sample SD over three seeds.

Correction	Linear NMSE	NMSE (dB)
Parent	0.0088397 ± 0.0000330	-22.4651 ± 0.0179
Current, refinement	$\mathbf{0.0088027 \pm 0.0000297}$	-22.4884 ± 0.0138
Current, both	0.0088116 ± 0.0000193	$\mathbf{-22.4912 \pm 0.0116}$
FIR, refinement	0.0088176 ± 0.0000363	-22.4749 ± 0.0182
FIR, both	0.0088185 ± 0.0000251	-22.4850 ± 0.0212

12 of 15 best checkpoints occur at 26,000 updates; the remainder are at 25,500. The study compares matched finite budgets, not demonstrated convergence. FIR-minus-current refinement PSNR is negative for all three seeds; the both-branch contrast has one positive and two negative seed differences. These observations provide no stable history-specific gain in this setting, and do not establish statistical equivalence or a universal disadvantage of history. The FIR location contrast remains mixed across seeds.

Timing scope. MTT S5000 model-only timing uses batch 64, 20 warmups and 100 synchronized forwards. Per-run medians span 6.388–6.648 ms (parent), 7.321–9.509 (current refinement), 7.358–7.467 (current both), 7.632–7.810 (FIR refinement) and 7.587–7.736 (FIR both). GPU4/7 were

initially idle, but continuous isolation was not verified. Preprocessing, transfer and stitching are excluded. The subsequent RAM-observation reconstruction measurements and finite-prefix comparison are in Appendix V.

S CURRENT-INPUT APPROXIMATION OF TRAINED FIR CORRECTIONS

Train-only calibration with matched sharing. For each of the six trained FIR checkpoints, we sample 4,096 calibration patches from the existing 400-image BSD training split using seed 20260907. All backbone weights remain fixed. Let $r_{k,i}$ be the current increment and $d_{k,i}$ the actual FIR correction along its original trajectory. We fit four shared coefficients, not six freely chosen stage gains:

$$\min_{\|c\|_1 \le 0.5} \sum_{k,i} \|d_{k,i} - a_k(c) r_{k,i}\|_2^2, \qquad a_k(c) = \sum_{j=0}^{\min(k,3)} c_j, \quad k = 0, \ldots, 5.$$

This preserves the four-tap sharing, gain budget and startup schedule of the matched current-input control. A projected-gradient solution converges for every checkpoint; all six solutions are interior to the constraint and agree within 10^{-8} with an independently solved normal equation. No validation or COCO target is used to fit the coefficients.

Correction approximation and actual replacement. On validation patches (64 per image, fixed seed 20260908), we compute $\eta_k = \sum_i \|d_{k,i} - a_k r_{k,i}\|^2 / \sum_i \|d_{k,i}\|^2$ on the original FIR trajectory. Zero denominators would be reported separately; none occur. Stage one admits only the current tap, so its near-zero residual is expected and does not show that later history is redundant. We then remove the history buffer and use the fitted $a_k r_k$ from initialization through all six stages. These complete replacement trajectories are evaluated on validation50; we do not reuse the original FIR trajectory to infer reconstruction quality.

Table 60: Full-trajectory replacement on BSD validation50. Mean ± sample SD over three seeds; differences are replacement minus original FIR. No equivalence margin or external deployment validation is claimed.

Location	Replacement PSNR	ΔPSNR	ΔSSIM
Refinement	$\mathbf{29.0648 \pm 0.0181}$	-0.00017 ± 0.00044	0.000008 ± 0.000015
Both branches	29.0475 ± 0.0234	-0.02757 ± 0.00304	-0.001174 ± 0.000121

Table 61: Validation normalized correction residual η_k after train-only calibration, mean ± sample SD across three seeds. Lower means closer approximation, not necessarily better reconstruction.

Stage	Refinement	Both branches
1	0.0000 ± 0.0000	0.0000 ± 0.0000
2	0.2299 ± 0.1869	0.2666 ± 0.0140
3	0.1123 ± 0.0808	0.1923 ± 0.0328
4	0.0782 ± 0.0646	0.1225 ± 0.0124
5	0.1118 ± 0.0860	0.1506 ± 0.0117
6	0.1437 ± 0.1030	0.2450 ± 0.0415

The refinement-only replacement changes PSNR by -0.00017 ± 0.00044 dB; its three changes have mixed signs. Both-branch replacement loses 0.02757 ± 0.00304 dB, with losses in every seed. Thus the two injection locations differ in sensitivity to this constrained history-free approximation. These results do not establish mathematical equivalence, universal replaceability or a new compression algorithm. Both-branch FIR depends on its trained historical corrections, yet the separately retrained current-input control has higher mean PSNR. Dependence of one trained solution and added value over a retrained alternative are different questions. This is a post-hoc diagnostic on development images; the opened COCO test is not reused to select or validate the replacement.

T FROZEN CROSS-DATASET CONFIRMATION ON COCO5000

Scope and freezing. We evaluate the 15 BSD-trained MR25 six-stage checkpoints from Appendix R on all 5,000 official COCO val2017 images. This is cross-dataset evaluation without adaptation, not evaluation of the 252 COCO-trained models. All five arms and seeds 0, 1, 2 enter the test. Model hashes, image identities, the released sensing matrix and dictionary, checkpoint selection and reporting rules were frozen before scoring. There is no retraining, calibration, noise sweep, subset selection or test-based checkpoint replacement. The primary metric is PSNR under clean observations; SSIM and both NMSE representations are retained as auxiliary metrics.

Protocol and provenance. Grayscale conversion, nonoverlapping 16×16 patches, padding, mean removal/restoration and clipping follow the BSD runner. Patch batch is 64; measurements are generated once per image and shared by all 15 models. Two accelerators process disjoint, interleaved halves of the sorted image list; each half uses all models. Both jobs exit successfully: 37,500 metric rows each, 75,000 total, with all expected model–image keys present exactly once and all four metrics finite. A development-only smoke comparison checks all four metrics against the original evaluator within 10^{-5} across 45 model–image evaluations. Images are averaged within each seed before reporting mean $\pm$ sample SD over the three training seeds. Image count does not increase the number of training replicates.

The COCO registration, paper ledger and inspected COCO/AMP audit records reserve official val2017 for final testing and use train2017-derived images for development. This is a records-based provenance audit, not proof that undocumented access never occurred elsewhere. No exact file-content overlap with any of the 500 BSD images is found; near-duplicate or semantic overlap has not been audited. These limits qualify the cross-dataset interpretation.

Table 62: Clean MR25 cross-dataset PSNR and SSIM on the full COCO5000 test, mean $\pm$ sample SD across three seeds. Bold marks the highest mean, not statistical significance.

Method	PSNR (dB)	SSIM
Parent	28.1269 ± 0.0248	0.817129 ± 0.000753
Current, refinement	28.1475 ± 0.0152	0.817945 ± 0.000378
Current, both	$\mathbf{28.1585 \pm 0.0114}$	$\mathbf{0.818032 \pm 0.000253}$
FIR, refinement	28.1444 ± 0.0096	0.817542 ± 0.000595
FIR, both	28.1438 ± 0.0330	0.818009 ± 0.001068

Table 63: All prespecified same-seed PSNR contrasts on COCO5000. The signs are ordered by seeds 0, 1, 2; SD is not a confidence interval.

Contrast	ΔPSNR (dB)	Seed signs
Current, both − Parent	0.0316 ± 0.0337	$-, +, +$
FIR, both − Parent	0.0168 ± 0.0150	$+, +, +$
Current, refinement − Parent	0.0206 ± 0.0317	$-, +, +$
FIR, refinement − Parent	0.0175 ± 0.0272	$-, +, +$
FIR, both − Current, both	-0.0148 ± 0.0439	$+, -, -$
FIR, refinement − Current, refinement	-0.0031 ± 0.0058	$-, -, +$

FIR-both improves on the parent in every seed, but its mean advantage is only 0.0168 ± 0.0150 dB. Current-both has the highest mean PSNR and SSIM; its parent contrast is 0.0316 ± 0.0337 dB with mixed signs. FIR-minus-current is -0.0148 ± 0.0439 dB for both-branch and -0.0031 ± 0.0058 dB for refinement-only, with mixed signs in both comparisons. The external panel therefore preserves the distinction between improving a parent and establishing an additional benefit from history. It does not prove equivalence or a universal disadvantage of FIR.

Additional metrics and adoption boundary. The fixed-size latency probes in Appendix V are separate from the original-size COCO quality evaluation; this test does not newly measure COCO end-to-end throughput. In the measured reference implementation, current-both costs about 15.6%

Table 64: COCO5000 auxiliary reconstruction errors, mean ± sample SD across three seeds. Bold marks the lowest mean in each column.

Method	Linear NMSE	Per-image NMSE (dB)
Parent	0.0105307 ± 0.0000465	-22.0718 ± 0.0248
Current, refinement	0.0105019 ± 0.0000399	-22.0924 ± 0.0152
Current, both	$\mathbf{0.0104759 \pm 0.0000383}$	$\mathbf{-22.1034 \pm 0.0114}$
FIR, refinement	0.0104947 ± 0.0000130	-22.0893 ± 0.0096
FIR, both	0.0105157 ± 0.0000540	-22.0886 ± 0.0330

more than the parent, while FIR-both costs about 20.5% more. Thus the current-input control is also a quality–speed trade-off, not a cost-free improvement. The total FIR–parent difference must not be attributed entirely to history maintenance: the current-input implementation already incurs most of that difference. Neither these measurements nor the external quality results isolate the runtime of individual mechanisms. Current-both's higher mean PSNR and lower measured latency than FIR-both support a limited implementation-level comparison, not a proof of statistically inferior FIR reconstruction.

The completed clean external confirmation closes its registered evaluation route. Subsequent development-only mechanism and inference diagnostics (Appendices S–U) do not reuse this test for selection. We retain all outcomes. No claim of FIR replaceability, noise robustness or verified convergence is inferred from this experiment.

U Uniform inference simplification and repeated timing

Common transformation and output checks. All 15 checkpoints are evaluated both with the original implementation and with an inference-only simplification. The symmetry-loss convolution branch is omitted for every arm. Current-input controls additionally replace their unused historical tensor list by an integer startup counter; the four shared coefficients, active-tap schedule and numerical correction are unchanged. FIR retains its actual historical tensors. The optimized reconstruction outputs agree exactly (zero tolerances) with the original on random measurement batches of sizes 1, 64 and 97 for every checkpoint (45 checks). The omitted training loss is intentionally not preserved; this is not a replacement training implementation. There is no retraining or test-based selection.

Matched repeated measurement. The same MTT S5000 GPU4, FP32, RAM-observation boundary, patch batch64, image sizes $256^2/512^2$, 20 warmups and five rounds of 30 measurements are used. Methods are rotated by round and original/optimized order alternates. All 300 blocks and 9,000 timed reconstructions complete. Latency is summarized by each seed's median of 150 samples, then mean ± sample SD over the three seeds. This is a new paired timing run, not a modification of old measurements. No training is scheduled on GPU4; host CPU is shared with the mechanism diagnostic and a brief environment check. Continuous third-party monitoring is not available. Absolute allocated memory includes ten resident original/optimized models per seed; workspace below reports the maximum increment above the pre-call allocation and is not comparable to the earlier absolute-memory table.

Table 65: Repeated full-depth 512^2 reconstruction latency (ms). Negative relative changes mean faster optimized inference. Relative changes are computed within each seed, then averaged; all methods are retained. Bold marks the lowest mean latency in each implementation column.

Method	Original	Optimized	Change (%)
Parent	$\mathbf{115.60 \pm 0.25}$	$\mathbf{104.20 \pm 0.10}$	-9.86 ± 0.22
Current, refinement	131.63 ± 0.43	119.90 ± 0.50	-8.91 ± 0.34
Current, both	131.92 ± 0.27	119.90 ± 0.18	-9.11 ± 0.28
FIR, refinement	137.18 ± 0.54	125.86 ± 0.48	-8.25 ± 0.43
FIR, both	137.48 ± 0.33	125.88 ± 0.43	-8.43 ± 0.12

Table 66: Companion 256^2 latency (ms) and maximum incremental device workspace (MiB) at 512^2. Workspace entries are maxima, not three-seed means.

Method	256^2 latency		512^2 workspace	
	Original	Optimized	Original	Optimized
Parent	$\mathbf{29.30 \pm 0.10}$	$\mathbf{26.58 \pm 0.25}$	**24.79**	**14.71**
Current, refinement	34.67 ± 2.41	31.64 ± 2.39	25.54	14.96
Current, both	34.89 ± 2.80	31.26 ± 1.65	25.54	14.96
FIR, refinement	34.69 ± 0.25	34.54 ± 4.51	25.54	15.46
FIR, both	34.72 ± 0.08	31.88 ± 0.07	25.54	15.46

These timings separate removable implementation work from the measured cost of each remaining complete solver. They do not isolate history maintenance as the cause of every FIR–parent difference. Reconstruction quality is unchanged by the tested inference transformation; it does not produce a new image-accuracy result or establish the absent stable history-specific gain in that image configuration. Prior reference-code measurements remain in Appendix V.

Separate synthetic evidence. The formerly pending synthetic comparison is complete. A corrected matched nine-model protocol, including a gain-range-matched current control, is reported in Section 5.2 and Appendix K. It is not a faithful rerun of the legacy TensorFlow optimization trajectory. Its positive final-NMSE result and the image quality–cost results answer different questions and are not combined into a cross-task efficiency claim.

V Measured reconstruction cost and finite-prefix quality

Frozen evaluation and verification. All 15 checkpoints from Appendix R enter this panel; no weights are retrained and no seed is selected by outcome. Five methods, three seeds and six prefixes yield 90 validation50 evaluations (4,500 model–image reconstructions). Every stage-six PSNR mean agrees with the original full-model result within 10^{-4} dB. All image-level PSNR, SSIM, linear NMSE and log NMSE means were independently recomputed. Prefixes retain the trained six-stage dictionary routing; they are not separately trained shallow networks. The stage-six forward was checked against the original implementation before evaluation. External confirmation is reported separately in Appendix T.

Timing boundary and hardware. On an MTT S5000 GPU, FP32 timing starts from synthetic RAM-resident measurement arrays and the protocol's known patch means, and ends with a clipped RAM-resident reconstructed image. It includes transfer, patch batching, stitching and mean restoration, but excludes acquisition, disk I/O, model loading and metrics. Image batch is one, patch batch is 64, and the two output sizes are 256^2 and 512^2. For every seed, size and prefix, five rotated method-order rounds use 20 warmups and 30 measured reconstructions per method, with MUSA synchronization. All 900 blocks and 27,000 measurements completed successfully. GPU4 had no other process at launch and no concurrent job was scheduled there by us; continuous external process monitoring was not recorded. GPU7 quality evaluation shared the host CPU during the first part of timing.

The per-seed statistic is the median of 150 measurements; Table 67 reports mean $\pm$ sample SD of these three medians, not the SD of 450 independent training runs. Within-seed timing IQRs are reported separately. Peak memory is the maximum allocated-device high-water mark across measured blocks, not total device occupancy or reserved allocator memory. These are reference-implementation costs: the symmetry path is still executed, and the matched current-input control retains a history list even though it uses only the current increment. The measurements do not establish the cost of an optimized history-free deployment. A four-tensor history buffer of 64×512 FP32 entries occupies 0.5 MiB within a reconstruction; feedback state is reset between calls, not persisted across images.

At 512^2, FIR refinement/both takes 138.54/138.55 ms, versus 132.55/132.93 ms for the corresponding current-input controls and 114.99 ms for the parent. Four additional parameters do not imply

Table 67: Full six-stage validation quality and RAM-observation reconstruction cost. Quality and latency are mean ± sample SD over three seeds. Bold marks the best mean in each quality, latency or memory column. Memory is a measured maximum, not mean ± SD.

Method	PSNR (dB)	256^2 (ms)	512^2 (ms)	Peak MiB
Parent	29.055 ± 0.018	$\mathbf{29.21 \pm 0.12}$	$\mathbf{114.99 \pm 0.68}$	**65.47**
Current, refinement	29.078 ± 0.014	33.54 ± 0.22	132.55 ± 0.84	70.29
Current, both	$\mathbf{29.081 \pm 0.012}$	33.49 ± 0.14	132.93 ± 0.63	70.29
FIR, refinement	29.065 ± 0.018	34.94 ± 0.08	138.54 ± 0.99	70.29
FIR, both	29.075 ± 0.021	34.82 ± 0.12	138.55 ± 0.73	70.29

negligible runtime: FIR adds approximately 20.5% relative to the parent here. The full-depth FIR configurations have lower mean PSNR and higher mean latency than the corresponding current-input controls; this is an observed PSNR–latency comparison, not a statistical equivalence claim or an SSIM dominance claim.

Finite-depth quality and information–location interaction. Table 68 retains every prefix, rather than choosing an exit only for FIR. FIR-both has the highest PSNR mean at stages two and five, while the parent leads at stages three and four. Intermediate gains therefore do not establish a uniform acceleration benefit. In particular, five-stage FIR-both reaches 28.377 dB at 116.26 ms for 512^2; the six-stage parent reaches 29.055 dB at 114.99 ms. A same-stage gain can thus disappear when compared at measured cost. Quality is evaluated on original validation image sizes; the timing sizes are separate fixed probes, not resized quality-test images.

The paired information–location interaction is

$$I_s = (P_{\mathrm{FIR,both},s} - P_{\mathrm{current,both},s}) - (P_{\mathrm{FIR,ref},s} - P_{\mathrm{current,ref},s}).$$

Its seed values are $0.02912, -0.03101, 0.02378$ dB, giving 0.00730 ± 0.03329 dB. A positive mean indicates a smaller FIR disadvantage under both-branch injection; the mixed signs do not demonstrate FIR superiority.

Within-seed latency variability. The following table reports IQR in milliseconds for each seed's 150 full-depth measurements. These quantify repeated-timing variability, not image or training uncertainty. Raw per-round latencies and all intermediate-stage memory measurements are retained in the evaluation archive.

Remaining evidence boundary. The subsequent shared-parameter current-input diagnostic is reported in Appendix S. The subsequent clean COCO5000 confirmation is in Appendix T. No claim of FIR replaceability or noise robustness is inferred from these panels.

W FROZEN-MODEL PER-CHANNEL RGB EXTENSION

Chronology and scope. This supplementary evaluation was specified after inspecting the grayscale results. It reuses all 15 frozen BSD-trained MR25 six-stage checkpoints (five methods, three seeds) and the same 5,000 COCO val2017 images as Appendix T. No weights, checkpoints, hyperparameters or seeds were selected using RGB scores. All $5 \times 3 \times 5000 = 75{,}000$ model–image outcomes are retained. These are additional color-protocol measurements, not a second external dataset, additional independent training replicates or evidence of learned cross-channel relationships. Channel-wise reuse is available to all five methods, not unique to DASH.

Acquisition and independent reconstruction state. Pillow decodes each original image to RGB at its stored native resolution, without resizing. Each channel is processed separately using the same 16×16 patch partition, zero boundary padding, per-patch mean removal and normalization by 255 as the grayscale runner. The frozen 64×256 sensing matrix is identical across channels and methods; observations for each channel are computed once and shared across all models. The nominal rate is $3 \cdot 64/(3 \cdot 256) = 25\%$. Per-channel patch means are additional side information, as in the grayscale protocol; their transmission is not included in this nominal rate. Each model forward call initializes

Table 68: PSNR (dB) for prefixes of full-depth trained models on validation50, mean ± sample SD over three seeds. Bold marks the highest PSNR and lowest latency at each stage.

Stage	Method	PSNR (dB)	512^2 latency (ms)
1	Parent	23.246 ± 0.073	$\mathbf{23.40 \pm 0.15}$
1	Current, refinement	23.273 ± 0.040	26.41 ± 0.26
1	Current, both	22.932 ± 0.055	26.67 ± 0.50
1	FIR, refinement	$\mathbf{23.300 \pm 0.041}$	26.50 ± 0.17
1	FIR, both	23.244 ± 0.093	26.53 ± 0.14
2	Parent	24.586 ± 0.046	$\mathbf{41.85 \pm 0.20}$
2	Current, refinement	24.526 ± 0.102	47.89 ± 0.06
2	Current, both	24.429 ± 0.015	47.80 ± 0.20
2	FIR, refinement	24.394 ± 0.110	48.63 ± 0.25
2	FIR, both	$\mathbf{24.675 \pm 0.090}$	48.89 ± 0.51
3	Parent	$\mathbf{25.860 \pm 0.053}$	$\mathbf{60.36 \pm 0.32}$
3	Current, refinement	25.695 ± 0.107	69.06 ± 0.23
3	Current, both	25.503 ± 0.041	69.26 ± 0.35
3	FIR, refinement	25.700 ± 0.071	70.96 ± 0.43
3	FIR, both	25.826 ± 0.040	70.96 ± 0.25
4	Parent	$\mathbf{27.169 \pm 0.146}$	$\mathbf{78.62 \pm 0.45}$
4	Current, refinement	27.064 ± 0.037	90.23 ± 0.79
4	Current, both	26.699 ± 0.036	90.20 ± 0.50
4	FIR, refinement	27.150 ± 0.150	93.17 ± 0.42
4	FIR, both	27.150 ± 0.021	93.28 ± 0.40
5	Parent	28.348 ± 0.025	$\mathbf{97.49 \pm 1.53}$
5	Current, refinement	28.201 ± 0.102	111.47 ± 0.89
5	Current, both	28.113 ± 0.075	112.15 ± 1.59
5	FIR, refinement	28.259 ± 0.087	116.35 ± 0.92
5	FIR, both	$\mathbf{28.377 \pm 0.035}$	116.26 ± 0.82
6	Parent	29.055 ± 0.018	$\mathbf{114.99 \pm 0.68}$
6	Current, refinement	29.078 ± 0.014	132.55 ± 0.84
6	Current, both	$\mathbf{29.081 \pm 0.012}$	132.93 ± 0.63
6	FIR, refinement	29.065 ± 0.018	138.54 ± 0.99
6	FIR, both	29.075 ± 0.021	138.55 ± 0.73

Table 69: Within-seed timing IQRs (ms), ordered as seeds 0, 1, 2.

Method	256^2	512^2
Parent	0.72, 0.35, 0.32	1.50, 1.21, 0.86
Current, refinement	0.69, 0.48, 0.60	1.14, 1.51, 1.68
Current, both	2.51, 0.32, 0.40	2.91, 1.80, 2.51
FIR, refinement	0.50, 0.47, 0.51	1.75, 2.60, 1.79
FIR, both	0.49, 0.47, 0.28	1.27, 1.27, 1.99

both the sparse code to zero and the feedback state to `None`. History is therefore reset for every independent patch batch and channel, never carried from R to G or B or between images. Batch size is 64. Predictions are rescaled, their own channel means restored, assembled, cropped to the original extent, clipped to $[0, 255]$, then stacked as RGB. No original chroma is copied into the output.

Metrics and paired statistics. For an image of size $H \times W$, we compute

$$\mathrm{MSE}_{\mathrm{RGB}} = \frac{1}{3HW} \sum_{c \in \{R,G,B\}} \sum_{i,j} (\hat{x}_{ijc} - x_{ijc})^2, \qquad \mathrm{PSNR}_{\mathrm{RGB}} = 10 \log_{10} \frac{255^2}{\mathrm{MSE}_{\mathrm{RGB}}}.$$

Squared errors are accumulated in float64 from clipped floating-point reconstructions; this is not the average of three channel PSNR values. RGB SSIM is the arithmetic mean of three channel-

wise SSIM values, each using the original valid 11×11 Gaussian window (standard deviation 1.5), dynamic range 255 and constants $(0.01 \cdot 255)^2, (0.03 \cdot 255)^2$. No quantization is applied before metric calculation. For each seed, we average the 5,000 image scores with equal image weight; tables then report the mean and sample SD of three such seed means. Each paired contrast subtracts matched seed means before computing its sample SD. We do not infer paired SD from marginal SDs or reuse grayscale SDs.

Verification and cost boundary. All five evaluation workers pass the RGB wrapper's repeated-grayscale identity check against the original evaluator. The frozen checkpoint and input identities are checked, all workers exit normally, and local verification confirms unique model–image keys, full image coverage per model, finite metrics, PSNR–MSE consistency and recomputed seed means. The longest five-way shard took 7,265 seconds (approximately two hours), which is an execution record, not a method-level efficiency comparison. Recorded native-size reconstruction-section timings are observational; we do not combine them or the earlier grayscale fixed-size latency probe with this table to claim an RGB quality–speed advantage.

