CONTENT-AWARE SCANNING IN VISION STATE SPACE MODELS: IS GROUPING ENOUGH?

Anonymous authors
Paper under double-blind review

ABSTRACT

Content-aware scanning for vision state space models (SSMs) rests on a proximity hypothesis: placing informative tokens next to each other shields their interaction from the recurrence's decay. We test it in a controlled SSM restoration testbed (1,108 trained models) that varies the grouping of informative tokens independently of the information an order carries, and the hypothesis fails in both directions. Sorting tokens by training-free importance scores, such as constant false alarm rate (CFAR) detection, raises grouping without adding information and performs like a random permutation (within 0.1 dB), even when the scores find the targets with AUC 0.98. Ground-truth orders that place informative tokens farther apart than a raster scan still help, in all 72 runs, so "oracle" orders measure leakage rather than an attainable gain. In a factorial design that sets the precision of an order (the informative fraction of the tokens it selects) and its grouping independently, raising precision from 0 to 1 adds 3.5 dB at fixed grouping, and grouping helps only in proportion to precision: grouping alone explains 0.62 of the variance, grouping and precision with their interaction 0.96. The same information delivered as an input channel instead of an order helps as much or more ($+9.82$ against $+8.03$ dB for the ground-truth support; $+0.78$ against $+0.69$ dB for a supervised $O(N)$ CNN scorer). Finally, because the recurrence has no history at the start of a scan, position is readable there: a ground-truth block placed first gains 3–6 dB more than mid-sequence in a causal scan, also with the scan convolution removed and with a variable number of targets. We release a protocol with region-aware metrics and random-permutation, locality, leakage, and input-channel controls.

1 INTRODUCTION

State space models (SSMs) process a sequence with a linear recurrence whose state decays over time (Gu et al., 2022; Gu & Dao, 2023; Dao & Gu, 2024). To apply them to images, vision SSMs flatten the two-dimensional token grid into a one-dimensional scan (Zhu et al., 2024a; Liu et al., 2024; Guo et al., 2024). That the scan order matters is by now well established. Locality-preserving paths such as zigzag, windowed, and nested S-shaped scans improve over naive traversals (Hu et al., 2024; Huang et al., 2024; Li et al., 2025). Large benchmarks find contiguous rasters ahead of disjointed paths (Hardan et al., 2025; Zhu et al., 2024b), and a learned, dataset-level patch order improves several long-sequence vision backbones, including a Mamba variant (Kutscher et al., 2025).

A newer line of work goes further and reorders tokens *per image, by content*: it groups pixels by learned semantic category (Guo et al., 2025), partitions windows by estimated locality scores (Xie et al., 2024), or plans scan paths from a ViT importance map (Hu et al., 2025; Wang et al., 2025a). The stated motivation is proximity: if an order places related tokens next to each other in the sequence, their interaction survives the recurrence's decay (Guo et al., 2025). We call this the *proximity hypothesis*. The hypothesis is the natural reading of the SSM recurrence, and it is untested. The gains reported for content-aware reordering are small: semantic grouping adds 0.03–0.04 dB in the ablation of Guo et al. (2025). To our knowledge, no vision study compares a content-aware order against a random permutation or separates what an order *groups* from what it *reveals*.

We test the proximity hypothesis in a controlled restoration testbed whose only spatial mixing is the scan (Section 3). The testbed sets two properties of an order independently: how tightly it groups the informative tokens, and how much information about the target it carries. Across 1,108 trained

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

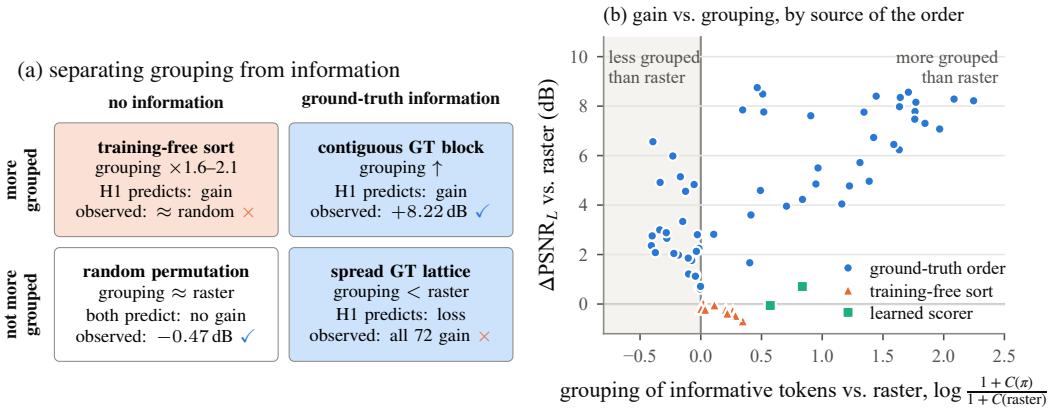

Figure 1: **Grouping is not enough.** (a) The proximity hypothesis (H1) predicts that an order helps in proportion to how tightly it groups the informative tokens. Separating grouping from information refutes it in both cells where the two come apart (marked ×): grouping without information performs like a random permutation, and ground-truth (GT) information helps even when a spread lattice places the informative tokens farther apart than raster does. (b) Ground-truth, training-free, and learned orders from the grouping-controlled experiments, placed by their grouping (x) and their gain in region-balanced PSNR (y); one point per experimental cell. Grouping does not separate the outcomes; the source of the order does. Blue: ground-truth-derived orders; orange: training-free importance sorts; green: learned scorers.

models, the hypothesis fails where grouping and information come apart (Figure 1). Sorting by a training-free importance score raises the grouping of informative tokens by a factor of 1.6–2.1 and yet performs like a random permutation, because the sort scatters spatial neighbours (Sections 4 and 5). Orders built from the ground truth help even when they *separate* the informative tokens, because the order itself carries the answer. A factorial experiment that crosses the precision of an order with its grouping shows that grouping matters only as an amplifier of information, and the same information delivered as an input channel, with no grouping at all, helps as much (Section 6). Position is read most easily at the start of each scan direction, where the recurrence has no history (Section 7).

Contributions.

- A testbed and protocol that separate grouping, information, locality, and position of a scan order (Section 3).
- Evidence against the proximity hypothesis: training-free importance sorts behave like a random permutation, and ground-truth orders help without grouping (Sections 4–5).
- An information account tested by a precision × grouping factorial and input-channel controls, and a start-of-scan anchor that survives removing the scan convolution and varying the number of targets (Sections 6–7).
- Two measurement pitfalls with remedies: "oracle" orders measure leakage, and whole-image PSNR reverses the sign of a quarter of small effects on sparse targets (Section 9).

2 Background: Scan Order, Grouping, and the Proximity Hypothesis

Scans as permutations. A vision SSM maps an image to N tokens and visits them in the order of a permutation π, so that it processes token i at step $\pi(i)$. It runs the recurrence $h_t = a_t \odot h_{t-1} + B_t x_t$ along this order and scatters the outputs back to the grid. With decay $a_t \approx e^{-1/\tau}$, the influence of token j on the output at token i falls roughly as $e^{-|\pi(i)-\pi(j)|/\tau}$ (Wang et al., 2025c); bidirectional scans run the recurrence in both directions (Zhu et al., 2024a). A *fixed* scan uses the same π for every image; a *content-aware* scan computes π from a per-token score $s(\boldsymbol{y})$ of the input $\boldsymbol{y}$, for example by sorting tokens by score.

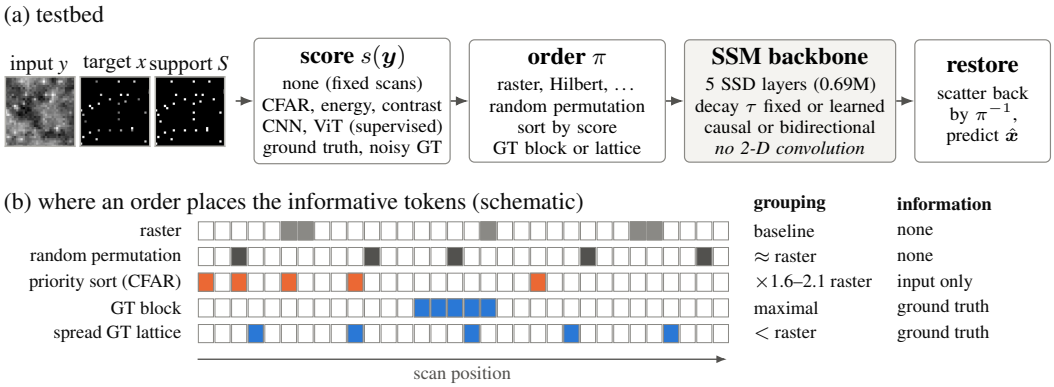

Figure 2: **Testbed.** (a) A noisy sparse image $\boldsymbol{y}$ is flattened by an order π that is fixed, random, computed from a score of $\boldsymbol{y}$, or computed from the target; with no 2-D convolution, all spatial mixing passes through the scan. (b) Orders differ in how tightly they group the informative tokens (filled) and in how much their placement reveals about the target; our experiments vary the two independently.

Grouping. Let $\mathbb{S}$ be the set of K informative tokens of an image (in our tasks, the support of the target). We measure how tightly an order groups them by the realised coupling

$$C(\pi) = \frac{1}{K} \sum_{i \in \mathbb{S}} \sum_{j \in \mathbb{S},\, j \neq i} e^{-|\pi(i) - \pi(j)|/\tau}, \tag{1}$$

the expected number of informative partners within reach of an informative token. A raster scan of a sparse image has small C; sorting the informative tokens into one block maximises it.

Hypothesis H1 (Proximity)**.** The benefit of an order over raster is an increasing function of the grouping gain $\log \frac{1+C(\pi)}{1+C(\text{raster})}$, independent of how π was obtained and of where the informative tokens sit in the sequence.

We designed this study to test H1, which formalises the motivation of content-aware scanning. A quantitative version (benefit linear in the grouping gain, with a single fitted constant) was part of the study design before we ran the main experiments. H1 makes three testable predictions: (a) raising grouping without adding information should help; (b) an order that groups the informative tokens *less* than raster should not help; (c) among orders carrying the same information, tighter grouping should help more.

Information and leakage. An order can also help because π itself tells the backbone something about the target that it does not extract by itself. An order computed from the ground truth is leakage in the classical sense: information about the target that is not legitimately available at test time (Kaufman et al., 2012; Kapoor & Narayanan, 2023). An order computed from the input by a fixed function adds no information in the information-theoretic sense, although it may present the input in a form the backbone reads more easily; a scorer trained with target labels also carries what it learned from the labels.

Locality. Every reordering also changes which *spatial* neighbours are adjacent in the sequence (Hu et al., 2024; Li et al., 2025; Liang et al., 2024). We measure this by the neighbour retention $P(\pi)$, the mean of $e^{-|\pi(u)-\pi(v)|/16}$ over all 4-adjacent pixel pairs (u, v): 0.54 for raster, 0.74 for a Hilbert curve, and 0.03 for a random permutation.

3 Testbed and Protocol

Backbone. We implement a Mamba-2-style backbone based on state space duality (SSD; Dao & Gu, 2024) in pure PyTorch, with five layers of width 96, state size 64, and 0.69M parameters. The decay is either fixed to a chosen horizon $\tau \in \{4, \ldots, 1024\}$ or learned. The backbone has *no*

two-dimensional convolution: all spatial mixing passes through the scan. As in Mamba, each block also applies a width-4 depthwise convolution along the scan; it pads the start by two positions and sees one token ahead, so our causal variant is causal in the recurrence only. Scans are bidirectional unless stated otherwise; Sections 4 and 7 repeat the key experiments without the convolution.

Tasks. The main task is sparse restoration on a 32×32 grid: $K = sN$ informative tokens (the support; s from 0.005 to 0.35) carry blurred sources whose width and gain are drawn per image, observed with additive noise. Because every sample has exactly K support tokens, all region metrics below are exact. A second domain is simulated frequency-modulated continuous-wave (FMCW) radar range–Doppler maps ($s \approx 0.004$), where targets are easy to detect and the difficulty lies in the background. The appendix also reports natural-image denoising and sparse-view CT.

Orders. We compare four families (Figure 2). *Fixed* scans: raster, snake, Hilbert, and the nested S-shaped scan (NSS) of Li et al. (2025); plus a *random* permutation per image. *Training-free importance orders*: tokens scored by a cell-averaging CFAR detector (cf. Rohling, 1983), local energy, or local contrast, then sorted (priority), reverse-sorted, partitioned in $O(N)$, or sorted by window. *Learned scorers*: a three-layer CNN and a four-layer ViT trained with an auxiliary loss towards the target's importance map; the sort is not differentiated, and at test time they see only $\boldsymbol{y}$. *Ground-truth (GT) orders*: the support sorted first (oracle); a contiguous support block at a fixed or random position; the support on a lattice of stride σ (spread when σ is maximal); a block of *non*-support tokens (decoy); a noisy-GT scorer whose AUC we control; and the selection lattices and input-channel controls of Section 6.

Training, metrics, statistics. We train every model for 1,800 steps (radar: 3,000) with a Charbonnier loss that weights the support and the background equally; arms that differ only in the order see byte-identical batches and the same evaluation set. We report four metrics: PSNR on the whole image (W), the informative region (S), the background (B), and the *region-balanced* PSNR $L = -10\log_{10}\big(\frac{1}{2}\mathrm{MSE}_S + \frac{1}{2}\mathrm{MSE}_B\big)$, which weights the regions as the training objective does (its error is Charbonnier; the two rank 53 arms alike, Spearman 0.99). We use L as the primary metric; Section 9 shows why W alone misleads. The unit of replication is the seed (2–4 per cell); every effect is a paired difference to the raster run of the same cell and seed, with a 95% hierarchical bootstrap interval (seeds, then evaluation samples). The study comprises 908 trained models in 14 experiments and a second set of 200 models, designed after the main analysis to test H2–H4 (Appendix H).

Hypotheses and their provenance. Table 1 lists the four hypotheses. We specified H1 before the experiments and formulated H2–H4 after it failed. We then tested them on runs not used to formulate them, including the second set, whose predictions we recorded before running it. Appendix G gives the full record.

4 The proximity hypothesis fails when grouping and information are separated

(a) Grouping without information does not help. Sorting tokens by their CFAR score places high-scoring tokens together. The score correlates with the support (AUC 0.62), so the sort raises the grouping of the informative tokens by a factor of 1.6–2.1 relative to raster. H1 predicts a gain. Instead the sort changes L by -0.50 dB $[-0.64, -0.36]$, indistinguishable from a random permutation (-0.47 dB $[-0.60, -0.34]$; Figure 3a). Paired by seed on the same evaluation images, the two differ by at most 0.05 dB in L and 0.08 dB in S across 11 pairs; reversing the sort (anti-priority) changes nothing. The same holds in the scorer study (CFAR -0.46, uninformative random scores -0.41 dB), on radar, where CFAR finds the targets with AUC 0.98 (sort -4.45, random permutation -4.55 dB), without the scan convolution (sort -0.41, random -0.33 dB), and with three times longer training (sort -0.63, random -0.59 dB).

