WHEN MOTION ACCURACY MISLEADS DOWNSTREAM IMAGING: EVALUATING LEARNED INVERSE MODELS UNDER SLOW-TIME NEAR-EQUIVALENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Learned inverse models often predict an intermediate physical state that a known downstream operator consumes, yet they are commonly trained, validated, and selected by state-space accuracy rather than downstream performance. We ask whether state-space accuracy preserves model ranking for the downstream task when the sampled measurement physics admits approximate equivalences between distant states. We study this in a controlled dynamic Wi-Fi synthetic-aperture simulator: a CNN infers target motion from complex channel-state information and coherent backprojection consumes the inferred velocity. At a 31-Hz diagnostic setting, a state-supervised CNN has lower velocity RMSE than a coherence-regularised CNN (0.149 versus 0.332 m/s) yet higher mean image-localisation error (0.474 versus 0.303 m), in all three training seeds. A sampling-rate diagnostic links the large-error modes to approximate bistatic range-rate branches, and an alias-aware range-rate error agrees better with downstream ranking across the tested estimators, although in a sealed selection experiment it chooses the same model as velocity RMSE. The reversal persists under a fixed update-indexed schedule without validation-based selection, for two architectures (paired image gains of 0.07 and 0.09 m on a sealed 3,000-scene test), under image-grid changes, on a disjoint-support speed holdout, and in an independent ray-traced digital twin for phase-dominated measurements; its size depends on the coherence weight. It weakens or disappears in geometries where few scenes approach the first sampled branch, with more in-distribution training data, at a fixed 100-Hz operating point under the stated speed prior, on a direction holdout, and when deterministic amplitude variation or 0-dB static multipath is present; oracle-motion and static-subtraction controls identify a separate clutter-dominated imaging boundary. These controlled synthetic results support a conditional evaluation principle: when sampled physics admits distant near-equivalent states, Euclidean latent-state accuracy should be reported alongside ambiguity-aware and downstream-task evaluation.

1 INTRODUCTION

Many learned inverse pipelines have the form $x \xrightarrow{f_\phi} \hat{z} \xrightarrow{G} \hat{y}$: a network maps a measurement x to an intermediate physical state $\hat{z}$, and a known operator G turns that state into the quantity of interest $\hat{y}$. A motion or pose estimate can calibrate a reconstruction, a latent physical parameter can drive a simulator, and a state estimate can initialise a solver or determine a phase correction (Zhu et al., 2024; Bhatia et al., 2025; Chen et al., 2025). Such models are usually trained, validated, and selected with a state-space distance $d_z(\hat{z}, z)$, whereas deployment cares about $d_y(G(\hat{z}), y)$. The two criteria need not order models alike when the measurement physics induces approximate equivalence classes in z, so that distant states produce similar measurements and similar downstream outputs: a physically inaccurate estimate may still be useful to G, while a smaller error in a sensitive direction may be harmful. This paper studies that possibility empirically, in one controlled physical instantiation in which every ingredient is known.

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.

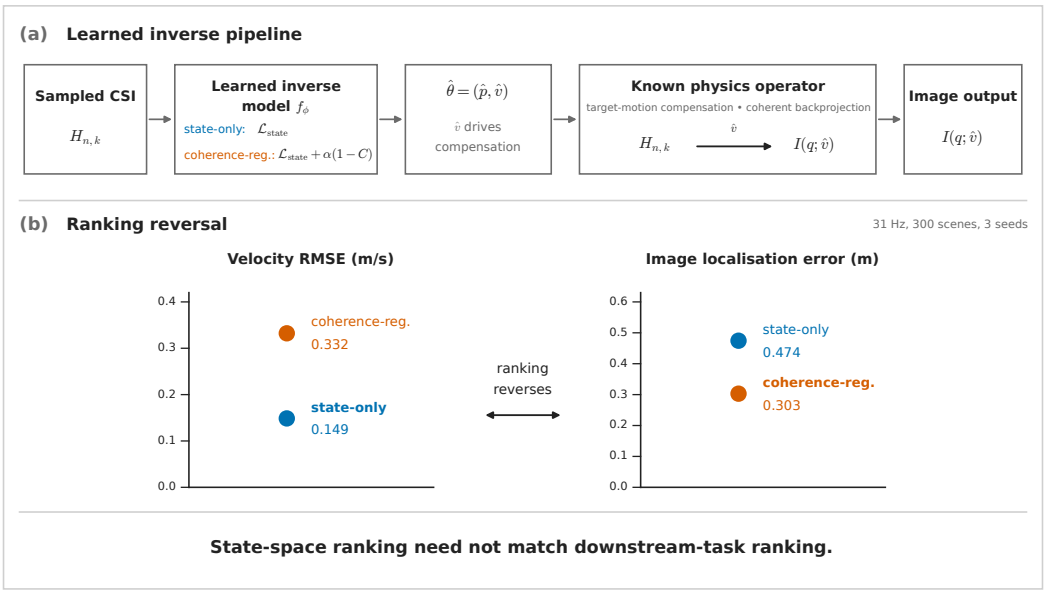


Figure 1: **State-space ranking and downstream-task ranking can disagree.** (a) A learned inverse model f_ϕ predicts an intermediate state $(\hat{\mathbf{p}}, \hat{\mathbf{v}})$ from sampled CSI; only $\hat{\mathbf{v}}$ enters the known physics operator (target-motion compensation, then coherent backprojection over a scanned grid $\mathcal{Q}$), and $\hat{\mathbf{p}}$ is not inserted as the image coordinate. The two estimators differ only in the objective. (b) At 31 Hz (300 scenes, three seeds), the state-only CNN ranks best on velocity RMSE (0.149 vs. 0.332 m/s) but worst on mean image-localisation error (0.474 vs. 0.303 m; Table 1): the two rankings cross.

The instantiation is a wave-imaging problem (Figure 1). A fixed transmitter illuminates a translating point target; a moving receiver with a known trajectory records complex channel-state information (CSI). A CNN predicts target position and velocity, and coherent backprojection uses its velocity to form an image. Unlike conventional SAR platform-motion compensation, the receiver trajectory is given and the unknown is the additional phase history induced by *target motion*; the setting is dynamic bistatic synthetic-aperture imaging, not an inverse-SAR rotation model or a receiver-calibration task.

The central observation is a reversal of model ranking. On 300 paired test scenes at 31 Hz, the state-supervised CNN has lower velocity RMSE than the coherence-regularised CNN (0.149 versus 0.332 m/s), but higher mean image-localisation error (0.474 versus 0.303 m; Table 1). Both directions hold in each of the three training seeds (Appendix D). The state-only model is better at recovering physical velocity under squared Euclidean error; the coherence model is better for the specified imaging objective. Our claim concerns *model ranking for that downstream objective*; physical-state RMSE remains necessary, and a benefit from changing the selection metric must be tested separately.

The pipeline permits these metrics to disagree: the imager scans a spatial grid rather than using the predicted position, and the predicted velocity sets the phase compensation at every candidate position, so a velocity error can move or defocus the peak in a way not proportional to its Cartesian size. Coherence regularisation rewards phase consistency, including at distant approximate branches. The diagnostics support this mechanism for some error modes, but do not attribute every imaging gain or failure to branch selection.

Our contribution is an empirical evaluation study with three parts. (1) *Phenomenon:* a reproducible ranking reversal between physical-state accuracy and downstream coherent imaging, consistent across seeds, a fixed optimisation budget, and two architectures. (2) *Mechanism and diagnostic:* a derivation of the local sampled phase-branch structure (Appendix A), a sampling-rate diagnostic in which large-error modes move with the predicted branch spacing, and an ambiguity-aware projected range-rate error that tracks downstream rankings better than Euclidean velocity error in the ambiguous regime, with its limitations and a sealed selection experiment in which it chooses the same model as velocity RMSE. (3) *Boundaries:* controlled tests of which conditions preserve or

remove the reversal: coherence weight, image grid and field of view, training scale, sampling rate, transmitter geometry, disjoint-support holdouts, static clutter, and an independent ray-traced digital twin (Appendices G–J). We do not propose a universally superior loss, a universal ambiguity metric, or a clutter-robust imaging system.

2 Related work

Hybrid learned/physics pipelines and latent physical variables. Recent work embeds learned components inside physics operators: Latent Intuitive Physics infers hidden fluid properties that a particle simulator consumes (Zhu et al., 2024), PRDP trains through a progressively refined differentiable solver (Bhatia et al., 2025), and PINP estimates latent physical quantities inside a physics-informed predictor (Chen et al., 2025). These works evaluate the end task. Our complementary question, not studied there, is whether state-space error preserves operator-level model ranking when the intermediate state is a point estimate and the operator is deterministic and known; we do not claim that these systems exhibit the reversal reported here.

Task-aware and decision-focused learning. Decision-focused learning trains a predictor through the downstream optimisation it feeds, so that its error is measured by decision quality rather than by a generic regression loss (Donti et al., 2017; Wilder et al., 2019; Elmachtoub & Grigas, 2022; Mandi et al., 2024), typically by differentiating through an optimisation layer (Amos & Kolter, 2017); TaskMet instead learns a task-driven metric for the prediction loss itself (Bansal et al., 2023). These works share our premise that upstream accuracy and downstream utility can disagree. They differ in the object of study: their downstream stage is an optimisation problem whose solution map is relaxed or differentiated, whereas ours is a deterministic physical operator, and the disagreement we examine is induced by sampled near-equivalence of the measurement physics rather than by an asymmetric decision cost. We characterise when and how the two rankings disagree; we do not propose a universal task-aware loss.

CSI learning and downstream outputs. CSI fingerprinting systems such as PhaseFi and ConFi learn position from channel measurements and appropriately evaluate localisation as the endpoint (Wang et al., 2016; Chen et al., 2017). More generally, a learned output may be intermediate rather than terminal. We investigate that distinction when an estimated physical motion state supplies the phase correction for a deterministic coherent imager. Our concern is task alignment of evaluation, not superiority over measured-CSI localisation systems. Practical COTS CSI additionally contains carrier/sampling offsets, packet-dependent phase terms, and hardware-dependent gain/phase effects (Halperin et al., 2011; Ma et al., 2019); prior Wi-Fi sensing systems use calibration, cross-antenna processing, correlation, or CSI-ratio constructions to recover Doppler-sensitive structure (Kotaru et al., 2015; Xie et al., 2015; Qian et al., 2018; Zeng et al., 2019). These acquisition effects are not modelled here.

Target motion and coherent imaging. Target-motion-induced defocus, motion compensation, and ambiguity analysis are established SAR topics (Perry et al., 1999; Xu et al., 2018; Gaibel & Boag, 2021). Image focus has also been used as an optimisation objective, including minimum-entropy ISAR compensation (Wang et al., 2022). Thus neither motion ambiguity nor the use of image-quality measures is new; our study concerns the ordering of *learned* estimators under physical-state and downstream-image errors when sampled motion branches are nearly equivalent, and cites ISAR work as precedent rather than as our acquisition geometry.

Non-unique learned inverses. Invertible networks model multiple parameter states compatible with a measurement (Ardizzone et al., 2019); conditional invertible registration models represent substantially different plausible poses (Trofimova et al., 2020). Posterior sampling addresses reconstruction uncertainty and diversity (Jalal et al., 2021); DDRM and diffusion posterior sampling extend generative priors to linear and noisy nonlinear reconstruction (Kawar et al., 2022; Chung et al., 2023). The perception–distortion tradeoff further illustrates that different output-quality criteria need not agree (Blau & Michaeli, 2018), although perceptual realism is not our endpoint. The gap we address is whether model ranking under intermediate physical-state error survives a deterministic coherent downstream operator when sampled measurements admit near-equivalent states, studied with point-valued motion estimates used for coherent imaging.