Fixed qualitative examples. These eight IDs were selected in the earlier grayscale review, not for representative or favorable RGB performance. Of 5,000 images, 24 had positive FIR-both PSNR differences against both comparators in all three seeds; the eight largest mean FIR-minus-parent differences were retained. Each fixed 64×64 crop maximizes reference gradient energy on a 16-pixel grid, without using reconstruction errors. All panels show seed 0 with identical 8-bit display conversion and nearest-neighbor enlargement; metrics use floating-point reconstructions. No sharpening, generative processing or chroma replacement is applied. Full-image increments retain three-seed paired SD and both signs, including SSIM. Original error maps use a common 0–40 intensity scale and remain in the evidence materials.

Table 70: Frozen per-channel RGB reconstruction on all 5,000 COCO images. Mean ± sample SD across three seed-specific dataset means; higher is better. Bold indicates the highest mean, not statistical significance.

Method	RGB PSNR (dB)	RGB SSIM
HLISTA-CP	27.8503 ± 0.0241	0.809194 ± 0.000751
Current increment, refinement	27.8704 ± 0.0137	0.810037 ± 0.000335
Current increment, both	$\mathbf{27.8802 \pm 0.0102}$	$\mathbf{0.810076 \pm 0.000284}$
FIR, refinement	27.8671 ± 0.0086	0.809584 ± 0.000590
FIR, both	27.8655 ± 0.0326	0.810063 ± 0.001094

Table 71: Same-seed RGB differences, with PSNR and SSIM in separate panels. Positive favors FIR. All three signed seed differences and their sample SD are retained.

FIR minus comparator	Seed 0	Seed 1	Seed 2	Mean ± SD
RGB PSNR (dB)				
Refinement − parent	-0.0093	$+0.0425$	$+0.0171$	$+0.0168 \pm 0.0259$
Both − parent	$+0.0183$	-0.0014	$+0.0287$	$+0.0152 \pm 0.0153$
Refinement − matched current	-0.0017	-0.0092	$+0.0011$	-0.0033 ± 0.0053
Both − matched current	$+0.0234$	-0.0608	-0.0067	-0.0147 ± 0.0427
RGB SSIM				
Refinement − parent	-0.000380	$+0.001086$	$+0.000464$	$+0.000390 \pm 0.000736$
Both − parent	$+0.000891$	-0.000246	$+0.001960$	$+0.000869 \pm 0.001103$
Refinement − matched current	-0.000170	-0.000263	-0.000924	-0.000452 ± 0.000411
Both − matched current	$+0.001015$	-0.001128	$+0.000073$	-0.000013 ± 0.001074

Near-zero mean differences do not establish statistical equivalence.

Quality and seed variability across image evaluations. Figure 14 places the BSD validation, COCO grayscale and channel-wise RGB scores on their original metric axes. Rows use the same method order across panels; the error bars show variation across the three trained seeds. The corresponding same-seed contrasts are reported separately above and in Appendix T.

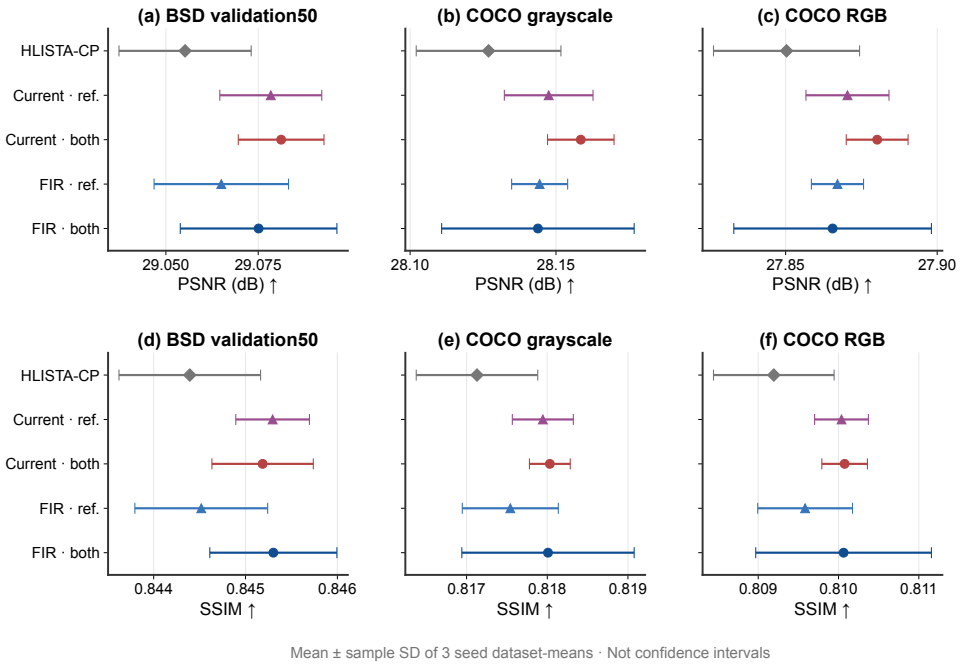
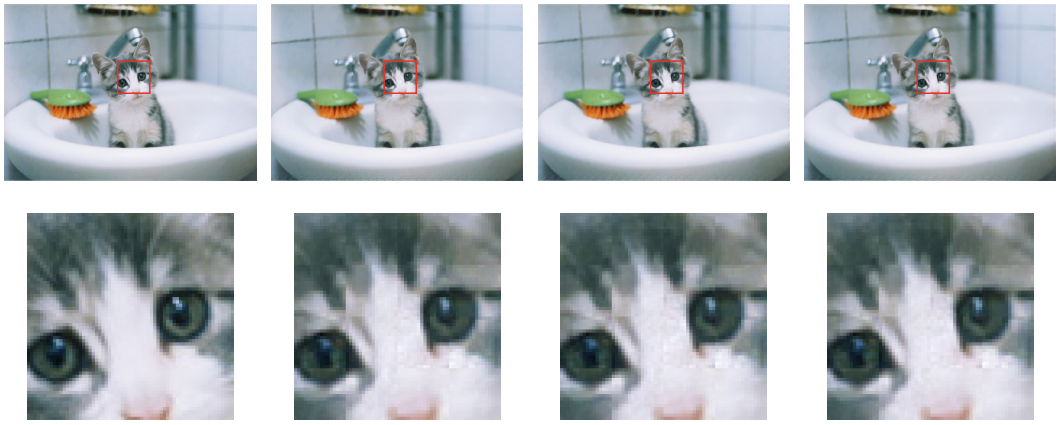


Figure 14: Image reconstruction quality for the five frozen BSD-trained MR25, six-stage methods. Panels (a–c) show PSNR in dB and (d–f) show SSIM; higher is better. Points are means of three seed-specific dataset means, and bars are their sample standard deviations, not confidence intervals or standard deviations of paired method differences. Ref. denotes refinement-only injection. COCO grayscale and RGB reuse the same 5,000 images and checkpoints; RGB reconstructs each channel separately. These panels summarize the existing evaluations without additional training.

COCO 000000209747 | MR25 | frozen RGB | seed 0

Reference	HLISTA-CP	Current increment	DASH (FIR)
PSNR (dB) / SSIM	38.100 / 0.9652	38.103 / 0.9652	38.251 / 0.9665



FIR - parent: whole-image paired mean +/- SD

PSNR +0.1909 +/- 0.2061

SSIM +0.001656 +/- 0.001575

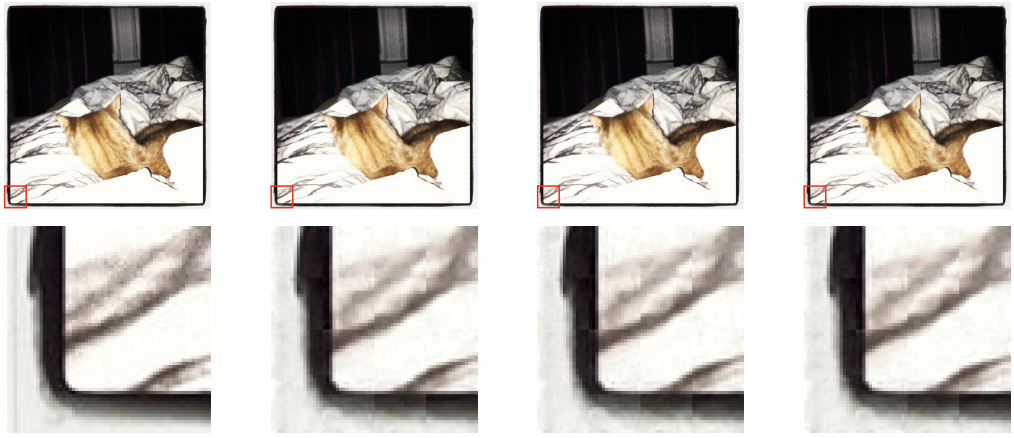
FIR - current: whole-image paired mean +/- SD

PSNR +0.1054 +/- 0.0479

SSIM +0.000679 +/- 0.000568

COCO 000000202445 | MR25 | frozen RGB | seed 0

Reference	HLISTA-CP	Current increment	DASH (FIR)
PSNR (dB) / SSIM	34.065 / 0.9333	34.085 / 0.9334	34.133 / 0.9339



FIR - parent: whole-image paired mean +/- SD

PSNR +0.1714 +/- 0.1913

SSIM +0.000950 +/- 0.001187

FIR - current: whole-image paired mean +/- SD

PSNR +0.0496 +/- 0.0025

SSIM +0.000177 +/- 0.000370

Figure 15: Frozen per-channel RGB examples, COCO 209747 and 202445. All methods share the displayed seed, original image and fixed 64×64 crop. Full-image PSNR/SSIM appear above the reconstructions; paired differences are mean $\pm$ sample SD over three seeds. Positive is better. Examples were selected in the earlier grayscale review; they do not represent full-set performance. Full error-map panels remain in the reproduction archive.

COCO 000000095786 | MR25 | frozen RGB | seed 0

Reference	HLISTA-CP	Current increment	DASH (FIR)
PSNR (dB) / SSIM	38.832 / 0.9420	38.824 / 0.9417	38.965 / 0.9431



FIR - parent: whole-image paired mean +/- SD

PSNR +0.1514 +/- 0.0831

SSIM +0.000881 +/- 0.000426

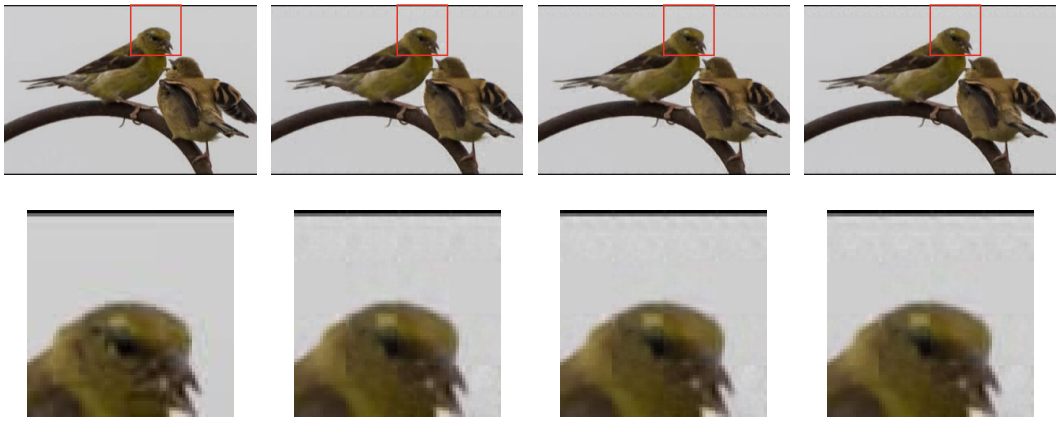
FIR - current: whole-image paired mean +/- SD

PSNR +0.0963 +/- 0.0426

SSIM +0.000675 +/- 0.000690

COCO 000000098261 | MR25 | frozen RGB | seed 0

Reference	HLISTA-CP	Current increment	DASH (FIR)
PSNR (dB) / SSIM	36.322 / 0.9592	36.314 / 0.9586	36.356 / 0.9582



FIR - parent: whole-image paired mean +/- SD

PSNR +0.1124 +/- 0.1340

SSIM +0.001067 +/- 0.002408

FIR - current: whole-image paired mean +/- SD

PSNR +0.0628 +/- 0.0182

SSIM +0.001904 +/- 0.002177

Figure 16: Frozen per-channel RGB examples, COCO 95786 and 98261. All methods share the displayed seed, original image and fixed 64×64 crop. Full-image PSNR/SSIM appear above the reconstructions; paired differences are mean $\pm$ sample SD over three seeds. Positive is better. Examples were selected in the earlier grayscale review; they do not represent full-set performance. Full error-map panels remain in the reproduction archive.

COCO 000000262048 | MR25 | frozen RGB | seed 0

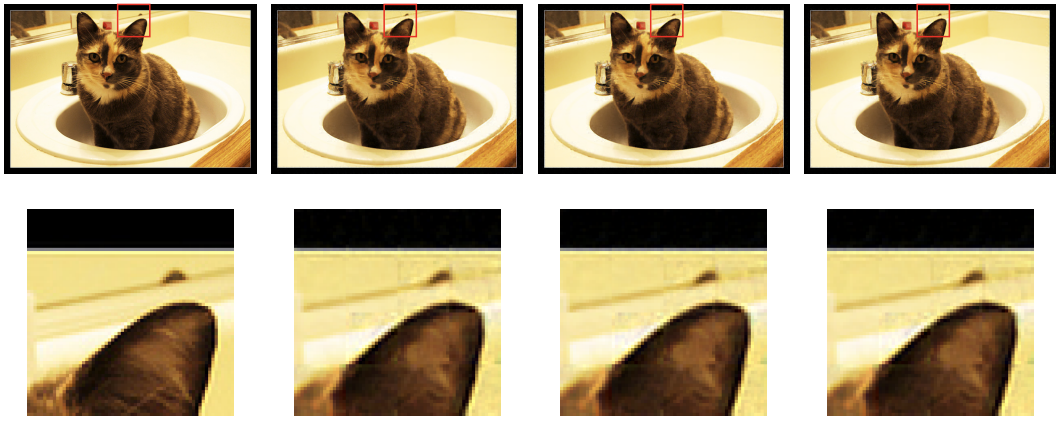
Reference	HLISTA-CP	Current increment	DASH (FIR)
PSNR (dB) / SSIM	33.026 / 0.9128	33.067 / 0.9131	33.105 / 0.9138



FIR - parent: whole-image paired mean +/- SD	FIR - current: whole-image paired mean +/- SD
PSNR +0.0626 +/- 0.0827	PSNR +0.0354 +/- 0.0284
SSIM +0.000798 +/- 0.001222	SSIM +0.000692 +/- 0.000395

COCO 000000094336 | MR25 | frozen RGB | seed 0

Reference	HLISTA-CP	Current increment	DASH (FIR)
PSNR (dB) / SSIM	31.756 / 0.8901	31.752 / 0.8908	31.807 / 0.8918

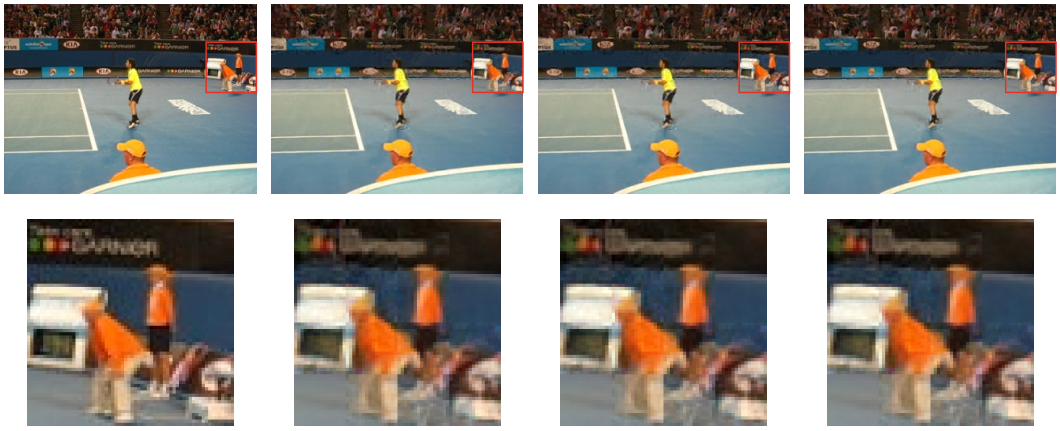


FIR - parent: whole-image paired mean +/- SD	FIR - current: whole-image paired mean +/- SD
PSNR +0.0665 +/- 0.0435	PSNR +0.0146 +/- 0.0355
SSIM +0.002049 +/- 0.002368	SSIM +0.001266 +/- 0.000641

Figure 17: Frozen per-channel RGB examples, COCO 262048 and 94336. All methods share the displayed seed, original image and fixed 64×64 crop. Full-image PSNR/SSIM appear above the reconstructions; paired differences are mean $\pm$ sample SD over three seeds. Positive is better. Examples were selected in the earlier grayscale review; they do not represent full-set performance. Full error-map panels remain in the reproduction archive.

COCO 000000005586 | MR25 | frozen RGB | seed 0

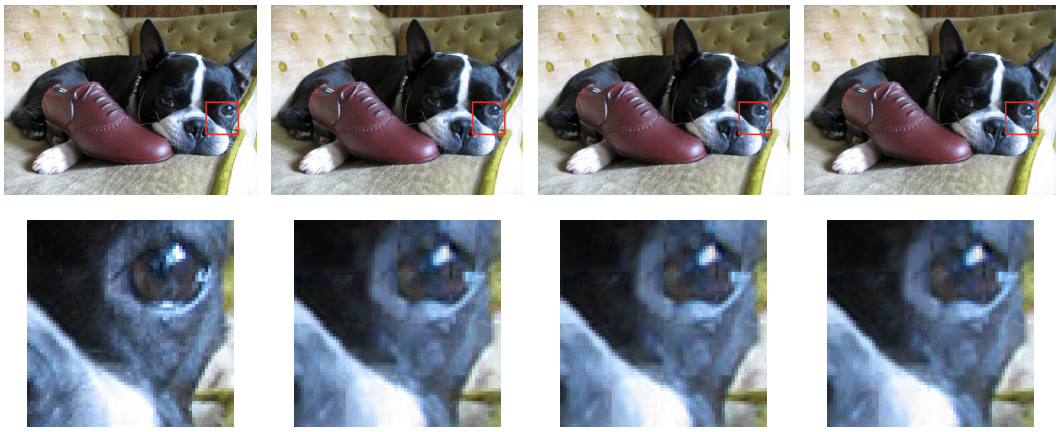
Reference	HLISTA-CP	Current increment	DASH (FIR)
PSNR (dB) / SSIM	28.863 / 0.8855	28.868 / 0.8857	28.917 / 0.8865



FIR - parent: whole-image paired mean +/- SD	FIR - current: whole-image paired mean +/- SD
PSNR +0.0565 +/- 0.0424	PSNR +0.0252 +/- 0.0350
SSIM +0.001243 +/- 0.001384	SSIM +0.000283 +/- 0.000551

COCO 000000399655 | MR25 | frozen RGB | seed 0

Reference	HLISTA-CP	Current increment	DASH (FIR)
PSNR (dB) / SSIM	31.305 / 0.8759	31.302 / 0.8760	31.345 / 0.8767



FIR - parent: whole-image paired mean +/- SD	FIR - current: whole-image paired mean +/- SD
PSNR +0.0442 +/- 0.0532	PSNR +0.0141 +/- 0.0269
SSIM +0.001089 +/- 0.001314	SSIM +0.000394 +/- 0.000516

Figure 18: Frozen per-channel RGB examples, COCO 5586 and 399655. All methods share the displayed seed, original image and fixed 64×64 crop. Full-image PSNR/SSIM appear above the reconstructions; paired differences are mean $\pm$ sample SD over three seeds. Positive is better. Examples were selected in the earlier grayscale review; they do not represent full-set performance. Full error-map panels remain in the reproduction archive.

Additional application coverage. Channel-wise RGB reuse on COCO gives FIR-both -0.0147 ± 0.0427 dB versus current-both (Appendix W). Across 26 FIR configurations, 19 improve their parent, but 13 of these do not beat the tested MLP, sized to LSTM rather than FIR (Appendix O). Full location, module, AMP and noise/operator results remain in Appendices Z, AA, AE and Y.

X PROCESSING AND MEASUREMENT CONVENTIONS

Image roles. The BSD500 manifests retain the official training, validation and test roles and reject image-ID, path or declared-hash overlap. These checks do not constitute a perceptual near-duplicate audit. The training split contains 200 source images, validation contains 100 and test contains 200; a smaller development run must report the actual selected counts. Do not merge the validation or test images into training while retaining the same claims. Optional COCO training/selection images come from the training source and held-out images from the validation source. A selected subset must be fixed before comparison; calling a cached archive a completed benchmark is invalid.

Preprocessing. Pillow decodes the stored image, converts it to YCbCr and extracts the Y channel, scaled to $[0, 1]$. The implementation does not resize or apply EXIF rotation. Training crops depend on image ID, seed and epoch; validation crops are centered. Small images receive symmetric edge padding if a crop requires it. Final test evaluation sets the crop size to none. Non-overlapping patches receive edge padding only on the bottom/right to reach a patch multiple, then reconstructed images are cropped back to their original dimensions. The configuration records patch size, crop size, measurement rate, operator seed, depth, regularization, step scale and every training setting.

Operators and optimization. The adaptation uses a fixed orthogonal DCT dictionary and separately labelled row-orthonormal Gaussian, i.i.d. Gaussian with variance $1/M$, or pixel subsampling operators. I.i.d. operator construction is checked on CPU, while an i.i.d. AMP fixture has CPU/MUSA coverage; neither is a matched benchmark. The finite stored matrix is logged by hash. The numerical step bound uses the maximum absolute row sum of $AA^\top$, computed in float64 with a recorded safety margin. This is a numerical implementation audit, not interval arithmetic or proof about unbounded floating-point execution. The separate fixed-reference solver and its execution audits are documented in the reproduction archive.

ISTA-Net+ fits its least-squares initial map from training images only, before optimization; the engine records its training manifest and matrix hash. The fitted map is fixed during training. The loss adds 0.01 times the sum of per-phase unthresholded symmetry penalties to pixel MSE. These implemented choices do not reproduce the original training protocol or establish accuracy.

Frozen-weight robustness stress. The registered CP stress audit uses 24 accepted LISTA-CP checkpoints: no feedback, C3, S3M and LSTM, each with three training seeds, at 25% sampling under row-orthonormal Gaussian and IID Gaussian operators. For each operator family we evaluate three measurement-noise levels, $\sigma_y \in \{1, 5, 10\}/255$, and two new operator seeds. The resulting $2 \times 5 \times 12 = 120$ model–condition positions contain 12,000 rows from the validation images. All four feedback choices in a condition use the same image IDs and the same measurement realization, identified by an observation hash. The new-operator conditions use the explicit `frozen_learned` policy: learned backprojections and normalization buffers are reused without retraining, refitting, test-image selection or budget changes. Each seed is first reduced over the 100 validation images; the table reports the mean and sample SD across the three training seeds for PSNR, SSIM, NMSE and NMSEdB. The audit recomputes metrics and hashes independently from the recorded evaluation source archived with the paper evidence. It is a validation stress test, not a held-out result, convergence certificate or claim that a checkpoint should transfer to another operator.

Image metrics. Predictions are not clipped before the reported metrics. With dynamic range one, full-image mean squared error determines PSNR, and full-image error energy divided by target energy determines NMSE. The implementation floors logarithm inputs and the target-energy denominator by the dtype machine epsilon. Thus perfect float32 logarithmic metrics saturate near ± 69.24

dB; this convention must not be compared numerically with a legacy implementation using a different floor. Linear NMSE still equals zero at exact reconstruction. SSIM uses an 11×11 Gaussian window with standard deviation 1.5, valid convolution, population moments and constants 0.01^2 and 0.03^2, without pre-downsampling (Wang et al., 2004). Its finite count is reported; undersized inputs are not silently resized to obtain a score.

Uncertainty and compute. The paired bootstrap resamples complete image pairs and rejects mismatched image-ID sets or missing selected metrics. Its percentile interval concerns image variability conditional on the trained checkpoints, not variability across training seeds. Independent seed summaries must therefore be retained separately. Timing requires device synchronization, declared warm-up and repetition counts, batch size, software/hardware identity and a clear distinction between I/O and reconstruction. Measured image latency and memory are now reported in Appendices V and U; these measurements do not certify unmeasured synthetic efficiency.

Y IMAGE PROTOCOLS, LOCATIONS AND EXTERNAL ALTERNATIVES

The DCT protocol preserves BSD500 train200/validation100/test200 splits, with training crops only, QR or IID sensing and common matrix seeds. The module/location experiment uses seven arms, six operator/rate settings and three seeds (126 outcomes), each with 13,000 updates including a 650-update parent and continuation with fresh Adam. Validation selects checkpoints; it is not an external test. Figure 11 shows every paired PSNR contrast. The all-rate absolute and paired tables are retained once in Appendix Z.