(b) Information without grouping helps. The spread order uses the ground truth to place the informative tokens on a maximally spaced lattice. It groups them *less* than raster in all 72 of 72 runs, so H1 predicts no gain. Every one of these runs gains (from $+0.53$ to $+9.52$ dB; Figure 3b; 66 of 72 also on the informative region alone, Appendix B). A decoy block of non-informative tokens, placed

Table 1: Hypotheses, the predictions that distinguish them from their competitors, and the outcome. L: region-balanced PSNR gain over raster.

Hypothesis	Distinguishing prediction	Outcome
H1 proximity (a priori)	(a) grouping without information helps	× priority sort: grouping ×1.6–2.1, L −0.50 vs. random −0.47
	(b) less grouping than raster does not help	× spread: 72/72 runs gain (+0.53 to +9.52)
	(c) given information, tighter grouping helps	✓ stride 1→32: +8.22→+2.08 at equal locality
H2 locality cost	importance sorting scatters neighbours like a random permutation and performs like one; the same scores as an input channel cost nothing	✓ P: 0.05 vs. 0.03; $\lvert L_{\text{sort}} - L_{\text{random}} \rvert \leq 0.05$; CFAR as channel −0.07 vs. as order −0.45
H3 information	at fixed grouping, gain rises with the precision of the order; the same information as an input channel helps as much	✓ +3.5 dB per unit precision [2.8, 4.1]; R^2: grouping 0.62, +precision 0.84; channel vs. order: GT +9.82/+8.03, CNN +0.78/+0.69
H4 start-of-scan anchor	causal scan: start only; bidirectional: both ends; also without the scan convolution and with variable K	✓ causal +9.97/+5.99/+6.23, no conv. +9.94/+3.83/+2.88; bidir. no conv. +10.00/+7.29/+10.01 (start/middle/end)

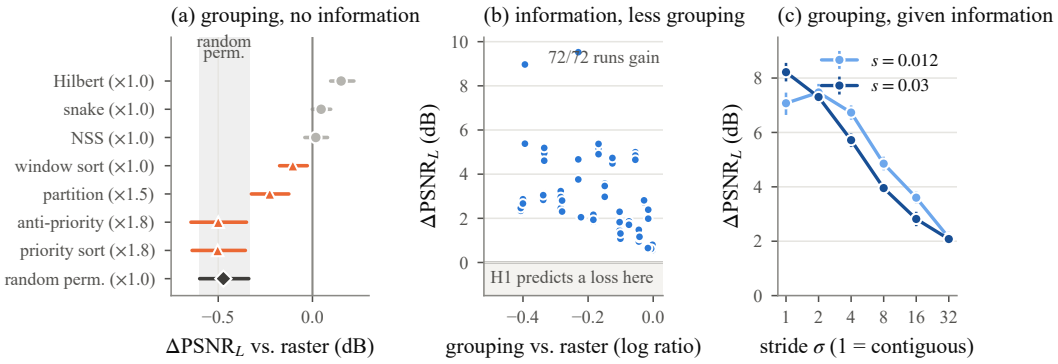

Figure 3: **Testing the three predictions of H1.** (a) Grouping without information: importance sorts raise the grouping of informative tokens (in brackets, relative to raster) yet do no better than a random permutation; mean and 95% interval over 9 seed–sparsity units. (b) Information with *less* grouping than raster: all 72 spread runs gain, where H1 predicts a loss. (c) With information and locality fixed, tighter grouping (smaller stride) helps more, as H1 predicts.

where the ground-truth block would sit, carries no information about the support and stays within −0.05 to +0.01 dB of raster. The spread order also gains without the scan convolution (+1.36 dB) and with three times longer training (+3.17 dB). It is informative because we computed it from the answer: it leaks the target. Ground-truth orders therefore do not bound the attainable benefit of an order; the oracle sort gains +9.57 to +10.10 dB here, and +20.96 dB in the scorer study, whose oracle also ranks the support by amplitude (Appendix A).

(c) Given information, grouping helps. H1 is right where grouping and information move together. Placing the ground-truth support on a lattice of stride σ with a random phase keeps the information and the spatial locality fixed ($P = 0.51$ at every stride) and varies only the grouping: the gain falls from +8.22 (σ=1) to +2.08 dB (σ=32; Figure 3c). Section 6 shows that this effect scales with the information the order carries.

Takeaway. Grouping informative tokens helps only when the order already carries information about them. Without information it buys nothing; with information it amplifies the gain.

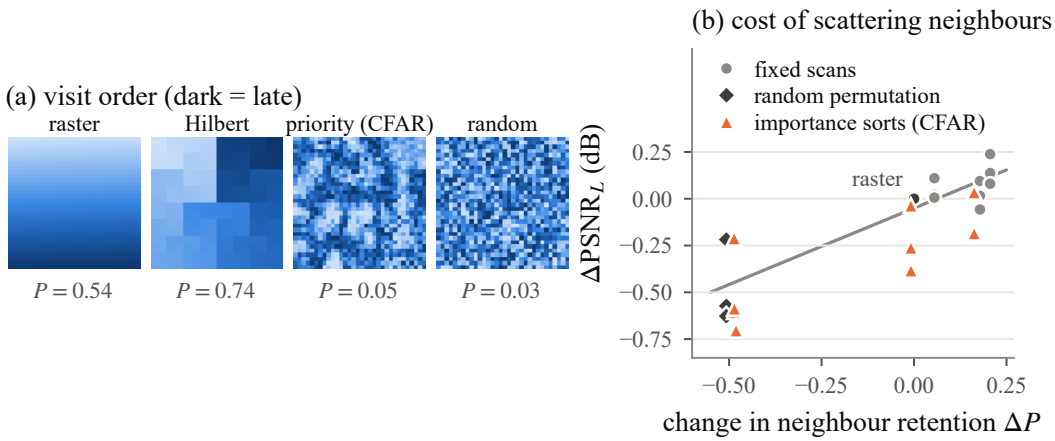

Figure 4: **Importance sorting scatters spatial neighbours as thoroughly as a random permutation.** (a) Visit order of one image under four scans; P is the neighbour retention. (b) Gain against the change in neighbour retention, synthetic task, three sparsities. The line fits the fixed scans and the random permutation; the CFAR-based orders (orange) scatter around it, on average slightly below: their grouping adds nothing beyond what their locality predicts.

5 Why grouping without information fails: the locality cost

Hypothesis H2 (Locality cost)**.** A content-aware order changes two things at once: it groups the high-scoring tokens and it moves every other token away from its spatial neighbours. When the scores carry no information the backbone lacks, only the second change matters.

A priority sort orders tokens by score, and the scores of neighbouring pixels follow no spatial order, so the sort destroys the spatial adjacency that a raster scan provides. Its neighbour retention is 0.05, against 0.54 for raster and 0.03 for a random permutation (Figure 4a). H2 therefore predicts that the sort behaves like a random permutation, which is what Section 4 found. It also predicts that variants keeping spatial order among the non-selected tokens lose less: the partition (0.53) and window sort (0.70) beat the random permutation by $+0.18$ to $+0.44$ dB, but neither reliably beats raster; and the content-aware orders sit on the locality trend of the fixed scans (Figure 4b). Most directly, the same scores delivered without reordering should cost nothing. Given the CFAR map as an extra input channel with raster order, the backbone changes L by -0.07 dB, against -0.45 dB for the CFAR sort (difference $+0.38$ $[+0.31, +0.46]$; Section 6): the sort's whole loss is its locality cost, and the CFAR map itself is worth nothing to the backbone. On natural images, where locality matters more, a contrast sort and a random permutation lose -3.56 and -4.02 dB (Appendix E).

Takeaway. A training-free importance sort is, to the backbone, a random permutation of its spatial neighbourhoods. A content-aware order must earn back this locality cost with information before it can help.

6 What makes an order useful: information

Hypothesis H3 (Information)**.** The benefit of an order increases with the information its placement gives the backbone beyond what the backbone extracts itself. Grouping amplifies this information and has no effect without it.

A fidelity sweep in the first set of runs traces a smooth dose–response in scorer accuracy (Appendix E), but grouping rises with accuracy along it, and its scores also rank the support by amplitude. We therefore test H3 with a factorial design (Figure 5a). Each order selects K tokens, a fraction q of them informative (the *precision*; $q = 0$ selects background tokens only), and places the selected set on a lattice of stride σ with a random phase; all other tokens keep raster order. Only membership is used; q sets the information, σ the grouping.

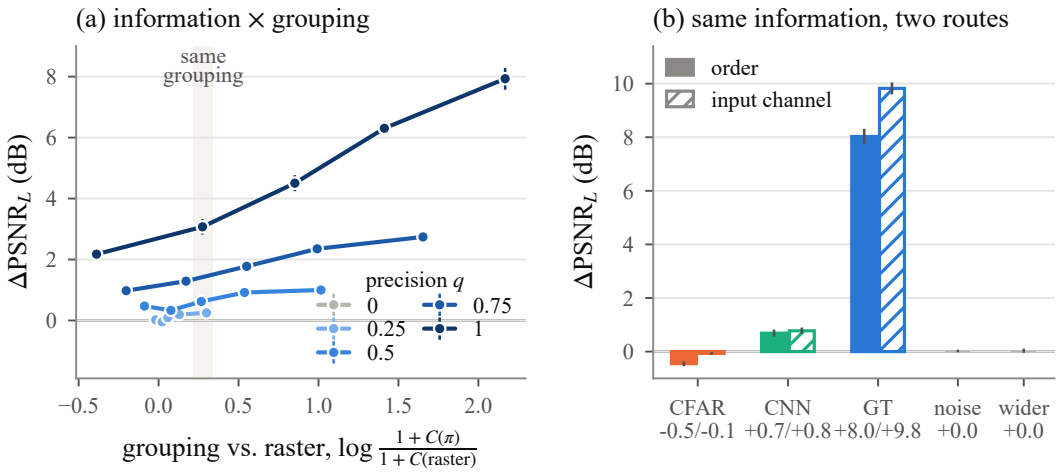

Figure 5: **Information, not grouping, sets the gain.** (a) Selection orders: K tokens, a fraction q of them informative, placed on a lattice of stride σ; one line per precision q, points ordered by grouping (stride 1 at the right). At the same grouping (shaded), precision separates the outcomes; without information ($q = 0$) grouping does nothing. (b) The same maps delivered as the order or as an input channel with raster order; noise channel and a wider raster backbone as controls. Mean and 95% interval over 3 (a) and 4 (b) seeds.

Without information, grouping does nothing: at $q = 0$ a block and a spread lattice stay within 0.1 dB of raster (-0.06 and -0.08 dB). At matched grouping, precision decides. At a grouping gain near 0.3, the gain rises from $+0.25$ ($q = 0.25$) to $+0.62$ ($q = 0.5$) and $+3.07$ dB ($q = 1$); near 1.0, raising q from 0.5 to 0.75 adds $+1.35$ dB $[+1.20, +1.50]$ within seeds. Across the grid, grouping alone explains $R^2 = 0.62$ of the gain, grouping and precision 0.84, and with their interaction 0.96; at fixed grouping, precision adds 3.5 dB per unit [2.8, 4.1]. Grouping helps in proportion to the information: its slope is 2.35 dB per unit of log-grouping at $q = 1$ and at most 1.00 for $q \leq 0.75$.

If grouping only makes information easier to read, the same information delivered by another route should serve as well. We give the backbone the CFAR map, the supervised CNN scorer's map, or the ground-truth support as a second input channel, with raster order (Figure 5b). The CNN helps equally as an order ($+0.69$ dB) and as a channel ($+0.78$ dB; difference $+0.09$ $[-0.03, +0.22]$). The ground-truth support helps more as a channel ($+9.82$ dB) than as a contiguous block ($+8.03$ dB; difference $+1.79$ $[+1.59, +2.02]$), although the channel groups nothing. A noise channel and a raster backbone widened to 104 channels change nothing ($+0.03$ and $+0.03$ dB), so neither the extra input nor extra compute explains the CNN's gain. The CNN adds 0.93% to the model's FLOPs, and the $O(N^2)$ ViT scorer gains nothing (-0.07 dB): what limits content-aware scanning here is the information a scorer adds.

Takeaway. An order helps through the information it makes readable. Grouping amplifies that information, does nothing without it, and is unnecessary when the same information arrives through an input channel. "Oracle" orders measure what the answer is worth.

7 Where the information is read: a start-of-scan anchor

H1 says the position of a block should not matter once its grouping is fixed. In a causal scan with the ground-truth support as one contiguous block (grouping $C = 8.44$ at every position), the block gains $+5.91$ to $+6.44$ dB anywhere in the interior and $+6.23$ dB at the end, but $+9.70$ dB at the start (Figure 6b). The start helps on the informative region (S: $+8.20$ against $+5.57$ to $+6.22$ dB) and most on the background, which holds 3% of the remaining squared error with the block first against 86% in the interior: once the block has passed, every later token is background.

Hypothesis H4 (Start-of-scan anchor)**.** At the start of each scan direction the recurrence has no history, which makes absolute position readable there, as zero padding does for convolutional networks (Islam et al., 2020; Kayhan & van Gemert, 2020) and causal attention for transformers

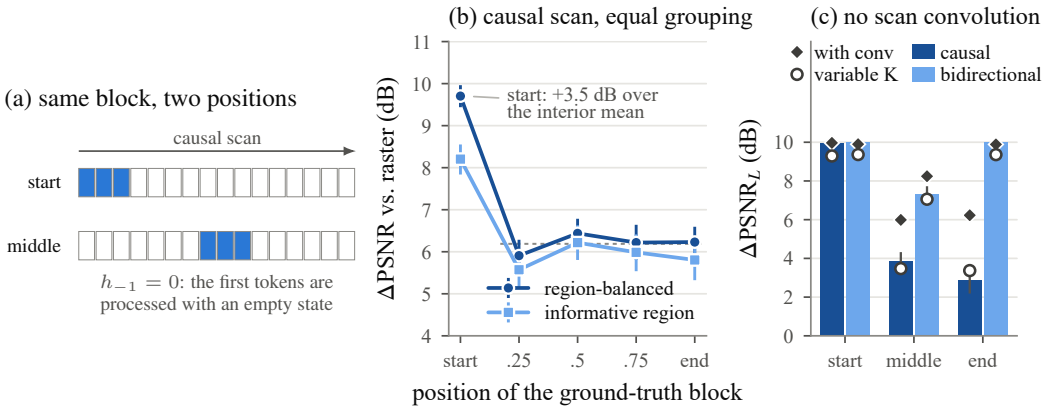

Figure 6: **A start-of-scan anchor.** (a) The same contiguous block of ground-truth tokens, placed at the start or in the middle of a causal scan; grouping is identical. (b) Gain by block position in the experiment where the effect was found (causal scan, s=0.03, 3 seeds). (c) Tests on runs not used to formulate H4 (s=0.02): without the scan convolution (bars, 4 seeds), with it (diamonds, 2 seeds), and without it with a variable number of targets (circles, 3 seeds). In a causal scan only the start is special; in a bidirectional scan, whose backward pass starts at the end, both ends are.

without position encodings (Haviv et al., 2022). An informative block is therefore read more easily at the start of a scan direction than elsewhere.