3 PHYSICAL MODEL, ESTIMATORS, AND EVALUATION

3.1 A LEARNED STATE CONSUMED BY A COHERENT OPERATOR

Let $\theta = (\mathbf{p}_0, \mathbf{v}) \in \mathbb{R}^4$ denote reference position and constant target velocity. At known times τ_n, transmitter $\mathbf{t}$ and receiver $\mathbf{r}_n$ define

$$\mathbf{p}_n = \mathbf{p}_0 + \mathbf{v}(\tau_n - \tau_{\mathrm{ref}}), \qquad L_n(\theta) = \|\mathbf{p}_n - \mathbf{t}\|_2 + \|\mathbf{p}_n - \mathbf{r}_n\|_2, \tag{1}$$

$$a_{nk}(\theta) = \exp[-j2\pi f_k L_n(\theta)/c], \qquad H_{nk} = \gamma a_{nk}(\theta) + \epsilon_{nk}. \tag{2}$$

The complex reflectivity γ is unknown and shared across samples; α below denotes a loss weight. In clutter experiments we add $\sum_{\ell=1}^{8} \beta_\ell a_{nk}(\mathbf{c}_\ell, \mathbf{0})$. Known geometry and ideal coherent CSI are assumptions; no measured hardware data are used.

For an estimate $\hat{\mathbf{v}}$, the spatial image and localisation are

$$I(\mathbf{q}; \hat{\mathbf{v}}) = \left| \frac{1}{NK} \sum_{n,k} H_{nk} \overline{a_{nk}(\mathbf{q}, \hat{\mathbf{v}})} \right|, \qquad \hat{\mathbf{q}} = \arg\max_{\mathbf{q} \in \mathcal{Q}} I(\mathbf{q}; \hat{\mathbf{v}}). \tag{3}$$

Static SAR uses zero velocity; oracle-motion imaging uses true velocity. The predicted position is used in training and state-space refinement, but is *not* inserted as the final image location: back-projection scans $\mathcal{Q}$. We therefore compare velocity RMSE to $\|\hat{\mathbf{q}} - \mathbf{p}_0\|_2$, and report position error separately. Image quality in our main claims means *global-peak localisation*, not perceptual quality or guaranteed image sharpness.

3.2 TWO MATCHED-ARCHITECTURE LEARNED ESTIMATORS

The CNN takes $2 \times N \times K$ real/imaginary CSI channels, after label-free global-phase alignment and RMS normalisation, and outputs $(\hat{x}, \hat{y}, \hat{v}_x, \hat{v}_y)$. Both estimators minimise normalised supervised state loss

$$\mathcal{L}_s = \tfrac{1}{2}\|D_p^{-1}(\hat{\mathbf{p}} - \mathbf{p}_0)\|_2^2 + \tfrac{1}{2}\|D_v^{-1}(\hat{\mathbf{v}} - \mathbf{v})\|_2^2. \tag{4}$$

The second estimator, which we call the *coherence CNN*, adds a measurement-consistency term $\alpha(1 - C)$ with $\alpha = 0.5$ in the main comparison, where

$$C(H, \hat{\theta}) = \frac{|\langle a(\hat{\theta}), H \rangle|}{\|a(\hat{\theta})\|_2 \|H\|_2}. \tag{5}$$

This auxiliary term does not use labels, but the *full training procedure is supervised*. The coherence term is not the downstream localisation loss: it measures consistency of the predicted state with the measured complex field, whereas the downstream endpoint is the spatial argmax of a separately scanned coherent image. Eliminating one unconstrained complex amplitude gives normalised residual $1 - C^2$ (Appendix A); the implemented penalty is $1 - C$. These have the same coherence ordering but are not interchangeable when added to state loss. Architecture, initial learning rate, and scene streams are matched. Originally, each model's own validation objective controls scheduling and checkpoint selection; the comparison therefore concerns training-and-selection procedures. Section 5.3 uses a fixed update schedule, no validation-based selection, and a second architecture.

3.3 SLOW-TIME NEAR-EQUIVALENCE

Writing $t_n = \tau_n - \tau_{\mathrm{ref}}$ and $\mathbf{g}_n = \mathbf{u}_{\mathrm{TX,n}} + \mathbf{u}_{\mathrm{RX,n}}$ for unit vectors from the sensors to the true target, local state linearisation gives

$$\Delta L_n \simeq \mathbf{g}_n^\top \Delta\mathbf{p} + \mathbf{g}_n^\top \Delta\mathbf{v}\, t_n. \tag{6}$$

Freezing $\mathbf{g}_n \simeq \mathbf{g}_0$ and replacing f_k by f_c, the phase increment caused by $\Delta\mathbf{v}$ is $-2\pi\mathbf{g}_0^\top \Delta\mathbf{v}/(\lambda \mathrm{PRF})$. Integer cycles suggest candidate offsets

$$\mathbf{g}_0^\top \Delta\mathbf{v} \simeq mP, \qquad P = \lambda\,\mathrm{PRF}, \quad m \in \mathbb{Z}, \quad \lambda = c/f_c. \tag{7}$$

Here P is a spacing in *target-induced bistatic path-rate difference*, not speed or a Cartesian component. Exact equivalence additionally requires one common phase offset modulo integer cycles at *every* time–frequency sample. Geometry variation, frequency diversity, and nonlinear state dependence generally break the approximate periodicity; Eq. (7) is not a global four-dimensional state symmetry or an identifiability result (Appendix A). Strong distant alternatives need not predict which branch a CNN selects.

3.4 STATE, AMBIGUITY-AWARE, AND DOWNSTREAM METRICS

Ordinary velocity error is $e_v = \|\hat{\mathbf{v}} - \mathbf{v}\|_2$. With $d = \mathbf{g}_0^\top(\hat{\mathbf{v}} - \mathbf{v})$, we evaluate

$$e_{\mathrm{RR}} = |d|, \qquad m = \mathrm{round}(d/P), \qquad e_{\text{AA-RR}} = |d - mP|. \tag{8}$$

The geometry factor is evaluated on the true trajectory (central-gradient convention in Appendix B). Thus d projects velocity error at truth, rather than subtracting two full nonlinear path rates. Folding measures its distance to $P\mathbb{Z}$. It is a labelled *offline diagnostic*, not deployable confidence. We aggregate state errors by RMSE and image localisation by mean; unfolded range-rate error separates projection from folding.

Folding is a diagnostic, not a full-state metric: it ignores cross-range velocity and position errors and may assign a small value to an unsuitable high-order branch, so we always retain ordinary state error and downstream localisation. The $m \neq 0$ flag is called *alias-like*; it does not by itself prove high measurement coherence, and severe clutter extrapolation can trigger it without a meaningful branch match. Predicted-displacement scores are supporting analyses because their method-ranking agreement is less stable across geometries.

4 EXPERIMENTAL PROTOCOL

The carrier is 5.32 GHz with 32 frequency samples spanning 80 MHz. A receiver traverses $x \in [-0.75, 0.75]$ m at $y = 0$ over $[-0.5, 0.5]$ s. Reference target position is uniform in $[-0.6, 0.6] \times [2.4, 3.6]$ m, speed in $[0, 0.7]$ m/s, direction in $[-150, 150]$ degrees, and target-referenced SNR in $[15, 35]$ dB. Backprojection uses a 65×65 grid on $[-0.8, 0.8] \times [2, 4]$ m. The original TX is at $(-1.5, 0)$; the replication shifts it to $(-4, 0)$.

We use seeds 101, 202, and 303 with distinct training, validation, and evaluation streams. A 300-scene condition contains 100 per seed; the 600-scene scaling evaluation has 200 per seed. Test states are paired across relevant conditions, but reused diagnostic scenes are not independent confirmations. The original runs select checkpoints on validation and permit up to 120 epochs with adaptive scheduling and early stopping. The controlled bank below instead uses fixed endpoints. Cap-reaching runs are not called converged. Appendix C gives architecture, optimisation, split ranges, and selection details.

Diagnostic and boundary experiments. We use matched-rate models at 15/31/62/124 Hz, with 16/32/63/125 samples over the same second, and 2,048 training scenes. Per-sample SNR is fixed; integrated SNR and sample count therefore change with PRF. Cluster movement tests the sampling mechanism, but absolute accuracy changes cannot be attributed solely to alias removal. At 31 Hz, the data-size study uses 2k/8k/32k scenes; the second-geometry replication uses 2k. We use 31 Hz as a mechanism-revealing diagnostic regime because the first sampled range-rate branch predicted by Eq. (7) intersects the support induced by the stated motion prior; it is not presented as a recommended Wi-Fi operating rate. The 100-Hz operating-point study uses 101 samples and 2k/65k training scenes in the original geometry; it is a representative configured rate, not a Wi-Fi standard. The second geometry was tested at 31 Hz only.

Controlled training bank and sealed splits. For Section 5.3, 24 trajectories (the CNN and a residual CNN, ResCNN, with 0.73 M parameters and 39× the multiply–accumulates; $\alpha \in \{0, 0.1, 0.5, 1\}$; three seeds) are trained at 31 Hz for exactly 11,520 updates with a learning rate fixed by update count, no early stopping or plateau scheduling, and checkpoints at 3,840 and 11,520 updates of the same run. Within a seed all candidates share data and minibatch order, and within an architecture the initial parameters (verified by hashes). Two new scene roles, 512 selection-validation and 1,000 sealed final-test scenes per seed, are hash-verified disjoint from every earlier stream. The candidate pool, selection rules, tie rule, and primary statistic were frozen in writing before any outcome was evaluated (Appendix G).

Clutter and classical controls. At 100 Hz, eight static scatterers have random phase and positions in $[-1.5, 1.5] \times [1.5, 4.5]$ m. We sweep no clutter and nominal CIR $-10, 0, +10, +20$ dB, recording the realised ratio $10 \log_{10}(\mathbb{E}|H_c|^2/\mathbb{E}|H_t|^2)$ before noise. Clean-trained and clutter-trained 65k CNNs are evaluated; the latter use a log-uniform amplitude-factor mixture corresponding nominally

Table 1: **Matched comparisons and regime boundaries.** Each row contains 300 scenes (three seeds). Position/velocity columns are vector RMSE; image is mean global-peak localisation error. Both 31-Hz geometries use 2k training scenes.

Condition	Model	Position (m)	Velocity (m/s)	Image (m)	Alias-like (%)
31 Hz, original	State-only	0.329	0.149	0.474	0.0
31 Hz, original	Coherence	0.320	0.332	0.303	14.3
31 Hz, shifted TX	State-only	0.343	0.153	0.461	0.0
31 Hz, shifted TX	Coherence	0.329	0.174	0.334	3.0
100 Hz, 2k	State-only	0.317	0.099	0.312	0.0
100 Hz, 2k	Coherence	0.313	0.084	0.265	0.0
100 Hz, 65k	State-only	0.043	0.011	0.052	0.0
100 Hz, 65k	Coherence	0.053	0.016	0.054	0.0

to -10 through $+20$ dB. Classical finite-grid search, local refinement, static SAR, and oracle-motion imaging are controls, and a branch test that refines three initial velocities separated by Eq. (7) is an exploratory correction, not a proposed algorithm.

Uncertainty. We report seed-specific outcomes and paired differences; intervals use 4,000 paired resamples within each seed with fixed seed weights and quantify scene uncertainty conditional on the three trained models, not uncertainty over training runs. Across-estimator rank correlations are descriptive: variants share architectures and are not independent replications.