The separate author-aligned image CS experiment uses fixed released matrices, a dictionary and a BSD400/50/50 repartition with 13,000 end-to-end updates, not the original stagewise training. Its complete test50 results appear once in Table 83; the earlier PSNR-only and repeated appendix versions are archived. This within-corpus repartition is distinct from the DCT protocol and the 26,000-update five-arm controls.

External method adaptations and failures. External DCT tables are validation-only, with no external test200 ranking. AMP/LAMP retain Onsager corrections; LDAMP uses full-image DnCNN and Monte Carlo Onsager. Classical image SD differs from learned-model seed SD, and fixed versus selected depth remains separated. LAMP QR10 seed2 initially had zero gradients; a registered diagnostic recovery changed its early-stop audit within the original update budget. SVD-LVAMP uses a validation-selected seed0/r25 learning rate transferred unchanged to other cells; IID50 remains unstable (17.8998 ± 10.4070 dB). Source adaptations and incomplete legacy runs are disclosed in the reproduction archive (Appendix AO).

Y.1 FROZEN NOISE AND OPERATOR STRESS

The 24 frozen-checkpoint stress audit uses declared noisy observations and new operators without learned-backprojection adaptation. It tests frozen operator transfer, not a new trained protocol. All signs are retained below.

Z LONG-BUDGET VALIDATION GROUPS

The independently audited snapshot contains 275 distinct results and 90 complete three-seed groups. The original 143-run continuation and 126-run Hybrid-CP queues are terminal. The separate SCNet/SCL queue has six accepted results and three recorded timeouts. Only complete groups enter the tables; QR ISTA-Net+ entries are omitted from this display. The complete source archive, incomplete groups and failures are retained. Each seed first averages the same 100 full validation images. Tables give mean and sample SD across seeds 0, 1 and 2, not across image patches. No training or image inference was repeated to construct these tables. These historical tables are validation-only and contain no test200 evaluations. HCP denotes Hybrid-LISTA-CP; ref. denotes refinement-only feedback. The 13,000 updates comprise 650 parent updates and 12,350 additional updates, with model-only warm starts and fresh Adam between phases. SCNet/SCL uses 39,000 updates and a different self-supervised objective, with validation ground truth for checkpoint selection; it is not a same-cost comparison. The separate progressive protocol appears in Appendix AO.

Table 72: Frozen-weight validation stress test at 25% sampling for LISTA-CP. Each entry reports the no-feedback PSNR mean ± seed SD, or the paired PSNR difference (module minus no feedback) with its seed SD. Bold compares module differences against the zero difference of no feedback within each row. Each row contains 12 models (four feedback choices and three seeds) evaluated on the same 100 validation images. New-operator rows deliberately reuse the learned checkpoint buffers; they are a transfer stress test, not retraining.

Op.	Condition	None	C3 dPSNR	S3M dPSNR	LSTM dPSNR
QR Gaussian	noise 1	26.7042 ± 0.0050	$+0.0620 \pm 0.0026$	$+0.0262 \pm 0.0196$	$\mathbf{+0.0819 \pm 0.0073}$
QR Gaussian	noise 5	26.1604 ± 0.0053	$+0.0411 \pm 0.0017$	$+0.0204 \pm 0.0226$	$\mathbf{+0.0590 \pm 0.0064}$
QR Gaussian	noise 10	24.8371 ± 0.0052	$\mathbf{+0.0257 \pm 0.0011}$	$+0.0080 \pm 0.0236$	$+0.0190 \pm 0.0053$
QR Gaussian	new A 101	-46.0609 ± 1.5849	$+3.5657 \pm 4.9205$	$\mathbf{+8.5376 \pm 3.2586}$	-1.7038 ± 4.0736
QR Gaussian	new A 102	-48.5627 ± 0.9231	$+2.0429 \pm 2.7313$	$\mathbf{+8.0214 \pm 0.3991}$	-5.0203 ± 2.1697
IID Gaussian	noise 1	26.7776 ± 0.0017	-0.0529 ± 0.0024	$+0.0039 \pm 0.0061$	$\mathbf{+0.0285 \pm 0.0066}$
IID Gaussian	noise 5	26.5819 ± 0.0030	-0.0600 ± 0.0026	$+0.0027 \pm 0.0058$	$\mathbf{+0.0258 \pm 0.0067}$
IID Gaussian	noise 10	26.0124 ± 0.0046	-0.0682 ± 0.0033	$+0.0004 \pm 0.0054$	$\mathbf{+0.0205 \pm 0.0078}$
IID Gaussian	new A 101	$\mathbf{-26.6177 \pm 1.2964}$	-15.5060 ± 2.8489	-4.0093 ± 2.8537	-6.4965 ± 4.8191
IID Gaussian	new A 102	-31.6610 ± 1.3949	-12.0095 ± 4.5083	$\mathbf{+0.8907 \pm 1.1287}$	-0.4406 ± 2.1741

Table 73: QR, 10%, 13,000 updates: validation100 mean ± sample SD across three training seeds.

Method	PSNR (dB)	SSIM	NMSE	NMSE (dB)
ALISTA	20.2432 ± 0.0093	0.408451 ± 0.000719	0.057272 ± 0.000161	-13.4823 ± 0.0093
ALISTA-SS	19.5973 ± 0.0886	0.374148 ± 0.001996	0.069483 ± 0.005473	-12.8364 ± 0.0886
CP + C3	23.8900 ± 0.0041	0.637056 ± 0.000325	0.026164 ± 0.000024	-17.1291 ± 0.0041
CP + LSTM	23.9078 ± 0.0122	0.638173 ± 0.000436	0.026088 ± 0.000059	-17.1469 ± 0.0122
LISTA-CP	23.8797 ± 0.0085	0.636824 ± 0.000619	0.026220 ± 0.000040	-17.1188 ± 0.0085
CP + S3M	23.9061 ± 0.0137	0.638070 ± 0.000596	0.026088 ± 0.000066	-17.1452 ± 0.0137
HCP + C3 (both)	23.9688 ± 0.0197	0.639943 ± 0.000893	0.025768 ± 0.000079	-17.2079 ± 0.0197
HCP + C3 (ref.)	23.9517 ± 0.0099	0.639493 ± 0.000580	0.025848 ± 0.000049	-17.1908 ± 0.0099
HCP + LSTM (both)	23.9642 ± 0.0132	0.640157 ± 0.000357	0.025785 ± 0.000062	-17.2033 ± 0.0132
HCP + LSTM (ref.)	23.9464 ± 0.0154	0.639132 ± 0.000630	0.025870 ± 0.000083	-17.1855 ± 0.0154
HLISTA-CP	23.9401 ± 0.0101	0.638899 ± 0.000259	0.025902 ± 0.000052	-17.1792 ± 0.0101
HCP + S3M (both)	$\mathbf{23.9927 \pm 0.0064}$	$\mathbf{0.641712 \pm 0.000238}$	$\mathbf{0.025650 \pm 0.000028}$	$\mathbf{-17.2317 \pm 0.0064}$
HCP + S3M (ref.)	23.9568 ± 0.0050	0.639585 ± 0.000344	0.025816 ± 0.000025	-17.1959 ± 0.0050

AA ADDITIONAL DATASET AND MATCHED-CS EVIDENCE

All tables are recomputed from existing completed records. The accompanying evidence file retains seed differences and trained FIR coefficients. The CS test50 protocol uses 13,000 updates and the separate 400/50/50 within-corpus split with an overcomplete dictionary. Same-seed DASH minus HLISTA-CP PSNR differences at MR4/10/25/50 are respectively $+0.0157 \pm 0.0025$, $+0.0226 \pm 0.0018$, $+0.0229 \pm 0.0112$ and $+0.0629 \pm 0.0327$ dB (mean ± sample SD over three seeds). The absolute scores are presented once in Table 83.

Table 74: QR, 25%, 13,000 updates: validation100 mean ± sample SD across three training seeds.

Method	PSNR (dB)	SSIM	NMSE	NMSE (dB)
ALISTA	23.8276 ± 0.0290	0.636745 ± 0.001597	0.025954 ± 0.000160	-17.0667 ± 0.0290
ALISTA-SS	21.9040 ± 0.0942	0.487614 ± 0.011396	0.039280 ± 0.001258	-15.1431 ± 0.0942
CP + C3	26.7901 ± 0.0044	0.777662 ± 0.000159	0.014014 ± 0.000010	-20.0292 ± 0.0044
CP + LSTM	26.8100 ± 0.0027	0.777609 ± 0.000149	0.013958 ± 0.000006	-20.0491 ± 0.0027
LISTA-CP	26.7266 ± 0.0049	0.775719 ± 0.000138	0.014183 ± 0.000013	-19.9656 ± 0.0049
CP + S3M	26.7534 ± 0.0163	0.776476 ± 0.000149	0.014101 ± 0.000041	-19.9924 ± 0.0163
HCP + C3 (both)	26.9604 ± 0.0217	0.783452 ± 0.000591	0.013588 ± 0.000050	-20.1995 ± 0.0217
HCP + C3 (ref.)	26.9822 ± 0.0109	0.784036 ± 0.000238	0.013531 ± 0.000023	-20.2212 ± 0.0109
HCP + LSTM (both)	26.9240 ± 0.0063	0.782488 ± 0.000231	0.013672 ± 0.000019	-20.1631 ± 0.0063
HCP + LSTM (ref.)	$\mathbf{26.9903 \pm 0.0139}$	$\mathbf{0.784239 \pm 0.000404}$	$\mathbf{0.013498 \pm 0.000031}$	$\mathbf{-20.2294 \pm 0.0139}$
HLISTA-CP	26.9157 ± 0.0085	0.782271 ± 0.000260	0.013690 ± 0.000021	-20.1548 ± 0.0085
HCP + S3M (both)	26.9138 ± 0.0144	0.782324 ± 0.000422	0.013695 ± 0.000036	-20.1529 ± 0.0144
HCP + S3M (ref.)	26.9267 ± 0.0048	0.782555 ± 0.000148	0.013661 ± 0.000011	-20.1658 ± 0.0048
LAMP	26.2063 ± 0.0123	0.756800 ± 0.000694	0.015933 ± 0.000045	-19.4454 ± 0.0123

Table 75: QR, 50%, 13,000 updates: validation100 mean ± sample SD across three training seeds.

Method	PSNR (dB)	SSIM	NMSE	NMSE (dB)
ALISTA	27.7784 ± 0.0043	0.820474 ± 0.000162	0.011067 ± 0.000003	-21.0175 ± 0.0043
ALISTA-SS	27.6130 ± 0.0683	0.812704 ± 0.001869	0.011493 ± 0.000103	-20.8521 ± 0.0683
CP + C3	$\mathbf{31.0282 \pm 0.0013}$	$\mathbf{0.901067 \pm 0.000036}$	$\mathbf{0.005636 \pm 0.000003}$	$\mathbf{-24.2673 \pm 0.0013}$
CP + LSTM	30.5814 ± 0.0049	0.893428 ± 0.000077	0.006124 ± 0.000008	-23.8205 ± 0.0049
LISTA-CP	30.9070 ± 0.0047	0.898856 ± 0.000090	0.005772 ± 0.000007	-24.1461 ± 0.0047
CP + S3M	30.8929 ± 0.0145	0.898600 ± 0.000278	0.005786 ± 0.000017	-24.1320 ± 0.0145
HCP + C3 (both)	30.9160 ± 0.1626	0.899187 ± 0.002681	0.005751 ± 0.000175	-24.1551 ± 0.1626
HCP + C3 (ref.)	30.9838 ± 0.1132	0.900208 ± 0.001953	0.005678 ± 0.000121	-24.2228 ± 0.1132
HCP + LSTM (both)	30.9764 ± 0.0566	0.900016 ± 0.000994	0.005681 ± 0.000060	-24.2155 ± 0.0566
HCP + LSTM (ref.)	30.9025 ± 0.0374	0.898725 ± 0.000596	0.005764 ± 0.000044	-24.1416 ± 0.0374
HLISTA-CP	30.8628 ± 0.1366	0.898296 ± 0.002198	0.005806 ± 0.000149	-24.1018 ± 0.1366
HCP + S3M (both)	30.8846 ± 0.1446	0.898647 ± 0.002271	0.005785 ± 0.000157	-24.1236 ± 0.1446
HCP + S3M (ref.)	30.8093 ± 0.0146	0.897345 ± 0.000224	0.005864 ± 0.000013	-24.0484 ± 0.0146
LAMP	30.5316 ± 0.0012	0.891898 ± 0.000019	0.006168 ± 0.000002	-23.7707 ± 0.0012

Table 76: QR, 25%, 39,000 updates: validation100 mean ± sample SD across three training seeds.

Method	PSNR (dB)	SSIM	NMSE	NMSE (dB)
SCNet/SCL	27.9597 ± 0.0901	0.814149 ± 0.002517	0.011256 ± 0.000193	-21.1988 ± 0.0901

Table 77: IID, 10%, 13,000 updates: validation100 mean ± sample SD across three training seeds.

Method	PSNR (dB)	SSIM	NMSE	NMSE (dB)
ALISTA	19.9815 ± 0.0090	0.397274 ± 0.000481	0.060111 ± 0.000130	-13.2206 ± 0.0090
ALISTA-SS	19.7331 ± 0.1124	0.394085 ± 0.004681	0.064945 ± 0.002593	-12.9722 ± 0.1124
CP + C3	23.8373 ± 0.0062	0.636508 ± 0.000307	0.026498 ± 0.000024	-17.0764 ± 0.0062
CP + LSTM	23.9351 ± 0.0060	0.640283 ± 0.000034	0.026002 ± 0.000024	-17.1742 ± 0.0060
LISTA-CP	23.9105 ± 0.0052	0.640320 ± 0.000444	0.026157 ± 0.000020	-17.1496 ± 0.0052
CP + S3M	23.8973 ± 0.0046	0.638518 ± 0.000334	0.026209 ± 0.000013	-17.1364 ± 0.0046
HCP + C3 (both)	24.0428 ± 0.0461	0.647169 ± 0.001766	0.025478 ± 0.000215	-17.2819 ± 0.0461
HCP + C3 (ref.)	24.0662 ± 0.0433	0.648157 ± 0.001712	0.025377 ± 0.000207	-17.3053 ± 0.0433
HCP + LSTM (both)	24.0439 ± 0.0329	0.647280 ± 0.001289	0.025476 ± 0.000145	-17.2830 ± 0.0329
HCP + LSTM (ref.)	24.0576 ± 0.0300	0.647900 ± 0.001144	0.025418 ± 0.000142	-17.2967 ± 0.0300
HLISTA-CP	24.0482 ± 0.0459	0.647469 ± 0.001883	0.025457 ± 0.000212	-17.2873 ± 0.0459
HCP + S3M (both)	24.0596 ± 0.0264	0.647790 ± 0.001255	0.025394 ± 0.000120	-17.2986 ± 0.0264
HCP + S3M (ref.)	24.0456 ± 0.0175	0.647437 ± 0.000680	0.025474 ± 0.000086	-17.2847 ± 0.0175
ISTA-Net+	$\mathbf{24.4493 \pm 0.0397}$	$\mathbf{0.663976 \pm 0.001567}$	$\mathbf{0.023741 \pm 0.000178}$	$\mathbf{-17.6884 \pm 0.0397}$
LAMP	22.6138 ± 0.0613	0.562590 ± 0.007864	0.033982 ± 0.000388	-15.8528 ± 0.0613

Table 78: IID, 25%, 13,000 updates: validation100 mean ± sample SD across three training seeds.

Method	PSNR (dB)	SSIM	NMSE	NMSE (dB)
ALISTA	23.8278 ± 0.0284	0.639419 ± 0.001576	0.026009 ± 0.000133	-17.0669 ± 0.0284
ALISTA-SS	23.4501 ± 0.2214	0.621282 ± 0.019240	0.028342 ± 0.000974	-16.6892 ± 0.2214
CP + C3	26.7325 ± 0.0018	0.778210 ± 0.000022	0.014160 ± 0.000008	-19.9716 ± 0.0018
CP + LSTM	26.8139 ± 0.0051	0.779975 ± 0.000113	0.013948 ± 0.000018	-20.0530 ± 0.0051
LISTA-CP	26.7853 ± 0.0016	0.779074 ± 0.000088	0.014037 ± 0.000009	-20.0244 ± 0.0016
CP + S3M	26.7891 ± 0.0044	0.779382 ± 0.000207	0.014019 ± 0.000012	-20.0282 ± 0.0044
HCP + C3 (both)	26.9398 ± 0.0381	0.784365 ± 0.001057	0.013632 ± 0.000094	-20.1789 ± 0.0381
HCP + C3 (ref.)	27.0091 ± 0.0373	0.786113 ± 0.001032	0.013462 ± 0.000088	-20.2481 ± 0.0373
HCP + LSTM (both)	26.9330 ± 0.0361	0.783926 ± 0.001181	0.013639 ± 0.000076	-20.1720 ± 0.0361
HCP + LSTM (ref.)	26.9516 ± 0.0265	0.784504 ± 0.000697	0.013597 ± 0.000061	-20.1906 ± 0.0265
HLISTA-CP	26.9316 ± 0.0375	0.784016 ± 0.001104	0.013648 ± 0.000095	-20.1707 ± 0.0375
HCP + S3M (both)	26.9654 ± 0.0316	0.784754 ± 0.000930	0.013557 ± 0.000068	-20.2045 ± 0.0316
HCP + S3M (ref.)	26.9395 ± 0.0314	0.784169 ± 0.000838	0.013629 ± 0.000076	-20.1786 ± 0.0314
ISTA-Net+	$\mathbf{28.2076 \pm 0.0348}$	$\mathbf{0.817118 \pm 0.000777}$	$\mathbf{0.010847 \pm 0.000074}$	$\mathbf{-21.4467 \pm 0.0348}$
LAMP	25.4783 ± 0.0178	0.728740 ± 0.000733	0.018679 ± 0.000068	-18.7174 ± 0.0178

Table 79: IID, 50%, 13,000 updates: validation100 mean ± sample SD across three training seeds.

Method	PSNR (dB)	SSIM	NMSE	NMSE (dB)
ALISTA	27.8232 ± 0.0335	0.821723 ± 0.000639	0.011021 ± 0.000057	-21.0623 ± 0.0335
ALISTA-SS	27.3752 ± 0.1679	0.801392 ± 0.009511	0.011977 ± 0.000321	-20.6143 ± 0.1679
CP + C3	30.5337 ± 0.0149	0.892971 ± 0.000277	0.006184 ± 0.000017	-23.7728 ± 0.0149
CP + LSTM	30.5386 ± 0.0066	0.892915 ± 0.000109	0.006177 ± 0.000009	-23.7777 ± 0.0066
LISTA-CP	30.5317 ± 0.0031	0.892781 ± 0.000051	0.006190 ± 0.000002	-23.7708 ± 0.0031
CP + S3M	30.5253 ± 0.0109	0.892657 ± 0.000183	0.006196 ± 0.000011	-23.7644 ± 0.0109
HCP + C3 (both)	30.6644 ± 0.0025	0.895149 ± 0.000062	0.006031 ± 0.000003	-23.9035 ± 0.0025
HCP + C3 (ref.)	30.6940 ± 0.0113	0.895427 ± 0.000104	0.005994 ± 0.000015	-23.9331 ± 0.0113
HCP + LSTM (both)	30.6870 ± 0.0101	0.895509 ± 0.000164	0.006005 ± 0.000013	-23.9261 ± 0.0101
HCP + LSTM (ref.)	30.6511 ± 0.0426	0.894796 ± 0.000673	0.006040 ± 0.000043	-23.8902 ± 0.0426
HLISTA-CP	30.6470 ± 0.0109	0.894852 ± 0.000249	0.006047 ± 0.000015	-23.8861 ± 0.0109
HCP + S3M (both)	30.7211 ± 0.0241	0.895939 ± 0.000283	0.005964 ± 0.000029	-23.9602 ± 0.0241
HCP + S3M (ref.)	30.6498 ± 0.0186	0.894867 ± 0.000255	0.006044 ± 0.000023	-23.8889 ± 0.0186
ISTA-Net+	$\mathbf{32.6455 \pm 0.0617}$	$\mathbf{0.922656 \pm 0.000970}$	$\mathbf{0.004130 \pm 0.000045}$	$\mathbf{-25.8846 \pm 0.0617}$
LAMP	29.2330 ± 0.0362	0.861480 ± 0.000873	0.008375 ± 0.000072	-22.4721 ± 0.0362

Table 80: Plain CP with feedback minus plain CP without feedback. Paired ΔPSNR is mean ± sample SD of three same-seed differences. All models use 13,000 updates. No significance test is implied.

Method	Operator	Rate	ΔPSNR (dB)
CP + C3	QR	10%	0.0103 ± 0.0064
CP + LSTM	QR	10%	$\mathbf{0.0281 \pm 0.0063}$
CP + S3M	QR	10%	0.0264 ± 0.0056
CP + C3	QR	25%	0.0636 ± 0.0025
CP + LSTM	QR	25%	$\mathbf{0.0835 \pm 0.0072}$
CP + S3M	QR	25%	0.0268 ± 0.0192
CP + C3	QR	50%	$\mathbf{0.1212 \pm 0.0059}$
CP + LSTM	QR	50%	-0.3256 ± 0.0045
CP + S3M	QR	50%	-0.0141 ± 0.0188
CP + C3	IID	10%	-0.0732 ± 0.0010
CP + LSTM	IID	10%	$\mathbf{0.0247 \pm 0.0069}$
CP + S3M	IID	10%	-0.0132 ± 0.0043
CP + C3	IID	25%	-0.0528 ± 0.0023
CP + LSTM	IID	25%	$\mathbf{0.0286 \pm 0.0065}$
CP + S3M	IID	25%	0.0038 ± 0.0060
CP + C3	IID	50%	0.0020 ± 0.0145
CP + LSTM	IID	50%	$\mathbf{0.0069 \pm 0.0035}$
CP + S3M	IID	50%	-0.0064 ± 0.0132

Table 81: Hybrid-CP both-branch feedback minus no feedback. Paired ΔPSNR is mean ± sample SD of three same-seed differences. All models use 13,000 updates. No significance test is implied.

Method	Operator	Rate	ΔPSNR (dB)
HCP + C3 (both)	QR	10%	0.0287 ± 0.0095
HCP + LSTM (both)	QR	10%	0.0241 ± 0.0031
HCP + S3M (both)	QR	10%	$\mathbf{0.0526 \pm 0.0037}$
HCP + C3 (both)	QR	25%	$\mathbf{0.0447 \pm 0.0133}$
HCP + LSTM (both)	QR	25%	0.0083 ± 0.0028
HCP + S3M (both)	QR	25%	-0.0018 ± 0.0064
HCP + C3 (both)	QR	50%	0.0533 ± 0.0279
HCP + LSTM (both)	QR	50%	$\mathbf{0.1137 \pm 0.1583}$
HCP + S3M (both)	QR	50%	0.0218 ± 0.0119
HCP + C3 (both)	IID	10%	-0.0054 ± 0.0142
HCP + LSTM (both)	IID	10%	-0.0044 ± 0.0143
HCP + S3M (both)	IID	10%	$\mathbf{0.0113 \pm 0.0195}$
HCP + C3 (both)	IID	25%	0.0082 ± 0.0236
HCP + LSTM (both)	IID	25%	0.0013 ± 0.0318
HCP + S3M (both)	IID	25%	$\mathbf{0.0338 \pm 0.0326}$
HCP + C3 (both)	IID	50%	0.0174 ± 0.0088
HCP + LSTM (both)	IID	50%	0.0400 ± 0.0167
HCP + S3M (both)	IID	50%	$\mathbf{0.0741 \pm 0.0349}$

Table 82: Hybrid-CP both-branch feedback minus refinement-only feedback. Paired ΔPSNR is mean ± sample SD of three same-seed differences. All models use 13,000 updates. No significance test is implied.

Method	Operator	Rate	ΔPSNR (dB)
HCP + C3 (both)	QR	10%	0.0171 ± 0.0123
HCP + LSTM (both)	QR	10%	0.0178 ± 0.0022
HCP + S3M (both)	QR	10%	0.0358 ± 0.0036
HCP + C3 (both)	QR	25%	-0.0218 ± 0.0108
HCP + LSTM (both)	QR	25%	-0.0663 ± 0.0201
HCP + S3M (both)	QR	25%	-0.0128 ± 0.0122
HCP + C3 (both)	QR	50%	-0.0677 ± 0.0534
HCP + LSTM (both)	QR	50%	0.0739 ± 0.0448
HCP + S3M (both)	QR	50%	0.0753 ± 0.1373
HCP + C3 (both)	IID	10%	-0.0234 ± 0.0235
HCP + LSTM (both)	IID	10%	-0.0138 ± 0.0079
HCP + S3M (both)	IID	10%	0.0139 ± 0.0092
HCP + C3 (both)	IID	25%	-0.0692 ± 0.0020
HCP + LSTM (both)	IID	25%	-0.0186 ± 0.0347
HCP + S3M (both)	IID	25%	0.0259 ± 0.0248
HCP + C3 (both)	IID	50%	-0.0297 ± 0.0134
HCP + LSTM (both)	IID	50%	0.0359 ± 0.0420
HCP + S3M (both)	IID	50%	0.0713 ± 0.0299

Table 83: All four metrics for the separate CS test50 panel; mean ± sample SD across three training seeds. NMSE is linear; the last column is in dB.