We formulated H4 on the experiment above and tested it on a separate experiment that varies the scan direction and that we had not analysed when forming the hypothesis. H4 predicts that the effect follows the start of each scan direction, not the ends of the sequence. In a causal scan only the first position should be special; in a bidirectional scan, whose backward pass starts at the end, both ends should be. Both hold (Figure 6c). With the convolution, the causal scan gains $+9.97$, $+5.99$, and $+6.23$ dB at the start, middle, and end ($C = 6.38$), and the bidirectional scan $+9.90$, $+8.24$, and $+9.89$ dB. Two further tests, with predictions recorded in advance, address what the convolution and the fixed number of targets left open. Without the convolution, which pads the start more than the end, the causal start still beats the middle by $+6.11$ dB $[+5.64, +6.60]$ and the end by $+7.06$ dB $[+6.39, +7.65]$, and the bidirectional scan gains equally at both ends ($+10.00$ and $+10.01$ dB) and less in the middle ($+7.29$ dB); removing the convolution lowers the end, not the start. With the number of targets drawn per image, so that the block's extent cannot be counted from the start, the causal start still beats the middle by $+5.82$ dB $[+5.46, +6.21]$. The anchor therefore rests on the empty recurrent state, like the first-token position signals of attention models (Xiao et al., 2024) and Mamba-2 (Jiang & Zhang, 2026).

Takeaway. Where an informative block sits matters, but only at the start of a scan direction. Content-aware orders that sort important tokens first place them there, which makes the start of the scan a leakage channel for ground-truth orders and a design choice for learned ones.

8 Related Work

Scan order in vision SSMs. Vision Mamba backbones flatten images with raster, bidirectional, or multi-directional scans (Zhu et al., 2024a; Liu et al., 2024). Many works design locality-preserving paths (Hu et al., 2024; Huang et al., 2024; Li et al., 2025; Hsieh et al., 2026), as space-filling curves do for point clouds (Liang et al., 2024), and empirical studies favour contiguous paths (Hardan et al., 2025; Zhu et al., 2024b) or learn a dataset-level order (Kutscher et al., 2025). Content-aware methods compute the order per image from learned semantics (Guo et al., 2025), locality scores (Xie et al., 2024), ViT importance maps (Hu et al., 2025), or adaptive scan scores (Wang et al., 2025a); in learned image compression, a content-adaptive permutation that makes similar tokens contiguous improves on a raster scan, without a random-permutation control (Chen et al., 2026); a graph SSM sorts nodes by degree, which adds little beyond permutation-based training (Wang et al., 2024). We do not evaluate these methods; we test the proximity rationale they share, in a setting that separates

grouping, information, locality, and position. Yu & Wang (2025) ask more broadly when an SSM mixer is needed in vision.

Position from boundaries and first tokens. Zero padding lets convolutional networks encode absolute position (Islam et al., 2020; Kayhan & van Gemert, 2020), and causal transformers without positional encodings recover it (Haviv et al., 2022; Kazemnejad et al., 2023). First tokens act as attention sinks (Xiao et al., 2024; Gu et al., 2025), Mamba-2 exhibits a state sink on boundary tokens (Jiang & Zhang, 2026), and both ViTs and vision Mamba repurpose background tokens as registers (Darcet et al., 2024; Wang et al., 2025b). SSMs are biased towards recent tokens (Wang et al., 2025c) and limited by their fixed state (Jelassi et al., 2024); they can be read as implicit attention (Ali et al., 2025). Hsieh et al. (2026) alternate scan directions to reduce a start-index bias without quantifying it; our anchor is the scan-order counterpart of these effects.

Leakage and evaluation. Leakage, the use of target information that is unavailable at test time, inflates results across machine learning (Kaufman et al., 2012; Kapoor & Narayanan, 2023), and models exploit such shortcuts readily (Geirhos et al., 2020). Ground-truth orders are leakage that is easy to overlook, because the permutation looks like part of the architecture. We report interval estimates over few runs (Agarwal et al., 2021) and heed the limits of MSE-based fidelity (Wang & Bovik, 2009). Order sensitivity predates SSMs (Vinyals et al., 2016).

9 Discussion and Limitations

Measurement pitfalls. Two pitfalls would have reversed our conclusions. First, ground-truth orders measure leakage (Section 4): an "oracle" order bounds what the answer is worth, not what a reordering can attain. Second, the background dominates whole-image PSNR on sparse targets. The raster reference scores 11.5 dB on the whole image, below an all-zero output (18.0) and the noisy input (15.9), yet restores the informative region (7.4 against 2.8 and 4.1 dB; Appendix B). Among the 51 comparisons that involve no ground truth, 13 change sign between the whole-image and the informative-region metric.

A protocol for content-aware scans. From these results we recommend that evaluations of content-aware orders (i) include a random permutation and a locality-matched order computed from uninformative scores, (ii) report grouping and neighbour retention alongside accuracy, (iii) never use ground-truth orders as upper bounds without leakage controls (a spread lattice and a randomly placed block), (iv) deliver the order's information through a second route, such as an input channel, to separate information from reordering, (v) control where the informative tokens sit relative to the start of each scan direction, and (vi) report region-aware metrics, trivial baselines, and seed-level intervals.

Limitations. The backbone is small (0.69M parameters), has no 2-D convolution by design, and trains briefly. Without ground-truth information it learns the synthetic task only weakly (training objective 0.26 against 0.353 for an all-zero output and 0.10 with ground-truth orders), and three times longer training changes neither this (0.262 to 0.255) nor any conclusion (Appendix H). Whether training-free sorts stay equivalent to random permutations in larger models is untested; on radar, where the reference learns well, the equivalence holds, and on natural images a contrast sort loses nearly as much as a random permutation. The synthetic and radar data are simulated, and there are 2–4 seeds per cell. We formulated H2–H4 after the main experiments and tested them on pre-planned, held-out, and prospectively specified runs; confirmation on production vision SSMs is the natural next step. Appendix D lists four design issues of the first set that we report rather than repair.

Conclusion. Separating grouping from information shows that, in our testbed, proximity is not the mechanism of content-aware scanning: an order helps through the information it makes readable, grouping amplifies that information, and position is read most easily at the start of each scan direction. Evaluations of content-aware scans should control all four properties.

AI use statement

In this work, we used generative AI tools (a large language model assistant) for the following tasks with required disclosure: proposing and refining hypotheses (H2–H4 were formulated with

the assistant after H1 failed), helping develop the conceptual framework (the four properties of a scan order), providing feedback on research methodology (the analysis protocol, the provenance labelling, the held-out test of H4, and the design of the second set of runs), implementing the new experimental arms of the second set (selection lattices, input-channel controls) and generating its variable-K synthetic data, and interpreting results. We have not used generative AI tools to clean or reformat datasets or for qualitative or thematic data analysis; formal mathematical claims, proofs, and translation are not applicable to this work. We also used generative AI tools for tasks with recommended disclosure: creating and editing the analysis and plotting code, creating figures, identifying and summarising related literature, suggesting the structure and title, and drafting and editing text. We have reviewed all AI-assisted work: the analysis code reads only the stored run records, every number in the paper is generated by a script from those records (supplementary material), and every reference was checked against its original source. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

Ethics statement

This work uses synthetic data, simulated radar maps, and public image datasets; it involves no human subjects or personal data. We report negative and unfavourable results, including design issues in our own experiments, to support research integrity.

Reproducibility statement

Section 3 and Appendix A describe the backbone, tasks, orders, training, and metrics. The supplementary material contains the full code, the per-run records of all 1,148 runs, and the scripts that regenerate every number (`numbers.tex`), table, and figure of this paper from those records.

References

Rishabh Agarwal, Max Schwarzer, Pablo Samuel Castro, Aaron Courville, and Marc G. Bellemare. Deep reinforcement learning at the edge of the statistical precipice. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Ameen Ali, Itamar Zimerman, and Lior Wolf. The hidden attention of Mamba models. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2025.

Yunuo Chen, Zezheng Lyu, Bing He, Hongwei Hu, Qi Wang, Yuan Tian, Li Song, Wenjun Zhang, and Guo Lu. Content-aware Mamba for learned image compression. In *International Conference on Learning Representations (ICLR)*, 2026. arXiv:2508.02192.

Tri Dao and Albert Gu. Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In *International Conference on Machine Learning (ICML)*, 2024.

Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision transformers need registers. In *International Conference on Learning Representations (ICLR)*, 2024.

Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A. Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2020. doi: 10.1038/s42256-020-00257-z.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations (ICLR)*, 2022.

Xiangming Gu, Tianyu Pang, Chao Du, Qian Liu, Fengzhuo Zhang, Cunxiao Du, Ye Wang, and Min Lin. When attention sink emerges in language models: An empirical view. In *International Conference on Learning Representations (ICLR)*, 2025.

Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. MambaIR: A simple baseline for image restoration with state-space model. In *European Conference on Computer Vision (ECCV)*, 2024.

Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. MambaIRv2: Attentive state space restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Osama Hardan, Omar Elshenhabi, Tamer Khattab, and Mohamed Mabrok. Flatten wisely: How patch order shapes Mamba-powered vision for MRI segmentation. *arXiv preprint arXiv:2507.13384*, 2025.

Adi Haviv, Ori Ram, Ofir Press, Peter Izsak, and Omer Levy. Transformer language models without positional encodings still learn positional information. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, 2022.

Yi-Kuan Hsieh, Kuan-Chuan Peng, Xin Li, Ming-Ching Chang, Yu-Chee Tseng, and Jun-Wei Hsieh. Partial ring scan: Revisiting scan order in vision state space models. *arXiv preprint arXiv:2602.04170*, 2026.

Han Hu, Zhuoran Zheng, Liang Li, and Chen Lyu. VAMamba: An efficient visual adaptive Mamba for image restoration. *arXiv preprint arXiv:2509.23601*, 2025.

Vincent Tao Hu, Stefan Andreas Baumann, Ming Gui, Olga Grebenkova, Pingchuan Ma, Johannes Schusterbauer, and Björn Ommer. ZigMa: A DiT-style zigzag Mamba diffusion model. In *European Conference on Computer Vision (ECCV)*, 2024.

Tao Huang, Xiaohuan Pei, Shan You, Fei Wang, Chen Qian, and Chang Xu. LocalMamba: Visual state space model with windowed selective scan. *arXiv preprint arXiv:2403.09338*, 2024.

Md Amirul Islam, Sen Jia, and Neil D. B. Bruce. How much position information do convolutional neural networks encode? In *International Conference on Learning Representations (ICLR)*, 2020.

Samy Jelassi, David Brandfonbrener, Sham M. Kakade, and Eran Malach. Repeat after me: Transformers are better than state space models at copying. In *International Conference on Machine Learning (ICML)*, 2024.

Yuhang Jiang and Bowen Zhang. A circuit, not the circuit: Non-unique causal localisation of the Mamba-2 state sink. *arXiv preprint arXiv:2606.00930*, 2026.

Sayash Kapoor and Arvind Narayanan. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9):100804, 2023. doi: 10.1016/j.patter.2023.100804.

Shachar Kaufman, Saharon Rosset, and Claudia Perlich. Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data*, 6(4), 2012. doi: 10.1145/2382577.2382579.

Osman Semih Kayhan and Jan C. van Gemert. On translation invariance in CNNs: Convolutional layers can exploit absolute spatial location. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

Amirhossein Kazemnejad, Inkit Padhi, Karthikeyan Natesan Ramamurthy, Payel Das, and Siva Reddy. The impact of positional encoding on length generalization in transformers. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Declan Kutscher, David M. Chan, Yutong Bai, Trevor Darrell, and Ritwik Gupta. REOrdering patches improves vision models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Boyun Li, Haiyu Zhao, Wenxin Wang, Peng Hu, Yuanbiao Gou, and Xi Peng. MaIR: A locality- and continuity-preserving Mamba for image restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Dingkang Liang, Xin Zhou, Wei Xu, Xingkui Zhu, Zhikang Zou, Xiaoqing Ye, Xiao Tan, and Xiang Bai. PointMamba: A simple state space model for point cloud analysis. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. VMamba: Visual state space model. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

H. Rohling. Radar CFAR thresholding in clutter and multiple target situations. *IEEE Transactions on Aerospace and Electronic Systems*, AES-19:608–621, 1983. doi: 10.1109/TAES.1983.309350.

Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order matters: Sequence to sequence for sets. In *International Conference on Learning Representations (ICLR)*, 2016.

Bo Wang, Mengyuan Xu, Yue Yan, Yuqun Yang, Kechen Shu, Wei Ping, Xu Tang, Wei Jiang, and Zheng You. ASM-UNet: Adaptive scan Mamba integrating group commonalities and individual variations for fine-grained segmentation. *arXiv preprint arXiv:2508.07237*, 2025a.

Chloe Wang, Oleksii Tsepa, Jun Ma, and Bo Wang. Graph-Mamba: Towards long-range graph sequence modeling with selective state spaces. *arXiv preprint arXiv:2402.00789*, 2024.

Feng Wang, Jiahao Wang, Sucheng Ren, Guoyizhe Wei, Jieru Mei, Wei Shao, Yuyin Zhou, Alan Yuille, and Cihang Xie. Mamba-Reg: Vision Mamba also needs registers. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025b.

Peihao Wang, Ruisi Cai, Yuehao Wang, Jiajun Zhu, Pragya Srivastava, Zhangyang Wang, and Pan Li. Understanding and mitigating bottlenecks of state space models through the lens of recency and over-smoothing. In *International Conference on Learning Representations (ICLR)*, 2025c.

Zhou Wang and Alan C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009.

Guangxuan Xiao, Yuandong Tian, Beidi Chen, Song Han, and Mike Lewis. Efficient streaming language models with attention sinks. In *International Conference on Learning Representations (ICLR)*, 2024.

Fei Xie, Weijia Zhang, Zhongdao Wang, and Chao Ma. QuadMamba: Learning quadtree-based selective scan for visual state space model. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Weihao Yu and Xinchao Wang. MambaOut: Do we really need Mamba for vision? In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision Mamba: Efficient visual representation learning with bidirectional state space model. In *International Conference on Machine Learning (ICML)*, 2024a.