5 Results

5.1 The 31-Hz reversal holds in all three seeds

Table 1 reports one consistent 300-scene evaluation. Coherence training raises velocity RMSE from 0.149 to 0.332 m/s while reducing mean image error from 0.474 to 0.303 m. Both directions occur in each training seed (Appendix D). The paired image difference is -0.171 m, with a conditional 95% interval $[-0.209, -0.131]$. Position RMSE is similar (0.329 versus 0.320 m), but is not the downstream image location. This establishes the mismatch for the stated task without treating the worse velocity estimate as physically correct.

Figure 6 (Appendix D) shows a saved scene in which the coherence estimate has a much larger velocity error but a less displaced image peak, still less accurate than oracle-motion imaging: near-equivalence explains utility despite physical error, not interchangeability with the true state.

5.2 Error modes follow sampling; folded range rate improves ranking

The observed first alias-like error cluster moves from 0.813 m/s at 15 Hz to 1.704 m/s at 31 Hz, compared with predicted periods 0.845 and 1.747 m/s (Figure 2; cluster medians pool the tested estimators with $|m| = 1$ and are not speed estimates). At 62/124 Hz, the two primary CNNs show no alias-like failures. For the coherence CNN, the corresponding alias-like fractions are 56.0%, 14.3%, 0%, and 0%. The movement of the error mode, rather than a raw improvement in accuracy, is the mechanism evidence.

Across the original 16-estimator diagnostic inventory, ordinary velocity RMSE agrees poorly with mean image-error ranking (Spearman $\rho_s = 0.09$ at 15 Hz, 0.13 at 31 Hz), whereas alias-aware range-rate RMSE yields 0.92 and 0.98. The unfolded range-rate comparison in Appendix B shows that this is not merely a projection onto a sensitive direction. Figure 3 illustrates method-level alignment; these scores complement direct imaging evaluation and cannot replace it or guarantee focus.

5.3 The reversal survives fixed budgets, architecture, and grid changes

Figure 4 and Appendix G report the fixed-update bank on the sealed test. At 11,520 updates, $\alpha = 0.5$ raises velocity RMSE from 0.138 to 0.332 m/s (CNN) and 0.096 to 0.216 m/s (residual CNN)

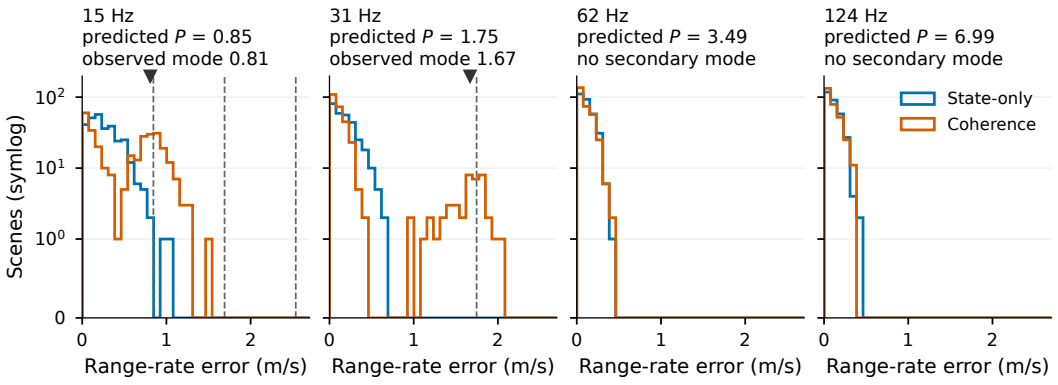


Figure 2: **Sampling diagnostic.** Absolute bistatic range-rate errors of two matched-rate CNNs, 300 scenes per panel. Dashed lines mark integer multiples of $P = \lambda \mathrm{PRF}$ where visible. At 15 and 31 Hz the first secondary mode moves with the predicted branch spacing; no secondary mode is observed for the primary CNNs at 62/124 Hz, where the first period lies beyond the displayed range.

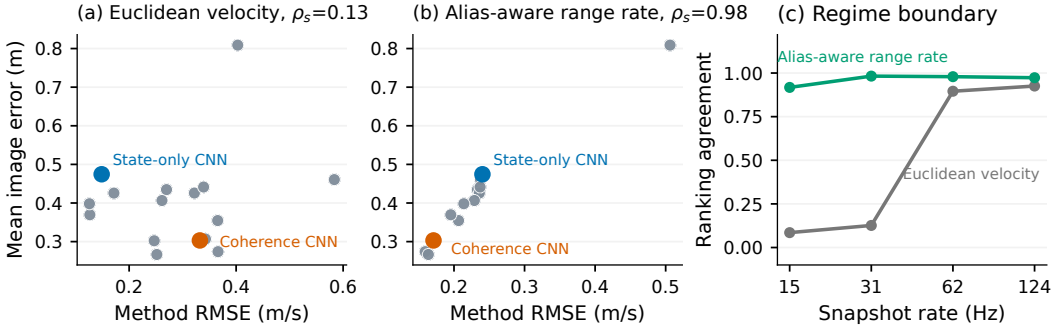


Figure 3: **Evaluation alignment.** Each point in (a,b) is one estimator on the same 31-Hz scene set; highlighted points are the two CNNs. Panel (c) uses the rate-specific inventory of 16 non-oracle estimators, including static SAR and excluding truth-selected hypotheses.

while lowering mean image error from 0.438 to 0.370 m and 0.313 to 0.220 m. Paired image differences are -0.068 m $[-0.078, -0.057]$ and -0.093 m $[-0.101, -0.085]$, with image gains in every seed; the aggregate reversal also holds at 3,840 updates. It therefore does not require early stopping, plateau scheduling, model-specific checkpoint criteria, or the original architecture, although its magnitude is smaller (controlled CNN gap -0.059 m rather than -0.171 m on the diagnostic scenes); since budget, schedule, and selection changed together, this difference cannot be attributed to checkpoint selection specifically. Weight also matters: $\alpha = 0.1$ improves both velocity RMSE (0.126 and 0.075 m/s) and image error (0.392 and 0.240 m), with alias-like rates of 0.13% and 0%, so the reversal occurs at the stronger tested weights; it is neither universal nor proof of a branch-selection threshold, and no run is called converged.

Re-imaging saved predictions on a 129×129 grid with the same field of view retains the original-geometry paired difference (-0.173 m); doubling each spatial extent at the original spacing increases it to -0.216 m $[-0.271, -0.163]$ (Figure 4c). The original grid has 11.3% state-only boundary peaks versus 1.7% for coherence; widening increases aggregate state-only error. The shifted-geometry gaps are $-0.128/-0.129/-0.127$ m. Oracle mean error falls from 0.031 to 0.014 m on the finer grid (discretisation sensitivity, not a resolution floor).

At 11,520 updates, selecting among eight candidates per seed on 512 labelled selection scenes by velocity RMSE, folded range-rate RMSE, or direct image error chooses the same model, with zero paired sealed-test image difference and zero regret to the hindsight test-best candidate; unfolded range rate chooses a different candidate in one seed (regret 0.115 m). Folded range rate has stronger agreement with sealed-test image rankings (Spearman 0.90–0.98 per seed) than velocity RMSE

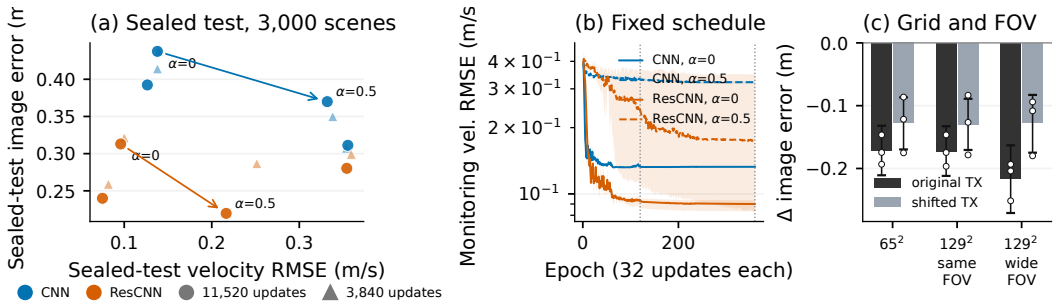


Figure 4: **Controlled optimisation, architecture, and grid.** (a) Velocity RMSE against mean image error on the sealed test for every candidate and endpoint; arrows join $\alpha = 0$ to $\alpha = 0.5$ at 11,520 updates. (b) Monitoring-validation velocity RMSE under the fixed schedule, seed mean and seed range for $\alpha \in \{0, 0.5\}$; all 24 trajectories are in Figure 9. (c) Paired coherence-minus-state-only image difference of the saved 31-Hz predictions under three image grids, 300 scenes per bar, 95% intervals from 4,000 within-seed resamples; white markers are seeds.

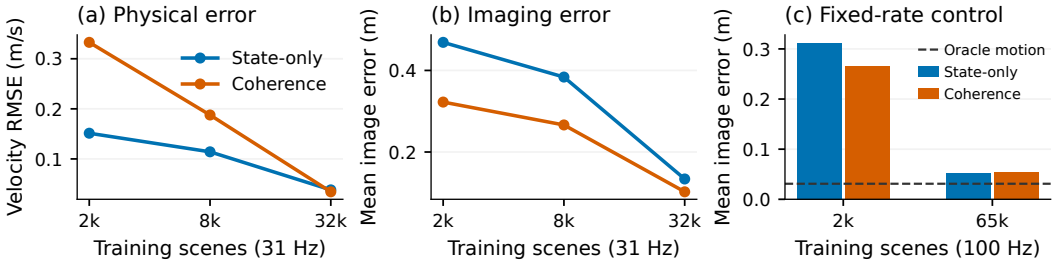


Figure 5: **The mismatch is conditional.** At 31 Hz, larger training sets improve both physical and image errors on 600 paired scenes. At 100 Hz, 2k/65k models are evaluated on a distinct 300-scene set; the dashed line is oracle-motion imaging. Comparisons across these panels are not a factorial isolation of rate and data scale.

(0.07–0.36). Ranking agreement therefore improves without a selection advantage in this pool; the selected models have low, but not uniformly zero, alias-like rates (Appendix G).

5.4 Geometry and data scale constrain the claim

Moving the TX to $(-4, 0)$ under the original protocol retains a weaker reversal (Table 1; folded range-rate ranking agreement $0.98/0.93$ against $-0.02/-0.32$ for velocity, Appendix B). Retraining the fixed-update bank in that geometry and in a lateral geometry (TX at $(3, 3)$ m, receiver path rotated $30°$) bounds the claim (Appendix I): the shifted-geometry CNN reversal is marginal (velocity 0.136 to 0.145 m/s, image -0.027 m), and in every other case both metrics improve, with alias-like fractions below 1%. The fraction of scenes whose true bistatic rate exceeds $P/2$ falls from 16.5% to 12.6% to 3.2% across the three geometries: the reversal appears where the prior reaches an approximate branch, and not otherwise.

On disjoint-support holdouts of the frozen bank (Appendix H) the reversal grows on a speed strip (CNN image error 0.81 to 0.53 m while velocity RMSE rises from 0.51 to 0.85 m/s), persists on a position strip, and is absent on a direction strip, where the coherence term lowers both errors (stress tests, not out-of-distribution accuracy claims).