MR (%)	Method	PSNR	SSIM	NMSE	NMSE (dB)
4	ISTA-Net+	21.8464 ± 0.0279	0.5511 ± 0.0003	0.0364 ± 0.0002	-15.2442 ± 0.0279
4	HLISTA-CP	22.1214 ± 0.0114	0.5630 ± 0.0005	0.0346 ± 0.0001	-15.5193 ± 0.0114
4	DASH-HLISTA-CP	$\mathbf{22.1371 \pm 0.0090}$	$\mathbf{0.5633 \pm 0.0005}$	$\mathbf{0.0345 \pm 0.0001}$	$\mathbf{-15.5349 \pm 0.0090}$
10	ISTA-Net+	23.7164 ± 0.0003	0.6486 ± 0.0002	0.0243 ± 0.0000	-17.1142 ± 0.0003
10	HLISTA-CP	24.2344 ± 0.0117	0.6666 ± 0.0006	0.0221 ± 0.0001	-17.6322 ± 0.0117
10	DASH-HLISTA-CP	$\mathbf{24.2570 \pm 0.0106}$	$\mathbf{0.6676 \pm 0.0005}$	$\mathbf{0.0220 \pm 0.0000}$	$\mathbf{-17.6548 \pm 0.0106}$
25	ISTA-Net+	26.9632 ± 0.0451	0.7909 ± 0.0012	0.0121 ± 0.0001	-20.3610 ± 0.0451
25	HLISTA-CP	27.7035 ± 0.0085	0.8123 ± 0.0005	$\mathbf{0.0105 \pm 0.0000}$	-21.1013 ± 0.0085
25	DASH-HLISTA-CP	$\mathbf{27.7264 \pm 0.0028}$	$\mathbf{0.8131 \pm 0.0003}$	$\mathbf{0.0105 \pm 0.0000}$	$\mathbf{-21.1242 \pm 0.0028}$
50	ISTA-Net+	30.5887 ± 0.0477	0.8946 ± 0.0010	0.0055 ± 0.0001	-23.9865 ± 0.0477
50	HLISTA-CP	31.8448 ± 0.0296	0.9156 ± 0.0004	0.0043 ± 0.0000	-25.2426 ± 0.0296
50	DASH-HLISTA-CP	$\mathbf{31.9078 \pm 0.0093}$	$\mathbf{0.9164 \pm 0.0001}$	$\mathbf{0.0042 \pm 0.0000}$	$\mathbf{-25.3056 \pm 0.0093}$

Table 84: Learned FIR coefficients from the twelve frozen DASH CS checkpoints. b_0 weights the current increment; b_1, b_2, b_3 weight past increments. Past share is $\sum_{j=1}^{3} |b_j| / \sum_{j=0}^{3} |b_j|$, not a reconstruction contribution.

MR (%)	b_0	b_1	b_2	b_3	Past share
4	0.155 ± 0.001	0.122 ± 0.002	0.096 ± 0.015	0.065 ± 0.013	0.645 ± 0.006
10	0.252 ± 0.000	0.177 ± 0.020	0.059 ± 0.031	0.011 ± 0.013	0.496 ± 0.001
25	0.277 ± 0.002	0.193 ± 0.010	0.029 ± 0.009	-0.001 ± 0.000	0.446 ± 0.003
50	0.296 ± 0.003	0.158 ± 0.002	0.040 ± 0.002	-0.006 ± 0.001	0.408 ± 0.006

Table 85: Complete COCO QR development validation5000 (from train2017) paired improvements: DASH minus parent for PSNR/SSIM, parent minus DASH for NMSE and NMSE(dB). Every uncertainty is the sample SD of three seed differences. Linear NMSE is sensitive to nearly zero-energy images.

Backbone	MR (%)	ΔPSNR	ΔSSIM	NMSE impr.	NMSE(dB) impr.
HALISTA	10	$+0.0709 \pm 0.1321$	$+0.0003 \pm 0.0062$	$+123.3812 \pm 346.7150$	$+0.0709 \pm 0.1321$
HALISTA	25	$+0.4954 \pm 0.1745$	$+0.0213 \pm 0.0079$	-3.0438 ± 4.6286	$+0.5001 \pm 0.1693$
HALISTA	50	$+0.1198 \pm 0.0258$	$+0.0039 \pm 0.0007$	-0.0092 ± 0.0166	$+0.1195 \pm 0.0258$
HCISTA	10	$+0.0962 \pm 0.0137$	$+0.0026 \pm 0.0000$	$+0.0135 \pm 0.0019$	$+0.0962 \pm 0.0137$
HCISTA	25	$+0.2430 \pm 0.1612$	$+0.0075 \pm 0.0059$	$+0.0041 \pm 0.0025$	$+0.2430 \pm 0.1612$
HCISTA	50	$+0.0308 \pm 0.0149$	$+0.0002 \pm 0.0007$	$+0.0001 \pm 0.0001$	$+0.0308 \pm 0.0149$
HELISTA	10	-0.0046 ± 0.0244	-0.0002 ± 0.0020	$+11.4494 \pm 23.3883$	-0.0057 ± 0.0232
HELISTA	25	$+0.0216 \pm 0.0321$	$+0.0007 \pm 0.0010$	-4.2823 ± 6.2884	$+0.0216 \pm 0.0321$
HELISTA	50	-0.0166 ± 0.0816	-0.0004 ± 0.0017	-0.3943 ± 0.5473	-0.0167 ± 0.0818
HGLISTA-gain	10	-0.0209 ± 0.0493	-0.0007 ± 0.0026	-0.0185 ± 0.0392	-0.0153 ± 0.0399
HGLISTA-gain	25	-0.0145 ± 0.2150	-0.0008 ± 0.0094	-0.0005 ± 0.0007	-0.0196 ± 0.2090
HGLISTA-gain	50	-0.2512 ± 0.4617	-0.0061 ± 0.0110	-0.0184 ± 0.0318	-0.2518 ± 0.4629
HGLISTA-overshoot	10	-0.0413 ± 0.0355	-0.0021 ± 0.0025	-0.0045 ± 0.0036	-0.0467 ± 0.0375
HGLISTA-overshoot	25	$+0.0686 \pm 0.1769$	$+0.0027 \pm 0.0070$	-0.0022 ± 0.0046	$+0.0650 \pm 0.1784$
HGLISTA-overshoot	50	$+0.1410 \pm 0.4429$	$+0.0037 \pm 0.0106$	$+0.0046 \pm 0.0108$	$+0.1372 \pm 0.4584$
HLISTA-CP	10	$+0.0061 \pm 0.0012$	$+0.0002 \pm 0.0001$	$+4.6322 \pm 8.6017$	$+0.0061 \pm 0.0012$
HLISTA-CP	25	$+0.0114 \pm 0.0032$	$+0.0004 \pm 0.0001$	-0.3011 ± 0.3764	$+0.0114 \pm 0.0032$
HLISTA-CP	50	$+0.0304 \pm 0.0140$	$+0.0006 \pm 0.0003$	$+1.6751 \pm 1.3937$	$+0.0304 \pm 0.0140$
HLISTA-CPSS	10	$+0.0058 \pm 0.0016$	$+0.0004 \pm 0.0001$	$+5.7782 \pm 7.0274$	$+0.0058 \pm 0.0016$
HLISTA-CPSS	25	$+0.0137 \pm 0.0034$	$+0.0005 \pm 0.0001$	$+0.7778 \pm 1.1709$	$+0.0136 \pm 0.0034$
HLISTA-CPSS	50	$+0.0137 \pm 0.0153$	$+0.0003 \pm 0.0003$	$+0.1988 \pm 0.8152$	$+0.0137 \pm 0.0153$

Table 86: Complete COCO IID development validation5000 (from train2017) paired improvements: DASH minus parent for PSNR/SSIM, parent minus DASH for NMSE and NMSE(dB). Every uncertainty is the sample SD of three seed differences. Linear NMSE is sensitive to nearly zero-energy images.

Backbone	MR (%)	ΔPSNR	ΔSSIM	NMSE impr.	NMSE(dB) impr.
HALISTA	10	$+0.0028 \pm 0.0227$	-0.0012 ± 0.0004	$+213.4477 \pm 813.7180$	$+0.0028 \pm 0.0227$
HALISTA	25	$+0.0180 \pm 0.0167$	$+0.0007 \pm 0.0007$	-3.3663 ± 25.0263	$+0.0127 \pm 0.0258$
HALISTA	50	-0.0006 ± 0.0126	$+0.0004 \pm 0.0002$	-1.3342 ± 1.1307	-0.0011 ± 0.0123
HCISTA	10	$+0.0302 \pm 0.0054$	-0.0006 ± 0.0000	$+0.0058 \pm 0.0010$	$+0.0302 \pm 0.0054$
HCISTA	25	$+0.0747 \pm 0.0081$	$+0.0002 \pm 0.0006$	$+0.0091 \pm 0.0009$	$+0.0747 \pm 0.0081$
HCISTA	50	$+0.1941 \pm 0.0072$	$+0.0042 \pm 0.0007$	$+0.0099 \pm 0.0003$	$+0.1941 \pm 0.0072$
HELISTA	10	$+0.0067 \pm 0.0049$	$+0.0003 \pm 0.0004$	$+0.0095 \pm 0.0382$	$+0.0070 \pm 0.0059$
HELISTA	25	$+0.0147 \pm 0.0060$	$+0.0006 \pm 0.0002$	$+0.2880 \pm 0.6271$	$+0.0176 \pm 0.0070$
HELISTA	50	$+0.0390 \pm 0.0342$	$+0.0009 \pm 0.0009$	$+0.5862 \pm 0.4649$	$+0.0402 \pm 0.0344$
HGLISTA-gain	10	$+0.0028 \pm 0.0012$	$+0.0001 \pm 0.0000$	-0.0538 ± 0.1205	$+0.0027 \pm 0.0012$
HGLISTA-gain	25	$+0.0073 \pm 0.0014$	$+0.0003 \pm 0.0000$	-0.0094 ± 0.0165	$+0.0064 \pm 0.0010$
HGLISTA-gain	50	$+0.0137 \pm 0.0112$	$+0.0003 \pm 0.0002$	-0.0000 ± 0.0000	$+0.0137 \pm 0.0111$
HGLISTA-overshoot	10	$+0.0041 \pm 0.0009$	$+0.0001 \pm 0.0000$	-0.0275 ± 0.0249	$+0.0041 \pm 0.0008$
HGLISTA-overshoot	25	$+0.0039 \pm 0.0039$	$+0.0002 \pm 0.0001$	$+0.0031 \pm 0.0056$	$+0.0038 \pm 0.0042$
HGLISTA-overshoot	50	$+0.0077 \pm 0.0061$	$+0.0002 \pm 0.0001$	-0.0000 ± 0.0001	$+0.0078 \pm 0.0055$
HLISTA-CP	10	$+0.0024 \pm 0.0049$	$+0.0001 \pm 0.0003$	-0.7874 ± 5.1673	$+0.0024 \pm 0.0049$
HLISTA-CP	25	$+0.0213 \pm 0.0264$	$+0.0009 \pm 0.0012$	$+0.5230 \pm 1.2696$	$+0.0212 \pm 0.0264$
HLISTA-CP	50	$+0.0106 \pm 0.0082$	$+0.0003 \pm 0.0002$	-0.0982 ± 0.2514	$+0.0106 \pm 0.0082$
HLISTA-CPSS	10	$+0.0041 \pm 0.0039$	$+0.0002 \pm 0.0001$	$+0.7528 \pm 4.7890$	$+0.0041 \pm 0.0039$
HLISTA-CPSS	25	$+0.0168 \pm 0.0113$	$+0.0006 \pm 0.0004$	$+0.2598 \pm 0.7802$	$+0.0168 \pm 0.0113$
HLISTA-CPSS	50	$+0.0211 \pm 0.0104$	$+0.0005 \pm 0.0002$	$+0.4712 \pm 0.3787$	$+0.0211 \pm 0.0104$

Table 87: Frozen-history validation diagnostics: all four metrics, three-seed mean ± sample SD. These are within-checkpoint counterfactuals, not separately trained baselines.

Input	PSNR	SSIM	NMSE	NMSE (dB)
All four taps	$\mathbf{28.8000 \pm 0.0115}$	0.8383 ± 0.0003	$\mathbf{0.0092 \pm 0.0000}$	$\mathbf{-22.2099 \pm 0.0115}$
Current tap only	28.7427 ± 0.0162	0.8360 ± 0.0004	0.0093 ± 0.0000	-22.1527 ± 0.0162
Gain-matched current	28.7827 ± 0.0106	$\mathbf{0.8385 \pm 0.0003}$	0.0093 ± 0.0000	-22.1927 ± 0.0106
No feedback	28.5475 ± 0.0183	0.8280 ± 0.0005	0.0095 ± 0.0000	-21.9575 ± 0.0183

AB CROSS-BACKBONE VALIDATION AND HELD-OUT TESTING

All 13 groups, seven arms and three seeds are retained. Validation100 and one-shot test200 are distinct partitions; PSNR and SSIM are reported with same-seed contrasts. Auxiliary NMSE tables remain in the evidence archive. The current-input MLP has width 109 and 328 feedback parameters, matching LSTM width 8, not four-tap FIR. It applies a scalar-input two-layer tanh network independently per coordinate and carries no temporal state. Fixed momentum uses $q_0 = 1$, $q_{k+1} = (1 + \sqrt{1 + 4q_k^2})/2$ and $m_k = ((q_k - 1)/q_{k+1})r_k$: this is a prescribed gain on the current increment, not the FISTA iterate update. The frozen selected13 configuration factory supplies these choices; equal parameter counts do not imply equal compute. These tables retain both signs.

AB.1 VALIDATION100

Table 88: Confirmatory selected13 PSNR (dB). Each seed is first averaged over validation100; entries are mean ± sample SD over training seeds 0/1/2. No-feedback is the matched backbone baseline. Branches are shown on separate rows for legibility. MLP matches LSTM parameters.

Branch	Parent	FIR	Momentum	MLP
IID / HALISTA / 10%				
Both	21.475 ± 0.021	21.652 ± 0.041	21.405 ± 0.073	$\mathbf{21.680 \pm 0.040}$
Refinement	21.475 ± 0.021	21.603 ± 0.146	21.264 ± 0.010	$\mathbf{21.659 \pm 0.137}$
IID / HALISTA / 25%				
Both	$\mathbf{24.892 \pm 0.011}$	24.877 ± 0.005	24.827 ± 0.016	24.702 ± 0.324
Refinement	24.892 ± 0.011	$\mathbf{24.904 \pm 0.031}$	24.781 ± 0.005	24.783 ± 0.040
IID / HCISTA / 25%				
Both	10.119 ± 0.005	10.755 ± 0.001	10.931 ± 0.005	$\mathbf{22.153 \pm 0.092}$
Refinement	10.119 ± 0.005	10.328 ± 0.027	10.121 ± 0.003	$\mathbf{11.117 \pm 0.021}$
IID / HGLISTA-gain / 25%				
Both	26.707 ± 0.010	26.705 ± 0.011	26.709 ± 0.002	$\mathbf{26.722 \pm 0.006}$
Refinement	26.707 ± 0.010	$\mathbf{26.708 \pm 0.008}$	26.705 ± 0.009	26.705 ± 0.011
IID / HLISTA-CPSS / 10%				
Both	24.045 ± 0.013	24.053 ± 0.016	24.035 ± 0.013	$\mathbf{24.067 \pm 0.002}$
Refinement	24.045 ± 0.013	24.050 ± 0.015	24.050 ± 0.010	$\mathbf{24.073 \pm 0.007}$
IID / HLISTA-CPSS / 25%				
Both	26.943 ± 0.004	26.939 ± 0.018	26.944 ± 0.003	$\mathbf{26.989 \pm 0.026}$
Refinement	26.943 ± 0.004	26.954 ± 0.022	26.974 ± 0.018	$\mathbf{27.035 \pm 0.021}$
QR / HALISTA / 10%				
Both	21.866 ± 0.007	$\mathbf{21.867 \pm 0.036}$	21.526 ± 0.035	21.675 ± 0.209
Refinement	21.866 ± 0.007	21.913 ± 0.023	21.719 ± 0.104	$\mathbf{21.918 \pm 0.067}$
QR / HALISTA / 25%				
Both	$\mathbf{24.878 \pm 0.015}$	24.812 ± 0.023	24.780 ± 0.020	24.783 ± 0.090
Refinement	$\mathbf{24.878 \pm 0.015}$	24.854 ± 0.010	24.762 ± 0.041	24.803 ± 0.048
QR / HCISTA / 25%				
Both	19.814 ± 0.003	22.606 ± 0.021	22.632 ± 0.031	$\mathbf{23.733 \pm 0.175}$
Refinement	19.814 ± 0.003	20.325 ± 0.216	19.670 ± 0.260	$\mathbf{21.590 \pm 0.030}$
QR / HGLISTA-gain / 25%				
Both	$\mathbf{26.546 \pm 0.045}$	26.495 ± 0.015	26.418 ± 0.044	25.631 ± 0.800
Refinement	$\mathbf{26.546 \pm 0.045}$	26.513 ± 0.024	26.545 ± 0.028	26.390 ± 0.141
QR / HGLISTA-overshoot / 25%				
Both	26.503 ± 0.048	$\mathbf{26.504 \pm 0.046}$	26.230 ± 0.261	26.213 ± 0.138
Refinement	26.503 ± 0.048	$\mathbf{26.513 \pm 0.030}$	26.505 ± 0.031	26.197 ± 0.212
QR / HLISTA-CPSS / 10%				
Both	23.971 ± 0.014	23.985 ± 0.016	24.001 ± 0.018	$\mathbf{24.004 \pm 0.007}$
Refinement	23.971 ± 0.014	23.975 ± 0.021	23.974 ± 0.006	$\mathbf{23.988 \pm 0.011}$
QR / HLISTA-CPSS / 25%				
Both	26.899 ± 0.005	$\mathbf{26.906 \pm 0.007}$	$\mathbf{26.906 \pm 0.016}$	26.890 ± 0.002
Refinement	26.899 ± 0.005	26.912 ± 0.007	26.933 ± 0.011	$\mathbf{27.028 \pm 0.011}$

Table 89: Confirmatory selected13 SSIM. Each seed is first averaged over validation100; entries are mean ± sample SD over training seeds 0/1/2. No-feedback is the matched backbone baseline. Branches are shown on separate rows for legibility. MLP matches LSTM parameters.

Branch	Parent	FIR	Momentum	MLP
IID / HALISTA / 10%				
Both	0.4937 ± 0.0004	$\mathbf{0.5099 \pm 0.0019}$	0.4864 ± 0.0052	0.5039 ± 0.0013
Refinement	0.4937 ± 0.0004	0.5006 ± 0.0050	0.4839 ± 0.0030	$\mathbf{0.5070 \pm 0.0079}$
IID / HALISTA / 25%				
Both	$\mathbf{0.6901 \pm 0.0003}$	0.6889 ± 0.0005	0.6853 ± 0.0009	0.6798 ± 0.0155
Refinement	0.6901 ± 0.0003	$\mathbf{0.6905 \pm 0.0016}$	0.6839 ± 0.0002	0.6842 ± 0.0022
IID / HCISTA / 25%				
Both	0.1257 ± 0.0003	0.1515 ± 0.0003	0.1607 ± 0.0003	$\mathbf{0.5484 \pm 0.0055}$
Refinement	0.1257 ± 0.0003	0.1264 ± 0.0009	0.1257 ± 0.0002	$\mathbf{0.1547 \pm 0.0015}$
IID / HGLISTA-gain / 25%				
Both	0.7783 ± 0.0004	0.7783 ± 0.0005	0.7784 ± 0.0002	$\mathbf{0.7788 \pm 0.0002}$
Refinement	0.7783 ± 0.0004	$\mathbf{0.7784 \pm 0.0003}$	0.7783 ± 0.0004	0.7783 ± 0.0004
IID / HLISTA-CPSS / 10%				
Both	0.6463 ± 0.0007	0.6469 ± 0.0008	0.6465 ± 0.0009	$\mathbf{0.6474 \pm 0.0004}$
Refinement	0.6463 ± 0.0007	0.6465 ± 0.0007	0.6465 ± 0.0008	$\mathbf{0.6475 \pm 0.0003}$
IID / HLISTA-CPSS / 25%				
Both	0.7844 ± 0.0003	0.7842 ± 0.0007	0.7843 ± 0.0002	$\mathbf{0.7856 \pm 0.0007}$
Refinement	0.7844 ± 0.0003	0.7847 ± 0.0007	0.7850 ± 0.0006	$\mathbf{0.7864 \pm 0.0004}$
QR / HALISTA / 10%				
Both	$\mathbf{0.5187 \pm 0.0027}$	0.5163 ± 0.0035	0.4990 ± 0.0010	0.5077 ± 0.0114
Refinement	0.5187 ± 0.0027	$\mathbf{0.5197 \pm 0.0031}$	0.5107 ± 0.0049	0.5180 ± 0.0073
QR / HALISTA / 25%				
Both	$\mathbf{0.6877 \pm 0.0006}$	0.6833 ± 0.0015	0.6810 ± 0.0016	0.6821 ± 0.0045
Refinement	$\mathbf{0.6877 \pm 0.0006}$	0.6861 ± 0.0003	0.6811 ± 0.0020	0.6845 ± 0.0029
QR / HCISTA / 25%				
Both	0.4751 ± 0.0011	0.5902 ± 0.0007	0.5892 ± 0.0014	$\mathbf{0.6310 \pm 0.0069}$
Refinement	0.4751 ± 0.0011	0.4865 ± 0.0076	0.4716 ± 0.0077	$\mathbf{0.5482 \pm 0.0013}$
QR / HGLISTA-gain / 25%				
Both	$\mathbf{0.7709 \pm 0.0018}$	0.7691 ± 0.0012	0.7659 ± 0.0023	0.7200 ± 0.0441
Refinement	$\mathbf{0.7709 \pm 0.0018}$	0.7693 ± 0.0014	$\mathbf{0.7709 \pm 0.0012}$	0.7631 ± 0.0067
QR / HGLISTA-overshoot / 25%				
Both	0.7690 ± 0.0021	$\mathbf{0.7692 \pm 0.0017}$	0.7549 ± 0.0148	0.7537 ± 0.0076
Refinement	0.7690 ± 0.0021	$\mathbf{0.7696 \pm 0.0011}$	0.7689 ± 0.0015	0.7537 ± 0.0112
QR / HLISTA-CPSS / 10%				
Both	0.6410 ± 0.0008	0.6417 ± 0.0011	$\mathbf{0.6429 \pm 0.0009}$	0.6427 ± 0.0004
Refinement	0.6410 ± 0.0008	0.6413 ± 0.0013	0.6411 ± 0.0004	$\mathbf{0.6417 \pm 0.0007}$
QR / HLISTA-CPSS / 25%				
Both	0.7822 ± 0.0002	$\mathbf{0.7823 \pm 0.0002}$	0.7820 ± 0.0005	0.7816 ± 0.0000
Refinement	0.7822 ± 0.0002	0.7825 ± 0.0003	0.7829 ± 0.0003	$\mathbf{0.7850 \pm 0.0003}$

Table 90: Same-seed paired selected13 ΔPSNR (dB) (DASH arm minus no feedback), reported as mean ± sample SD over seeds 0/1/2. Positive favors the displayed arm for PSNR/SSIM. Branches are shown on separate rows for legibility. MLP matches LSTM parameters.