Qinfeng Zhu, Yuan Fang, Yuanzhi Cai, Cheng Chen, and Lei Fan. Rethinking scanning strategies with vision Mamba in semantic segmentation of remote sensing imagery: An experimental study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024b. doi: 10.1109/JSTARS.2024.3472296.

A TESTBED DETAILS

Backbone. A 1×1 convolutional stem (patch size 1) maps each pixel to a 96-dimensional token. Five pre-norm blocks follow, each a state space mixer and a gated MLP (expansion 2) with layer scale 0.1. Inside the mixer the tokens are gathered in scan order, projected (expansion 2, four heads, state size 64), passed through a width-4 depthwise convolution along the scanned sequence (zero padding of two positions, output truncated to the sequence length, so each output sees two tokens back and one ahead; removed in the conv-free runs of Appendix H), and processed by a chunked state space duality (SSD) scan (Dao & Gu, 2024) implemented in pure PyTorch and verified against the sequential recurrence (relative error below $2 \cdot 10^{-4}$). With a fixed horizon, every head decays by $a = e^{-1/\tau}$ per step; with a learned horizon, the step size is input-dependent as in Mamba (Gu & Dao, 2023). A bidirectional mixer runs a second recurrence on the reversed sequence and adds the two outputs; the depthwise convolution is shared. Outputs are scattered back to the grid and a 1×1 head predicts a residual correction to the input. There is no 2-D convolution and no down- or up-sampling, so all spatial mixing passes through the scan.

Training. AdamW (learning rate $2 \cdot 10^{-3}$, $\beta = (0.9, 0.95)$, weight decay 0.02), 100 warm-up steps, cosine decay to 2% of the peak, batch 32, gradient clipping at 1.0, bfloat16 autocast on one A100 GPU. Steps: 1,800 (E1–E3, E5–E8), 1,500 (E4, E14), 2,000 (E9, E12, E13), 3,000 (E10, E11). The loss is a Charbonnier loss that weights the support and the background equally, $\ell = \frac{1}{2}\,\overline{\rho}_S + \frac{1}{2}\,\overline{\rho}_B$ with $\rho(e) = \sqrt{e^2 + 10^{-6}}$ and $\overline{\rho}_R$ the mean over region R; it raises the cost of the all-zero output about sixteen-fold at $s = 0.03$ relative to an unweighted loss, which otherwise traps every arm at predicting nothing. Arms that differ only in the order share the per-step training batches (seeded by step) and the evaluation set (192 images for most experiments, drawn once with a fixed seed).

Synthetic sparse restoration. On a 32×32 grid, $K = \mathrm{round}(sN)$ support pixels are drawn uniformly (or as $\sqrt{g}\times\sqrt{g}$ blobs in E4) and given amplitudes in $[0.4, 1.0]$; the target $\boldsymbol{x}$ is this sparse amplitude map. The observation $\boldsymbol{y}$ blurs the sources with a Gaussian whose width is drawn per image from five values in $[0.7, 2.2]$, multiplies by a per-image gain in $[0.6, 1.4]$, and adds a smooth background (amplitude 0.15) and white noise ($\sigma = 0.05$). In E6's per-source condition every source draws its own width and gain from the same marginals. Restoring $\boldsymbol{x}$ therefore requires deblurring, background removal and denoising; every sample has exactly K support pixels.

Simulated radar. Range–Doppler maps (32×32) are the magnitude (dB, normalised) of the 2-D FFT of a windowed FMCW beat signal containing a few point targets (count set by s), zero-Doppler clutter with a range taper, and complex Gaussian noise whose power is drawn per image. The support is the target cells; targets are strong, so the difficulty lies in cleaning the background.

Other domains (appendix only). Natural images: CIFAR-10 in grey scale at 32×32 with Gaussian noise ($\sigma = 0.15$); the support is the half of the pixels whose local contrast is at least the per-image median. Sparse-view CT: 128×128 organ slices (MedMNIST) reconstructed by filtered back-projection from 24, 48 or 96 views, tokenised with patch size 4; the support is the 15% highest-contrast pixels.

Orders. All content-aware orders are expressed relative to the raster base order so that ties and unselected tokens stay in raster order. *Priority*: stable sort by descending score; *anti-priority*: ascending. *Partition*: the top 1.5% of tokens by score first, both groups in base order ($O(N)$ with a threshold). *Window sort*: 8×8 windows ordered by their maximum score, raster order inside each window. *Oracle*: priority sort on the true support mask (E3, E10). In the scorer study (E7) the oracle and noisy-ground-truth scores are the target amplitude $|\boldsymbol{x}|$, so their orders also rank the support tokens by amplitude, which reveals more than membership. *GT block*: the K support tokens form one contiguous block starting at a fixed position ($N/2$ unless a relative position p is given; "fixed position" in the tables) or at a position drawn uniformly per image ("random position"); the remaining tokens fill the other positions in raster order. *GT lattice*: the support tokens are placed every σ positions starting at a fixed position p or at a phase drawn per image ("random phase"); $\sigma = 1$ is the GT block, and the *spread* lattice has the maximal spacing. *Decoy*: a block of K non-support tokens at the contiguous block's position. *Noisy ground truth*: the score $m\,r + (1 - m)\,u$ with r the rank of the target amplitude $|\boldsymbol{x}|$, u uniform noise, and $m \in [0, 1]$ controlling the AUC. The selection lattices of Section 6 use membership only (Appendix H).

Scorers. CFAR: cell-averaging, guard radius 2, training radius 6, the ratio (dB) of the cell's power to the mean of the training ring. Energy: local mean of y^2 (3×3). Contrast: gradient magnitude plus local standard deviation (radius 2). CNN: three 3×3 convolutions of width 24 with GELU and a 1×1 head ($O(N)$). ViT: patch size 2, width 96, four layers, four heads, full attention with learned positional embeddings ($O(N^2)$). The CNN and ViT are trained with a binary cross-entropy loss (weight 0.1) towards the normalised target map; the sort is not differentiated, so the backbone receives nothing from the scorer except the order.

Grouping measure. $C(\pi)$ (Eq. 1) is computed on every evaluation image from the realised permutation, with τ the run's horizon (the effective horizon for learned decay); in causal scans only earlier partners count, which halves C.

Table 2: Experiments and run counts (all runs are included in the released records).

experiment	question	runs
E1	grouping, leakage and decodability controls over $s \times \tau$	312
E2	gain vs. τ/N at three sequence lengths	144
E3	fixed, random, and CFAR-based orders; causal arrival test	99
E4	clustered supports; NSS vs. random-position GT block	48
E5	stride sweep (grouping at fixed information); causal block-position sweep	60
E6	shared vs. per-source nuisance	36
E7	noisy-GT fidelity sweep; CFAR, energy, contrast, CNN, ViT scorers	33
E8	state size	16
E9	learned vs. fixed decay	48
E10	simulated radar: fixed, random, content-aware, learned, GT orders	46
E11	sparse-view CT (content-aware arms failed)	12
E12	natural-image denoising	9
E13	compute-matched widened raster vs. scorer arms	21
E14	block position and stride, causal vs. bidirectional	24
gates, pilot	pre-flight checks	40

B METRICS AND TRIVIAL BASELINES

For every evaluation image the run records hold the whole-image PSNR W and the support PSNR S (data range 1). With support fraction $f = K/N$, $\mathrm{MSE}_W = f\,\mathrm{MSE}_S + (1-f)\,\mathrm{MSE}_B$, so the background PSNR B and the region-balanced PSNR $L = -10\log_{10}(\frac{1}{2}\mathrm{MSE}_S + \frac{1}{2}\mathrm{MSE}_B)$ follow exactly on the synthetic task, where f is the same for every image. On radar the number of target cells varies slightly and we use the mean f, so B and L are approximate there; on CT and natural images we report only W and S. We chose L as the primary metric because it weights the two regions as the training loss does; the loss itself uses a Charbonnier error, and over the 53 arms of the second set, where the objective is recorded per sample, L and the objective rank the arms alike (Spearman 0.99; the sign of the change agrees in 98%). The conclusions of Sections 4–7 hold on S as well (Tables 6–19), with one qualification: 66 of the 72 spread runs gain on S, and the other six (at $\tau = 512$ and $s \geq 0.07$) lie within 0.2 dB of zero (minimum -0.19 dB).

Figure 7 shows why W alone is misleading on sparse targets. At $s = 0.03$ an all-zero output scores 18.0 dB on the whole image and the noisy input itself 15.9 dB, while the trained raster reference scores 11.5 dB: the balanced loss trades background accuracy for support accuracy. On the support the ordering is reversed (2.8, 4.1, and 7.4 dB). Across the 51 comparisons without ground-truth information, 13 change sign between W and S.

C STATISTICS

We pair each arm run with the raster run of the same experimental cell and seed. The cell includes the scan direction, so causal arms are never paired with bidirectional references; a pairing error of this kind affected the original analysis scripts, and we correct it here. The per-image differences of a pair form one unit. We report the mean over seeds and a 95% hierarchical bootstrap interval: 4,000 replicates that resample seeds with replacement and then images within each resampled seed. With two or three seeds these intervals are wide and reflect seed variation; we therefore also report how many seeds share the sign of the mean, and we avoid per-image significance tests, which would treat images as independent replicates. Pooled statements (e.g. over sparsities) resample seed–cell units in the same way.

D DATA INTEGRITY, TRAINING DYNAMICS, AND DESIGN ISSUES

Completeness. All 948 runs (including 40 pre-flight gate and pilot runs) produced result records, and all are released. Of these, 3 were stopped early by the wall-clock budget (an E10 raster reference at 2,000 of 3,000 steps, an E11 raster run at 2,000 of 3,000, and an E7 priority run at 1,200 of 1,800); they are included and marked.

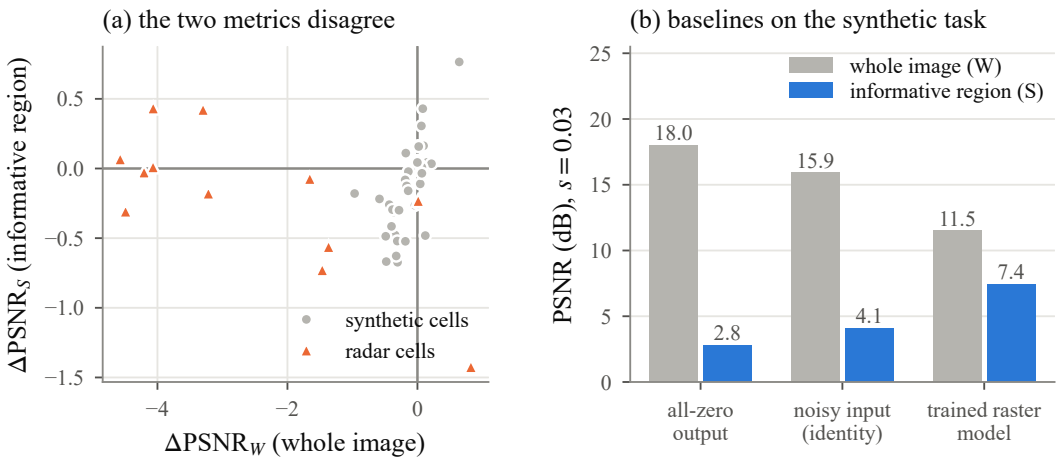

Figure 7: (a) Gain on the whole image against gain on the informative region for every comparison that uses no ground truth; points off the diagonal quadrants change sign between the two metrics. (b) Trivial baselines and the trained raster reference on the synthetic task ($s = 0.03$).

Training dynamics. Figure 8 shows the held-out whole-image PSNR and the training objective during training at $s = 0.03$, $\tau = 16$. Orders without ground-truth information barely lower the objective: the median final value is 0.26 for raster runs, against 0.353 for an all-zero output, 0.347 for the noisy input, and 0.10 for ground-truth orders. The whole-image PSNR of 72% of raster runs peaks early and then declines by more than 0.3 dB, as the balanced loss shifts effort to the support. On radar the reference learns well (median final objective 0.026 against 0.178 for the noisy input). Our conclusions about training-free orders on the synthetic task are therefore drawn in a regime where the reference learns weakly; they replicate on radar, where it does not.

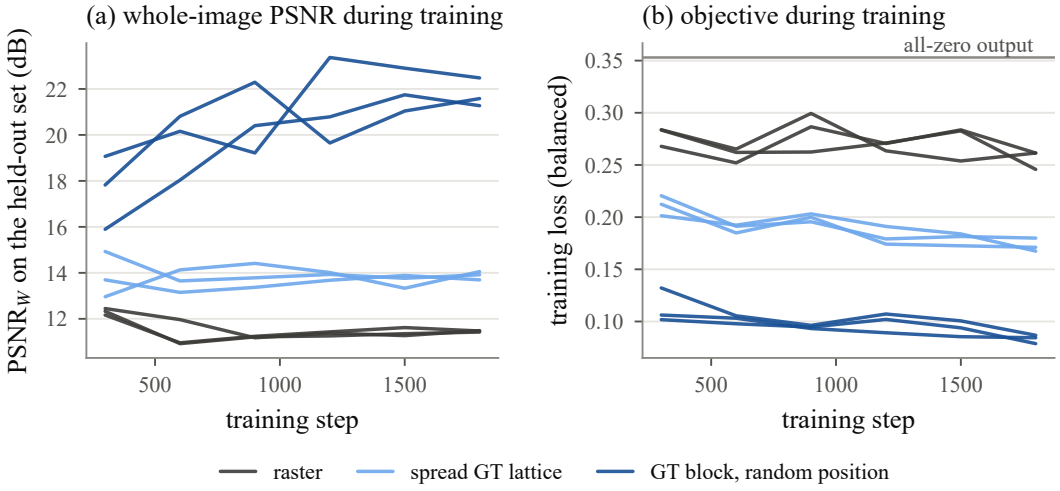

Figure 8: Training dynamics at $s = 0.03$, $\tau = 16$, three seeds per arm. (a) Held-out whole-image PSNR. (b) Training objective; the grey line is the objective of an all-zero output.

Design issues. We report four issues in the experiments rather than repair them after the fact. (i) In E4 the number of blobs is rounded from K/g; for $g = 64$ both sparsities yield a single blob, so the two conditions are identical, and for $g \geq 16$ the support size differs from sN. (ii) In E13 the width of the compute-matched raster backbone is rounded to a multiple of 8, which for the CFAR and CNN arms returns the unwidened backbone; only the ViT arm is compute-matched. (iii) In E11 the score maps of the content-aware arms were not pooled to the patch grid, so these 12 runs failed and

CT contributes no content-aware comparison. (iv) The original analysis scripts paired causal and bidirectional runs that share a sparsity (E3, E5); all numbers in this paper use corrected pairing.

E ADDITIONAL RESULTS

Position readout grows with the memory horizon. Figure 9a compares a ground-truth block at a fixed mid-sequence position with the same block at a random position (E1). With $\tau = 4$ the two differ by at most 0.16 dB; with $\tau = 512$ the fixed block gains $+0.82$ to $+8.34$ dB more, most at low sparsity. A fixed position becomes readable once the memory spans the distance to a sequence boundary, consistent with H4.