Data scale also changes the result (Figure 5): on the separate 600-scene scaling set at 31 Hz, coherence alias-like fractions decline from 14.5% at 2k to 3.8% at 8k and 0% at 32k. At 32k, state-only/coherence velocity RMSE is 0.038/0.034 m/s and image error is 0.134/0.102 m. This supports resolvability in distribution, not exact non-identifiability; larger datasets also supply more optimiser updates under the epoch-capped protocol, confounding sample diversity with optimisation budget.

5.5 At 100 Hz, no alias-like estimates occur under the stated prior

With $N = 101$ over 1 s, $P = 5.635$ m/s in bistatic range rate; at speed at most 0.7 m/s the maximal first-order range rate is 1.38 m/s, below $P/2 = 2.818$ m/s, and no alias-like estimates occur in the clean 100-Hz evaluations. At 2k, coherence improves image error from 0.312 to 0.265 m. At 65k, image errors are 0.052/0.054 m, near the oracle's 0.031 m; their paired difference is 0.002 m with interval $[-0.003, 0.007]$. This is no resolved image advantage in this sample, not a formal equivalence test; an evaluation-only high-speed extension is in Appendix F.

5.6 Clutter failure can persist even with true motion

Figure 8 separates estimator failure from failure of the chosen backprojection/global-peak pipeline. At -10 and 0 dB, clutter-trained coherence models reduce mean image error relative to clutter-trained state-only models by approximately 0.059 and 0.080 m, with the same direction in all seeds. This is a relative gain: failure above 0.2 m still occurs in roughly 62% and 73% of coherence images. Clean-trained networks transfer poorly even to -10 dB clutter, and mixture training worsens clean-scene accuracy.

At +10/+20 dB, oracle-motion imaging fails in 62%/91% of scenes: correct velocity is insufficient for this pipeline, and no motion ambiguity is needed to explain these failures. A CLEAN-style control that repeatedly fits and subtracts static-scene peaks reduces the +20-dB oracle median error from 0.83 to 0.14 m but leaves 42% failures, uses the simulated scatterer count, and without an adaptive stopping rule damages clean scenes (oracle median 0.55 m); clutter is not solved.

Slow-time mean subtraction does not provide the same benefit, because a static scatterer is not DC in slow time when the receiver moves (no resolved improvement at +10 dB), and after CLEAN clean-trained CNNs remain inaccurate because subtraction changes their input distribution. Appendix E separates global failure, clutter-associated false peaks, and the oracle-conditioned metric.

6 Discussion and conclusion

The evidence supports a narrow but consequential finding: under sampled near-equivalence, two learned inverse estimators can be ordered differently by physical-state accuracy and downstream coherent localisation. State RMSE remains the right criterion when physical recovery is the objective; when coherent imaging is the endpoint, it alone can choose the worse image-producing model. Folding range-rate error explains part of this mismatch, but does not determine all image failures.

Our controls also constrain the mechanism. A local branch test reduces original-geometry alias-like errors but *increases* them in the shifted geometry (3% to 7%), and strong distant branches do not predict which scenes a network aliases on; branch correction, separability, and observability are therefore diagnostics, not methods.

The study is limited to ideal synthetic CSI, one translating point target, known receiver motion, uniform timing, three geometries (one for the 100-Hz and clutter studies), and two CNN architectures in the fixed-budget control. The reversal's size depends on training protocol, coherence weight, and geometry (Sections 5.3 and 5.4). An independent SensingSP ray-traced digital twin reproduces the reversal for phase-dominated measurements, while same-noise controls indicate that deterministic slow-time amplitude variation provides an additional disambiguating cue; wall multipath further bounds the effect (Appendix J). This is cross-simulator, not measured-CSI, validation. We have no theorem connecting folded range-rate error to localisation. The controls show that the original selection procedure and architecture are not necessary for the reversal, without isolating every cause, and the PRF and data-size controls retain information/compute confounds. Within these limits, physical-state accuracy, ambiguity diagnostics, and downstream outcomes should be reported together.

Reproducibility statement

Appendices specify the forward model, architecture, losses, splits, and all evaluation conventions. This manuscript's companion source package contains frozen plot inputs, source-file hashes, numerical audits, and the script regenerating all included vector figures and tables. These files reproduce

the reported manuscript summaries without retraining; they are not a release of the full training environment or remote checkpoints. Conditional confidence intervals are explicitly distinguished from training-run uncertainty.

AI USE STATEMENT

Generative AI tools were used primarily for manuscript editing and polishing, and for limited assistance with analysis and figure-preparation scripts. All scientific content, analyses, results, and conclusions were reviewed and verified by the authors, who take full responsibility for the final manuscript.

ETHICS STATEMENT

The experiments use synthetic point scatterers and contain no human-participant or measured personal data. Future Wi-Fi sensing deployments may raise privacy concerns; this study does not validate human sensing or deployment.

REFERENCES

Brandon Amos and J. Zico Kolter. OptNet: Differentiable optimization as a layer in neural networks. In *International Conference on Machine Learning*, volume 70 of *PMLR*, pp. 136–145, 2017.

Lynton Ardizzone, Jakob Kruse, Sebastian Wirkert, Daniel Rahner, Eric W. Pellegrini, Ralf S. Klessen, Lena Maier-Hein, Carsten Rother, and Ullrich Köthe. Analyzing inverse problems with invertible neural networks. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1808.04730.

Dishank Bansal, Ricky T. Q. Chen, Mustafa Mukadam, and Brandon Amos. TaskMet: Task-driven metric learning for model learning. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://arxiv.org/abs/2312.05250.

Kanishk Bhatia, Felix Köhler, and Nils Thuerey. PRDP: Progressively refined differentiable physics. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=9Fh0z1JmPU.

Yochai Blau and Tomer Michaeli. The Perception-Distortion Tradeoff. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6228–6237, 2018. doi: 10.1109/cvpr.2018.00652.

Hao Chen, Yifan Zhang, Wei Li, Xiaofeng Tao, and Ping Zhang. ConFi: Convolutional Neural Networks Based Indoor Wi-Fi Localization Using Channel State Information. *IEEE Access*, 5: 18066–18074, 2017. doi: 10.1109/access.2017.2749516.

Huaguan Chen, Yang Liu, and Hao Sun. PINP: Physics-informed neural predictor with latent estimation of fluid flows. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=vAuodZOQEZ.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.14687.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68 (1):9–26, 2022. doi: 10.1287/mnsc.2020.3922.

Ariel Gaibel and Amir Boag. Backprojection Imaging of Moving Objects. *IEEE Transactions on Antennas and Propagation*, 69(8):4944–4954, 2021. doi: 10.1109/tap.2020.3045500.

Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM Computer Communication Review*, 41 (1):53, 2011. doi: 10.1145/1925861.1925870.

Ajil Jalal, Sushrut Karmalkar, Alex Dimakis, and Eric Price. Instance-optimal compressed sensing via posterior sampling. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 4709–4720. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/jalal21a.html.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/95504595b6169131b6ed6cd72eb05616-Abstract-Conference.html.

Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. SpotFi: Decimeter level localization using WiFi. In *ACM SIGCOMM*, 2015. doi: 10.1145/2829988.2787487.

Yongsen Ma, Gang Zhou, and Shuangquan Wang. WiFi sensing with channel state information: A survey. *ACM Computing Surveys*, 52(3):46:1–46:36, 2019. doi: 10.1145/3310194.

Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Victor Bucarey, Tias Guns, and Ferdinando Fioretto. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 80:1623–1701, 2024. doi: 10.1613/jair.1.15320.

R. P. Perry, R. C. DiPietro, and R. L. Fante. SAR imaging of moving targets. *IEEE Transactions on Aerospace and Electronic Systems*, 35(1):188–200, 1999. doi: 10.1109/7.745691.

Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, Zheng Yang, and Yunhao Liu. Widar2.0: Passive human tracking with a single Wi-Fi link. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2018. doi: 10.1145/3210240.3210314.

Darya Trofimova, Tim Adler, Lisa Kausch, Lynton Ardizzone, Klaus Maier-Hein, Ulrich Köthe, Carsten Rother, and Lena Maier-Hein. Representing ambiguity in registration problems with conditional invertible neural networks. *arXiv preprint arXiv:2012.08195*, 2020. URL https://arxiv.org/abs/2012.08195.

Jiadong Wang, Yachao Li, Ming Song, Pingping Huang, and Mengdao Xing. Noise Robust High-Speed Motion Compensation for ISAR Imaging Based on Parametric Minimum Entropy Optimization. *Remote Sensing*, 14(9):2178, 2022. doi: 10.3390/rs14092178.

Xuyu Wang, Lingjun Gao, and Shiwen Mao. CSI Phase Fingerprinting for Indoor Localization With a Deep Learning Approach. *IEEE Internet of Things Journal*, 3(6):1113–1123, 2016. doi: 10.1109/jiot.2016.2558659.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *AAAI Conference on Artificial Intelligence*, 2019. doi: 10.1609/aaai.v33i01.33011658.

Yaxiong Xie, Zhenjiang Li, and Mo Li. Precise power delay profiling with commodity WiFi. In *ACM International Conference on Mobile Computing and Networking (MobiCom)*, pp. 53–64, 2015. doi: 10.1145/2789168.2790124.

Gang Xu, Yanyang Liu, Jialian Sheng, Guobin Jing, Xin Lin, and Junli Chen. Parametric Sparse SAR Imaging of Ground Moving Targets Integrated With Motion Compensation. *IEEE Access*, 6:73457–73464, 2018. doi: 10.1109/access.2018.2884233.

Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):121:1–121:26, 2019. doi: 10.1145/3351279.

Xiangming Zhu, Huayu Deng, Haochen Yuan, Yunbo Wang, and Xiaokang Yang. Latent intuitive physics: Learning to transfer hidden physics from a 3d video. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=WZu4gUGN13.

A FROM SAMPLED PHASE TO APPROXIMATE MOTION BRANCHES

This appendix separates an exact property of sampled measurements from the chain of approximations used to interpret estimator errors. It concerns the clean, single-target model of Eq. (2); the clutter controls test a separate limitation of the imaging pipeline. A "branch" is a candidate region with a similar phase history, not a proven equivalence class of physical states. The assumption ladder is: (i) exact measurement equivalence up to one complex amplitude (Section A.1); (ii) local linearisation of the path length in the state increment (Section A.2); (iii) a frozen reference gradient and (iv) a single carrier frequency, which together give the slow-time branch spacing (Section A.3); (v) the residual terms that make the full multicarrier model only approximately periodic (Section A.4); and (vi) the empirically observed alias-like error modes of Section 5.3, which are consistent with (iii)–(iv) but do not certify (i).

A.1 FORWARD MODEL, NUISANCE AMPLITUDE, AND EXACT EQUIVALENCE

Nuisance-amplitude elimination. Flatten the $N \times K$ steering samples into $a(\theta)$ and use $\langle a, h\rangle = a^H h$. For nonzero data h, the least-squares amplitude and the normalised residual are

$$\hat{\gamma}(\theta) = \frac{a(\theta)^H h}{a(\theta)^H a(\theta)}, \qquad \min_{\gamma\in\mathbb{C}} \frac{\|h - \gamma a(\theta)\|_2^2}{\|h\|_2^2} = 1 - \frac{|a(\theta)^H h|^2}{\|a(\theta)\|_2^2 \|h\|_2^2} = 1 - C(h,\theta)^2, \tag{9}$$

obtained by substituting $\hat{\gamma}$ into the quadratic. This identity uses a *single* amplitude across both time and frequency, as in the simulator; one free amplitude per sample or per subcarrier would change the equivalence problem. The training penalty $1 - C$ and the variable-projection residual $1 - C^2$ order coherence identically, but have different gradients ($-\partial C$ versus $-2C\,\partial C$) and different relative weights when combined with the supervised state loss; the two are not interchangeable in a loss.