Branch	FIR	Momentum	MLP
IID / HALISTA / 10%			
Both	$+0.177 \pm 0.047$	-0.070 ± 0.053	$\mathbf{+0.205 \pm 0.034}$
Refinement	$+0.128 \pm 0.145$	-0.212 ± 0.029	$\mathbf{+0.184 \pm 0.125}$
IID / HALISTA / 25%			
Both	$\mathbf{-0.015 \pm 0.006}$	-0.064 ± 0.020	-0.190 ± 0.314
Refinement	$\mathbf{+0.012 \pm 0.020}$	-0.110 ± 0.008	-0.108 ± 0.052
IID / HCISTA / 25%			
Both	$+0.636 \pm 0.004$	$+0.811 \pm 0.003$	$\mathbf{+12.033 \pm 0.087}$
Refinement	$+0.209 \pm 0.022$	$+0.002 \pm 0.002$	$\mathbf{+0.997 \pm 0.022}$
IID / HGLISTA-gain / 25%			
Both	-0.001 ± 0.002	$+0.002 \pm 0.009$	$\mathbf{+0.015 \pm 0.006}$
Refinement	$\mathbf{+0.002 \pm 0.004}$	-0.002 ± 0.002	-0.001 ± 0.021
IID / HLISTA-CPSS / 10%			
Both	$+0.008 \pm 0.005$	-0.009 ± 0.001	$\mathbf{+0.023 \pm 0.011}$
Refinement	$+0.006 \pm 0.010$	$+0.006 \pm 0.008$	$\mathbf{+0.029 \pm 0.012}$
IID / HLISTA-CPSS / 25%			
Both	-0.004 ± 0.016	$+0.002 \pm 0.002$	$\mathbf{+0.046 \pm 0.029}$
Refinement	$+0.011 \pm 0.019$	$+0.031 \pm 0.014$	$\mathbf{+0.092 \pm 0.020}$
QR / HALISTA / 10%			
Both	$\mathbf{+0.001 \pm 0.030}$	-0.340 ± 0.028	-0.191 ± 0.203
Refinement	$+0.047 \pm 0.016$	-0.147 ± 0.099	$\mathbf{+0.052 \pm 0.071}$
QR / HALISTA / 25%			
Both	$\mathbf{-0.065 \pm 0.032}$	-0.098 ± 0.006	-0.095 ± 0.076
Refinement	$\mathbf{-0.023 \pm 0.020}$	-0.116 ± 0.045	-0.075 ± 0.062
QR / HCISTA / 25%			
Both	$+2.792 \pm 0.019$	$+2.818 \pm 0.033$	$\mathbf{+3.919 \pm 0.177}$
Refinement	$+0.511 \pm 0.219$	-0.144 ± 0.260	$\mathbf{+1.776 \pm 0.032}$
QR / HGLISTA-gain / 25%			
Both	$\mathbf{-0.051 \pm 0.055}$	-0.128 ± 0.023	-0.915 ± 0.783
Refinement	-0.033 ± 0.066	$\mathbf{-0.001 \pm 0.052}$	-0.156 ± 0.127
QR / HGLISTA-overshoot / 25%			
Both	$\mathbf{+0.001 \pm 0.035}$	-0.273 ± 0.300	-0.290 ± 0.143
Refinement	$\mathbf{+0.010 \pm 0.042}$	$+0.002 \pm 0.061$	-0.306 ± 0.164
QR / HLISTA-CPSS / 10%			
Both	$+0.014 \pm 0.004$	$+0.029 \pm 0.005$	$\mathbf{+0.032 \pm 0.019}$
Refinement	$+0.004 \pm 0.007$	$+0.002 \pm 0.007$	$\mathbf{+0.017 \pm 0.014}$
QR / HLISTA-CPSS / 25%			
Both	$\mathbf{+0.007 \pm 0.001}$	$+0.006 \pm 0.015$	-0.010 ± 0.004
Refinement	$+0.013 \pm 0.003$	$+0.033 \pm 0.007$	$\mathbf{+0.128 \pm 0.007}$

Table 91: Same-seed paired selected13 ΔSSIM (DASH arm minus no feedback), reported as mean ± sample SD over seeds 0/1/2. Positive favors the displayed arm for PSNR/SSIM. Branches are shown on separate rows for legibility. MLP matches LSTM parameters.

Branch	FIR	Momentum	MLP
IID / HALISTA / 10%			
Both	$\mathbf{+0.0162 \pm 0.0015}$	-0.0073 ± 0.0051	$+0.0103 \pm 0.0013$
Refinement	$+0.0070 \pm 0.0053$	-0.0098 ± 0.0026	$\mathbf{+0.0133 \pm 0.0083}$
IID / HALISTA / 25%			
Both	$\mathbf{-0.0012 \pm 0.0007}$	-0.0049 ± 0.0010	-0.0103 ± 0.0151
Refinement	$\mathbf{+0.0004 \pm 0.0014}$	-0.0062 ± 0.0002	-0.0059 ± 0.0025
IID / HCISTA / 25%			
Both	$+0.0258 \pm 0.0003$	$+0.0351 \pm 0.0002$	$\mathbf{+0.4228 \pm 0.0053}$
Refinement	$+0.0007 \pm 0.0007$	$+0.0000 \pm 0.0004$	$\mathbf{+0.0291 \pm 0.0018}$
IID / HGLISTA-gain / 25%			
Both	-0.0000 ± 0.0001	$+0.0001 \pm 0.0003$	$\mathbf{+0.0004 \pm 0.0002}$
Refinement	$\mathbf{+0.0001 \pm 0.0002}$	-0.0000 ± 0.0001	$+0.0000 \pm 0.0008$
IID / HLISTA-CPSS / 10%			
Both	$+0.0006 \pm 0.0001$	$+0.0002 \pm 0.0001$	$\mathbf{+0.0011 \pm 0.0010}$
Refinement	$+0.0002 \pm 0.0004$	$+0.0002 \pm 0.0002$	$\mathbf{+0.0012 \pm 0.0005}$
IID / HLISTA-CPSS / 25%			
Both	-0.0002 ± 0.0006	-0.0001 ± 0.0002	$\mathbf{+0.0012 \pm 0.0010}$
Refinement	$+0.0003 \pm 0.0005$	$+0.0006 \pm 0.0003$	$\mathbf{+0.0020 \pm 0.0005}$
QR / HALISTA / 10%			
Both	$\mathbf{-0.0023 \pm 0.0043}$	-0.0196 ± 0.0024	-0.0110 ± 0.0132
Refinement	$\mathbf{+0.0010 \pm 0.0058}$	-0.0079 ± 0.0071	-0.0007 ± 0.0094
QR / HALISTA / 25%			
Both	$\mathbf{-0.0044 \pm 0.0018}$	-0.0066 ± 0.0012	-0.0056 ± 0.0039
Refinement	$\mathbf{-0.0016 \pm 0.0008}$	-0.0066 ± 0.0020	-0.0032 ± 0.0035
QR / HCISTA / 25%			
Both	$+0.1151 \pm 0.0012$	$+0.1141 \pm 0.0025$	$\mathbf{+0.1558 \pm 0.0079}$
Refinement	$+0.0113 \pm 0.0087$	-0.0036 ± 0.0073	$\mathbf{+0.0730 \pm 0.0020}$
QR / HGLISTA-gain / 25%			
Both	$\mathbf{-0.0018 \pm 0.0028}$	-0.0050 ± 0.0014	-0.0509 ± 0.0431
Refinement	-0.0015 ± 0.0031	$\mathbf{-0.0000 \pm 0.0025}$	-0.0078 ± 0.0062
QR / HGLISTA-overshoot / 25%			
Both	$\mathbf{+0.0002 \pm 0.0015}$	-0.0141 ± 0.0165	-0.0153 ± 0.0078
Refinement	$\mathbf{+0.0006 \pm 0.0020}$	-0.0001 ± 0.0028	-0.0153 ± 0.0092
QR / HLISTA-CPSS / 10%			
Both	$+0.0007 \pm 0.0003$	$\mathbf{+0.0019 \pm 0.0002}$	$+0.0016 \pm 0.0009$
Refinement	$+0.0002 \pm 0.0005$	$+0.0001 \pm 0.0004$	$\mathbf{+0.0007 \pm 0.0007}$
QR / HLISTA-CPSS / 25%			
Both	$\mathbf{+0.0001 \pm 0.0001}$	-0.0002 ± 0.0005	-0.0006 ± 0.0002
Refinement	$+0.0003 \pm 0.0002$	$+0.0007 \pm 0.0003$	$\mathbf{+0.0028 \pm 0.0001}$

AB.2 FROZEN TEST200

Table 92: Frozen selected13 held-out test200 PSNR (dB). Each seed is first averaged over the same 200 images; cells are mean ± sample SD over seeds 0/1/2 (ddof=1). Baseline is the matched no-feedback backbone. Branches are shown on separate rows for legibility. MLP matches LSTM parameters.

Branch	Parent	FIR	Momentum	MLP
IID / HALISTA / 10%				
Both	21.434 ± 0.018	21.600 ± 0.038	21.367 ± 0.066	$\mathbf{21.630 \pm 0.043}$
Refinement	21.434 ± 0.018	21.558 ± 0.144	21.230 ± 0.010	$\mathbf{21.607 \pm 0.130}$
IID / HALISTA / 25%				
Both	$\mathbf{25.016 \pm 0.012}$	24.997 ± 0.007	24.943 ± 0.017	24.808 ± 0.352
Refinement	25.016 ± 0.012	$\mathbf{25.029 \pm 0.035}$	24.887 ± 0.004	24.897 ± 0.043
IID / HCISTA / 25%				
Both	10.089 ± 0.005	10.724 ± 0.001	10.899 ± 0.005	$\mathbf{22.106 \pm 0.096}$
Refinement	10.089 ± 0.005	10.299 ± 0.027	10.092 ± 0.003	$\mathbf{11.082 \pm 0.022}$
IID / HGLISTA-gain / 25%				
Both	26.930 ± 0.010	26.931 ± 0.012	26.936 ± 0.000	$\mathbf{26.950 \pm 0.008}$
Refinement	26.930 ± 0.010	$\mathbf{26.932 \pm 0.009}$	26.929 ± 0.010	$\mathbf{26.932 \pm 0.011}$
IID / HLISTA-CPSS / 10%				
Both	24.116 ± 0.011	24.126 ± 0.017	24.110 ± 0.012	$\mathbf{24.141 \pm 0.003}$
Refinement	24.116 ± 0.011	24.120 ± 0.017	24.120 ± 0.011	$\mathbf{24.145 \pm 0.005}$
IID / HLISTA-CPSS / 25%				
Both	27.194 ± 0.006	27.191 ± 0.020	27.199 ± 0.003	$\mathbf{27.247 \pm 0.030}$
Refinement	27.194 ± 0.006	27.204 ± 0.022	27.226 ± 0.020	$\mathbf{27.296 \pm 0.022}$
QR / HALISTA / 10%				
Both	$\mathbf{21.833 \pm 0.012}$	$\mathbf{21.833 \pm 0.037}$	21.467 ± 0.032	21.610 ± 0.208
Refinement	21.833 ± 0.012	$\mathbf{21.881 \pm 0.026}$	21.667 ± 0.107	21.873 ± 0.072
QR / HALISTA / 25%				
Both	$\mathbf{24.993 \pm 0.017}$	24.921 ± 0.028	24.887 ± 0.024	24.892 ± 0.101
Refinement	$\mathbf{24.993 \pm 0.017}$	24.966 ± 0.011	24.860 ± 0.051	24.916 ± 0.051
QR / HCISTA / 25%				
Both	19.830 ± 0.001	22.631 ± 0.020	22.676 ± 0.028	$\mathbf{23.759 \pm 0.187}$
Refinement	19.830 ± 0.001	20.362 ± 0.223	19.679 ± 0.264	$\mathbf{21.627 \pm 0.030}$
QR / HGLISTA-gain / 25%				
Both	$\mathbf{26.749 \pm 0.050}$	26.687 ± 0.015	26.623 ± 0.045	25.761 ± 0.861
Refinement	26.749 ± 0.050	26.715 ± 0.023	$\mathbf{26.750 \pm 0.028}$	26.579 ± 0.151
QR / HGLISTA-overshoot / 25%				
Both	$\mathbf{26.702 \pm 0.050}$	$\mathbf{26.702 \pm 0.049}$	26.421 ± 0.282	26.393 ± 0.146
Refinement	26.702 ± 0.050	$\mathbf{26.714 \pm 0.033}$	26.707 ± 0.031	26.372 ± 0.230
QR / HLISTA-CPSS / 10%				
Both	24.029 ± 0.014	24.045 ± 0.017	24.065 ± 0.020	$\mathbf{24.066 \pm 0.006}$
Refinement	24.029 ± 0.014	24.034 ± 0.021	24.031 ± 0.008	$\mathbf{24.047 \pm 0.012}$
QR / HLISTA-CPSS / 25%				
Both	27.146 ± 0.008	$\mathbf{27.154 \pm 0.009}$	27.151 ± 0.017	27.135 ± 0.002
Refinement	27.146 ± 0.008	27.160 ± 0.010	27.183 ± 0.012	$\mathbf{27.287 \pm 0.011}$

Table 93: Frozen selected13 held-out test200 SSIM. Each seed is first averaged over the same 200 images; cells are mean ± sample SD over seeds 0/1/2 (ddof=1). Baseline is the matched no-feedback backbone. Branches are shown on separate rows for legibility. MLP matches LSTM parameters.

Branch	Parent	FIR	Momentum	MLP
IID / HALISTA / 10%				
Both	0.5180 ± 0.0002	$\mathbf{0.5323 \pm 0.0019}$	0.5080 ± 0.0049	0.5269 ± 0.0011
Refinement	0.5180 ± 0.0002	0.5242 ± 0.0048	0.5071 ± 0.0027	$\mathbf{0.5297 \pm 0.0071}$
IID / HALISTA / 25%				
Both	$\mathbf{0.7136 \pm 0.0003}$	0.7121 ± 0.0006	0.7082 ± 0.0008	0.7031 ± 0.0163
Refinement	0.7136 ± 0.0003	$\mathbf{0.7142 \pm 0.0019}$	0.7063 ± 0.0002	0.7075 ± 0.0022
IID / HCISTA / 25%				
Both	0.1217 ± 0.0003	0.1482 ± 0.0003	0.1579 ± 0.0003	$\mathbf{0.5631 \pm 0.0059}$
Refinement	0.1217 ± 0.0003	0.1227 ± 0.0008	0.1218 ± 0.0002	$\mathbf{0.1509 \pm 0.0017}$
IID / HGLISTA-gain / 25%				
Both	0.8011 ± 0.0004	0.8011 ± 0.0005	0.8013 ± 0.0002	$\mathbf{0.8017 \pm 0.0003}$
Refinement	0.8011 ± 0.0004	$\mathbf{0.8012 \pm 0.0002}$	0.8011 ± 0.0003	$\mathbf{0.8012 \pm 0.0004}$
IID / HLISTA-CPSS / 10%				
Both	0.6743 ± 0.0007	0.6750 ± 0.0008	0.6746 ± 0.0009	$\mathbf{0.6753 \pm 0.0005}$
Refinement	0.6743 ± 0.0007	0.6743 ± 0.0008	0.6744 ± 0.0007	$\mathbf{0.6753 \pm 0.0003}$
IID / HLISTA-CPSS / 25%				
Both	0.8080 ± 0.0004	0.8079 ± 0.0007	0.8079 ± 0.0003	$\mathbf{0.8094 \pm 0.0008}$
Refinement	0.8080 ± 0.0004	0.8082 ± 0.0008	0.8085 ± 0.0007	$\mathbf{0.8103 \pm 0.0005}$
QR / HALISTA / 10%				
Both	$\mathbf{0.5418 \pm 0.0027}$	0.5394 ± 0.0037	0.5199 ± 0.0012	0.5287 ± 0.0115
Refinement	0.5418 ± 0.0027	$\mathbf{0.5431 \pm 0.0030}$	0.5321 ± 0.0051	0.5410 ± 0.0075
QR / HALISTA / 25%				
Both	$\mathbf{0.7113 \pm 0.0008}$	0.7066 ± 0.0016	0.7042 ± 0.0016	0.7056 ± 0.0049
Refinement	$\mathbf{0.7113 \pm 0.0008}$	0.7096 ± 0.0003	0.7035 ± 0.0025	0.7086 ± 0.0033
QR / HCISTA / 25%				
Both	0.5021 ± 0.0010	0.6106 ± 0.0006	0.6110 ± 0.0014	$\mathbf{0.6504 \pm 0.0080}$
Refinement	0.5021 ± 0.0010	0.5147 ± 0.0075	0.4968 ± 0.0085	$\mathbf{0.5693 \pm 0.0011}$
QR / HGLISTA-gain / 25%				
Both	$\mathbf{0.7939 \pm 0.0018}$	0.7916 ± 0.0012	0.7888 ± 0.0022	0.7424 ± 0.0441
Refinement	$\mathbf{0.7939 \pm 0.0018}$	0.7924 ± 0.0012	$\mathbf{0.7939 \pm 0.0012}$	0.7861 ± 0.0068
QR / HGLISTA-overshoot / 25%				
Both	$\mathbf{0.7920 \pm 0.0020}$	0.7919 ± 0.0017	0.7780 ± 0.0146	0.7767 ± 0.0075
Refinement	0.7920 ± 0.0020	$\mathbf{0.7925 \pm 0.0012}$	0.7920 ± 0.0015	0.7765 ± 0.0113
QR / HLISTA-CPSS / 10%				
Both	0.6681 ± 0.0009	0.6690 ± 0.0012	$\mathbf{0.6702 \pm 0.0010}$	0.6696 ± 0.0003
Refinement	0.6681 ± 0.0009	0.6684 ± 0.0014	0.6682 ± 0.0004	$\mathbf{0.6688 \pm 0.0007}$
QR / HLISTA-CPSS / 25%				
Both	0.8059 ± 0.0002	$\mathbf{0.8061 \pm 0.0003}$	0.8058 ± 0.0005	0.8054 ± 0.0000
Refinement	0.8059 ± 0.0002	0.8062 ± 0.0004	0.8067 ± 0.0004	$\mathbf{0.8091 \pm 0.0003}$

Table 94: Same-seed paired selected13 held-out test200 PSNR (dB) improvement (DASH minus baseline), followed by cross-seed mean ± sample SD (ddof=1). Positive favors DASH for every displayed improvement. Branches are shown on separate rows for legibility. MLP matches LSTM parameters.

Branch	FIR	Momentum	MLP
IID / HALISTA / 10%			
Both	$+0.166 \pm 0.049$	-0.067 ± 0.051	$\mathbf{+0.196 \pm 0.044}$
Refinement	$+0.124 \pm 0.137$	-0.204 ± 0.028	$\mathbf{+0.173 \pm 0.114}$
IID / HALISTA / 25%			
Both	$\mathbf{-0.018 \pm 0.007}$	-0.073 ± 0.022	-0.208 ± 0.342
Refinement	$\mathbf{+0.013 \pm 0.024}$	-0.128 ± 0.009	-0.118 ± 0.055
IID / HCISTA / 25%			
Both	$+0.635 \pm 0.004$	$+0.810 \pm 0.003$	$\mathbf{+12.017 \pm 0.091}$
Refinement	$+0.210 \pm 0.022$	$+0.002 \pm 0.002$	$\mathbf{+0.993 \pm 0.023}$
IID / HGLISTA-gain / 25%			
Both	$+0.000 \pm 0.003$	$+0.006 \pm 0.010$	$\mathbf{+0.020 \pm 0.003}$
Refinement	$\mathbf{+0.002 \pm 0.004}$	-0.001 ± 0.003	$\mathbf{+0.002 \pm 0.021}$
IID / HLISTA-CPSS / 10%			
Both	$+0.010 \pm 0.007$	-0.006 ± 0.002	$\mathbf{+0.025 \pm 0.014}$
Refinement	$+0.004 \pm 0.008$	$+0.004 \pm 0.007$	$\mathbf{+0.029 \pm 0.010}$
IID / HLISTA-CPSS / 25%			
Both	-0.004 ± 0.019	$+0.005 \pm 0.005$	$\mathbf{+0.053 \pm 0.036}$
Refinement	$+0.009 \pm 0.016$	$+0.031 \pm 0.015$	$\mathbf{+0.102 \pm 0.023}$
QR / HALISTA / 10%			
Both	$\mathbf{+0.000 \pm 0.025}$	-0.366 ± 0.021	-0.223 ± 0.198
Refinement	$\mathbf{+0.048 \pm 0.015}$	-0.165 ± 0.098	$+0.040 \pm 0.076$
QR / HALISTA / 25%			
Both	$\mathbf{-0.072 \pm 0.039}$	-0.106 ± 0.007	-0.101 ± 0.083
Refinement	$\mathbf{-0.027 \pm 0.022}$	-0.133 ± 0.054	-0.078 ± 0.068
QR / HCISTA / 25%			
Both	$+2.801 \pm 0.019$	$+2.846 \pm 0.029$	$\mathbf{+3.929 \pm 0.187}$
Refinement	$+0.532 \pm 0.223$	-0.151 ± 0.265	$\mathbf{+1.797 \pm 0.031}$
QR / HGLISTA-gain / 25%			
Both	$\mathbf{-0.062 \pm 0.059}$	-0.126 ± 0.023	-0.988 ± 0.843
Refinement	-0.034 ± 0.070	$\mathbf{+0.001 \pm 0.054}$	-0.170 ± 0.136
QR / HGLISTA-overshoot / 25%			
Both	$\mathbf{-0.001 \pm 0.037}$	-0.282 ± 0.323	-0.309 ± 0.153
Refinement	$\mathbf{+0.012 \pm 0.048}$	$+0.004 \pm 0.063$	-0.330 ± 0.180
QR / HLISTA-CPSS / 10%			
Both	$+0.016 \pm 0.004$	$+0.036 \pm 0.007$	$\mathbf{+0.037 \pm 0.018}$
Refinement	$+0.004 \pm 0.008$	$+0.002 \pm 0.006$	$\mathbf{+0.018 \pm 0.014}$
QR / HLISTA-CPSS / 25%			
Both	$\mathbf{+0.008 \pm 0.003}$	$+0.006 \pm 0.017$	-0.010 ± 0.005
Refinement	$+0.014 \pm 0.004$	$+0.037 \pm 0.006$	$\mathbf{+0.141 \pm 0.004}$

Table 95: Same-seed paired selected13 held-out test200 SSIM improvement (DASH minus baseline), followed by cross-seed mean ± sample SD (ddof=1). Positive favors DASH for every displayed improvement. Branches are shown on separate rows for legibility. MLP matches LSTM parameters.

Branch	FIR	Momentum	MLP
IID / HALISTA / 10%			
Both	$\mathbf{+0.0143 \pm 0.0018}$	-0.0100 ± 0.0048	$+0.0089 \pm 0.0009$
Refinement	$+0.0062 \pm 0.0050$	-0.0109 ± 0.0025	$\mathbf{+0.0117 \pm 0.0072}$
IID / HALISTA / 25%			
Both	$\mathbf{-0.0014 \pm 0.0007}$	-0.0053 ± 0.0009	-0.0105 ± 0.0160
Refinement	$\mathbf{+0.0006 \pm 0.0017}$	-0.0072 ± 0.0004	-0.0060 ± 0.0025
IID / HCISTA / 25%			
Both	$+0.0265 \pm 0.0002$	$+0.0362 \pm 0.0001$	$\mathbf{+0.4414 \pm 0.0056}$
Refinement	$+0.0009 \pm 0.0007$	$+0.0000 \pm 0.0004$	$\mathbf{+0.0292 \pm 0.0018}$
IID / HGLISTA-gain / 25%			
Both	$+0.0001 \pm 0.0001$	$+0.0002 \pm 0.0004$	$\mathbf{+0.0006 \pm 0.0002}$
Refinement	$\mathbf{+0.0001 \pm 0.0002}$	-0.0000 ± 0.0001	$\mathbf{+0.0001 \pm 0.0007}$
IID / HLISTA-CPSS / 10%			
Both	$+0.0007 \pm 0.0002$	$+0.0003 \pm 0.0002$	$\mathbf{+0.0011 \pm 0.0010}$
Refinement	$+0.0001 \pm 0.0003$	$+0.0001 \pm 0.0002$	$\mathbf{+0.0011 \pm 0.0004}$
IID / HLISTA-CPSS / 25%			
Both	-0.0002 ± 0.0006	-0.0001 ± 0.0003	$\mathbf{+0.0013 \pm 0.0011}$
Refinement	$+0.0002 \pm 0.0005$	$+0.0005 \pm 0.0003$	$\mathbf{+0.0022 \pm 0.0006}$
QR / HALISTA / 10%			
Both	$\mathbf{-0.0024 \pm 0.0040}$	-0.0219 ± 0.0023	-0.0131 ± 0.0130
Refinement	$\mathbf{+0.0012 \pm 0.0055}$	-0.0098 ± 0.0071	-0.0008 ± 0.0098
QR / HALISTA / 25%			
Both	$\mathbf{-0.0047 \pm 0.0019}$	-0.0071 ± 0.0009	-0.0057 ± 0.0042
Refinement	$\mathbf{-0.0017 \pm 0.0009}$	-0.0078 ± 0.0025	-0.0027 ± 0.0041
QR / HCISTA / 25%			
Both	$+0.1084 \pm 0.0010$	$+0.1089 \pm 0.0024$	$\mathbf{+0.1482 \pm 0.0089}$
Refinement	$+0.0125 \pm 0.0085$	-0.0053 ± 0.0082	$\mathbf{+0.0672 \pm 0.0018}$
QR / HGLISTA-gain / 25%			
Both	$\mathbf{-0.0022 \pm 0.0028}$	-0.0051 ± 0.0012	-0.0514 ± 0.0432
Refinement	-0.0015 ± 0.0029	$\mathbf{+0.0000 \pm 0.0023}$	-0.0077 ± 0.0062
QR / HGLISTA-overshoot / 25%			
Both	$\mathbf{-0.0000 \pm 0.0015}$	-0.0139 ± 0.0162	-0.0152 ± 0.0077
Refinement	$\mathbf{+0.0005 \pm 0.0020}$	-0.0000 ± 0.0027	-0.0155 ± 0.0093
QR / HLISTA-CPSS / 10%			
Both	$+0.0009 \pm 0.0003$	$\mathbf{+0.0021 \pm 0.0002}$	$+0.0015 \pm 0.0010$
Refinement	$+0.0002 \pm 0.0005$	$+0.0000 \pm 0.0005$	$\mathbf{+0.0007 \pm 0.0008}$
QR / HLISTA-CPSS / 25%			
Both	$\mathbf{+0.0001 \pm 0.0002}$	-0.0002 ± 0.0006	-0.0006 ± 0.0002
Refinement	$+0.0003 \pm 0.0003$	$+0.0007 \pm 0.0003$	$\mathbf{+0.0032 \pm 0.0001}$

AC TERMINAL AMP-FAMILY VALIDATION COMPARISON

All eight registered protocols have six accepted QR/IID–rate conditions. The tables keep deterministic image-level variability separate from learned-model training-seed variability and report all four registered metrics.