Memory spanning the sequence. The quantitative law predicted that grouping stops mattering once $\tau \geq N/2$. Across three sequence lengths the gain of a randomly placed ground-truth block declines by only 0.27 to 1.12 dB from $\tau/N = 0.125$ to 1 and stays at least $+2.76$ dB (Figure 9b).

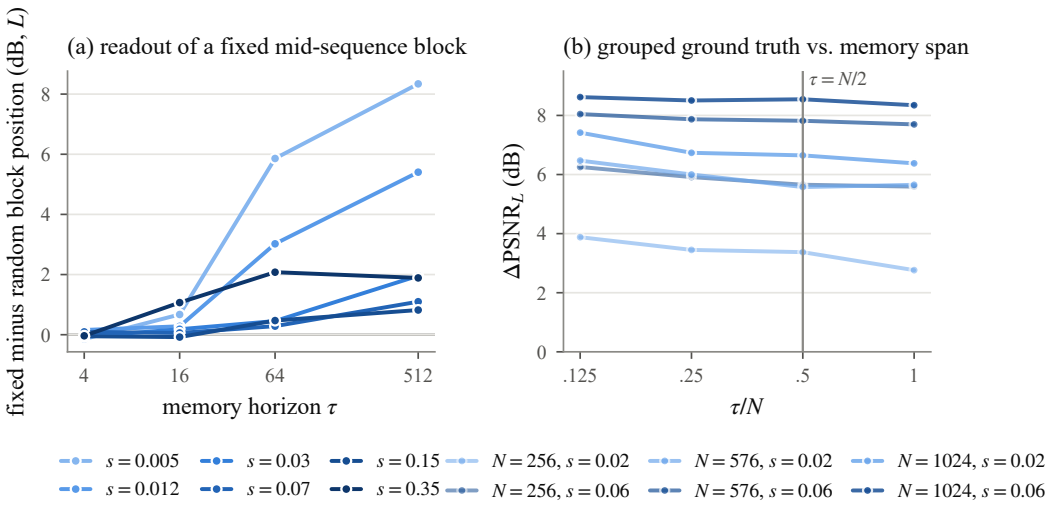

Figure 9: (a) Gain of a fixed-position ground-truth block over the same block at a random position, by memory horizon and sparsity. (b) Gain of a randomly placed ground-truth block against τ/N at three sequence lengths.

Coupling demand (E6). With a per-image nuisance, pooling evidence across sources is useful; with a per-source nuisance it is not. The gain of a randomly placed ground-truth block is larger with the per-source nuisance, but this comes from the reference: raster reaches 7.5 dB support PSNR with the shared nuisance and 6.1 dB with the per-source one, while the ground-truth block reaches 14.9 and 14.9 dB in both. The block's absolute performance is insensitive to whether pooling is useful, which is weak evidence against pooling as its mechanism.

State size and learned decay. Gains of a fixed ground-truth block vary by less than 0.25 dB across state sizes 16–128 ($+8.00$ to $+8.23$ dB, E8). With learned decay the fixed block gains more than with any fixed τ (Table 14), consistent with the readout effect above; the learned step size differs little between support and background tokens (ratio 1.04).

Fidelity sweep (E7). A noisy-ground-truth scorer mixes the target amplitude map with random scores; all its orders are priority sorts, so their locality stays near that of a random permutation (at most 0.06 up to AUC 0.89). The gain rises smoothly with AUC (Figure 10): -0.41 dB at AUC 0.50, $+0.02$ at 0.62, $+0.54$ at 0.76, $+2.79$ at 0.89, and $+5.42$ at 0.96, crossing zero near AUC 0.61 on L and 0.58 on S; only the whole-image metric shows an apparent cliff (near AUC 0.91). Two properties limit this sweep as a test of H3: grouping rises with AUC along it (Table 12), and its scores rank the support by amplitude. The selection lattices of Section 6 remove both. Real scorers fall near or below

the curve: training-free scorers reach AUC 0.58–0.62 and change L by -0.22 to -0.46 dB; the CNN reaches AUC 0.74 and gains $+0.72$ dB $[+0.60, +0.86]$, replicated in E13 ($+0.71$ dB).

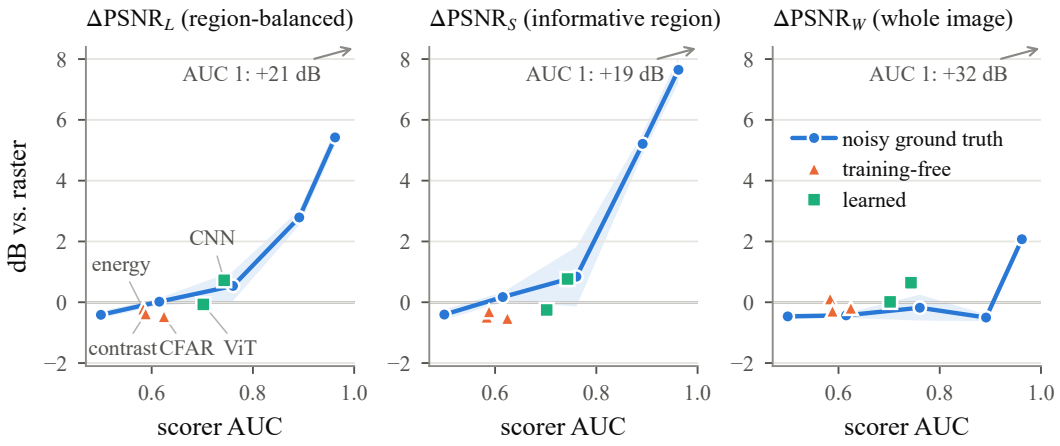

Figure 10: Fidelity sweep (E7): priority sorts driven by the target amplitude corrupted to a chosen AUC (blue; shaded: 95% interval), and by real scorers (markers).

Radar. Figure 11 shows all orders on radar. Fixed locality-preserving scans are near raster (snake) or below it (the nested S-shaped scan, possibly because its narrow stripes break the long rows of the map). Importance sorts by CFAR or energy match the random permutation. The ViT order and the ground-truth order improve the region-balanced metric by cleaning the background; the ViT order changes L by $+0.58$ dB but the target PSNR by -1.42 dB.

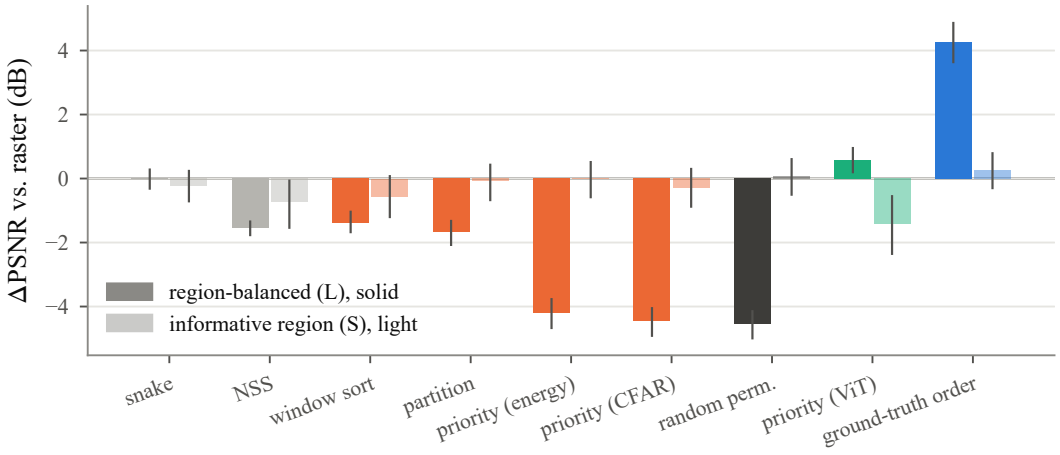

Figure 11: Radar range–Doppler maps: gain over raster on the region-balanced metric (solid) and on the target region (light), mean and 95% interval over three seeds.

Natural images and CT. On natural images, a priority sort by local contrast changes the whole-image PSNR by -3.56 dB and the PSNR of high-contrast pixels by -2.66 dB, and the nested S-shaped scan is within a few tenths of a dB of raster (Table 17). In the second set, a random permutation loses -4.02 dB on the same task; the contrast sort recovers $+0.46$ dB $[+0.43, +0.50]$ of this locality loss (Table 5). On CT only the nested S-shaped scan ran (Table 16); it is within 0.15 dB of raster.

Order transfer. On radar, models trained with scattered orders become insensitive to the order at inference: a model trained with the energy sort changes L by -4.75, -4.72, and -4.19 dB relative to the raster model when evaluated with raster, nested S-shaped, and its own order. The raster model,

in contrast, changes by -4.60 dB when evaluated with a priority order. Training with a scattered order teaches the backbone to ignore spatial context rather than to exploit the grouping.

F THE LOCALITY MEASURE

$P(\pi)$ averages $e^{-|\pi(u)-\pi(v)|/\tau_P}$ over the $2 \cdot 32 \cdot 31$ horizontally and vertically adjacent pixel pairs, with $\tau_P = 16$ matching the default horizon. Content-aware orders are measured on freshly sampled images with the same generator and scorer as the runs. Figure 12 shows that the ranking of the scans is the same for $\tau_P \in \{4, 16, 64\}$: the priority sort always sits with the random permutation, far below the fixed scans. The linear trend of Figure 4b is dominated by the random permutation; among locality-preserving scans the differences in L are within a few tenths of a dB and their slope is not robust (0.42 without the random scan against 0.82 with it). We therefore use the locality measure only for the coarse contrast between scattering and preserving neighbourhoods.

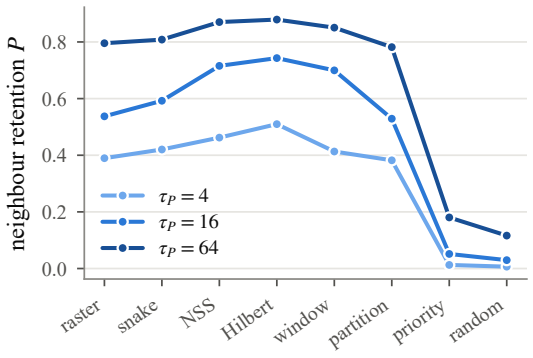

Figure 12: Neighbour retention of each scan for three choices of τ_P (synthetic task, $s = 0.03$).

G PROVENANCE OF THE HYPOTHESES AND STATUS OF THE ORIGINAL PREDICTIONS

We designed the study to test a quantitative form of H1: the gain in PSNR equals $\kappa \cdot 10 \log_{10} \frac{1+C(\pi)}{1+C(\text{raster})}$ with a single constant κ. Its predictions, the experiments designed for them, and the outcomes are listed in Table 3. Predictions marked "weak test" were confirmed, but in a way that competing accounts predict equally (for example, two orders that are both useless are trivially equal). Refitted on the region-balanced metric over the 276 runs whose block position is not fixed, the law explains 0.41 of the variance ($\kappa = 1.03$).

We formulated H2 while inspecting the complete table of results, in which the priority sort and the random permutation coincided in the scan panel (E3). Its agreement with the scorer study (E7, where an order from random scores is a random permutation) and with radar (E10) is a replication across three experiments, not a held-out test. We formulated H3 after H1(a) and (b) failed and tested it on the fidelity sweep, which we had designed in advance to trace a dose–response. We formulated H4 on the causal position sweep of E5 and tested it on E14, whose results we had not analysed at that time. Finally, we designed the second set of runs (Appendix H) after the main analysis and recorded, before running it, what each hypothesis and its competitor predicts (Table 4).

H SECOND SET OF RUNS

We designed a second set of 200 runs after the main analysis to test H2–H4 against the alternatives the first set could not exclude, and recorded the predictions of each hypothesis and its competitor before running them (Table 4). The runs use the original code unchanged (model, data generator, loss, training loop, paired data streams); new arms are added in a separate module. All 200 runs completed on one A100 in 14.8 GPU-hours; every family has raster references of its own cell, and effects are paired with them as in Appendix C. 15 runs repeat a configuration and seed of the first set (raster

Table 3: Predictions of the original quantitative law and their outcomes (L unless noted).

Prediction	Exp.	Result	Status
Given GT information, tighter grouping helps	E5, E14	stride $1 \to 32$: $+8.22 \to +2.08$ dB	held
Random-position GT block beats spread lattice	E1	in 21 of 24 cells	held (mostly)
Gain is non-monotone in sparsity	E1	$\tau = 16$: $+4.96 \to +8.35 \to +5.50$ dB	held
State size is irrelevant	E8	$+8.00$ to $+8.23$ dB	held
Block position is irrelevant at fixed grouping (interior)	E5	$+5.91$ to $+6.44$ dB	held
Anti-priority equals priority (causal)	E3	difference -0.01 dB (both -0.51)	weak test
$O(N)$ partition matches the sort	E3	both below raster	weak test
Cheap CFAR scorer matches the ViT scorer	E7, E13	both at or below raster	weak test
Content-aware orders gain with their grouping	E3, E7, E10	priority sort = random permutation	failed
Less grouping than raster gives no gain	E1	spread gains in 72/72 runs	failed
Block position is irrelevant at fixed grouping (start)	E5, E14	start $+9.70$ vs. interior $+6.44$ dB	failed
Gain requires a shared nuisance to pool	E6	GT block equal in both conditions (support PSNR)	failed
No gain once $\tau \geq N/2$	E1, E2	at least $+2.76$ dB at $\tau = N$	failed
Locality suffices when clusters are large	E4	$g = 64$: NSS -0.34 vs. GT block $+1.58$ dB	failed
One constant κ explains all gains	all	$R^2 = 0.41$	failed

references at $s = 0.02$, the CFAR and CNN orders for three seeds, and the natural-image raster and contrast sort); each reproduces its first-set result to within 10^{-6} dB, so training is deterministic given the configuration and seed. Table 5 lists every arm.

Families. *IG* (information $\times$ grouping, $s = 0.02$, bidirectional, 3 seeds): each order selects K tokens, $\mathrm{round}(qK)$ of them drawn from the support and the rest from the background, and places the selected set on a lattice of stride $\sigma \in \{1, 4, 8, 16, \text{spread}\}$ with a random phase; all other tokens keep raster order; $q \in \{0, 0.25, 0.5, 0.75, 1\}$ ($q = 0$ with $\sigma \in \{1, \text{spread}\}$ only). Only membership is used. *CH* (4 seeds): the CFAR map (z-scored per image), the sigmoid of the CNN scorer's output (trained with the same auxiliary loss as the CNN order, detached from the backbone), the ground-truth support mask, or uniform noise, appended to the input as a second channel with raster order; the same CFAR, CNN and ground-truth information as orders; and a raster backbone of width 104 ($+17\%$ parameters, unlike the rounded control of E13). *NCA* (4 seeds) and *VK* (3 seeds): the ground-truth block at the start, middle, or end of a causal or bidirectional scan with the scan convolution removed; in VK the number of sources is drawn per image, $K \sim \mathcal{U}\{0.01N, \ldots, 0.03N\}$. *NCC* (3 seeds): random permutation, CFAR sort, spread lattice and randomly placed block at $s = 0.03$ without the convolution. *LT* (3 seeds): the core comparisons with 5,400 instead of 1,800 steps. *NAT* (3 seeds): natural-image denoising with a random permutation and the contrast sort.