Exact noiseless equivalence. For two noiseless states let $\Delta L_n = L_n(\theta') - L_n(\theta)$. Since every steering entry has unit modulus, the states are measurement-equivalent up to the nuisance amplitude if and only if their steering vectors differ by one common unit-modulus scalar, that is,

$$\frac{f_k \Delta L_n}{c} = b + q_{nk} \quad \text{for all } n, k, \qquad b \in \mathbb{R}, \quad q_{nk} \in \mathbb{Z}: \tag{10}$$

one common real phase offset b and integers q_{nk} at *every* time–frequency sample. Eq. (10) gives $a(\theta') = e^{-j2\pi b} a(\theta)$, whose constant phase is absorbed by γ; conversely, proportional unit-modulus vectors require these congruences (equality in Cauchy–Schwarz). The noise-free coherence

$$C_*(\theta, \theta') = \left| \frac{1}{NK} \sum_{n,k} \exp\!\left(-j2\pi \frac{f_k \Delta L_n}{c}\right) \right| \tag{11}$$

equals one exactly under this condition. Three statements follow and are used throughout. High coherence is not exact equivalence: a value of 0.95 leaves a normalised noise-free residual of $1 - 0.95^2 = 0.0975$. Exact equivalence is not established by observing one large coherence value at a refined alternative. Whether a nonzero residual is distinguishable in noise depends on the acquisition and inference procedure; we do not establish statistical indistinguishability or a universal coherence threshold.

A.2 LOCAL STATE PERTURBATIONS, PATH RATE, AND RECEIVER MOTION

Let $t_n = \tau_n - \tau_{\mathrm{ref}}$, $\theta' = (\mathbf{p}_0 + \Delta\mathbf{p}, \mathbf{v} + \Delta\mathbf{v})$, and $\delta\mathbf{p}_n = \Delta\mathbf{p} + t_n \Delta\mathbf{v}$. At the true trajectory, define the unit vectors from the sensors to the target and their sum,

$$\mathbf{u}_{\mathrm{TX,n}} = \frac{\mathbf{p}_n - \mathbf{t}}{\|\mathbf{p}_n - \mathbf{t}\|}, \quad \mathbf{u}_{\mathrm{RX,n}} = \frac{\mathbf{p}_n - \mathbf{r}_n}{\|\mathbf{p}_n - \mathbf{r}_n\|}, \quad \mathbf{g}_n = \mathbf{u}_{\mathrm{TX,n}} + \mathbf{u}_{\mathrm{RX,n}}. \tag{12}$$

Taylor expansion of the two distances gives

$$\Delta L_n = \mathbf{g}_n^\top \delta\mathbf{p}_n + R_n = \mathbf{g}_n^\top \Delta\mathbf{p} + t_n \mathbf{g}_n^\top \Delta\mathbf{v} + R_n. \tag{13}$$

For each sensor distance r, the position Hessian is $(I - \mathbf{u}\mathbf{u}^\top)/r$, so R_n is second order in $\|\delta\mathbf{p}_n\|$ and bounded by $\|\delta\mathbf{p}_n\|^2 / \min(r_{\mathrm{TX}}, r_{\mathrm{RX}})$ as long as the segment between the two target positions

stays away from the sensors. The approximation therefore requires small state increments over the *whole aperture*, not only at $t = 0$, and it fails for large $|t_n|\|\Delta\mathbf{v}\|$ or near a sensor. Phase accuracy is stricter than distance accuracy: the neglected phase is $2\pi f_k R_n/c$, so a remainder of a few centimetres is already a substantial fraction of a cycle at 5.3 GHz. Substituting a distant branch into a local expansion does not certify its coherence; the exact kernel must be evaluated.

Receiver-motion contribution. For a differentiable known receiver path and a fixed transmitter, the actual path rate is

$$\dot{L}(t) = \mathbf{g}(t)^\top \mathbf{v} - \mathbf{u}_{\mathrm{RX}}(t)^\top \dot{\mathbf{r}}(t). \tag{14}$$

The signed phase frequency is $-\dot{L}/\lambda$ at the carrier. Thus $\mathbf{g}^\top\mathbf{v}$ is the *target-induced contribution*, not the complete Doppler of a moving-receiver measurement. When two target velocities are compared at the same reference state, $\mathbf{u}_{\mathrm{RX}}(0)$ and $\dot{\mathbf{r}}(0)$ are identical for both, so the receiver term cancels in the difference and $\Delta\dot{L}(0) = \mathbf{g}(0)^\top\Delta\mathbf{v}$; this is why the branch analysis can be phrased in target-induced rate alone. When position also changes, differentiating the linear term in Eq. (13) instead gives

$$\Delta\dot{L}(0) \simeq \dot{\mathbf{g}}(0)^\top\Delta\mathbf{p} + \mathbf{g}(0)^\top\Delta\mathbf{v}. \tag{15}$$

Position–velocity coupling is therefore omitted by a velocity-only projection. It matters for full four-dimensional refinement and for spatial imaging.

A.3 Slow-time phase increments and the carrier branch spacing

Two further approximations are made: freeze the local gradient at $\mathbf{g}_{\mathrm{ref}}$ ($\mathbf{g}_n \approx \mathbf{g}_{\mathrm{ref}}$) and use one carrier frequency ($f_k \approx f_c$). Then Eq. (13) becomes $\Delta L_n \simeq A + Bt_n$, with $A = \mathbf{g}_{\mathrm{ref}}^\top\Delta\mathbf{p}$ and $B = \mathbf{g}_{\mathrm{ref}}^\top\Delta\mathbf{v}$. For uniform times $t_n = t_{\min} + n\Delta t$, $n = 0, \dots, N-1$,

$$\frac{f_c\Delta L_n}{c} \simeq \frac{A + Bt_{\min}}{\lambda} + n\frac{B\Delta t}{\lambda}. \tag{16}$$

The first term is one common phase offset. The second contributes integer cycles at every sample when $B\Delta t/\lambda = m$. Therefore

$$\mathbf{g}_{\mathrm{ref}}^\top\Delta\mathbf{v} = mP, \qquad P = \frac{\lambda}{\Delta t} = \lambda\,\mathrm{PRF}, \qquad \mathrm{PRF} = \frac{N-1}{T}, \quad T = t_{N-1} - t_0. \tag{17}$$

This is exact periodicity of the *frozen-gradient, carrier-only linearised phase-increment model*, and only a candidate-spacing rule for the nonlinear multicarrier model. The reference time need not be an acquired sample: changing $t_{\min}$ only changes the common phase in Eq. (16).

Path rate, line-of-sight velocity, and Cartesian velocity. The units of P are metres of *total bistatic path per second*. For $\mathbf{g}_{\mathrm{ref}} \neq 0$, one minimum-norm velocity offset satisfying the scalar rule is

$$\Delta\mathbf{v}_m = mP\frac{\mathbf{g}_{\mathrm{ref}}}{\|\mathbf{g}_{\mathrm{ref}}\|^2}; \qquad \mathbf{g}_{\mathrm{ref}}^\top\mathbf{w} = 0 \implies \mathbf{g}_{\mathrm{ref}}^\top(\Delta\mathbf{v}_m + \mathbf{w}) = mP. \tag{18}$$

Thus the scalar constraint determines neither Cartesian velocity nor the four-dimensional state. The spacing along the unit bistatic bisector is $P/\|\mathbf{g}_{\mathrm{ref}}\|$, where $\|\mathbf{g}_{\mathrm{ref}}\| = 2\cos(\beta/2)$ and β is the bistatic angle between the two unit vectors. The monostatic limit ($\beta = 0$, $\|\mathbf{g}\| = 2$) gives $\lambda\mathrm{PRF}/2$ along the line of sight, the familiar radial-velocity ambiguity. There is no missing factor of two in Eq. (17): it describes path rate, not one-way radial speed. Near $\mathbf{g}_{\mathrm{ref}} = 0$ the bisector-velocity conversion becomes ill-conditioned and is not a useful branch construction.

A.4 Why the full measurements are only approximately periodic

Changing geometry and finite perturbations. Writing $\mathbf{g}(t) = \mathbf{g}(0) + t\dot{\mathbf{g}}(0) + O(t^2)$ in the linearised path gives a slope term $\dot{\mathbf{g}}(0)^\top\Delta\mathbf{p}$ and curvature terms including $t^2\dot{\mathbf{g}}(0)^\top\Delta\mathbf{v}$. Higher time derivatives and the nonlinear remainder R_n contribute further phase, and position–velocity coupling (Eq. (15)) lets a position change partially compensate a velocity change. A constant projected velocity offset therefore need not generate an integer ramp throughout the aperture. Adjusting position during refinement can absorb some phase variation, which is why a fixed-position branch initialiser need not locate the strongest distant alternative.

Frequency diversity. Write $f_k = f_c + \nu_k$ and, for any chosen b, m, define the carrier residual $\epsilon_n = f_c \Delta L_n / c - b - mn$. Then the *exact* phase difference in cycles is

$$\frac{f_k \Delta L_n}{c} = b + mn + \epsilon_n + \frac{\nu_k \Delta L_n}{c}. \tag{19}$$

Even an integer carrier ramp ($\epsilon_n = 0$) leaves the frequency-dependent term $\nu_k \Delta L_n / c$. A small fractional bandwidth alone does not bound its accumulated magnitude for large $|\Delta L_n|$. Exact equivalence requires all these residual phases to agree modulo integers; subtracting Eq. (10) for two frequencies requires $(f_k - f_{k'}) \Delta L_n / c \in \mathbb{Z}$. Consequently $\lambda\mathrm{PRF}$ is not an exact wideband ambiguity period. Frequency diversity and changing geometry generally break the simple P-periodicity, but this is not a proof that every finite sampled configuration is globally injective; special congruences or geometric symmetries are not ruled out.

Near-equivalence, admissibility, and evidence. The branch rule is a first-order initialiser, not a guarantee of a high-coherence alternative within the training prior. The bounded multi-start pilot of Appendix E.2 evaluates the *exact nonlinear kernel* at the alternatives it finds; it does not find "exactly equivalent" states or certify the global optimum. Median best-alternative coherences are 0.947 and 0.910 in the original and shifted geometries, with respective median position shifts of 0.40 and 0.15 m. Only 12% and 7% of the best alternatives lie inside the stated target prior. Their existence supports an approximate-branch interpretation; the negative failure-prediction result prevents equating branch strength with estimator failure risk. Better in-distribution recovery with more data likewise does not prove global identifiability or exact non-identifiability.

B Metric definitions, inventories, and local downstream approximation

B.1 Unfolded range rate and the folded diagnostic

For each labelled scene, evaluate $\mathbf{g}_n$ along the *true* target trajectory and use the central sample when N is odd, or the mean of the two central gradients when N is even. This sampled convention, denoted $\mathbf{g}_0$ in Eq. (8), approximates $\mathbf{g}(0)$ but is not exactly its continuous-time value for even N. The unfolded range-rate error is $e_{\mathrm{RR}} = |d|$ with $d = \mathbf{g}_0^\top (\hat{\mathbf{v}} - \mathbf{v})$, and with $P = (c/\bar{f})\mathrm{PRF}$, $\bar{f} = f_c$ for the symmetric frequency grid, the folded error is

$$e_{\text{AA-RR}} = \min_{m \in \mathbb{Z}} |d - mP| = |d - P\,\mathrm{round}(d/P)| \in [0, P/2]. \tag{20}$$

It is a local projected distance of the *velocity error* to the approximate carrier branch lattice $P\mathbb{Z}$ of Appendix A.3. It is not the full nonlinear rate difference between the true and predicted states (Eq. (15)), and we do not call it a quotient-space metric. The signed index is rounded to the nearest integer (ties-to-even); at a half-period tie either nearest integer gives the same folded error. We call $m \neq 0$ "alias-like", rather than evidence of exact aliasing or high coherence.