Table 96: QR deterministic AMP protocols on validation100. Entries are image-population mean ± sample SD; this SD is not training-seed variability.

Protocol	Rate	PSNR (dB)	SSIM	NMSE	NMSE (dB)
AMP depth16	10%	12.9543 ± 1.7579	0.17974 ± 0.07201	0.24552 ± 0.05275	-6.1934 ± 0.9012
	25%	23.1562 ± 2.9342	0.58712 ± 0.10346	0.02962 ± 0.02238	-16.3952 ± 3.2899
	50%	27.1358 ± 3.4609	0.79391 ± 0.08172	0.01290 ± 0.01207	-20.3748 ± 3.7921
AMP validation-selected depth	10%	$\mathbf{18.7781 \pm 2.8974}$	$\mathbf{0.41785 \pm 0.14263}$	$\mathbf{0.08017 \pm 0.05380}$	$\mathbf{-12.0172 \pm 3.3101}$
	25%	$\mathbf{23.5581 \pm 3.1717}$	$\mathbf{0.62451 \pm 0.11874}$	$\mathbf{0.02797 \pm 0.02247}$	$\mathbf{-16.7972 \pm 3.5622}$
	50%	$\mathbf{27.1893 \pm 3.4832}$	$\mathbf{0.79585 \pm 0.08175}$	$\mathbf{0.01279 \pm 0.01204}$	$\mathbf{-20.4284 \pm 3.8130}$

Table 97: QR learned AMP-family protocols on validation100. Each seed is first averaged over 100 images; entries are mean ± sample SD over training seeds 0/1/2.

Protocol	Rate	PSNR (dB)	SSIM	NMSE	NMSE (dB)
LAMP 13k	10%	22.9809 ± 0.1237	0.58402 ± 0.00847	0.03177 ± 0.00090	-16.2200 ± 0.1237
	25%	26.2063 ± 0.0123	0.75680 ± 0.00069	0.01593 ± 0.00004	-19.4454 ± 0.0123
	50%	$\mathbf{30.5316 \pm 0.0012}$	$\mathbf{0.89190 \pm 0.00002}$	$\mathbf{0.00617 \pm 0.00000}$	$\mathbf{-23.7707 \pm 0.0012}$
progressive tied-LAMP	10%	22.1527 ± 0.4333	0.53545 ± 0.02061	0.03759 ± 0.00356	-15.3918 ± 0.4333
	25%	25.1723 ± 0.3902	0.69705 ± 0.02058	0.01936 ± 0.00136	-18.4113 ± 0.3902
	50%	29.9261 ± 0.2162	0.87975 ± 0.00474	0.00675 ± 0.00021	-23.1652 ± 0.2162
progressive untied-LAMP	10%	$\mathbf{23.5431 \pm 0.0010}$	$\mathbf{0.62967 \pm 0.00010}$	$\mathbf{0.02794 \pm 0.00000}$	$\mathbf{-16.7822 \pm 0.0010}$
	25%	$\mathbf{26.3919 \pm 0.0533}$	$\mathbf{0.76833 \pm 0.00194}$	$\mathbf{0.01504 \pm 0.00017}$	$\mathbf{-19.6310 \pm 0.0533}$
	50%	29.6122 ± 0.7874	0.87155 ± 0.02149	0.00748 ± 0.00127	-22.8513 ± 0.7874
LDAMP original depth	10%	10.0588 ± 2.2656	0.21990 ± 0.02104	0.62157 ± 0.31489	-3.2979 ± 2.2656
	25%	7.7228 ± 1.8190	0.10425 ± 0.11028	0.93664 ± 0.42234	-0.9619 ± 1.8190
	50%	24.8210 ± 3.9270	0.70216 ± 0.19204	0.04859 ± 0.06181	-18.0600 ± 3.9270
LDAMP validation-selected depth	10%	23.2170 ± 0.0660	0.57671 ± 0.00650	0.02949 ± 0.00040	-16.4561 ± 0.0660
	25%	24.9269 ± 0.0484	0.66943 ± 0.01175	0.02022 ± 0.00028	-18.1660 ± 0.0484
	50%	24.0297 ± 3.8685	0.65504 ± 0.14211	0.07321 ± 0.10002	-17.2688 ± 3.8685
LVAMP stabilized	10%	18.4661 ± 0.2533	0.36284 ± 0.00842	0.08247 ± 0.00462	-11.7051 ± 0.2533
	25%	23.2258 ± 0.0054	0.57827 ± 0.00051	0.02916 ± 0.00003	-16.4649 ± 0.0054
	50%	26.5612 ± 0.0123	0.74512 ± 0.00073	0.01356 ± 0.00003	-19.8003 ± 0.0123

Table 98: IID deterministic AMP protocols on validation100. Entries are image-population mean ± sample SD; this SD is not training-seed variability.

Protocol	Rate	PSNR (dB)	SSIM	NMSE	NMSE (dB)
AMP depth16	10%	18.3791 ± 2.7683	0.40572 ± 0.13712	0.08582 ± 0.05479	-11.6182 ± 3.1382
	25%	$\mathbf{23.2645 \pm 3.1746}$	$\mathbf{0.61720 \pm 0.11663}$	$\mathbf{0.02993 \pm 0.02402}$	$\mathbf{-16.5036 \pm 3.5596}$
	50%	27.0613 ± 3.4785	0.79573 ± 0.07291	$\mathbf{0.01313 \pm 0.01223}$	-20.3004 ± 3.8072
AMP validation-selected depth	10%	$\mathbf{18.4223 \pm 2.7868}$	$\mathbf{0.40829 \pm 0.13820}$	$\mathbf{0.08520 \pm 0.05468}$	$\mathbf{-11.6614 \pm 3.1632}$
	25%	$\mathbf{23.2645 \pm 3.1746}$	$\mathbf{0.61720 \pm 0.11663}$	$\mathbf{0.02993 \pm 0.02402}$	$\mathbf{-16.5036 \pm 3.5596}$
	50%	$\mathbf{27.0758 \pm 3.4990}$	$\mathbf{0.79649 \pm 0.07355}$	$\mathbf{0.01313 \pm 0.01229}$	$\mathbf{-20.3149 \pm 3.8245}$

The registered LVAMP learning-rate search retained 1 failed unused candidate(s): anonymized compute host gaussian_cs_lvamp_stabilized_r25_lr1e-3_s0: FloatingPointError: Nonfinite active gradient: log_noise_variance. Each had a nonzero return code and was excluded only by the registered successful-candidate selection rule; no rate-by-seed result was replaced.

The QR 10% LAMP seed-2 entry is the pre-registered diagnostic stop-rule recovery; its original zero-gradient failure remains retained and it is not described as an untouched or official reproduction. LVAMP uses the registered seed-0/r25 validation-only learning-rate selection and transfers that choice to all rates and seeds. This AMP-family evidence is validation-only and contains no matched external-baseline test200 ranking; the later selected13 held-out evaluation is a separate internal comparison.

Table 99: IID learned AMP-family protocols on validation100. Each seed is first averaged over 100 images; entries are mean ± sample SD over training seeds 0/1/2.

Protocol	Rate	PSNR (dB)	SSIM	NMSE	NMSE (dB)
LAMP 13k	10%	22.6138 ± 0.0613	0.56259 ± 0.00786	0.03398 ± 0.00039	-15.8528 ± 0.0613
	25%	25.4783 ± 0.0178	0.72874 ± 0.00073	0.01868 ± 0.00007	-18.7174 ± 0.0178
	50%	29.2330 ± 0.0362	0.86148 ± 0.00087	0.00838 ± 0.00007	-22.4721 ± 0.0362
progressive tied-LAMP	10%	22.6192 ± 0.1316	0.55362 ± 0.00801	0.03388 ± 0.00080	-15.8583 ± 0.1316
	25%	25.2786 ± 0.3011	0.69993 ± 0.02027	0.01923 ± 0.00091	-18.5177 ± 0.3011
	50%	29.8462 ± 0.2757	0.87743 ± 0.00833	0.00694 ± 0.00043	-23.0853 ± 0.2757
progressive untied-LAMP	10%	$\mathbf{23.5726 \pm 0.0031}$	$\mathbf{0.63313 \pm 0.00045}$	$\mathbf{0.02779 \pm 0.00002}$	$\mathbf{-16.8117 \pm 0.0031}$
	25%	$\mathbf{26.4659 \pm 0.0585}$	$\mathbf{0.77149 \pm 0.00177}$	$\mathbf{0.01481 \pm 0.00017}$	$\mathbf{-19.7050 \pm 0.0585}$
	50%	29.8844 ± 0.3599	0.87946 ± 0.00907	0.00703 ± 0.00051	-23.1235 ± 0.3599
LDAMP original depth	10%	19.5650 ± 1.3198	0.39810 ± 0.07403	0.06032 ± 0.01468	-12.8040 ± 1.3198
	25%	25.1743 ± 1.7365	0.69145 ± 0.09621	4.29869 ± 7.38570	-18.4134 ± 1.7365
	50%	30.0872 ± 1.1441	0.88242 ± 0.01666	0.00752 ± 0.00245	-23.3263 ± 1.1441
LDAMP validation-selected depth	10%	22.5041 ± 0.0929	0.50876 ± 0.00348	0.03360 ± 0.00032	-15.7432 ± 0.0929
	25%	25.1743 ± 1.7365	0.69145 ± 0.09621	4.29869 ± 7.38570	-18.4134 ± 1.7365
	50%	$\mathbf{30.3804 \pm 0.2324}$	$\mathbf{0.88303 \pm 0.00851}$	$\mathbf{0.00643 \pm 0.00041}$	$\mathbf{-23.6195 \pm 0.2324}$
LVAMP stabilized	10%	19.0646 ± 0.0552	0.39096 ± 0.00650	0.07428 ± 0.00113	-12.3037 ± 0.0552
	25%	23.5404 ± 0.0057	0.61439 ± 0.00084	0.02776 ± 0.00004	-16.7795 ± 0.0057
	50%	17.8998 ± 10.4070	0.40143 ± 0.40043	0.35720 ± 0.55719	-11.1389 ± 10.4070

AD CLASSICAL ISTA/FISTA COMPLEMENTS

FISTA uses accelerated proximal-gradient updates (Beck & Teboulle, 2009). Classical fixed-depth and validation-selected-depth results use image SD; AMP/LDAMP entries are presented only in the terminal AMP panel. No outcomes are removed from the archive.

Table 100: Classical complements on validation100. SD is across images, not training seeds. Fixed-depth and validation-selected-depth protocols remain separate; every solver depth is shown.

Method	Rate	Depth	PSNR (dB)	SSIM	NMSE	NMSE (dB)
QR, fixed depth						
FISTA	10%	16	$\mathbf{11.3501 \pm 2.3414}$	$\mathbf{0.19803 \pm 0.09031}$	$\mathbf{0.35303 \pm 0.06121}$	$\mathbf{-4.5892 \pm 0.7783}$
ISTA	10%	16	8.7631 ± 2.1415	0.09204 ± 0.04054	0.63168 ± 0.03624	-2.0022 ± 0.2509
FISTA	25%	16	$\mathbf{22.2066 \pm 2.4035}$	$\mathbf{0.51850 \pm 0.12687}$	$\mathbf{0.03525 \pm 0.02354}$	$\mathbf{-15.4457 \pm 2.9928}$
ISTA	25%	16	18.7858 ± 1.6037	0.40587 ± 0.14454	0.07129 ± 0.03585	-12.0249 ± 2.2827
FISTA	50%	16	$\mathbf{24.8433 \pm 3.4515}$	$\mathbf{0.71105 \pm 0.09341}$	$\mathbf{0.01913 \pm 0.01322}$	$\mathbf{-18.0824 \pm 2.9340}$
ISTA	50%	16	23.7485 ± 2.4677	0.61385 ± 0.10405	0.02381 ± 0.01527	-16.9876 ± 2.6326
IID, fixed depth						
FISTA	10%	16	$\mathbf{7.8245 \pm 2.1073}$	$\mathbf{0.05621 \pm 0.03983}$	$\mathbf{0.78307 \pm 0.02138}$	$\mathbf{-1.0636 \pm 0.1191}$
ISTA	10%	16	7.3944 ± 2.0559	0.03694 ± 0.02281	0.86433 ± 0.00970	-0.6335 ± 0.0488
FISTA	25%	16	$\mathbf{11.4186 \pm 2.4913}$	$\mathbf{0.20352 \pm 0.10049}$	$\mathbf{0.34553 \pm 0.04791}$	$\mathbf{-4.6577 \pm 0.6168}$
ISTA	25%	16	9.0990 ± 2.1602	0.11631 ± 0.06470	0.58421 ± 0.02446	-2.3381 ± 0.1827
FISTA	50%	16	$\mathbf{16.3523 \pm 2.7626}$	$\mathbf{0.37930 \pm 0.11258}$	$\mathbf{0.11396 \pm 0.02994}$	$\mathbf{-9.5914 \pm 1.2075}$
ISTA	50%	16	10.9300 ± 2.2066	0.20690 ± 0.08014	0.38366 ± 0.02405	-4.1691 ± 0.2739
QR, validation-selected depth						
FISTA	10%	512	$\mathbf{18.8387 \pm 2.9354}$	0.42552 ± 0.14582	$\mathbf{0.07962 \pm 0.05449}$	$\mathbf{-12.0778 \pm 3.3573}$
ISTA	10%	2048	18.7471 ± 2.7248	$\mathbf{0.42630 \pm 0.14158}$	0.07970 ± 0.05291	-11.9862 ± 3.1795
FISTA	25%	128	$\mathbf{23.4201 \pm 3.1993}$	0.62447 ± 0.11529	$\mathbf{0.02901 \pm 0.02361}$	$\mathbf{-16.6592 \pm 3.5861}$
ISTA	25%	2048	23.4198 ± 3.2000	$\mathbf{0.62450 \pm 0.11527}$	$\mathbf{0.02901 \pm 0.02363}$	-16.6589 ± 3.5871
FISTA	50%	128	$\mathbf{27.3434 \pm 3.4719}$	0.80408 ± 0.06973	$\mathbf{0.01236 \pm 0.01177}$	$\mathbf{-20.5825 \pm 3.8034}$
ISTA	50%	512	27.3392 ± 3.4702	$\mathbf{0.80409 \pm 0.06966}$	$\mathbf{0.01236 \pm 0.01175}$	-20.5783 ± 3.8012
IID, validation-selected depth						
FISTA	10%	512	$\mathbf{18.4909 \pm 2.7990}$	$\mathbf{0.41786 \pm 0.14078}$	$\mathbf{0.08392 \pm 0.05364}$	$\mathbf{-11.7300 \pm 3.1710}$
ISTA	10%	2048	18.2387 ± 2.6911	0.41537 ± 0.13956	0.08785 ± 0.05470	-11.4778 ± 3.0621
FISTA	25%	512	$\mathbf{23.3299 \pm 3.1950}$	$\mathbf{0.62308 \pm 0.11515}$	$\mathbf{0.02961 \pm 0.02408}$	$\mathbf{-16.5690 \pm 3.5806}$
ISTA	25%	2048	23.2058 ± 3.0264	0.60629 ± 0.11553	0.02987 ± 0.02347	-16.4449 ± 3.4383
FISTA	50%	512	$\mathbf{27.1896 \pm 3.4797}$	$\mathbf{0.80028 \pm 0.06958}$	$\mathbf{0.01276 \pm 0.01201}$	$\mathbf{-20.4287 \pm 3.7932}$
ISTA	50%	2048	26.7565 ± 3.2573	0.77494 ± 0.07333	0.01368 ± 0.01216	-19.9956 ± 3.6205

AE CLASSICAL AMP ON COCO DEVELOPMENT DATA

This is an absolute validation reference, not a held-out test ranking or a learned AMP comparison.

Table 101: Classical AMP on COCO2017 validation (mean ± image-level sample SD over 5,000 images; depth 16).

Operator	Rate	PSNR (dB)	SSIM	NMSE	NMSE (dB)
IID	10%	-26.6617±1.7941	0.0020±0.0181	1916.5963±402.1401	32.7401±0.8420
IID	25%	20.0948±3.1924	0.4633±0.0930	0.0530±0.0436	-14.0164±3.3898
IID	50%	27.3960±4.2555	0.8104±0.0822	0.0117±0.0125	-21.3177±4.3917
QR	10%	7.6943±1.9150	0.1791±0.0753	0.6905±0.0416	-1.6159±0.2533
QR	25%	22.3658±3.1423	0.5862±0.1333	0.0311±0.0252	-16.2874±3.3321
QR	50%	27.4535±4.2443	0.8106±0.0901	0.0115±0.0122	-21.3752±4.3813

METHOD DEFINITIONS AND CONDITIONAL PROOFS

AF MEMORY ENVELOPES AND ACCUMULATED FEEDBACK BOUNDS

AF.1 A SUFFICIENT ACCUMULATED-GAIN CONDITION

Definition 9 (Memory contract)**.** For zero-initialized causal memory, coordinate-wise zero response and a uniform history envelope require

$$r_{t,i} = 0 \ (t \le k) \Longrightarrow m_{k,i} = 0, \tag{28}$$

$$\|m_k\| \le C_m \sum_{t=0}^{k} \tau_m^{k-t} \|r_t\|, \quad C_m \ge 0, \ 0 \le \tau_m < 1. \tag{29}$$

The norm and constants cover the claimed horizon and input class; the sparse-recovery route uses the ℓ_1 norm.

Coordinate-wise zero response is stronger than a global zero-input test. FIR satisfies it, and for $0 < \tau_m < 1$ admits

$$C_m = \sup_k \max_{j<J_k} |b_{k,j}| \tau_m^{-j}. \tag{30}$$

Other module bounds and support conditions are in Appendix AK.

For complete state $\xi = (x, s)$, fixed observations and injection mode $\delta \in \{0, 1\}$, define

$$\mathcal{B}_k^{\delta}(\xi, d; y) = C_{\nu,k}(v + d, v + \delta d; c, y), \quad (v, c) = F_{\nu,k}(\xi; y). \tag{31}$$

On an invariant domain and in one common norm, assume uniformly

$$\begin{aligned} \|\mathcal{B}_k(\xi, 0) - \xi^\star\| &\le q\|\xi - \xi^\star\|, \quad q < 1, \\ \|\mathcal{B}_k(\xi, d) - \mathcal{B}_k(\xi, 0)\| &\le L_d\|d\|, \quad \|r_k\| \le C_r\|\xi_k - \xi^\star\|. \end{aligned} \tag{32}$$

Here $q \ge 0$, and the reference state's iterate component must equal the recovery target to interpret this as a recovery bound. With $0 \le \gamma_k \le \bar{\gamma}$, $\beta = L_d \bar{\gamma} C_m C_r$ and $e_k = \|\xi_k - \xi^\star\|$, the standard perturbation argument gives

$$e_{k+1} \le q e_k + \beta \sum_{t=0}^{k} \tau_m^{k-t} e_t, \qquad \beta < (1 - q)(1 - \tau_m). \tag{33}$$

For $\beta > 0$, this sufficient condition implies geometric decay on the horizon where the assumptions hold; infinite-iteration convergence needs them at every future step. Appendix AG.4 gives the rate and proof. This supplies the exponential-memory stability rate.

Scope of the CP sufficient condition. The general complete-state route uses contraction, injection sensitivity and a memory envelope on an invariant domain. The CP specialization uses $2\mu(2s-1) < 1$. Its dimensional lower bound is 8.8636973 at $M = 250, N = 500, s = 50$, identifying the hypothesis that separates this specialization from the main synthetic protocol. The training protocols select checkpoints by validation and enforce their stated tap constraints; contraction constants are not part of that selection. Full CP and noise calculations appear in Appendices AG and AN.1, and Proposition 19.

AG HYBRID-FAMILY MAPPING AND PROOF OBLIGATIONS

The mapping below follows Eqs. (5), (20), (31), (35)–(36), (41), (48) and (50)–(54) of Hybrid-ISTA (Zheng et al., 2023), using the arXiv:2204.11640v2 numbering. Its direct-branch mixing coefficient is the complement of ours. "Mapped" means an equation-level decomposition, not executed author-code equivalence or an inherited convergence guarantee after adding DASH.

AG.1 What the family-wide quantifier means

Let $\mathcal{V}$ index backbone variants, each with complete state $\xi_k = (x_k, s_k)$. Suppose variant $\nu \in \mathcal{V}$ has a causal factorization

$$\begin{aligned}(v_k, c_k) &= F_{\nu,k}(\xi_k; y),\\ B_{\nu,k}(\xi_k; y) &= C_{\nu,k}(v_k, v_k; c_k, y),\end{aligned} \tag{34}$$

where C returns the complete next state and c_k retains every backbone quantity required to finish this iteration. In particular, adaptive coefficients, unperturbed anchors and stationary flags must not disappear when an implementation exposes the two inputs.

Definition 10 (DASH-admissible Hybrid backbone)**.** A backbone is DASH-admissible on a domain $\mathcal{D}$ when Eq. (34) is causal and well typed, ξ_k contains its complete persistent state, c_k contains every within-iteration quantity required by completion, and $C_{\nu,k}(v_k, v_k; c_k, y)$ equals the original full update for every reachable state in $\mathcal{D}$. The feedback has a separate causal state, returns the same vector type as v_k, and keeps both completion inputs in their declared domains. For a stochastic backbone, the identity comparison uses coupled random draws and disabling feedback neither consumes an extra draw nor mutates feedback state.

Proposition 11 (Faithful embedding on finite prefixes)**.** *Every DASH-admissible backbone admits the update $C_{\nu,k}(v_k + d_k, v_k + d_k; c_k, y)$. With feedback disabled, the same parameters, initial state and coupled random draws recover the original backbone state at every finite iteration.*

Proof. The two-input completion defines the perturbed recurrence on its stated domain. For the identity claim, disabling feedback sets $d_k = 0$ and bypasses memory calls, so no additional random draws or state mutations are introduced. Equal states at iteration k give equal (v_k, c_k). The two completion inputs are then (v_k, v_k), hence the factorization gives the same entire next state. Induction from the common initial state proves equality for every finite prefix. □

This quantifies over every backbone satisfying an explicit state and domain contract, rather than only the experimentally tested modules. In particular, the present mapping covers every named Hybrid-ISTA variant in the source: HCISTA, HLISTA-CP, HLISTA-CPSS, HALISTA, gain and gain-plus- overshoot HGLISTA, and HELISTA. It is an interface identity, not a new convergence theorem. A future variant needs its own faithful factorization; enabling feedback requires separate stability/recovery assumptions, and a positive accuracy effect still requires a matched experiment. The mappings below instantiate this contract for the published named family. The quantifier is conditional on a fixed problem instance: an operator, learned backprojection, normalization buffer and any other data-dependent coefficient are part of the state/configuration being ported. Replacing the measurement matrix while freezing those quantities tests transfer, not the identity above. The validation stress audit in Sec. W therefore reports such transfer as a separate falsification check.

HCISTA. P is its first classical proximal step. Completion retains the adaptive step, regularization and mixture rules, including its special stationary-point rule. A separate adaptive port now implements these rules; the earlier fixed-step hybrid must not be relabeled as this algorithm. Its objective-decrease and relative-error argument must be re-established after direct-branch feedback.

HLISTA-CP. P is the first coupled learned proximal estimate. Completion retains the free network, second proximal step and mixture. The independent local PyTorch adaptation and paired module comparisons cover this structure; its oracle conditions are not enforced by training.