Metrics. The second set also records per-sample squared and Charbonnier errors on the support and the background, so L is exact for every sample (also with variable K), and the training objective can be compared with L directly.

Longer training. Tripling the number of steps barely changes the raster reference (final objective 0.262 at 1,800 steps, 0.255 at 5,400; whole-image PSNR 11.5 and 11.6 dB, against 19.9 dB for an all-zero output), and every conclusion holds: the CFAR sort matches the random permutation (-0.63 and -0.59 dB; difference -0.03), the spread lattice and the randomly placed block gain ($+3.17$ and $+8.32$ dB), and the CNN order gains $+0.61$ dB.

Table 4: Predictions recorded before the second set of runs, and outcomes (L, dB).

Family	Competing account predicts	Tested account predicts	Outcome
IG	gain depends on grouping only	at fixed grouping, precision adds gain; grouping helps only with information	precision +3.5 dB per unit [2.8, 4.1]; $q = 0$: $\sigma = 1$ -0.06, spread -0.08; R^2 0.62 → 0.96
CH	reordering itself matters	same information as a channel helps as much; CFAR channel ≥ CFAR order	channel/order: GT +9.82/+8.03, CNN +0.78/+0.69, CFAR −0.07/−0.45; noise channel +0.03; width 104 +0.03
NCA	start effect vanishes without the convolution	causal: start only; bidirectional: both ends	start/middle/end: causal +9.94/+3.83/+2.88; bidir. +10.00/+7.29/+10.01
VK	start effect vanishes when K varies	start effect persists	start/middle/end: causal +9.29/+3.47/+3.37; bidir. +9.36/+7.06/+9.36
NCC	—	CFAR sort ≈ random < raster; spread and block gain	CFAR −0.41, random −0.33, spread +1.36, block +7.08
LT	—	conclusions unchanged	CFAR −0.63, random −0.59, spread +3.17, block +8.32, CNN +0.61
NAT	—	random ≈ contrast sort < raster	random −4.02, contrast sort −3.56

Table 5: Second set of runs (Appendix H). ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for L also the 95% hierarchical bootstrap interval. lg: grouping gain $\log \frac{1+C}{1+C_{\text{raster}}}$; n: seeds.

arm	lg	n	W	S	L
IG: selection precision q × lattice stride σ ($s = 0.02$, bidirectional)					
$q = 0$, $\sigma = 1$	0.00	3	−0.04	−0.03	−0.06 [−0.13, +0.04]
$q = 0$, spread	0.00	3	−0.03	−0.03	−0.08 [−0.15, −0.01]
$q = 0.25$, $\sigma = 1$	+0.30	3	+1.33	−0.10	+0.25 [+0.16, +0.33]
$q = 0.25$, $\sigma = 4$	+0.13	3	+0.62	+0.03	+0.20 [+0.12, +0.26]
$q = 0.25$, $\sigma = 8$	+0.06	3	+0.43	0.00	+0.09 [+0.01, +0.18]
$q = 0.25$, $\sigma = 16$	+0.02	3	+0.65	−0.29	−0.05 [−0.12, +0.03]
$q = 0.25$, spread	−0.02	3	+0.51	−0.21	+0.02 [−0.06, +0.10]
$q = 0.5$, $\sigma = 1$	+1.02	3	+3.67	+0.24	+1.00 [+0.88, +1.12]
$q = 0.5$, $\sigma = 4$	+0.54	3	+2.14	+0.43	+0.92 [+0.80, +1.04]
$q = 0.5$, $\sigma = 8$	+0.27	3	+1.43	+0.31	+0.62 [+0.49, +0.75]
$q = 0.5$, $\sigma = 16$	+0.08	3	+1.29	−0.01	+0.33 [+0.20, +0.47]
$q = 0.5$, spread	−0.09	3	+1.10	+0.10	+0.47 [+0.36, +0.58]
$q = 0.75$, $\sigma = 1$	+1.65	3	+6.10	+1.79	+2.74 [+2.54, +2.95]
$q = 0.75$, $\sigma = 4$	+0.99	3	+3.56	+1.81	+2.35 [+2.17, +2.53]
$q = 0.75$, $\sigma = 8$	+0.55	3	+1.82	+1.75	+1.78 [+1.60, +1.94]
$q = 0.75$, $\sigma = 16$	+0.17	3	+1.70	+1.10	+1.29 [+1.12, +1.46]
$q = 0.75$, spread	−0.20	3	+2.09	+0.39	+0.98 [+0.84, +1.11]
$q = 1$, $\sigma = 1$	+2.17	3	+8.92	+7.56	+7.93 [+7.57, +8.28]
$q = 1$, $\sigma = 4$	+1.41	3	+4.47	+7.30	+6.31 [+6.09, +6.52]
$q = 1$, $\sigma = 8$	+0.85	3	+2.55	+5.81	+4.50 [+4.24, +4.77]
$q = 1$, $\sigma = 16$	+0.27	3	+2.39	+3.60	+3.07 [+2.82, +3.33]
$q = 1$, spread	−0.39	3	+3.14	+1.63	+2.17 [+1.98, +2.36]
CH: the same maps as order or as input channel ($s = 0.02$, bidirectional)					
channel: GT support	0.00	4	+21.20	+8.17	+9.82 [+9.60, +10.05]
channel: CFAR	0.00	4	+0.12	−0.11	−0.07 [−0.12, −0.02]
channel: CNN	0.00	4	+0.58	+0.70	+0.78 [+0.66, +0.90]
channel: noise	0.00	4	+0.04	0.00	+0.03 [−0.03, +0.07]
order: GT block, random pos.	+2.17	4	+9.63	+7.46	+8.03 [+7.73, +8.32]
order: priority (CFAR)	+0.29	4	−0.20	−0.52	−0.45 [−0.55, −0.36]
order: priority (CNN)	+0.83	4	+0.72	+0.67	+0.69 [+0.56, +0.82]
raster width 104	0.00	4	+0.05	−0.03	+0.03 [−0.05, +0.11]

arm	lg	n	W	S	L
NCA: GT block position, no scan convolution ($s = 0.02$)					
causal, GT block at $p = 0$	+1.75	4	+17.64	+8.51	+9.94 [+9.71, +10.17]
causal, GT block at $p = 0.5$	+1.75	4	+2.99	+4.49	+3.83 [+3.30, +4.32]
causal, GT block at $p = 1$	+1.75	4	+1.77	+4.09	+2.88 [+2.19, +3.56]
bidir., GT block at $p = 0$	+2.17	4	+20.66	+8.39	+10.00 [+9.81, +10.21]
bidir., GT block at $p = 0.5$	+2.17	4	+7.36	+7.37	+7.29 [+6.80, +7.73]
bidir., GT block at $p = 1$	+2.17	4	+20.82	+8.39	+10.01 [+9.80, +10.22]
VK: GT block position, no scan convolution, K drawn per image					
causal, GT block at $p = 0$	+1.72	3	+12.22	+8.42	+9.29 [+9.02, +9.54]
causal, GT block at $p = 0.5$	+1.72	3	+3.20	+3.83	+3.47 [+3.15, +3.78]
causal, GT block at $p = 1$	+1.72	3	+3.24	+3.95	+3.37 [+3.07, +3.66]
bidir., GT block at $p = 0$	+2.14	3	+12.82	+8.37	+9.36 [+9.12, +9.59]
bidir., GT block at $p = 0.5$	+2.14	3	+7.10	+7.18	+7.06 [+6.55, +7.55]
bidir., GT block at $p = 1$	+2.14	3	+12.96	+8.34	+9.36 [+9.12, +9.60]
NCC: core comparisons, no scan convolution ($s = 0.03$, bidirectional)					
GT block, random pos.	+2.25	3	+7.63	+7.11	+7.08 [+6.68, +7.45]
priority (CFAR)	+0.34	3	−0.06	−0.43	−0.41 [−0.50, −0.32]
random	+0.01	3	−0.36	−0.32	−0.33 [−0.41, −0.26]
spread GT lattice	−0.41	3	+2.31	+0.83	+1.36 [+0.98, +1.67]
LT: three times longer training ($s = 0.02$, 5,400 steps)					
GT block, random pos.	+2.17	3	+10.52	+7.64	+8.32 [+7.96, +8.65]
priority (CFAR)	+0.29	3	−0.24	−0.62	−0.63 [−0.73, −0.52]
priority (CNN)	+0.86	3	+0.80	+0.64	+0.61 [+0.49, +0.73]
random	+0.01	3	−0.53	−0.51	−0.59 [−0.71, −0.49]
spread GT lattice	−0.39	3	+4.95	+2.48	+3.17 [+2.69, +3.76]
NAT: natural-image denoising (learned decay)					
priority (contrast)	+1.31	3	−3.56	−2.66	−3.56 [−3.71, −3.41]
random	+1.31	3	−4.02	−3.08	−4.02 [−4.17, −3.87]

I FULL RESULTS

Tables 6–19 list every arm of every experiment on the whole-image (W), informative-region (S), and region-balanced (L) metrics; background PSNR and per-seed values are in the released records.

Table 6: E1: Grouping sweep over sparsity s and memory horizon τ, with leakage and decodability controls. ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

s	τ	order	C / C_{raster}	n	W	S	L
0.005	4	GT block, fixed pos.	2.50 / 0.03	3	+3.27	+5.48	+4.63 [+4.19, +5.06]
		GT block, random pos.	2.50 / 0.03	3	+3.10	+5.88	+4.77 [+4.35, +5.19]
		spread GT lattice	0.00 / 0.03	3	+4.80	+1.96	+2.80 [+2.49, +3.12]
	16	GT block, fixed pos.	3.54 / 0.13	3	+3.78	+6.56	+5.64 [+4.96, +6.29]
		GT block, random pos.	3.54 / 0.13	3	+3.33	+5.90	+4.96 [+4.51, +5.40]
		spread GT lattice	0.00 / 0.13	3	+3.67	+4.74	+4.56 [+4.22, +4.89]
	64	GT block, fixed pos.	3.88 / 0.53	3	+15.97	+8.30	+9.90 [+9.58, +10.23]
		GT block, random pos.	3.88 / 0.53	3	+2.75	+4.95	+4.04 [+3.57, +4.53]
		spread GT lattice	0.03 / 0.53	3	+8.63	+5.63	+6.56 [+5.11, +8.83]
	512	GT block, fixed pos.	3.98 / 2.33	3	+16.71	+8.35	+10.01 [+9.67, +10.33]
		GT block, random pos.	3.98 / 2.33	3	+1.62	+1.96	+1.67 [+1.13, +2.27]
		spread GT lattice	1.65 / 2.33	3	+9.65	+4.77	+5.98 [+3.80, +9.46]
0.012	4	GT block, fixed pos.	4.52 / 0.08	3	+6.38	+6.49	+6.39 [+5.97, +6.80]
		GT block, random pos.	4.52 / 0.08	3	+5.64	+6.62	+6.24 [+5.83, +6.65]
		spread GT lattice	0.00 / 0.08	3	+3.76	+0.79	+1.76 [+1.56, +1.97]