Projection and folding address different questions. Unfolded $|d|$ weights the velocity-error direction by the bistatic geometry. Folding additionally identifies scalar offsets separated by P; it can therefore remain small for a distant branch even when Euclidean velocity error is large. For each estimator, the reported aggregate is $\sqrt{M^{-1} \sum_{i=1}^{M} e_i^2}$ over the stated evaluation scenes. Spearman correlations compare these estimator-level aggregates with mean image error on the same scene set.

What the scalar ignores. The score ignores (i) velocity orthogonal to $\mathbf{g}_0$ (cross-range velocity), (ii) position error, (iii) geometry variation along the aperture, (iv) frequency-dependent phase mismatch, (v) branch admissibility under the prior and the branch order, (vi) clutter, and (vii) spatial image-grid effects. A zero score can occur without satisfying Eq. (10), and a nonzero branch index can reflect out-of-support extrapolation. It is thus not a metric on the full state space, a certified symmetry quotient, a motion-correction rule, or a bound on image localisation. The supporting folded vector error subtracts $mP\mathbf{g}_0 / \|\mathbf{g}_0\|^2$ from the velocity error and retains its orthogonal component, but still does not repair these limitations.

Both true velocity and true geometry enter the score. It is useful for *offline* evaluation and can score models on labelled validation scenes; it cannot supply deployment confidence without those labels.

In the frozen eight-candidate pool it improves ranking agreement, yet selects the same models as velocity RMSE and direct image error (Appendix G). Better ranking correlation is therefore not evidence of a realised selection benefit. Physical-state accuracy and downstream image error remain separate required outcomes.

The distinction also follows directly from the imager. In the noiseless single-target model,

$$I(\mathbf{q}; \hat{\mathbf{v}}) = |\gamma|\, C_*\big(\theta, (\mathbf{q}, \hat{\mathbf{v}})\big). \tag{21}$$

High coherence can occur at a displaced $\mathbf{q}$, so neither high matched amplitude nor low folded velocity error fixes the global maximiser's location. The CNN position is not substituted into this search. Noise, clutter, and the finite search grid further separate measurement fit from localisation.

B.2 Observed agreement and estimator inventory

Table 2: Descriptive Spearman agreement between state-error RMSE and mean image-error ranking in the saved 16-estimator diagnostic inventory. The unfolded and folded columns use the same reference-gradient projection, recomputed from saved signed range-rate errors.

PRF (Hz)	Euclidean velocity	Unfolded range rate	Folded range rate
15	0.09	−0.21	0.92
31	0.13	−0.03	0.98
62	0.90	0.92	0.98
124	0.93	0.97	0.97

At 62 Hz, one multi-hypothesis branch-refinement prediction is alias-like, although both primary CNNs have zero such events. This explains why folding can still change the full-inventory ranking at that rate. At 124 Hz all inventory predictions have zero branch index.

The original inventory excludes oracle and ground-truth-selected hypotheses, but includes static SAR, single-head state/coherence/displacement CNNs, a multi-hypothesis head, classical grid search, local-refinement variants, and branch rules. It is a diagnostic inventory, not 16 independent learning algorithms. The shifted geometry has fewer variants. To check that this inventory difference does not explain the finding, our audit additionally uses nine methods shared by both geometries: classical search; search plus physics; state-only; coherence; local refinement from each CNN; and the prior, image, and refined branch rules applied to coherence. In the original/shifted geometry this gives velocity rank agreement $-0.02/-0.32$ and folded range-rate agreement $0.98/0.93$. This is a post-hoc reanalysis of saved predictions and is labelled accordingly.

B.3 Supporting local displacement and phase-variance approximation

To specify the supporting calculation, flatten (n, k) into $M = NK$ rows and define $X_{nk,:} = (f_k/c)\mathbf{g}_n^\top$, $Y_{nk,:} = t_n X_{nk,:}$ and $z_{nk} = n$. Let $Q = I - \mathbf{1}\mathbf{1}^\top/M$ remove the one common phase offset. For a fixed velocity error and integer ramp m, the local unwrapped-phase fit is

$$\Delta\mathbf{p}_* = \arg\min_{\Delta\mathbf{p}} \|Q(X\Delta\mathbf{p} + Y\Delta\mathbf{v} - mz)\|_2^2 = J\Delta\mathbf{v} + m\mathbf{s}, \tag{22}$$
$$J = -(QX)^+ QY, \qquad \mathbf{s} = (QX)^+ Qz,$$

where $+$ is the Moore–Penrose pseudoinverse and the minimum-norm solution is used if necessary. This reproduces the saved displacement-map calculation, which sets $m = \mathrm{round}(d/P)$. It chooses a *candidate image position shift*, rather than using the CNN's position error. It is unconstrained by the image grid and minimises a quadratic phase surrogate, not the exact nonlinear global-peak error. Its norm is therefore a predicted local displacement, a supporting calculation, not a proven localisation error.

For centred residual cycles $r = Q(X\Delta\mathbf{p}_* + Y\Delta\mathbf{v} - mz)$, the code reports the proxy $\widetilde{C} = \exp(-\sigma_\phi^2/2)$, where $\sigma_\phi^2 = (2\pi)^2\|r\|^2/M$. The phase-only coherence of these linearised residuals would instead be $|M^{-1}\sum_i e^{-j2\pi r_i}|$. The exponential is a Gaussian-phase *approximation*; its small-residual expansion agrees to second order, but it is not that finite-sample sum, an exact-kernel

coherence, or a bound. Its method-ranking agreement varies with geometry, so neither this proxy nor predicted displacement is elevated to the primary diagnostic. No theorem connecting folded range-rate error to localisation error is claimed.

C NETWORK, OPTIMISATION, AND PROTOCOL DETAILS

CNN and normalisation. The input is globally phase-aligned using the largest-magnitude CSI element and divided by its RMS magnitude. Neither operation reads labels. Three 3×3 convolutions have 16, 32, and 64 channels and strides 1, 2, and 2, each followed by four-group GroupNorm and GELU. Adaptive average pooling produces a 4×4 map; a 64-unit GELU layer precedes the four-output linear head. Outputs are rescaled with centre $(0, 3, 0, 0)$ and scales $(0.6, 0.6, 0.7, 0.7)$. Output rescaling is not clipping; predictions can leave the training support. The model has 89,524 trainable parameters and 2.7×10^6 multiply–accumulates per scene. The residual CNN of Appendix G has 728,868 parameters and 1.05×10^8 multiply–accumulates. CUDA pooling uses deterministic bin averaging with the same adaptive-pooling bins; CPU and GPU training should not be assumed bitwise identical.

Original training and model selection. Training uses AdamW, learning rate 10^{-3}, weight decay 10^{-4}, batch size 64, and gradient-norm clipping at 5. A validation plateau scheduler halves the rate after five stale epochs, down to 10^{-5}. Early stopping uses patience 15, minimum 20 epochs, minimum improvement 10^{-4}, and maximum 120 epochs. The checkpoint is the best validation state loss for state-only, or best validation total objective for coherence. Validation sets have 256 scenes at 2k and 1,024 at larger sizes. The state and coherence models use identical architecture and stream definitions. The 31-Hz diagnostic reuses previously trained frozen models; the other rates and the shifted geometry use separately trained matched models. The fixed-update bank of Appendix G replaces this selection procedure entirely. More data entails more steps per epoch, so the scaling results compare the stated learning protocol, not a compute-matched sample-efficiency law. Measured trajectory times in the fixed-update bank are 87 s (CNN) and 203 s (residual CNN) on one V100.

Scene streams and reuse. The generator assigns an independent stream to each $(\text{seed}, \text{split}, \text{index})$ tuple. Diagnostic rates, shifted geometry, and clean 100 Hz use the same scene identities 212–311 per seed. Data scaling uses 312–511, or 200 scenes per seed. The clutter study uses 312–411 at 100 Hz: these target states overlap the scaling evaluation, with different measurement conditions, and are not an additional independent confirmation. The holdout streams (Appendix H) and the sealed selection and test streams (Appendix G) are separate streams whose states are hash-verified disjoint from every training, monitoring, and diagnostic set. Clutter-free and cluttered model comparisons are paired within a condition. No reported checkpoint was selected using any evaluation image.

Reflectivity and noise. Target amplitude is uniform in $[0.7, 1.3]$ and phase uniform in $[-\pi, \pi]$. Noise is circular complex Gaussian with power determined by target-only signal power and the sampled target-referenced SNR. Thus adding clutter does not increase thermal noise power. The phase-only propagation model omits path-loss variation, synchronisation drift, missing packets, target articulation, and hardware calibration errors. Frequency samples are uniformly spaced over the specified bandwidth; they are an idealised CSI model rather than a complete device specification.

Finite-grid classical controls. The classical search evaluates the prescribed position–velocity grid; "full" means the whole configured grid, not continuous global optimality. In the main diagnostic the grid has $13 \times 13 \times 9 \times 9$ points. Local refinement uses the shared variable-projection residual and L-BFGS with a 20-step cap. The three-branch control initialises $m \in \{-1, 0, 1\}$ along $\lambda \mathrm{PRF}\, \mathbf{g}(\hat{\mathbf{p}})/\|\mathbf{g}(\hat{\mathbf{p}})\|^2$, then selects the lowest locally fitted residual. It uses estimated geometry, unlike the truth-referenced evaluation metric. Search resolution, solver bounds, and optimisation budgets restrict comparisons. We make no state-of-the-art radar or wall-clock speedup claim from these baselines.

Procedures. Algorithm 1 states the evaluated pipeline; the learned model is the only trained component and the operator is fixed. Algorithm 2 states the sealed selection procedure of Appendix G, which was frozen in writing before any outcome was evaluated.