HLISTA-CPSS. Both proximal stages retain the threshold-aware support selection operator and its schedule. An independent optional port now implements both selections; it uses a floor count and index-stable tie breaking, with $p_n = \min(pn, p_{\max})$ for zero-based n. This declared convention is not author-code equivalence. Selected CPSS rows are included in the new-family image ledger, while its full registered matrix remains incomplete. Changing a selection set under perturbation needs its own analysis, as shown below.

HALISTA. The two backprojections are scalar multiples of the variant's fixed analytic matrix. Those parameters and the original threshold/mixing relations stay in the backbone. The new optional port fixes the second step to one and learns the first step and two thresholds. It solves the paper's Frobenius surrogate for the analytic matrix, not the minimax-coherence problem. Its configured first-step interval and learned thresholds are not asserted to equal the oracle conditions. Standalone ALISTA and ALISTA-SS comparisons are not HALISTA comparisons.

HGLISTA. Context carries the previous proximal estimates used by the gain gate. The gain-only and gain-plus-overshoot versions must be distinguished: the source analyzes the former and evaluates the latter empirically. A scalar layer-gain feedback control is not HGLISTA. Separate ports now implement both versions. Their gate state stores the previous unperturbed first and corrected proximal estimates and thresholds, separately from DASH memory. Both proximal stages use the same causal gain. Overshoot remains anchored at the original input, not the gain-expanded input. We resolve the index ambiguity in the source's Eq. (49) causally using Eq. (41). The first iteration has no gates, consistent with the stated first-iteration CP equivalence and Table 3's 15 gated iterations, despite the overshoot experiment's prose mentioning all 16 layers. These are explicit replication choices, not asserted author-code equivalence.

HELISTA. P comprises both pre-network extragradient stages and returns $v_{k+1/2}$; completion comprises the network, both post-network stages and mixture. Its multistage thresholds and four-step mixing constraint must be retained. Neither a two-step CP port nor ISTA-Net+ implements this variant. The additional optional port uses all four stages, a shared learned matrix, and the source's multistage thresholds. Its spectral-scaled initialization and shared residual refiner are declared adaptations. HELISTA image results are reported in the seven-port COCO2017 FIR-refinement validation panel (Appendix AA); HELISTA is absent from the BSD500 selected13 test grid. The historical module grid is archived separately as a superseded experiment scope.

The adapter owns only causal residual feedback and its injection location. Backbone and memory states are independent and initialized at the start of each solve. A port must preserve dimensions and the original zero-feedback composition, and expose all parameter-dependent operations to automatic differentiation. These requirements permit different internal algorithms; they do not prove the conditions below. Scoped checks now cover every named family above, with separate gain-only and overshoot HGLISTA ports. Historical factory-acceptance and incomplete-grid records remain in the reproduction archive (Appendix AO).

Adaptive and gated state obligations. For HCISTA, the port retains each row's adaptive regularization parameter and stationary termination. It does not reinterpret different per-layer objectives as one fixed objective. An active row with a numerically zero positive regularization bound fails explicitly instead of changing the recursion by epsilon clamping; stopped rows remain stopped after feedback. A local sensitivity argument requires a positive lower bound on active mixture denominators, regularization-state bounds, and separate treatment of the stationary branch. HGLISTA requires bounds on its complete history state and data-dependent gain/overshoot terms. Positive parameterization alone does not provide these bounds or the source's oracle inequalities. The fixed-mixture calculation below therefore does not automatically cover HCISTA's adaptive mixture or certify the gated variants.

AG.2 Support switches and a local sensitivity condition

The threshold-aware support map is not globally Lipschitz for positive thresholds. For example, with two coordinates, one bypass entry and $\theta = 1$, use index-stable ties. At $z = (2, 2)$ its output is $(2, 1)$, whereas at $z_\epsilon = (2, 2+\epsilon)$ it is $(1, 2+\epsilon)$ for $\epsilon > 0$. The input difference tends to zero and the output difference does not. A selected coordinate crossing its threshold gives another discontinuity: with one selected coordinate, $S(1) = 0$ but $S(1+\epsilon) = 1+\epsilon$ under this convention.

Lemma 12 (A sufficient local margin for support selection)**.** *Fix a threshold $\theta \geq 0$ and a bypass count $1 \leq K \leq N$. Let I be the top-K set of z, and let $a_1 \geq \cdots \geq a_N$ be its sorted magnitudes. Define*

$$g = a_K - a_{K+1}, \qquad h = \min_{i \in I} \big||z_i| - \theta\big|, \tag{35}$$

where $g = +\infty$ when $K = N$. If $g, h > 0$, the threshold-aware support map is 1-Lipschitz in every ℓ_r norm, $1 \le r \le \infty$, on the open ℓ_∞ ball about z of radius $\min(g/2, h)$. For $K = 0$, ordinary soft thresholding is globally 1-Lipschitz.

Proof. For a perturbation with infinity norm strictly below $g/2$, the smallest selected magnitude remains larger than the largest unselected magnitude; thus I is unchanged. A perturbation below h also preserves whether each selected magnitude is above or below θ. On this entire ball, each coordinate map is therefore fixed as either the identity, zero, or ordinary soft thresholding. Each is 1-Lipschitz on its domain, and taking the ℓ_r norm of the coordinate differences proves the claim. □

This lemma supplies a possible local ingredient, not an invariant region or a convergence proof. Both proximal stages and all perturbed iterates must remain in their required regions; data-dependent thresholds require additional bounds. The oracle support-induction argument in Appendix AL.1 is a different route and does not require a global Lipschitz claim for this map.

A completion bound for fixed mixture parameters. For a two-branch completion with a fixed $\alpha_k \in [0, 1]$, fixed backprojection $W_k^\top$ and L_N-Lipschitz network N_k, write $Q_k(u) = S(u + W_k^\top(y - Au))$ and $\kappa = \|I - W_k^\top A\|$ in the matching induced norm. If S is 1-Lipschitz on the compared pre-threshold points, then

$$\begin{aligned}\|H(v + d, v + d) - H(v, v)\| &\le [(1 - \alpha_k) + \alpha_k \kappa L_N]\|d\|,\\ \|H(v + d, v) - H(v, v)\| &\le \alpha_k \kappa L_N \|d\|.\end{aligned}$$

These follow directly by composition and the triangle inequality. Soft thresholding supplies the required property globally; the CPSS margin lemma supplies it only locally. Input-dependent mixtures and adaptive backprojections need their own sensitivity terms. Neither bound supplies the contraction $q < 1$ or the residual bound in Eq. (32). The extra direct branch increases this worst-case bound; its performance value remains an empirical question.

AG.3 HELISTA requires different sensitivity constants

For fixed $0 < \theta < \widehat{\theta}$, the multistage threshold $M_{\theta,\widehat{\theta}}$ in the source's Eq. (50) is continuous, with slopes 0, $\widehat{\theta}/(\widehat{\theta} - \theta)$ and 1 away from its knots. Consequently, its componentwise application is globally ℓ_r-Lipschitz with constant

$$L_M = \frac{\widehat{\theta}}{\widehat{\theta} - \theta} = 1 + \epsilon, \qquad \widehat{\theta} = (1 + 1/\epsilon)\theta. \tag{36}$$

This follows by integrating the bounded slope on each scalar interval and taking the ℓ_r norm. For coordinate-dependent thresholds, use the largest coordinate constant. At positive ϵ this operator is not nonexpansive: with knots $1, 2$, inputs $1.25, 1.5$ have outputs $0.5, 1$. Thus the soft-threshold completion bound cannot be transferred to HELISTA with constant one.

For a fixed layer, write $K = W^\top A$ and the post-network pair as

$$\begin{aligned}P_3(u) &= M_3(u + \gamma_3 W^\top (y - Au)),\\ E(u) &= M_4(u + \gamma_4 W^\top (y - AP_3(u))).\end{aligned}$$

Both second-stage updates remain anchored at u, not $P_3(u)$. In a consistent induced norm, composition and the triangle inequality give

$$\mathrm{Lip}(E) \le L_E := L_4\big(1 + |\gamma_4|\|K\|L_3\|I - \gamma_3 K\|\big). \tag{37}$$

Indeed, P_3 has Lipschitz constant at most $L_3\|I - \gamma_3 K\|$; subtracting the second-stage arguments contributes the sum of the input difference and $|\gamma_4|\|K\|$ times this intermediate difference. For fixed corrected-branch weight $\beta \in [0, 1]$ and L_N-Lipschitz refiner, this yields completion sensitivities $(1 - \beta) + \beta L_E L_N$ for both-branch feedback and $\beta L_E L_N$ for refinement-only feedback. These coarse bounds are not an acceleration theorem or a verification of the source's oracle conditions. In particular, they do not establish the base contraction or invariant domain required below.

AG.4 A CONDITIONAL SMALL-GAIN TRANSFER LEMMA

Lemma 13. *Suppose Eq.* (32) *holds on a domain containing every perturbed state and the corresponding zero-feedback comparisons. Assume the memory-gain envelope holds in the same norm, with zero initial memory state,* $0 \le \tau_m < 1$*, and* $0 \le \gamma_k \le \bar{\gamma}$*. For* $\beta = L_d \bar{\gamma} C_m C_r > 0$*, if* $\beta < (1-q)(1-\tau_m)$*, then* $e_k \le e_0 \rho^k$*, where* $\rho \in (\tau_m, 1)$ *is the larger root of the quadratic below. For* $\beta = 0$*,* $e_k \le q^k e_0$*.*

Proof. The triangle inequality gives $e_{k+1} \le q e_k + L_d \|d_k\|$. Substituting the feedback envelope and the residual bound yields Eq. (33). Let $f(t) = t^2 - (q + \beta + \tau_m)t + q\tau_m$. For $\tau_m > 0$, $f(\tau_m) = -\beta\tau_m < 0$ and $f(1) = (1-q)(1-\tau_m) - \beta > 0$, so the larger root lies between τ_m and 1. For $\tau_m = 0$, that root is $q + \beta \in (0, 1)$. The root relation is $\rho = q + \beta\rho/(\rho - \tau_m)$. If $e_t \le e_0 \rho^t$ for $0 \le t \le k$, then

$$\begin{aligned} e_{k+1} &\le e_0 \rho^k \left[q + \beta \sum_{j=0}^{k} (\tau_m/\rho)^j \right] \\ &\le e_0 \rho^k \left[q + \frac{\beta\rho}{\rho - \tau_m} \right] = e_0 \rho^{k+1}. \end{aligned}$$

Induction starts at $k = 0$. The zero-gain case follows without a memory bound. Finite-horizon assumptions give only the corresponding finite prefix; convergence requires the assumptions for all future iterations. □

Coordinate-wise zero response is not needed for this norm-only lemma; it is an additional obligation for the sparse-support argument. A support claim also needs the backbone maps to preserve the required subspace. For instance, a baseline contraction $B(x, 0) = x/2$ with $r = x$ and a memoryless correction $d = x$ satisfies a bounded memory-gain condition, but $B(x, d) = x/2 + d$ diverges for nonzero initialization. Thus a memory contract alone cannot certify the closed loop. Likewise, positive β enlarges this worst-case bound rather than proving acceleration. Actual gains must be established by paired comparisons within each backbone and by execution cost at matched quality.

AH CP SPECIALIZATION AND IMPLEMENTATION DETAILS

CP specialization. Write soft_θ for coordinate-wise soft thresholding. The original CP-style DASH recursion specializes Eq. (6) as

$$\begin{aligned} z_k &= x_k + \overline{W}_k^\top (y - A x_k), \\ v_k &= \mathrm{soft}_{\theta_{1,k}}(z_k), \qquad r_k = v_k - x_k, \\ m_k &= \mathcal{M}_k(r_0, \dots, r_k), \qquad (38) \\ \widehat{v}_k &= v_k + \gamma_k m_k, \\ u_k &= \widehat{v}_k + \mathrm{DNN}_k(\widehat{v}_k), \\ \overline{z}_k &= u_k + \widehat{W}_k^\top (y - A u_k), \\ w_k &= \mathrm{soft}_{\theta_{2,k}}(\overline{z}_k), \qquad (39) \\ x_{k+1} &= \alpha_k w_k + (1 - \alpha_k)\widehat{v}_k. \qquad (40) \end{aligned}$$

The gains $\gamma_k \ge 0$ and the mixing coefficient obey

$$0 \le \alpha_k \le 1, \qquad \alpha_k \theta_{2,k} \le (1 - \alpha_k)\theta_{1,k}. \qquad (41)$$

The original positive-threshold parameterization is $\alpha_k = \sigma(a_k)\theta_{1,k}/(\theta_{1,k} + \theta_{2,k})$. Zero feedback recovers the corresponding no-memory mixture. It does not equate independently trained models. Equation (41) is the CP specialization, not a universal mixing rule for every Hybrid variant.

Learned CP ports and fixed references. The implementation distinguishes `lista_cp`, with layer-wise learned residual backprojection and positive thresholds, from `hybrid_lista_cp`, with a second learned proximal step and a shared residual CNN. The latter retains the mixing bound $\alpha_k \le \theta_{1,k}/(\theta_{1,k}+\theta_{2,k})$. Learning these parameters does not establish their oracle values or coherence optimality. The complete-state family mappings and their local regularity requirements are in Appendix AG. The selected13 grid covers HALISTA, HLISTA-CPSS, HCISTA and both HGLISTA modes, but not HELISTA; its separate extragradient requirements remain in Appendix AG.3. The fixed proximal-gradient reference is another algorithm and does not transfer a certificate to learned CP parameters.

Coefficient–pixel routing in the matched image code. Let $D \in \mathbb{R}^{256\times 512}$ and $A = \Phi D$ be fixed, and let D_{final} be the learned synthesis matrix. In stage k, $D_k = D$ except at the last stage where $D_k = D_{\mathrm{final}}$. The code computes $p_k = D_k \widehat{v}_k$, followed by $g_k^{\mathrm{pix}} = p_k - t_k \Phi^\top(\Phi p_k - y)$ and $p_k^+ = g_k^{\mathrm{pix}} + R_k(g_k^{\mathrm{pix}})$. Here R_k is the six-convolution thresholded pixel residual network. The shared learned matrix $E \in \mathbb{R}^{512\times 256}$ (named `D_inv` in code) gives $u_k = Ep_k^+$; it is not an exact inverse constraint. The second consistency step uses $y - Au_k$ with the original fixed $A = \Phi D$, even at the last stage, and the final image is $D_{\mathrm{final}} x_K$. All five image-control arms preserve this routing. The unthresholded symmetry branch is a training penalty, not an additional inference output. These details correspond to the recorded image-CS implementation, not the separate DCT-coordinate engine.

AI Additional conditional analysis

AI.1 The original CP/CPSS recovery route

Let $\mathcal{X}(B_x, s)$ contain signals with exactly s nonzero entries, where $2 \le s \le N$ is an integer and $0 < B_x < \infty$ bounds their magnitudes. Assume $y = Ax^\star$ and $x_0 = 0$. The generalized coherence used by the original analysis is

$$\mu = \inf_{W:\, W_i^\top A_i = 1} \max_{i \ne j} |W_i^\top A_j|. \tag{42}$$

The theorem additionally requires coherence-optimal weights, finite oracle threshold envelopes over the signal class, and Eq. (41); Appendix AL states these conditions explicitly. They are not automatically satisfied by supervised training or by a stable memory cell.

Define $E_k = \sup_{x^\star \in \mathcal{X}(B_x,s)} \|x_k - x^\star\|_1$ and

$$a = 2\mu(2s-1), \quad \eta_v = 1 + \mu(2s-1), \quad \Gamma = \overline{\gamma} C_m \eta_v, \tag{43}$$

where $\overline{\gamma}$ bounds the feedback gains on the claimed horizon. The support argument and the memory contract yield

$$E_{k+1} \le aE_k + \Gamma \sum_{t=0}^{k} \tau_m^{k-t} E_t. \tag{44}$$

Theorem 14 (Conditional CP bound)**.** *Under the full assumptions in Appendix AL and Definition 9, the iterates and both proximal estimates have no false-positive support. If $\Gamma > 0$ and $(1-a)(1-\tau_m) > \Gamma$, the larger root ρ of*

$$\rho^2 - (a + \Gamma + \tau_m)\rho + a\tau_m = 0 \tag{45}$$

lies in $(\tau_m, 1)$ and $\|x_k - x^\star\|_2 \le E_k \le sB_x\rho^k$ on that horizon. If the feedback vanishes, the separate bound is $E_{k+1} \le aE_k$.

The proof is given in Appendix AL. The CPSS variant replaces a by a uniform support-aware upper bound and η_v by the corresponding increment bound. Its bypass sets must satisfy the threshold rule in Appendix AL.1; choosing top-magnitude coordinates alone is not a substitute for that assumption. Uniform infinite-iteration conclusions require the conditions for all future iterations, not merely the trained prefix. For the amplitude class used here, the uniform maximum beginning at $k = 0$ recovers the CP coefficients. Thus the stated CPSS corollary does not relax the CP small-gain condition; Appendix AL.1 gives the argument.

What the bound does not say. The extra term in Eq. (44) is nonnegative. This sufficient upper bound controls the cost of admissible feedback and does not prove that feedback accelerates the baseline. Nor does it show that a finite network has reached the true signal. In particular, a dimensional inequality gives

$$\mu \geq \sqrt{\frac{N-M}{M(N-1)}} \quad (M < N). \tag{46}$$

At $M = 250, N = 500, s = 50$, it implies $a \geq 8.8636973 > 1$, whereas the CP condition requires $a < 1$. Thus this sufficient condition cannot certify that typical experimental setting, even with zero memory gain. The derivation in Appendix AN.1 is independent of training. It is not a claim that the algorithm diverges, and does not settle a different CPSS, local or probabilistic analysis. A restricted nonzero-feedback witness with computable threshold envelopes is proved in Appendix AL.2; it changes the oracle-equality rule and does not certify the image implementation.

AI.2 Separate fixed-reference construction

The summably perturbed proximal-gradient reference changes the update and cannot certify learned DASH. Its standard derivation, objective gate and counterexamples remain in the reproduction archive (Appendix AO); they are not an independent contribution.

AJ Separating current gain, history and direct-branch correction

The current-gain decomposition isolates historical differences; the branch identity gives their exact same-state error criterion.

AJ.1 FIR gain–history decomposition

$$m_k = g_k r_k + \sum_{j=1}^{J_k-1} b_{k,j}(r_{k-j} - r_k), \qquad g_k = \sum_{j=0}^{J_k-1} b_{k,j}. \tag{47}$$

At iteration k, let $J_k = \min(J, k+1)$ and $m_k = \sum_{j=0}^{J_k-1} b_{k,j} r_{k-j}$. Adding and subtracting $\sum_{j=1}^{J_k-1} b_{k,j} r_k$ gives Eq. (47). Thus the gain-matched current-input counterfactual is $g_k r_k$, using only available taps. Equal updates in the available history make the second term zero. Historical benefit is the output effect of the second term, evaluated against the matched current-input correction. A trained control should preserve the available gain range, layer-sharing rule, initialization and training budget; functional capacity is specified by the correction map as well as its parameter count.

AJ.2 A same-state identity for fixed mixture

For one state and the same correction d, suppose the completion has form $H(a, b) = \alpha R(a) + (1 - \alpha)b$, with the same scalar $\alpha \in [0, 1]$ in both calls. Set $x^R = H(v + d, v)$ and $x^B = H(v + d, v + d)$. Subtraction gives $x^B - x^R = (1 - \alpha)d$. For any reference $x^\star$, expanding the square yields

$$\begin{aligned} \|x^B - x^\star\|_2^2 - \|x^R - x^\star\|_2^2 = 2(1-\alpha)\langle x^R - x^\star, d\rangle \\ + (1-\alpha)^2\|d\|_2^2. \end{aligned} \tag{48}$$

For $\alpha < 1$, dual-branch injection strictly improves this one-step error exactly when $\langle x^R - x^\star, d\rangle < -(1 - \alpha)\|d\|_2^2/2$. At $\alpha = 1$ the two updates coincide. The inequality specifies the error alignment needed for the additional injection. If the gate changes between calls, gate-difference terms enter the expansion. The common-state, fixed-mixture premises define this comparison; separately trained models are evaluated by their own complete trajectories. Ground truth supplies the offline alignment diagnostic; inference uses the trained mixture and measured observations.

AK Module contracts and implementation differences

All constructions below operate independently on each signal coordinate, using states initialized to zero. Sharing parameters across coordinates is allowed; unrestricted mixing of their hidden states is not covered by these zero-response arguments.

AK.1 Finite impulse-response feedback

First consider shared coefficients; layer-specific coefficients use the uniform bound in Eq. (30). For a length-J filter, take missing past inputs as zero and define

$$m_k = \sum_{j=0}^{\min(k,J-1)} b_j r_{k-j}. \tag{49}$$

For any chosen $0 < \tau < 1$, let $C = \max_{0 \le j < J} |b_j| \tau^{-j}$. The triangle inequality gives

$$\|m_k\|_1 \le C \sum_{t=0}^{k} \tau^{k-t} \|r_t\|_1. \tag{50}$$

Coordinate-wise zero response is immediate. The matched image protocol uses shared coefficients with absolute-sum cap 0.5. The synthetic protocol instead uses per-layer tanh coefficients without an additional absolute-sum cap. The image cap limits overall filter gain but is not by itself the complete closed-loop condition: the chosen (C, τ), the injected gain, and the backbone factor must still be checked. A length-four filter includes the current and three preceding increments, not four additional past increments.

AK.2 Stable state-space feedback

Let

$$h_{k+1}^{(i)} = P_k h_k^{(i)} + b_k r_{k,i}, \qquad m_{k,i} = c_k^\top h_{k+1}^{(i)}, \tag{51}$$

with $\|P_k\|_2 \le \tau < 1$, $\|b_k\|_2 \le B$ and $\|c_k\|_2 \le C_o$. Unrolling the zero-initialized state yields

$$|m_{k,i}| \le C_o B \sum_{t=0}^{k} \tau^{k-t} |r_{t,i}|. \tag{52}$$

Summation over coordinates establishes Definition 9 with $C_m = C_o B$. A zero coordinate history also gives a zero state and output. The legacy S3M parameterizes a diagonal continuous-time generator and discretizes it. The new adaptation instead directly uses shared diagonal decays $0.999\sigma(p_j)$, learned input weights and a readout. Its logged module bound uses $\tau = \max_j 0.999\sigma(p_j)$ and $C_m = \|b\|_2 \|c\|_2$. These are related state-space instances, not a checkpoint-equivalent parameterization. Stability of P_k alone does not bound an arbitrarily large readout or certify the whole DASH recursion.

AK.3 A constrained LSTM instance

For each coordinate, omit its index and use

$$\begin{aligned} c_{k+1} &= f_k \odot c_k + i_k \odot g_k, \\ h_{k+1} &= o_k \odot \tanh(c_{k+1}), \qquad m_k = q_k^\top h_{k+1}. \end{aligned} \tag{53}$$

Assume uniform bounds $\|f_k\|_\infty \le \bar{f} < 1$, $\|i_k\|_\infty \le \bar{i}$, $\|o_k\|_\infty \le \bar{o}$, $\|q_k\|_\infty \le \bar{q}$ and

$$\begin{aligned} &g_k = \phi_k(r_k, h_k), \quad \phi_k(0,0) = 0, \\ &\|\phi_k(r,h)\|_1 \le L_u |r| + L_h \|h\|_1. \end{aligned} \tag{54}$$

If $\tau = \bar{f} + \bar{i} L_h \bar{o} < 1$, then

$$\|c_{k+1}\|_1 \le \tau \|c_k\|_1 + \bar{i} L_u |r_k|. \tag{55}$$

Unrolling and using $|q_k^\top h_{k+1}| \le \bar{q}\bar{o}\|c_{k+1}\|_1$ gives $C_m = \bar{q}\bar{o}\bar{i}L_u$ in Eq. (29). If the coordinate history is zero, induction gives $c_k = h_k = g_k = 0$ even when the gates themselves have biases. An additive candidate bias, however, would invalidate this argument.

The implemented `contractive_lstm` sets the candidate bias to zero, caps the forget gate by 0.6, the candidate recurrent induced ℓ_1 norm by 0.1, the candidate input ℓ_1 norm by one and the readout ℓ_∞ norm by one. Taking $\bar{i} = \bar{o} = 1$ gives the valid module envelope $(C_m, \tau_m) = (1, 0.7)$. This is an input-to-output envelope from zero state, not a proof that arbitrary two-state LSTM dynamics are globally incrementally contractive. The unconstrained LSTM has no such enforced contract after training. Neither case alone certifies a learned backbone, oracle thresholds or the small-gain inequality.

AL Full assumptions and proof of the conditional DASH bound

This section restates the original CP/CPSS analytical route for Eqs. (38)–(40). It is not a theorem about every configuration of the new PyTorch adaptation. Its backbone argument follows the sparse-recovery and free-network analyses in (Chen et al., 2018; Zheng et al., 2023).