s	τ	order	C / C_{raster}	n	W	S	L
		decoy block (non-support)	0.08 / 0.08	3	-0.14	$+0.03$	-0.02 [-0.06, $+0.03$]
	16	GT block, fixed pos.	8.51 / 0.33	3	$+7.26$	$+7.28$	$+7.35$ [$+6.94$, $+7.74$]
		GT block, random pos.	8.51 / 0.33	3	$+6.38$	$+7.30$	$+7.07$ [$+6.64$, $+7.48$]
		spread GT lattice	0.01 / 0.33	3	$+3.40$	$+2.08$	$+2.66$ [$+2.30$, $+2.98$]
		decoy block (non-support)	0.33 / 0.33	3	-0.24	$+0.08$	-0.01 [-0.08, $+0.06$]
	64	GT block, fixed pos.	10.29 / 1.30	3	$+13.39$	$+8.31$	$+9.47$ [$+9.04$, $+9.86$]
		GT block, random pos.	10.29 / 1.30	3	$+5.48$	$+6.86$	$+6.45$ [$+6.02$, $+6.89$]
		spread GT lattice	0.54 / 1.30	3	$+4.00$	$+2.03$	$+2.75$ [$+2.55$, $+2.96$]
		decoy block (non-support)	1.30 / 1.30	3	-0.01	-0.08	-0.04 [-0.13, $+0.03$]
	512	GT block, fixed pos.	10.91 / 6.29	3	$+18.02$	$+8.48$	$+9.99$ [$+9.73$, $+10.25$]
		GT block, random pos.	10.91 / 6.29	3	$+4.09$	$+4.94$	$+4.59$ [$+4.12$, $+5.05$]
		spread GT lattice	5.57 / 6.29	3	$+4.23$	$+0.99$	$+1.86$ [$+1.45$, $+2.28$]
		decoy block (non-support)	6.27 / 6.29	3	-0.01	0.00	0.00 [-0.07, $+0.08$]
0.03	4	GT block, fixed pos.	6.02 / 0.20	3	$+8.67$	$+7.37$	$+7.71$ [$+7.36$, $+8.06$]
		GT block, random pos.	6.02 / 0.20	3	$+8.86$	$+7.42$	$+7.78$ [$+7.44$, $+8.12$]
		spread GT lattice	0.00 / 0.20	3	$+2.54$	$+1.50$	$+1.98$ [$+1.77$, $+2.20$]
		decoy block (non-support)	0.21 / 0.20	3	-0.21	$+0.07$	-0.04 [-0.10, $+0.01$]
	16	GT block, fixed pos.	16.88 / 0.89	3	$+10.75$	$+7.72$	$+8.40$ [$+8.07$, $+8.71$]
		GT block, random pos.	16.88 / 0.89	3	$+10.32$	$+7.62$	$+8.22$ [$+7.88$, $+8.54$]
		spread GT lattice	0.26 / 0.89	3	$+2.43$	$+2.25$	$+2.36$ [$+2.18$, $+2.54$]
		decoy block (non-support)	0.91 / 0.89	3	-0.31	$+0.19$	$+0.01$ [-0.04, $+0.06$]
	64	GT block, fixed pos.	25.55 / 3.52	3	$+10.94$	$+7.93$	$+8.61$ [$+8.35$, $+8.88$]
		GT block, random pos.	25.55 / 3.52	3	$+10.07$	$+7.64$	$+8.16$ [$+7.81$, $+8.47$]
		spread GT lattice	2.62 / 3.52	3	$+2.62$	$+1.76$	$+2.04$ [$+1.92$, $+2.17$]
		decoy block (non-support)	3.53 / 3.52	3	-0.18	$+0.01$	-0.05 [-0.11, $+0.01$]
	512	GT block, fixed pos.	29.38 / 17.09	3	$+17.52$	$+8.31$	$+9.72$ [$+9.49$, $+9.94$]
		GT block, random pos.	29.38 / 17.09	3	$+8.79$	$+7.58$	$+7.76$ [$+7.38$, $+8.13$]
		spread GT lattice	16.34 / 17.09	3	$+3.08$	$+0.51$	$+1.18$ [$+1.05$, $+1.31$]
		decoy block (non-support)	16.97 / 17.09	3	-0.16	0.00	-0.05 [-0.12, $+0.01$]
0.07	4	GT block, fixed pos.	6.60 / 0.48	3	$+9.61$	$+7.54$	$+8.07$ [$+7.79$, $+8.34$]
		GT block, random pos.	6.60 / 0.48	3	$+9.64$	$+7.43$	$+7.98$ [$+7.67$, $+8.26$]
		spread GT lattice	0.06 / 0.48	3	$+3.33$	$+2.69$	$+3.00$ [$+2.81$, $+3.16$]
	16	GT block, fixed pos.	23.98 / 2.10	3	$+10.40$	$+7.64$	$+8.34$ [$+8.10$, $+8.58$]
		GT block, random pos.	23.98 / 2.10	3	$+10.47$	$+7.58$	$+8.29$ [$+7.98$, $+8.55$]
		spread GT lattice	1.34 / 2.10	3	$+3.24$	$+2.66$	$+2.88$ [$+2.48$, $+3.23$]
	64	GT block, fixed pos.	50.16 / 8.26	3	$+11.68$	$+7.88$	$+8.85$ [$+8.60$, $+9.07$]
		GT block, random pos.	50.16 / 8.26	3	$+10.84$	$+7.79$	$+8.56$ [$+8.31$, $+8.81$]
		spread GT lattice	7.37 / 8.26	3	$+2.21$	$+0.73$	$+1.21$ [$+1.06$, $+1.36$]
	512	GT block, fixed pos.	67.74 / 40.29	3	$+15.19$	$+8.15$	$+9.59$ [$+9.36$, $+9.82$]
		GT block, random pos.	67.74 / 40.29	3	$+10.11$	$+7.93$	$+8.49$ [$+8.23$, $+8.73$]
		spread GT lattice	39.62 / 40.29	3	$+2.32$	-0.05	$+0.65$ [$+0.57$, $+0.73$]
0.15	4	GT block, fixed pos.	6.83 / 1.04	3	$+9.04$	$+7.17$	$+7.71$ [$+7.40$, $+7.99$]
		GT block, random pos.	6.83 / 1.04	3	$+8.99$	$+7.27$	$+7.76$ [$+7.46$, $+8.03$]
		spread GT lattice	0.46 / 1.04	3	$+4.63$	$+5.10$	$+4.92$ [$+4.62$, $+5.18$]
	16	GT block, fixed pos.	27.69 / 4.55	3	$+9.98$	$+7.47$	$+8.27$ [$+8.00$, $+8.54$]
		GT block, random pos.	27.69 / 4.55	3	$+10.17$	$+7.49$	$+8.35$ [$+8.12$, $+8.58$]
		spread GT lattice	3.78 / 4.55	3	$+3.17$	$+3.45$	$+3.34$ [$+2.99$, $+3.59$]
	64	GT block, fixed pos.	78.60 / 17.78	3	$+11.04$	$+7.79$	$+8.87$ [$+8.65$, $+9.09$]
		GT block, random pos.	78.60 / 17.78	3	$+10.24$	$+7.49$	$+8.40$ [$+8.15$, $+8.65$]
		spread GT lattice	16.97 / 17.78	3	$+1.78$	$+0.67$	$+1.12$ [$+0.89$, $+1.47$]
	512	GT block, fixed pos.	138.66 / 86.74	3	$+12.69$	$+8.17$	$+9.57$ [$+9.38$, $+9.76$]
		GT block, random pos.	138.66 / 86.74	3	$+10.61$	$+7.79$	$+8.75$ [$+8.57$, $+8.93$]
		spread GT lattice	86.19 / 86.74	3	$+1.70$	-0.06	$+0.61$ [$+0.51$, $+0.69$]
0.35	4	GT block, fixed pos.	6.95 / 2.44	3	$+3.96$	$+5.47$	$+4.19$ [$+3.64$, $+4.84$]
		GT block, random pos.	6.95 / 2.44	3	$+4.15$	$+5.06$	$+4.23$ [$+3.45$, $+5.16$]

s	τ	order	C / C_{raster}	n	W	S	L
		spread GT lattice	1.91 / 2.44	3	+5.20	+5.02	+5.15 [+4.92, +5.37]
	16	GT block, fixed pos.	29.58 / 10.65	3	+6.79	+6.63	+6.57 [+5.11, +8.65]
		GT block, random pos.	29.58 / 10.65	3	+5.52	+5.97	+5.50 [+5.18, +5.81]
		spread GT lattice	10.03 / 10.65	3	+4.89	+4.72	+4.83 [+4.65, +5.00]
	64	GT block, fixed pos.	104.21 / 41.54	3	+10.57	+8.29	+9.69 [+9.53, +9.86]
		GT block, random pos.	104.21 / 41.54	3	+8.00	+7.07	+7.61 [+7.23, +7.97]
		spread GT lattice	40.89 / 41.54	3	+2.37	+1.99	+2.25 [+2.00, +2.42]
	512	GT block, fixed pos.	286.32 / 202.51	3	+10.63	+8.32	+9.74 [+9.57, +9.90]
		GT block, random pos.	286.32 / 202.51	3	+8.21	+7.26	+7.85 [+7.63, +8.08]
		spread GT lattice	202.05 / 202.51	3	+1.03	+0.13	+0.71 [+0.56, +0.82]

Table 7: E2: Sequence length: the ratio τ/N. ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

$\sqrt{N}$	τ	s	order	C / C_{raster}	n	W	S	L
16	32	0.02	GT block, random pos.	3.76 / 0.85	3	+2.66	+4.79	+3.88 [+3.36, +4.38]
		0.06	GT block, random pos.	11.92 / 2.99	3	+6.23	+6.53	+6.26 [+5.81, +6.66]
	64	0.02	GT block, random pos.	3.88 / 1.48	3	+2.52	+4.16	+3.45 [+3.03, +3.87]
		0.06	GT block, random pos.	12.90 / 5.23	3	+5.74	+6.36	+5.91 [+5.52, +6.29]
	128	0.02	GT block, random pos.	3.94 / 2.25	3	+2.57	+4.06	+3.37 [+2.68, +4.08]
		0.06	GT block, random pos.	13.43 / 7.92	3	+5.35	+6.19	+5.65 [+5.24, +6.08]
	256	0.02	GT block, random pos.	3.97 / 2.93	3	+1.91	+3.50	+2.76 [+2.20, +3.28]
		0.06	GT block, random pos.	13.71 / 10.29	3	+5.59	+5.98	+5.59 [+5.15, +6.02]
24	72	0.02	GT block, random pos.	10.36 / 2.37	3	+5.77	+6.97	+6.47 [+5.90, +7.00]
		0.06	GT block, random pos.	28.96 / 7.40	3	+9.37	+7.65	+8.05 [+7.76, +8.31]
	144	0.02	GT block, random pos.	10.68 / 4.10	3	+5.49	+6.48	+6.00 [+5.59, +6.42]
		0.06	GT block, random pos.	31.33 / 12.79	3	+8.78	+7.62	+7.87 [+7.61, +8.12]
	288	0.02	GT block, random pos.	10.84 / 6.19	3	+5.23	+5.98	+5.58 [+5.11, +6.02]
		0.06	GT block, random pos.	32.63 / 19.26	3	+8.60	+7.67	+7.82 [+7.56, +8.08]
	576	0.02	GT block, random pos.	10.92 / 8.06	3	+5.52	+6.00	+5.65 [+5.09, +6.27]
		0.06	GT block, random pos.	33.30 / 24.99	3	+8.35	+7.61	+7.70 [+7.44, +7.95]
32	128	0.02	GT block, random pos.	18.00 / 4.20	3	+8.28	+7.29	+7.42 [+6.99, +7.85]
		0.06	GT block, random pos.	51.36 / 13.10	3	+10.95	+7.85	+8.62 [+8.35, +8.88]
	256	0.02	GT block, random pos.	18.49 / 7.24	3	+7.23	+6.80	+6.74 [+6.03, +7.38]
		0.06	GT block, random pos.	55.43 / 22.62	3	+10.47	+7.85	+8.51 [+8.27, +8.74]
	512	0.02	GT block, random pos.	18.74 / 10.86	3	+6.80	+6.88	+6.65 [+6.19, +7.05]
		0.06	GT block, random pos.	57.65 / 34.05	3	+10.41	+7.94	+8.55 [+8.30, +8.78]
	1024	0.02	GT block, random pos.	18.87 / 14.04	3	+6.69	+6.55	+6.38 [+5.47, +7.23]
		0.06	GT block, random pos.	58.81 / 44.14	3	+9.94	+7.81	+8.35 [+8.10, +8.57]

Table 8: E3: Scan panel with CFAR-based importance orders; causal arrival test at $s = 0.03$. ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

scan	s	order	AUC	C / C_{raster}	n	W	S	L
causal	0.03	priority (CFAR)	0.62	0.83 / 0.45	3	-0.40	-0.42	-0.51 [-0.61, -0.41]
		anti-priority (CFAR)	0.62	0.83 / 0.45	3	-0.38	-0.43	-0.52 [-0.61, -0.43]
bidir.	0.01	snake	–	0.27 / 0.26	3	$+0.22$	$+0.03$	$+0.11$ [$+0.03$, $+0.19$]
		NSS	–	0.29 / 0.26	3	$+0.01$	$+0.16$	$+0.09$ [0.00, $+0.19$]
		Hilbert	–	0.27 / 0.26	3	$+0.06$	$+0.31$	$+0.24$ [$+0.14$, $+0.33$]
		random permutation	–	0.29 / 0.26	3	-0.58	-0.22	-0.22 [-0.31, -0.13]
		priority (CFAR)	0.62	0.55 / 0.26	3	-0.38	-0.30	-0.22 [-0.35, -0.10]
		anti-priority (CFAR)	0.62	0.55 / 0.26	3	-0.43	-0.26	-0.21 [-0.33, -0.10]
		partition (CFAR)	0.62	0.41 / 0.26	3	-0.14	-0.02	-0.04 [-0.08, $+0.01$]
		window sort (CFAR)	0.62	0.29 / 0.26	3	-0.18	$+0.11$	$+0.03$ [-0.04, $+0.11$]
		oracle sort (GT)	–	7.22 / 0.26	3	$+24.02$	$+8.31$	$+10.10$ [$+9.84$, $+10.37$]
	0.03	snake	–	0.90 / 0.89	3	$+0.05$	$+0.02$	$+0.02$ [-0.04, $+0.08$]
		NSS	–	0.88 / 0.89	3	0.00	$+0.06$	$+0.02$ [-0.08, $+0.10$]
		Hilbert	–	0.89 / 0.89	3	$+0.09$	$+0.16$	$+0.14$ [$+0.07$, $+0.21$]
		random permutation	–	0.91 / 0.89	3	-0.49	-0.49	-0.57 [-0.67, -0.48]
		priority (CFAR)	0.62	1.66 / 0.89	3	-0.31	-0.52	-0.60 [-0.70, -0.51]
		anti-priority (CFAR)	0.62	1.66 / 0.89	3	-0.34	-0.48	-0.59 [-0.69, -0.49]
		partition (CFAR)	0.62	1.46 / 0.89	3	-0.19	-0.08	-0.26 [-0.34, -0.20]
		window sort (CFAR)	0.62	0.90 / 0.89	3	-0.17	-0.13	-0.16 [-0.23, -0.10]
		oracle sort (GT)	–	16.88 / 0.89	3	$+18.86$	$+8.17$	$+9.64$ [$+9.37$, $+9.91$]
	0.1	snake	–	3.02 / 3.00	3	$+0.07$	-0.02	$+0.01$ [-0.04, $+0.06$]
		NSS	–	3.02 / 3.00	3	$+0.04$	-0.11	-0.06 [-0.10, -0.02]
		Hilbert	–	3.02 / 3.00	3	$+0.15$	$+0.04$	$+0.08$ [$+0.04$, $+0.12$]
		random permutation	–	3.02 / 3.00	3	-0.48	-0.67	-0.63 [-0.73, -0.53]
		priority (CFAR)	0.65	4.66 / 3.00	3	-0.33	-0.63	-0.68 [-0.78, -0.59]
		anti-priority (CFAR)	0.65	4.66 / 3.00	3	-0.31	-0.67	-0.70 [-0.80, -0.61]
		partition (CFAR)	0.65	4.33 / 3.00	3	-0.37	-0.31	-0.38 [-0.46, -0.30]
		window sort (CFAR)	0.65	3.02 / 3.00	3	-0.04	-0.26	-0.18 [-0.22, -0.15]
		oracle sort (GT)	–	26.00 / 3.00	3	$+14.82$	$+7.99$	$+9.57$ [$+9.35$, $+9.78$]

Table 9: E4: Spatially clustered supports of cluster size g. ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

g	s	order	C / C_{raster}	n	W	S	L
1	0.02	NSS	0.57 / 0.57	2	$+0.07$	$+0.01$	$+0.08$ [0.00, $+0.16$]
		GT block, random pos.	12.75 / 0.57	2	$+9.34$	$+7.50$	$+8.03$ [$+7.60$, $+8.44$]
	0.08	NSS	2.41 / 2.39	2	0.00	$+0.04$	$+0.02$ [-0.06, $+0.10$]
		GT block, random pos.	24.81 / 2.39	2	$+10.36$	$+7.73$	$+8.38$ [$+8.06$, $+8.70$]
4	0.02	NSS	2.29 / 1.62	2	$+0.08$	$+0.43$	$+0.33$ [$+0.20$, $+0.46$]
		GT block, random pos.	12.33 / 1.62	2	$+8.05$	$+4.14$	$+5.23$ [$+4.82$, $+5.63$]
	0.08	NSS	3.87 / 3.29	2	-0.14	-0.16	-0.15 [-0.41, $+0.11$]
		GT block, random pos.	24.20 / 3.29	2	$+8.72$	$+4.92$	$+5.85$ [$+5.47$, $+6.20$]
16	0.02	NSS	6.40 / 3.43	2	$+0.02$	$+0.16$	$+0.10$ [-0.06, $+0.28$]
		GT block, random pos.	10.12 / 3.43	2	$+9.97$	$+1.81$	$+3.19$ [$+2.83$, $+3.57$]
	0.08	NSS	7.75 / 5.04	2	$+0.07$	-0.03	$+0.01$ [-0.15, $+0.17$]
		GT block, random pos.	23.99 / 5.04	2	$+8.50$	$+2.62$	$+3.99$ [$+3.69$, $+4.31$]
64	0.02	NSS	14.19 / 7.45	2	-0.97	-0.18	-0.34 [-0.49, -0.18]
		GT block, random pos.	21.79 / 7.45	2	$+5.74$	$+0.77$	$+1.58$ [$+1.35$, $+1.81$]

g	s	order	C / C_{raster}	n	W	S	L
	0.08	NSS	14.19 / 7.45	2	-0.97	-0.18	-0.34 [-0.49, -0.18]
		GT block, random pos.	21.79 / 7.45	2	$+5.74$	$+0.77$	$+1.58$ [$+1.37$, $+1.81$]