Algorithm 1 Learned motion estimation and coherent downstream imaging (one scene)

Require: complex CSI $H \in \mathbb{C}^{N\times K}$; transmitter $\mathbf{t}$ and receiver trajectory $\{\mathbf{r}_n\}$ (known); trained inverse model f_ϕ; imaging grid $\mathcal{Q}$; labels $(\mathbf{p}_0, \mathbf{v})$ used for evaluation only

1: $\tilde{H} \leftarrow$ align the global phase of H to its largest-magnitude element and divide by the RMS magnitude (label-free)
2: $(\hat{\mathbf{p}}, \hat{\mathbf{v}}) \leftarrow f_\phi(\tilde{H})$ {intermediate state; $\hat{\mathbf{p}}$ is not used in lines 3–7}
3: **for** each candidate position $\mathbf{q} \in \mathcal{Q}$ **do**
4: $\quad a_{nk}(\mathbf{q}, \hat{\mathbf{v}}) \leftarrow \exp[-j2\pi f_k L_n(\mathbf{q}, \hat{\mathbf{v}})/c]$ {target-motion compensation from $\hat{\mathbf{v}}$, Eq. (2)}
5: $\quad I(\mathbf{q}; \hat{\mathbf{v}}) \leftarrow \left|\frac{1}{NK}\sum_{n,k} H_{nk}\,\overline{a_{nk}(\mathbf{q}, \hat{\mathbf{v}})}\right|$ {coherent backprojection, Eq. (3)}
6: **end for**
7: $\hat{\mathbf{q}} \leftarrow \arg\max_{\mathbf{q}\in\mathcal{Q}} I(\mathbf{q}; \hat{\mathbf{v}})$ {global image peak is the downstream output}

Ensure: report separately: physical-state errors $\|\hat{\mathbf{v}} - \mathbf{v}\|_2$ and $\|\hat{\mathbf{p}} - \mathbf{p}_0\|_2$; ambiguity-aware diagnostic $e_{\text{AA-RR}}$ of Eq. (8); downstream image-localisation error $\|\hat{\mathbf{q}} - \mathbf{p}_0\|_2$

Algorithm 2 Sealed candidate selection and evaluation

Require: frozen candidate pool $\mathcal{C}$ (architectures $\times$ weights α $\times$ checkpoint endpoints) in a fixed tie-break order; selection scenes S_{sel} and sealed test scenes S_{test}, hash-verified disjoint from all training, monitoring, and diagnostic streams; selection rules $\rho_1, \ldots, \rho_4$ (velocity RMSE, unfolded range-rate RMSE, folded range-rate RMSE, mean image error); a written protocol fixing all of the above

1: train every candidate with the fixed update-indexed schedule; no validation-based stopping, scheduling, or checkpoint choice
2: evaluate every candidate on S_{sel}; for each rule choose $c_\rho \leftarrow \arg\min_{c\in\mathcal{C}} \rho(c)$, ties to the earliest candidate in the frozen order
3: write an immutable manifest (rules, choices, model and data hashes) *before* any test job is allowed to run
4: evaluate every candidate on S_{test} exactly once; report the sealed-test image error of each c_ρ, the paired scene-level difference between rules (primary: ρ_3 minus ρ_1), and the regret to the hindsight-best candidate

Ensure: no re-selection, no test-driven tuning; identical choices across rules are a valid (null) outcome

D SEED-LEVEL REVERSAL AND QUALITATIVE RECONSTRUCTIONS

Table 3: Seed-specific 31-Hz results. Δ image is coherence minus state-only mean localisation error. All numbers are computed from the same 100 paired scenes per seed.

Geometry	Seed	State-only velocity RMSE	Coherence velocity RMSE	Δ image (m)
Original	101	0.139	0.376	-0.174
Original	202	0.151	0.273	-0.193
Original	303	0.156	0.339	-0.147
Shifted TX	101	0.147	0.204	-0.086
Shifted TX	202	0.152	0.150	-0.175
Shifted TX	303	0.160	0.164	-0.122

For each seed, we pair methods on identical scene identities and resample those pairs with replacement 4,000 times using a fixed RNG seed. We average the three seed-level resampled differences with equal weight. This preserves the pairing and prevents training seeds with more rows from receiving greater weight. It does not estimate uncertainty over unseen training runs; three seeds are insufficient for precise claims of that kind. The main original-geometry image interval is $[-0.209, -0.131]$ m; shifted geometry is $[-0.170, -0.087]$ m. No multiple-testing significance claims are attached to the numerous exploratory comparisons.

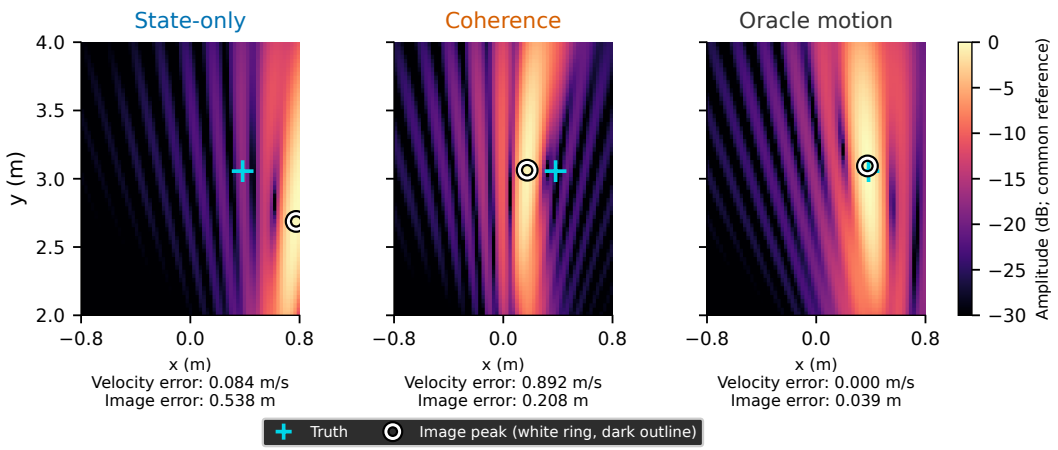


Figure 6: **A saved scene illustrating the reversal.** Common amplitude scale; cyan crosses mark the true reference position and white circles the global image peaks. This is the median image improvement among the 27 scenes with an alias-like coherence estimate, larger velocity error, and smaller image error than state-only (seed 101, index 227). The selection illustrates the mechanism; prevalence is assessed by the complete 300-scene table. Per-scene errors are norms, not dataset RMSE.

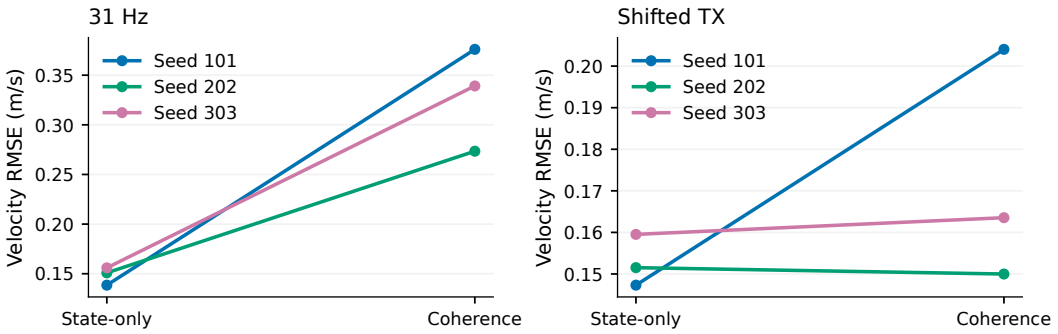


Figure 7: The original geometry has a velocity-RMSE penalty for coherence in every seed. The shifted geometry has an aggregate penalty but one seed does not exhibit it. This distinction is retained rather than calling the replication seed-universal.

E Clutter Failure Boundary and Negative Controls

E.1 Clutter Metrics and Controls

Power and scene definition. For eight scatterers, the per-scatterer amplitude factor for nominal level z is $\sqrt{10^{z/10}/8}$. Because coherent cross terms remain, realised CIR is measured from noise-free summed clutter CSI and target CSI, rather than equating this coefficient with power. Medians are $-10.021, -0.021, 9.979, 19.979$ dB. Clutter-aware training samples the factor log-uniformly between the -10 and $+20$ dB values. Its six runs plateau after 23–41 epochs under the same stopping rule; this mixture is not a clean-performance-preserving curriculum.

Global versus label-conditioned metrics. Global failure is image-peak error greater than 0.2 m. A clutter-associated false peak is a global maximum within 0.2 m of a static scatterer but farther than 0.2 m from the target. These are different events: not every failure lies near a known scatterer, and compensated static responses may be displaced. The target-local metric restricts the peak search to a 0.5-m radius around the *true* target. It is an oracle-conditioned diagnostic and must not be reported as autonomous target detection. Our main results use global image error; no truth-centred search is used to improve the reported localisation.

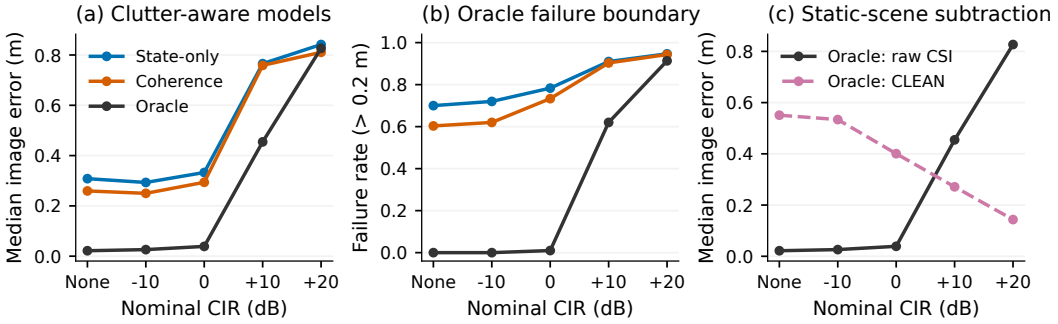


Figure 8: **Fixed-rate failure boundary, 100 Hz.** Panels (a,b) use clutter-trained CNNs and raw-CSI oracle-motion images, 300 paired scenes per level. Realised CIR medians differ from nominal levels by less than 0.1 dB. Panel (c) compares oracle images before/after eight-step static-scene CLEAN. It helps strong clutter but damages weak/absent clutter. CLEAN and slow-time mean subtraction are distinct controls.

Table 4: Oracle-motion controls on 300 paired scenes per level. FP denotes clutter-associated false-peak rate; failure denotes global error > 0.2 m. CLEAN uses eight iterations.

	Raw CSI			After static CLEAN		
Nominal CIR	Median (m)	Failure	FP	Median (m)	Failure	FP
None	0.022	0.000	0.000	0.551	0.817	0.000
-10 dB	0.026	0.000	0.000	0.534	0.813	0.133
0 dB	0.039	0.010	0.000	0.401	0.710	0.093
$+10$ dB	0.455	0.620	0.127	0.271	0.577	0.053
$+20$ dB	0.827	0.913	0.140	0.143	0.420	0.070

Static CLEAN is not DC removal. Each CLEAN iteration backprojects the residual at zero target velocity on a 129×129 grid spanning $[-1.6, 1.6] \times [1.4, 4.6]$ m, refines the strongest location on a 25×25 local grid over ± 0.03 m, fits its complex amplitude by projection, and subtracts the atom. Eight iterations reflect the true clutter count, an explicit advantage to this control. A fitted static atom can absorb moving-target energy, especially when clutter is absent. No learned model is retrained on CLEAN residuals.

DC removal instead subtracts the per-subcarrier mean over slow time, $H_{nk} - N^{-1} \sum_{n'} H_{n'k}$. With a moving receiver, static scatterers are generally not DC; this control is evaluated only without clutter and at +10 dB. Oracle median error at +10 dB changes from 0.455 to 0.421 m, and failure from 0.620 to 0.613; the saved paired intervals do not resolve an improvement. This must not be substituted for the +20-dB CLEAN result. Neither control establishes a deployable robust sensing pipeline.

E.2 Boundary and negative analyses

Branch correction is not uniformly beneficial. At 31 Hz in the original geometry, local branch fitting reduces coherence alias-like rate from 14.3% to 4.3%, velocity RMSE from 0.332 to 0.252 m/s, and mean image error from 0.303 to 0.267 m. In the shifted geometry it changes alias-like rate from 3.0% to 7.0% and velocity RMSE from 0.174 to 0.335 m/s, while image error improves from 0.334 to 0.292 m. This is another reason to report physical recovery and imaging separately. At clean 100 Hz it selects no corrective alias branch; any gain from the three-fit variant is reproduced by local refinement from the central coherence estimate.