Assumption 15 (Conditional sparse-recovery setting)**.** Fix $A \in \mathbb{R}^{M\times N}$, an integer $2 \le s \le N$, $0 < B_x < \infty$, and

$$\mathcal{X}(B_x, s) = \{x^\star \in \mathbb{R}^N : \|x^\star\|_0 = s,\ |x_i^\star| \le B_x\}. \tag{56}$$

For every signal in this class, $y = Ax^\star$, $x_0 = 0$, and memory starts at zero. Both $\overline{W}_k$ and $\widehat{W}_k$ attain the generalized coherence in Eq. (42): their diagonal products with A equal one and every off-diagonal product has magnitude at most μ. All required supremums are finite, and the thresholds are

$$E_k = \sup_{x^\star\in\mathcal{X}(B_x,s)} \|x_k - x^\star\|_1,$$

$$\theta_{1,k} = \mu E_k, \tag{57}$$

$$\theta_{2,k} = \sup_{x^\star\in\mathcal{X}(B_x,s)} \left(\mu\|u_k - x^\star\|_1 + \frac{\mu s}{s-1}\|u_k\|_1\right). \tag{58}$$

The mixing rule satisfies Eq. (41); if both thresholds are zero, take $\alpha_k = 0$. The same nonnegative bound $\overline{\gamma}$ covers γ_k on the stated horizon. The memory satisfies Definition 9.

The trajectory inside a supremum is the trajectory generated by that signal's own observations. The envelopes in Eqs. (57) and (58) depend on unknown recovery errors; they are analytical sufficient conditions, not a deployed recipe that accesses the test target. Arbitrary trained scalar thresholds are not substituted for them.

Lemma 16 (Support and uniform error recursion)**.** *Under Assumption 15, $x_k, v_k, \widehat{v}_k, w_k$ are supported on $S = \operatorname{supp}(x^\star)$. Equation* (44) *holds with Eq.* (43)*.*

Proof. Proceed by induction from $x_0 = 0$. For $i \notin S$, the first pre-threshold coordinate has magnitude at most $\mu\|x_k - x^\star\|_1 \le \theta_{1,k}$, because the diagonal product equals one. Hence $v_{k,i} = 0$ and $r_{k,i} = 0$. The same holds at every preceding step, so Eq. (28) gives $m_{k,i} = 0$ and $\widehat{v}_{k,i} = 0$. The refinement u_k need not be supported on S. Nevertheless its second pre-threshold coordinate obeys

$$|\overline{z}_{k,i}| \le \mu \sum_{j\ne i} |u_{k,j} - x_j^\star| \le \mu\|u_k - x^\star\|_1 \le \theta_{2,k}. \tag{59}$$

Therefore $w_{k,i} = 0$ and the mixture also has no off-support activity.

For the error bound, summing on S and using $|\operatorname{soft}_\theta(z) - x| \le |z - x| + \theta$ gives

$$\begin{aligned}\|v_k - x^\star\|_1 &\le \mu(s-1)\|x_k - x^\star\|_1 + s\theta_{1,k}\\ &\le \mu(2s-1)E_k,\\ \|r_k\|_1 &\le [1 + \mu(2s-1)]E_k = \eta_v E_k.\end{aligned} \tag{60}$$

Thus the injected correction is bounded by

$$\gamma_k\|m_k\|_1 \le \Gamma\sum_{t=0}^{k}\tau_m^{k-t}E_t =: G_k. \tag{61}$$

For the second proximal branch, the threshold envelope yields

$$\begin{aligned}\|w_k - x^\star\|_1 &\le \mu(s-1)\|u_k - x^\star\|_1 + \mu s\|u_k\|_1 + s\theta_{2,k}\\ &\le (2s-1)\theta_{2,k}.\end{aligned} \tag{62}$$

Put $r_s = 2s-1$. Combining the two branch bounds with $\alpha_k\theta_{2,k} \le (1-\alpha_k)\theta_{1,k}$ gives

$$\begin{aligned}\|x_{k+1} - x^\star\|_1 &\le \alpha_k r_s\theta_{2,k} + (1-\alpha_k)(\mu r_s E_k + G_k)\\ &\le 2\mu r_s E_k + G_k.\end{aligned} \tag{63}$$

The upper bound is uniform over the signal class; taking its supremum proves Eq. (44). □

Proof of Theorem 14. The support assertion is Lemma 16. Let $p(t) = t^2 - (a+\Gamma+\tau_m)t + a\tau_m$. If $\tau_m > 0$, $p(\tau_m) = -\Gamma\tau_m < 0$ and $p(1) = (1-a)(1-\tau_m) - \Gamma > 0$, so the larger root satisfies $\tau_m < \rho < 1$. For $\tau_m = 0$, the larger root is $a + \Gamma \in (0,1)$. Since $E_0 \le sB_x$, induction in Eq. (44) gives

$$\begin{aligned}E_{k+1} &\le sB_x\rho^k\left(a + \Gamma\sum_{j=0}^{k}(\tau_m/\rho)^j\right)\\ &\le sB_x\rho^k\left(a + \frac{\Gamma\rho}{\rho-\tau_m}\right) = sB_x\rho^{k+1}.\end{aligned} \tag{64}$$

The last equality is Eq. (45). If the correction vanishes, the memory term is zero and the separate recursion is $E_{k+1} \le aE_k$. This zero-feedback conclusion should not be obtained by dividing by $\rho - \tau_m$ in a degenerate root case. □

AL.1 Support-selection extension

Replace each soft-thresholding operator with

$$(\mathcal{S}^{\mathrm{ss}}_{\theta,p}(z))_i = \begin{cases} z_i, & i \in \mathrm{Top}_p(z),\ |z_i| > \theta,\\ \mathrm{soft}_\theta(z_i), & \text{otherwise,}\end{cases} \tag{65}$$

where Top_p selects the prescribed percentage of largest magnitudes, with a fixed tie convention. In particular, $|z_i| \le \theta$ still implies zero output. Let $\Psi_{v,k}$ and $\Psi_{w,k}$ be the true-support coordinates satisfying the bypass condition in the respective branches. Define uniform lower bounds

$$\begin{aligned}\underline{\psi}_{v,k} &= \inf_{x^\star\in\mathcal{X}(B_x,s)}|\Psi_{v,k}|,\\ \underline{\psi}_{*,k} &= \inf_{x^\star\in\mathcal{X}(B_x,s)}\min\{|\Psi_{v,k}|,|\Psi_{w,k}|\},\\ a_{\mathrm{ss},k} &= 4\mu s - 2\mu - 2\mu\underline{\psi}_{*,k},\\ \eta_{\mathrm{ss},k} &= 1 + \mu(2s-1-\underline{\psi}_{v,k}).\end{aligned} \tag{66}$$

These are uniform support-aware quantities, not a replacement by the observed support of one test signal. Set $\overline{a}_{\mathrm{ss}} = \max_k a_{\mathrm{ss},k}$, $\overline{\eta}_{\mathrm{ss}} = \max_k \eta_{\mathrm{ss},k}$ and $\Gamma_{\mathrm{ss}} = \overline{\gamma}C_m\overline{\eta}_{\mathrm{ss}}$ on the horizon.

Corollary 17 (Conditional CPSS bound)**.** *Under Assumption 15 with Eq.* (65)*, support is preserved and*

$$E_{k+1} \le a_{\mathrm{ss},k}E_k + \Gamma_{\mathrm{ss}}\sum_{t=0}^{k}\tau_m^{k-t}E_t. \tag{67}$$

If $\Gamma_{\mathrm{ss}} > 0$ *and* $(1-\overline{a}_{\mathrm{ss}})(1-\tau_m) > \Gamma_{\mathrm{ss}}$*, then* $E_k \le sB_x\rho_{\mathrm{ss}}^k$*, where* ρ_{ss} *is the larger root of Eq.* (45) *with* (a,Γ) *replaced by* $(\overline{a}_{\mathrm{ss}},\Gamma_{\mathrm{ss}})$*. When the correction vanishes,* $E_k \le sB_x\prod_{t=0}^{k-1}a_{\mathrm{ss},t}$*; a contraction interpretation additionally requires these factors to be less than one.*

Proof. The support induction is unchanged because Eq. (65) still zeros coordinates below threshold. In Eq. (60), only $s - |\Psi_{v,k}|$ true-support coordinates incur threshold shrinkage, giving $\mu(2s - 1 - \underline{\psi}_{v,k})E_k$ and increment bound $\eta_{\mathrm{ss},k}E_k$. Equation (62) similarly becomes $(2s - 1 - |\Psi_{w,k}|)\theta_{2,k}$. The mixing inequality and the minimum of the two bypass counts give $a_{\mathrm{ss},k}E_k$ for the backbone term. The history term uses the uniform $\overline{\eta}_{\mathrm{ss}}$. Bounding $a_{\mathrm{ss},k}$ by $\overline{a}_{\mathrm{ss}}$ permits the same polynomial and induction as Theorem 14. The zero-correction case follows by iterating its scalar recursion. □

Why this uniform CPSS condition does not relax the CP condition. For $\mu > 0$ and $B_x > 0$, the stated signal class has no positive lower bound on nonzero amplitudes. Since $x_0 = 0$, $E_0 = sB_x$ and $\theta_{1,0} = \mu s B_x > 0$. Fix any s-sparse direction and scale it by a sufficiently small positive number. The first pre-threshold vector scales linearly with this number, so every coordinate has magnitude at most $\theta_{1,0}$. No coordinate bypasses shrinkage in the first branch. Hence

$$\underline{\psi}_{v,0} = \underline{\psi}_{*,0} = 0, \qquad a_{\mathrm{ss},0} = a, \qquad \eta_{\mathrm{ss},0} = \eta_v. \tag{68}$$

Because all bypass counts are nonnegative, taking the defined maxima over a horizon including $k = 0$ gives $\overline{a}_{\mathrm{ss}} = a$ and $\overline{\eta}_{\mathrm{ss}} = \eta_v$. Consequently $\Gamma_{\mathrm{ss}} = \Gamma$: the uniform constant-rate condition in Corollary 17 has the same envelope as CP and cannot remove the dimensional obstruction in Section 4. This is a limitation of this uniform bound, not a claim that support selection cannot help an individual trajectory. A stronger support-selection claim would need an appropriately revised signal class, potentially with a specified positive amplitude lower bound, and a proved uniform identification or burn-in condition. Neither such an assumption nor such a guarantee is established here; observed bypass counts on finitely many examples cannot replace it.

AL.2 A RESTRICTED WITNESS WITH COMPUTABLE THRESHOLDS

This construction retains the CP update structure but replaces the unknown supremum equalities by explicit envelopes. It therefore requires the following separate induction, rather than a literal invocation of Theorem 14. Merely making arbitrary thresholds larger than the supremums is insufficient: constant thresholds of 100 would keep this bounded example at zero.

Take an orthonormal basis of $\mathbf{1}^{\perp} \subset \mathbb{R}^8$ as the rows of Q and set $A = \sqrt{8/7}\, Q \in \mathbb{R}^{7\times 8}$. Then

$$A^{\top}A = \tfrac{8}{7}I - \tfrac{1}{7}\mathbf{1}\mathbf{1}^{\top}, \qquad A\mathbf{1} = 0. \tag{69}$$

Every feasible $W_i^{\top}A_i = 1$ has $\sum_{j\neq i} W_i^{\top}A_j = -1$, so its maximum off-diagonal magnitude is at least $1/7$; $W = A$ attains it. Thus $\mu = 1/7$ is exact. Assume $y = Ax^{\star}$, $\|x^{\star}\|_0 \le s = 2$, $|x_i^{\star}| \le B_x = 1$ and zero initial iterates and memory. Set both weights to A, $\gamma_k = 1$, and $\mathrm{DNN}_k \equiv 0$, so $u_k = \widehat{v}_k$.

Two explicit coordinate-wise memories are

$$\begin{aligned} &h_{k+1} = \tfrac{1}{2}h_k + \tfrac{1}{10}r_k, \quad m_k = \tfrac{1}{10}h_{k+1}, \\ &\text{or} \quad m_k = \tfrac{1}{100}r_k + \tfrac{1}{200}r_{k-1} + \tfrac{1}{400}r_{k-2}. \end{aligned} \tag{70}$$

Missing history is zero. The state-space impulse response and FIR coefficients are bounded by $C\tau^j$ with $C = 1/100$ and $\tau = 1/2$; both satisfy the coordinate-wise zero-response condition. Write $\beta = 3/7$, $a = 6/7$, $\eta_v = 10/7$, $\Gamma = 1/70$ and $\rho = 9/10$. Let $b_k = 2\rho^k$ and $H_{-1} = 0$. The complete data-independent schedules are

$$\begin{aligned} &H_k = b_k + \tau H_{k-1}, \quad G_k = \Gamma H_k, \\ &U_k = \beta b_k + G_k, \quad \theta_{1,k} = \mu b_k, \\ &\theta_{2,k} = \mu U_k + \tfrac{\mu s}{s-1}(2 + U_k), \\ &\alpha_k = \frac{\theta_{1,k}}{2(\theta_{1,k} + \theta_{2,k})}. \end{aligned} \tag{71}$$

Use these values in Eqs. (38)–(40); no target or support mask enters the algorithm.

Proposition 18 (An instance with nonzero feedback)**.** *In exact arithmetic this rule satisfies* $\mathrm{supp}(x_k) \subseteq \mathrm{supp}(x^{\star})$ *and* $\|x_k - x^{\star}\|_1 \le 2(9/10)^k$ *for every* $k \ge 0$.

Proof. The assertion holds at zero. Suppose it holds through step k, and let $S = \operatorname{supp}(x^\star)$. Off-support first pre-threshold coordinates have magnitude at most $\mu b_k = \theta_{1,k}$; hence v_k, the increment history and memory are supported on S. The coordinate threshold inequality gives

$$\begin{aligned} \|v_k - x^\star\|_1 &\le \mu(s-1)b_k + s\theta_{1,k} = \beta b_k, \\ \|r_k\|_1 &\le \eta_v b_k, \qquad \|m_k\|_1 \le G_k. \end{aligned} \tag{72}$$

The last inequality uses the history envelope and $\gamma_k = 1$. Thus $\|u_k - x^\star\|_1 \le U_k$ and $\|u_k\|_1 \le 2 + U_k$, so the prescribed second threshold satisfies

$$\theta_{2,k} \ge \mu\|u_k - x^\star\|_1 + \tfrac{\mu s}{s-1}\|u_k\|_1. \tag{73}$$

It removes off-support entries of w_k. Summing the coordinate errors gives

$$\begin{aligned} &\|w_k - x^\star\|_1 \\ &\le \mu(s-1)\|u_k - x^\star\|_1 \\ &\quad + \mu s\|u_k\|_1 + s\theta_{2,k} \\ &\le (2s-1)\theta_{2,k}. \end{aligned} \tag{74}$$

The mixing constraint then yields $\|x_{k+1} - x^\star\|_1 \le ab_k + G_k$ and preserves support. Finally, $H_k \le b_k/(1-\tau/\rho)$ and the exact rational slack

$$\rho - a - \frac{\Gamma}{1-\tau/\rho} = \frac{3}{280} > 0 \tag{75}$$

give $ab_k + G_k \le \rho b_k = b_{k+1}$, closing the induction. □

Scope. The positive small-gain margin is $2/35$, but $\theta_{2,k} \to 4/7$ and $\alpha_k \to 0$: the witness is conservative and does not cover arbitrary refinement, noise, approximately sparse images or trained checkpoints. It does not resolve the 250×500, $s = 50$ obstruction or prove acceleration. The explicit infinite rule implies real-arithmetic convergence; a finite execution is only a prefix. The portable certificate script was independently rerun on 483 signals, six schedules and 100 steps: its rational scalar checks are exact, while trajectory checks passed their stated binary64 tolerance. That finite audit is not a substitute for this induction or an image-performance benchmark.

AM Conditional stability under observation perturbations

This extension concerns the general lifted-state interface, rather than the oracle-threshold support induction. All bounds below use one norm on an invariant domain containing both clean and perturbed trajectories. Fix the model parameters and let y_0 be a clean observation with reference state $z^\star$. For $y = y_0 + \eta$, assume uniformly in k that

$$\begin{aligned} \|B_k(z,0;y_0) - z^\star\| &\le q\|z - z^\star\|, \\ \|B_k(z,d;y) - B_k(z,0;y)\| &\le L_d\|d\|, \\ \|B_k(z,0;y) - B_k(z,0;y_0)\| &\le C_B\|\eta\|, \\ \|r_k(z;y)\| &\le C_r\|z - z^\star\| + C_\eta\|\eta\|, \end{aligned} \tag{76}$$

where $0 \le q < 1$ and all sensitivity constants are finite and nonnegative. These are additional assumptions; empirical noise robustness does not establish them. Assume zero-initialized memory with the history envelope of Eq. (29), and $0 \le \gamma_k \le \bar\gamma$. Put

$$\begin{aligned} \beta &= L_d\bar\gamma C_m C_r, \\ C_{\mathrm{obs}} &= C_B + \frac{L_d\bar\gamma C_m C_\eta}{1-\tau_m}, \\ \delta_\eta &= C_{\mathrm{obs}}\|\eta\|, \\ g &= 1 - q - \frac{\beta}{1-\tau_m}. \end{aligned} \tag{77}$$

Proposition 19 (Forced history recurrence)**.** *Suppose* $0 \le \tau_m < 1$ *and* $g > 0$*. If* $\beta > 0$*, let* ρ *be the larger root of* $\rho^2 - (q + \beta + \tau_m)\rho + q\tau_m = 0$ *(Appendix AG.4). Then*

$$\|z_k - z^\star\| \le \rho^k \|z_0 - z^\star\| + \frac{\delta_\eta}{g}. \tag{78}$$

Consequently $\limsup_k \|z_k - z^\star\| \le \delta_\eta/g$ *when the assumptions hold for all iterations. If* $\beta = 0$*, the sharper bound is* $q^k e_0 + \delta_\eta(1 - q^k)/(1 - q)$*, where* $e_0 = \|z_0 - z^\star\|$*.*

Proof. The triangle inequality applied to Eq. (76), followed by the memory envelope and $\sum_{j=0}^{k} \tau_m^j \le (1 - \tau_m)^{-1}$, gives

$$e_{k+1} \le q e_k + \beta \sum_{t=0}^{k} \tau_m^{k-t} e_t + \delta_\eta. \tag{79}$$

For $\beta > 0$, the small-gain condition gives $\tau_m < \rho < 1$ and $\rho = q + \beta/(1 - \tau_m/\rho)$. Let $E = \delta_\eta/g$ and $u_k = e_0\rho^k + E$. We have $e_0 \le u_0$. If $e_t \le u_t$ for $t \le k$, the nonnegative coefficients in Eq. (79) imply

$$\begin{aligned} e_{k+1} &\le e_0\rho^k \left(q + \frac{\beta}{1 - \tau_m/\rho}\right) \\ &\quad + \left(q + \frac{\beta}{1 - \tau_m}\right) E + \delta_\eta \\ &= e_0\rho^{k+1} + E = u_{k+1}. \end{aligned} \tag{80}$$

Induction and $\rho < 1$ prove the result. When $\beta = 0$, summing the ordinary scalar geometric recurrence proves the separate formula. □

Approximately sparse targets. Write a coefficient target as $x^\star = x_s + t$, where x_s is its chosen s-sparse approximation, and observe $y = Ax^\star + n$. Relative to $y_0 = Ax_s$, the effective perturbation is $\eta = At + n$; hence $\|\eta\| \le \|A\|\|t\| + \|n\|$ in the corresponding induced norm. For $\beta = 0$, set $\rho = q$ in the bound below. If Eq. (76) holds around a clean lifted state whose iterate is x_s, and the state norm dominates the iterate norm, then

$$\begin{aligned} \|x_k - x^\star\| &\le \rho^k e_0 + \|t\| \\ &\quad + \frac{C_{\mathrm{obs}}}{g}(\|A\|\|t\| + \|n\|). \end{aligned} \tag{81}$$

The existence of that clean reference state and the sensitivity bounds must be proved for the backbone in question. A biased regularized fixed point need not equal x_s. This extension establishes conditional perturbation stability, not noisy support preservation, acceleration, a PSNR ordering, or a certificate for the trained image checkpoints. It also does not cover changing the sensing operator while freezing operator-specific learned weights.

AN Dimensional scope of the CP condition

AN.1 Scope audit of the legacy CP sufficient condition

For the legacy generalized coherence

$$\mu(A) = \inf_{W:\, W_i^\top A_i = 1} \max_{i \ne j} |W_i^\top A_j|, \tag{82}$$

the following elementary dimensional obstruction applies independently of training or of the chosen memory generator.

Proposition 20 (A necessary dimensional condition for the legacy CP bound)**.** *Let* $A \in \mathbb{R}^{M \times N}$ *with* $M < N$ *and suppose the displayed feasible set is nonempty. Then*

$$\mu(A) \ge \sqrt{\frac{N - M}{M(N - 1)}}. \tag{83}$$

In particular, at $M = 250, N = 500, s = 50$*, the legacy factor* $a = 2\mu(A)(2s - 1)$ *satisfies* $a \ge 8.8636973 > 1$*.*

Proof. For any feasible W, put $H = W^\top A$ and $c = \max_{i \neq j} |H_{ij}|$. Then $\operatorname{rank} H \leq M$, $\operatorname{tr} H = N$, and every diagonal entry equals one. Therefore

$$\begin{aligned} N = |\operatorname{tr} H| \leq \|H\|_* \leq \sqrt{M}\|H\|_F, \\ \|H\|_F^2 \leq N + N(N-1)c^2. \end{aligned} \tag{84}$$

Combining these inequalities and taking the infimum over feasible W gives (83). Substitution yields the numerical bound. Neither symmetry nor positive semidefiniteness of H is needed. □

The legacy small-gain condition $(1-a)(1-\tau) > \Gamma \geq 0$ with $0 \leq \tau < 1$ requires $a < 1$. Under the dimensions above, even the most optimistic dimensional lower bound requires integer sparsity $s \leq 6$. This is a limitation of that sufficient certificate, not a divergence claim about the trained model and not a refutation of a distinct CPSS, local, or probabilistic analysis. The legacy support and linear-recovery claims must remain attached to their complete original assumptions; the new fixed-Lasso reference guarantee must not be used to certify unchanged learned DASH layers.

AO Reproduction archive and evidence index

The single companion `DASH-LISTA-supplementary.zip` includes this complete paper and supplement, a reading guide, machine-readable records and CPU recomputation entry points. All paths below resolve after extraction; `EVIDENCE_MAP.md` states coverage for each result family.

Original matched evidence: `records/core/` includes the 24-model index, original per-signal records, 26 cross-backbone contrasts, image/COCO/RGB records and 3,600 raw timing values. Its `recompute_core.py` reconstructs the matched synthetic and cross-backbone comparisons and the recorded timing summaries.

Fresh samples and directional effects: `records/stability/` contains the five-instance block/support totals and complete equal-norm direction arrays. The paired-support, direction and theory-accounting checks retain all signs and distinguish pooled energy identities from checkpoint execution. Full generated fresh/stress signal arrays are excluded; their packaged block totals support the reported aggregation and contribution decomposition.

Shared-cap evidence: `research/gain_budget_20260911/` and `research/reviewer_followup_20260913/gain_summary.json` retain original-instance summaries. `records/confirmation/` contains all nine independent-confirmation training histories and 15 per-signal endpoints, with saved prediction chronology. Its verification script recomputes 1,500,000 model–signal records without weights.

Image controls: `research/objective_control_20260913/` and `research/reviewer_followup_20260913/` support the fitting-objective and common-tolerance comparisons. `records/image_confirmation/` and `records/depth_curves/` retain independent-image and stage results.

Replacement budget and path effects: `records/replacement/` includes all 90 adaptation histories, 11,900 image records, 14,400 path records and 3,000,000 synthetic model–signal outcomes. Its verification script reconstructs both image splits, every checkpoint cohort and budget, and the optimization checks.

Tables and entry point: `tables/` contains current generated result tables. Run `verify_all.py` with Python and NumPy to check the manifest and advertised record reconstructions. No GPU, dataset download or checkpoint access is required for these checks.

Repeated numeric arrays are stored once with their original dtype, shape and byte hashes. The reversible storage transform in `array_store.py` changes neither precision nor values. `MANIFEST.json` records the packaged file hashes; `REDACTIONS.json` identifies files whose resource or identity metadata was anonymized. Original source hashes and saved file chronology are retained separately from redacted-code hashes.

The archive is self-contained for reading and the advertised record checks. It does not include dataset collections, model weights or a complete training environment. Original large-sample capped results remain summary-level in this package, and historical auxiliary tables do not all have full raw records. The evidence map records these distinctions. Author review is complete. The record checks are distinct from independent retraining or checkpoint replay.