Table 10: E5: Stride sweep and block-position sweep (causal); p: relative position in the scan (0 = start); lattices without p have a random phase. ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

scan	s	order	C / C_{raster}	n	W	S	L
causal	0.03	GT block at $p = 0$	8.44 / 0.45	3	$+19.08$	$+8.20$	$+9.70$ [$+9.44$, $+9.96$]
		GT block at $p = 0.25$	8.44 / 0.45	3	$+7.14$	$+5.57$	$+5.91$ [$+5.52$, $+6.29$]
		GT block at $p = 0.5$	8.44 / 0.45	3	$+7.25$	$+6.22$	$+6.44$ [$+6.09$, $+6.78$]
		GT block at $p = 0.75$	8.44 / 0.45	3	$+7.15$	$+5.99$	$+6.22$ [$+5.80$, $+6.64$]
		GT block at $p = 1$	8.44 / 0.45	3	$+8.23$	$+5.80$	$+6.23$ [$+5.86$, $+6.59$]
bidir.	0.012	GT block, random pos.	8.51 / 0.33	3	$+6.38$	$+7.30$	$+7.07$ [$+6.64$, $+7.46$]
		GT lattice $\sigma = 2$	6.74 / 0.33	3	$+6.77$	$+7.54$	$+7.47$ [$+7.13$, $+7.79$]
		GT lattice $\sigma = 4$	4.52 / 0.33	3	$+4.49$	$+7.68$	$+6.73$ [$+6.47$, $+6.99$]
		GT lattice $\sigma = 8$	2.43 / 0.33	3	$+2.30$	$+6.48$	$+4.85$ [$+4.61$, $+5.11$]
		GT lattice $\sigma = 16$	1.01 / 0.33	3	$+1.98$	$+4.66$	$+3.60$ [$+3.37$, $+3.82$]
		GT lattice $\sigma = 32$	0.28 / 0.33	3	$+1.87$	$+2.18$	$+2.13$ [$+1.91$, $+2.35$]
	0.03	GT block, random pos.	16.88 / 0.89	3	$+10.32$	$+7.62$	$+8.22$ [$+7.88$, $+8.56$]
		GT lattice $\sigma = 2$	10.98 / 0.89	3	$+6.15$	$+7.87$	$+7.30$ [$+7.08$, $+7.52$]
		GT lattice $\sigma = 4$	6.02 / 0.89	3	$+3.83$	$+7.03$	$+5.72$ [$+5.46$, $+5.96$]
		GT lattice $\sigma = 8$	2.83 / 0.89	3	$+1.92$	$+5.36$	$+3.95$ [$+3.74$, $+4.17$]
		GT lattice $\sigma = 16$	1.10 / 0.89	3	$+2.25$	$+3.14$	$+2.82$ [$+2.55$, $+3.08$]
		GT lattice $\sigma = 32$	0.30 / 0.89	3	$+2.00$	$+2.04$	$+2.08$ [$+1.89$, $+2.27$]

Table 11: E6: Shared vs. per-source nuisance (coupling demand). ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

s	nuisance	order	C / C_{raster}	n	W	S	L
0.01	per-source	GT block, random pos.	7.22 / 0.26	3	$+8.15$	$+8.44$	$+8.19$ [$+7.83$, $+8.50$]
	shared	GT block, random pos.	7.22 / 0.26	3	$+6.01$	$+7.12$	$+6.86$ [$+6.45$, $+7.25$]
0.03	per-source	GT block, random pos.	16.88 / 0.89	3	$+11.73$	$+9.06$	$+9.52$ [$+9.34$, $+9.67$]
	shared	GT block, random pos.	16.88 / 0.89	3	$+10.32$	$+7.62$	$+8.22$ [$+7.86$, $+8.55$]
0.1	per-source	GT block, random pos.	26.00 / 3.00	3	$+11.16$	$+9.06$	$+9.52$ [$+9.41$, $+9.63$]
	shared	GT block, random pos.	26.00 / 3.00	3	$+10.45$	$+7.59$	$+8.35$ [$+8.09$, $+8.60$]

Table 12: E7: Scorer fidelity (noisy ground truth) and real scorers. ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

s	order	AUC	C / C_{raster}	n	W	S	L
0.02	priority (CFAR)	0.62	1.10 / 0.57	3	-0.19	-0.52	-0.46 [-0.58, -0.36]
	priority (energy)	0.58	0.63 / 0.57	3	$+0.12$	-0.48	-0.22 [-0.30, -0.13]
	priority (contrast)	0.59	0.96 / 0.57	3	-0.28	-0.30	-0.37 [-0.48, -0.27]
	priority (CNN)	0.74	2.62 / 0.57	3	$+0.64$	$+0.77$	$+0.72$ [$+0.60$, $+0.86$]

s	order	AUC	C / C_{raster}	n	W	S	L
	priority (ViT)	0.70	1.78 / 0.57	3	$+0.01$	-0.25	-0.07 [-0.15, $+0.01$]
	noisy GT, $m = 0$	0.50	0.59 / 0.57	2	-0.47	-0.41	-0.41 [-0.53, -0.29]
	noisy GT, $m = 0.2$	0.62	0.73 / 0.57	2	-0.43	$+0.17$	$+0.02$ [-0.11, $+0.14$]
	noisy GT, $m = 0.4$	0.76	1.16 / 0.57	2	-0.18	$+0.84$	$+0.54$ [$+0.04$, $+1.02$]
	noisy GT, $m = 0.6$	0.89	2.13 / 0.57	2	-0.50	$+5.21$	$+2.79$ [$+2.53$, $+3.04$]
	noisy GT, $m = 0.8$	0.96	4.30 / 0.57	2	$+2.07$	$+7.65$	$+5.42$ [$+5.14$, $+5.68$]
	noisy GT, $m = 1$	1.00	12.75 / 0.57	2	$+32.05$	$+19.29$	$+20.94$ [$+20.44$, $+21.46$]
	oracle sort (GT)	–	12.75 / 0.57	3	$+32.08$	$+19.35$	$+20.96$ [$+20.55$, $+21.40$]

Table 13: E8: State size d_{state}. ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

d_{state}	order	C / C_{raster}	n	W	S	L
16	GT block, fixed pos.	12.75 / 0.57	2	$+10.29$	$+7.57$	$+8.23$ [$+7.79$, $+8.67$]
32	GT block, fixed pos.	12.75 / 0.57	2	$+9.52$	$+7.44$	$+8.00$ [$+7.61$, $+8.36$]
64	GT block, fixed pos.	12.75 / 0.57	2	$+9.78$	$+7.55$	$+8.15$ [$+7.78$, $+8.50$]
128	GT block, fixed pos.	12.75 / 0.57	2	$+9.41$	$+7.60$	$+8.10$ [$+7.71$, $+8.48$]

Table 14: E9: Learned decay (τ learned) vs. fixed horizons. ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

τ	s	order	C / C_{raster}	n	W	S	L
learned	0.01	GT block, fixed pos.	9.00 / 8.08	3	$+16.30$	$+7.76$	$+9.18$ [$+8.86$, $+9.52$]
	0.03	GT block, fixed pos.	29.99 / 29.47	3	$+15.29$	$+7.66$	$+8.94$ [$+8.51$, $+9.37$]
8	0.01	GT block, fixed pos.	5.90 / 0.13	3	$+6.62$	$+7.10$	$+7.04$ [$+6.56$, $+7.49$]
	0.03	GT block, fixed pos.	10.98 / 0.43	3	$+9.40$	$+7.47$	$+7.93$ [$+7.59$, $+8.27$]
16	0.01	GT block, fixed pos.	7.22 / 0.26	3	$+6.64$	$+7.29$	$+7.22$ [$+6.72$, $+7.65$]
	0.03	GT block, fixed pos.	16.88 / 0.89	3	$+11.35$	$+7.64$	$+8.44$ [$+8.11$, $+8.75$]
48	0.01	GT block, fixed pos.	8.35 / 0.81	3	$+11.38$	$+8.07$	$+8.96$ [$+8.44$, $+9.57$]
	0.03	GT block, fixed pos.	24.28 / 2.67	3	$+11.26$	$+7.93$	$+8.66$ [$+8.39$, $+8.93$]

Table 15: E10: Simulated radar range–Doppler maps (learned decay). ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

s	sweep	order	AUC	C / C_{raster}	n	W	S	L
0.002	sweep	priority (CFAR)	0.99	0.52 / 0.47	2	-3.22	-0.18	-3.19 [-3.76, -2.66]
0.004	panel	snake	–	1.94 / 1.91	3	$+0.01$	-0.23	0.00 [-0.35, $+0.31$]
		NSS	–	1.99 / 1.91	3	-1.46	-0.73	-1.55 [-1.80, -1.31]
		random permutation	–	1.93 / 1.91	3	-4.57	$+0.07$	-4.55 [-5.03, -4.11]
		priority (CFAR)	0.98	2.00 / 1.91	3	-4.49	-0.31	-4.45 [-4.95, -4.02]
		partition (CFAR)	0.98	1.98 / 1.91	3	-1.66	-0.07	-1.66 [-2.11, -1.29]
		window sort (CFAR)	0.98	1.96 / 1.91	3	-1.37	-0.56	-1.38 [-1.71, -1.00]
		priority (energy)	0.98	2.00 / 1.91	3	-4.21	-0.03	-4.19 [-4.71, -3.74]
		priority (ViT)	0.99	2.00 / 1.91	3	$+0.82$	-1.42	$+0.58$ [$+0.16$, $+0.99$]

s	sweep	order	AUC	C / C_{raster}	n	W	S	L
		oracle sort (GT)	–	2.00 / 1.91	3	$+4.27$	$+0.27$	$+4.26$ $[+3.60, +4.90]$
0.005	sweep	priority (CFAR)	0.98	2.58 / 1.87	2	-4.07	$+0.01$	-3.91 $[-4.21, -3.62]$
0.015	sweep	priority (CFAR)	0.97	9.66 / 8.05	2	-4.07	$+0.43$	-3.92 $[-4.12, -3.71]$
0.04	sweep	priority (CFAR)	0.94	28.73 / 24.86	2	-3.30	$+0.42$	-3.12 $[-3.27, -2.96]$

Table 16: E11: Sparse-view CT (content-aware arms failed; see Appendix D). ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

views	order	C / C_{raster}	n	W	S	L
24	NSS	303.71 / 325.90	2	$+0.08$	$+0.01$	–
48	NSS	331.71 / 334.44	2	$+0.11$	$+0.14$	–
96	NSS	316.50 / 344.12	2	$+0.04$	$+0.08$	–

Table 17: E12: Natural-image denoising. ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

σ_n	order	C / C_{raster}	n	W	S	L
0.15	NSS	257.16 / 137.04	3	-0.22	-0.34	–
	priority (contrast)	511.00 / 137.04	3	-3.56	-2.66	–

Table 18: E13: Compute-matched backbone (widened raster, width factor w) vs. scorer arms. ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

s	order	AUC	C / C_{raster}	n	W	S	L
0.02	raster, $w = 1.000$	–	0.57 / 0.57	3	0.00	0.00	0.00 $[0.00, 0.00]$
	raster, $w = 1.005$	–	0.57 / 0.57	3	0.00	0.00	0.00 $[0.00, 0.00]$
	raster, $w = 1.068$	–	0.57 / 0.57	3	-0.06	-0.01	-0.03 $[-0.10, +0.04]$
	priority (CFAR)	0.62	1.10 / 0.57	3	-0.51	-0.33	-0.46 $[-0.56, -0.36]$
	priority (CNN)	0.74	2.63 / 0.57	3	$+0.50$	$+0.84$	$+0.71$ $[+0.60, +0.83]$
	priority (ViT)	0.70	1.79 / 0.57	3	-0.16	-0.13	-0.06 $[-0.14, +0.02]$

Table 19: E14: Position and stride under causal and bidirectional scans (p: relative position, 0 = start). ΔPSNR (dB) relative to the raster run of the same cell and seed, mean over seeds; for the primary metric L also the 95% hierarchical bootstrap interval (intervals for W and S are in the released records). C: grouping of the arm and of raster; n: seeds.

scan	s	order	C / C_{raster}	n	W	S	L
causal	0.02	GT block at $p = 0$	6.38 / 0.29	2	$+20.86$	$+8.41$	$+9.97$ $[+9.67, +10.26]$
		GT block at $p = 0.5$	6.38 / 0.29	2	$+5.75$	$+6.08$	$+5.99$ $[+5.58, +6.38]$
		GT block at $p = 1$	6.38 / 0.29	2	$+7.43$	$+5.89$	$+6.23$ $[+5.76, +6.67]$
		GT lattice $\sigma = 8$ at $p = 0.5$	1.35 / 0.29	2	$+2.23$	$+3.92$	$+3.34$ $[+3.08, +3.59]$
		GT lattice $\sigma = 32$ at $p = 0.5$	0.15 / 0.29	2	$+1.91$	$+1.78$	$+2.01$ $[+1.76, +2.26]$

scan	s	order	C / C_{raster}	n	W	S	L
bidir.	0.02	GT block at $p = 0$	12.75 / 0.57	2	$+20.71$	$+8.30$	$+9.90$ $[+9.60, +10.19]$
		GT block at $p = 0.5$	12.75 / 0.57	2	$+9.52$	$+7.75$	$+8.24$ $[+7.89, +8.60]$
		GT block at $p = 1$	12.75 / 0.57	2	$+20.83$	$+8.29$	$+9.89$ $[+9.59, +10.18]$
		GT lattice $\sigma = 8$ at $p = 0.5$	2.69 / 0.57	2	$+2.30$	$+5.51$	$+4.26$ $[+3.95, +4.57]$
		GT lattice $\sigma = 32$ at $p = 0.5$	0.29 / 0.57	2	$+1.62$	$+3.90$	$+3.30$ $[+3.01, +3.58]$