In the noiseless single-target model the true state has zero projection residual, whereas a strictly near-equivalent state has positive residual. This fact does not guarantee the branch-fitting algorithm's success on noisy observations: noise can change the ordering, and finite local optimisation, bounds, or untested branches can prevent recovery of the best fit. Nor does a lower fitted residual guarantee a better global image location.

Branch existence versus branch selection. A bounded multi-start pilot initialised alternatives at $m \in \{-2, -1, 1, 2\}$ and refined all four coordinates for up to 40 L-BFGS steps against the noise-free truth kernel. The primary domain is $[-0.8, 0.8] \times [2, 4]$ m in position and $[-3, 3]^2$ m/s in velocity; a wider-position sensitivity uses $[-3, 3] \times [0.5, 6.5]$ m. Candidates with $|\mathbf{g}_0^\top \Delta\mathbf{v}| < P/2$ after refinement are excluded to avoid collapse into the central scalar cell. This exclusion does not classify all possible four-dimensional ambiguities. The score is the largest *exact-kernel coherence found*, not a certified global maximum. It failed to improve leave-one-seed-out alias prediction over speed and geometry: AUROC changed from 0.922 to 0.916 when added; alone it attained 0.509. These results do not validate scene-level failure prediction. They also do not prove that speed or the prior edge causes branch selection. Local-conditioning and residual-gap pilots likewise did not supply a sufficient explanation. In the pilot's taxonomy, alias-like cases account for 33 of 193 original-geometry coherence image failures and 9 of 172 shifted-geometry failures; the bulk of failures is not explained by that category.

Additional model and optimisation controls. The multi-hypothesis head did not improve the joint state/imaging trade-off enough to become a main method. Generic local refinement can lower residual while worsening physical recovery, depending on initialisation. These controls remain in the numerical record.

F HIGH-SPEED OPERATING-REGIME EXTENSION

The carrier scalar has a conventional centred interval $[-P/2, P/2)$; this is not an unambiguous region of the full state space. Keeping projected target rates within that interval excludes nonzero integer-period coincidences in the frozen scalar model, not all four-dimensional alternatives. The clean 100-Hz results combine a prior-based scale check with the separate empirical observation of zero alias-like predictions. Since network outputs are not clipped to the prior, the latter does not follow from the prior check alone.

The 100-Hz 2k/65k models were evaluated without retraining on 600 scenes with speeds uniform in $[1.3, 2.0]$ m/s and full-circle directions. This changes the target distribution beyond the trained speed and direction support. Some truth range rates cross the first-order half-period, but that alone does not establish measurement ambiguity within a prior or a selected branch. Learned estimates can extrapolate severely; a nonzero folded index is then not reliable evidence of coherent alias selection. For this reason we do not turn this stress test into a claim that aliasing is practically prevalent at 100 Hz, nor a claim of robust out-of-distribution inference. The controlled disjoint-support holdouts at 31 Hz are reported in Appendix H.

G CONTROLLED OPTIMISATION, ARCHITECTURE, GRID, AND SEALED SELECTION

Protocol. Settings were frozen in a written protocol before any comparison outcome was evaluated (Algorithm 2). The bank uses the original geometry, 31 Hz, 2,048 training scenes (train stream 0–2047), 256 monitoring-validation scenes (validation stream 0–255), batch 64, AdamW with weight decay 10^{-4}, and gradient clipping at 5. The learning rate is 10^{-3} for zero-indexed updates 0–1919, 5×10^{-4} for 1920–3839, 10^{-4} for 3840–7679, and 10^{-5} for 7680–11519. There are 32 updates per epoch; checkpoints at epochs 120 and 360 belong to the same trajectory. The objective is $\mathcal{L}_s + \alpha(1 - C)$ with $\alpha \in \{0, 0.1, 0.5, 1\}$, applied once. Monitoring validation never selected checkpoints or scheduled training. The CUDA smoke verified bit-exact stop/resume. The residual CNN keeps the input normalisation and output rescaling; it has a 3×3 stem (32 channels), three stages of two pre-activation residual blocks (GroupNorm–GELU–conv, identity or 1×1 projection skip) with 32, 64 (stride 2), and 128 (stride 2) channels, deterministic adaptive pooling to 2×2, and a 64-unit head. Equal update counts are not equal compute. All 24 full runs completed without resumes or non-finite gradients; Figure 9 shows every trajectory.

Sealed-test bank at 11,520 updates. Table 5 lists every candidate on the 3,000 sealed scenes. Paired $\alpha = 0.5$ minus $\alpha = 0$ image differences are -0.068 m $[-0.078, -0.057]$ (CNN) and -0.093 m $[-0.101, -0.085]$ (ResCNN); at 3,840 updates they are -0.088 m $[-0.100, -0.077]$ and -0.035

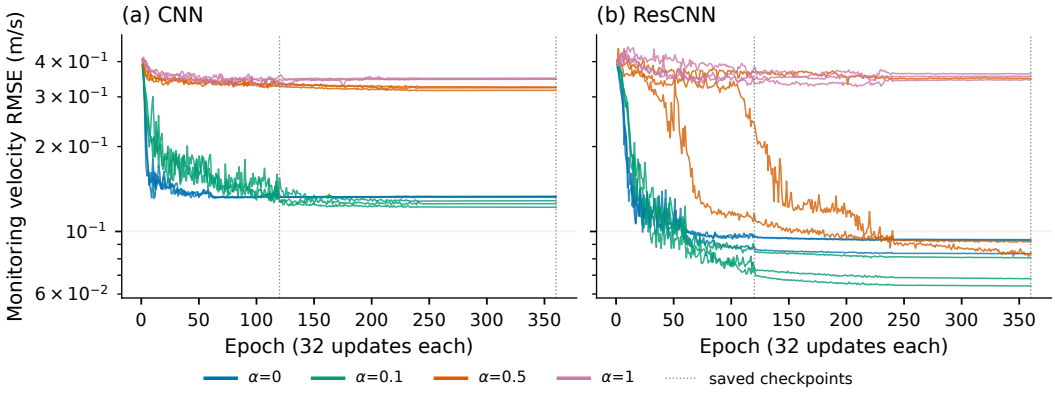


Figure 9: Monitoring-validation velocity RMSE of all 24 fixed-schedule trajectories, (a) CNN and (b) ResCNN, three seeds per weight; dotted lines mark the two saved checkpoints. The $\alpha \geq 0.5$ curves still drift over the last 60 epochs; no run is called converged.

m $[-0.043, -0.026]$. On the 300 diagnostic scenes the corresponding 11,520-update differences are -0.059 m $[-0.094, -0.024]$ and -0.066 m $[-0.091, -0.040]$. In the ResCNN, alias-like selection at $\alpha = 0.5$ occurs mainly in one seed (14.2% against 1.1% and 0.4%), which is why its velocity penalty is seed-dependent while its image gain is not. The monitoring loss of the $\alpha \geq 0.5$ runs still changes over the last 60 epochs and rises slightly for the CNN, so these runs are not called converged. Because budget, schedule, and selection changed together relative to the original comparison, the smaller controlled gap cannot be attributed to checkpoint selection specifically.

Table 5: Frozen candidate bank on the sealed test (1,000 fresh scenes per seed, 11,520 updates, 65×65 grid). Velocity is vector RMSE; image is mean global-peak error; alias-like is the fraction with nonzero folded index. Per-seed columns list seeds 101 / 202 / 303.

Model	α	Vel. (m/s)	Image (m)	Alias (%)	Vel. 101 / 202 / 303	Image 101 / 202 / 303
CNN	0	0.138	0.438	0.2	0.139 / 0.136 / 0.138	0.447 / 0.433 / 0.433
CNN	0.1	0.126	0.392	0.1	0.125 / 0.126 / 0.128	0.393 / 0.397 / 0.387
CNN	0.5	0.332	0.370	13.5	0.342 / 0.327 / 0.328	0.377 / 0.356 / 0.377
CNN	1	0.355	0.311	15.3	0.354 / 0.347 / 0.365	0.305 / 0.306 / 0.323
ResCNN	0	0.096	0.313	0.1	0.096 / 0.089 / 0.102	0.316 / 0.296 / 0.327
ResCNN	0.1	0.075	0.240	0.0	0.070 / 0.068 / 0.086	0.221 / 0.223 / 0.277
ResCNN	0.5	0.216	0.220	5.2	0.121 / 0.347 / 0.075	0.227 / 0.270 / 0.162
ResCNN	1	0.354	0.280	14.9	0.351 / 0.346 / 0.366	0.278 / 0.278 / 0.285

What the two saved checkpoints establish. For the controlled CNN pair, both metrics were evaluated on the same final-test scenes at epochs 120 and 360. The coherence model's velocity RMSE decreases from 0.338 to 0.332 m/s, while mean image error increases from 0.350 to 0.370 m; these directions hold in all three seeds. Among these two endpoints, velocity therefore favours epoch 360 and imaging favours epoch 120. State-only favours epoch 360 on both aggregate metrics, but its mean image improvement is only 0.00049 m and does not hold in every seed. The cross-model ranking reversal already exists at epoch 120. These are retrospective endpoint comparisons, not validation-based checkpoint choices; they do not identify the onset of reversal or the optimum over intervening epochs.

Image grid and field of view. The saved 31-Hz predictions were re-imaged after regenerating the CSI from the saved configuration; original-grid errors reproduced the saved values to 10^{-16} m. Mean image error (state-only / coherence / oracle) and fraction of global peaks on the outer row or column (state-only / coherence): original geometry, 65×65: 0.474 / 0.303 / 0.031 m, 11.3% / 1.7%; 129×129 same field of view: 0.474 / 0.301 / 0.014 m, 10.7% / 1.7%; 129×129 wide field of view $[-1.6, 1.6] \times [1, 5]$ m: 0.529 / 0.313 / 0.031 m, 2.7% / 0.3%. Shifted geometry: 0.461 / 0.334 / 0.032

m and 9.7% / 4.3%; 0.461 / 0.332 / 0.015 m; 0.477 / 0.350 / 0.032 m and 0.7% / 0.3%. Boundary cases are kept in every primary number. A secondary interior subset (neither method's original-grid peak on the boundary; 261 and 263 scenes) keeps the sign with smaller magnitude (-0.120 and -0.094 m). This is a post-hoc sensitivity analysis of previously used scenes, not an independent confirmation.

Sealed selection. For each seed and endpoint the eight candidates were scored on the 512 selection scenes by (1) vector velocity RMSE, (2) unfolded range-rate RMSE with the reference gradient of Eq. (8), (3) folded range-rate RMSE, and (4) mean image error; exact ties would go to the earlier candidate in the frozen order. An immutable selection manifest was written before the final-test jobs were allowed to run. At 11,520 updates, rules 1, 3, and 4 choose ResCNN $\alpha = 0.1$ in seeds 101 and 202 and ResCNN $\alpha = 0.5$ in seed 303 (sealed-test image error 0.221, 0.223, 0.162 m; velocity RMSE 0.070, 0.068, 0.075 m/s; alias-like 0.1%, 0.0%, 0.4%); rule 2 chooses ResCNN $\alpha = 0.1$ in seed 303 (0.277 m). At 3,840 updates all four rules choose ResCNN $\alpha = 0.1$ in every seed. The primary statistic (rule 3 minus rule 1, sealed-test mean image error) is 0.000 m in every seed; identical choices are a valid null outcome, and the study was not extended to search for a positive one. Figure 10 shows every candidate's sealed-test error with the selected models marked. Bootstrap intervals in this appendix use 4,000 within-seed paired resamples with a recorded RNG seed and condition on the three trained models.

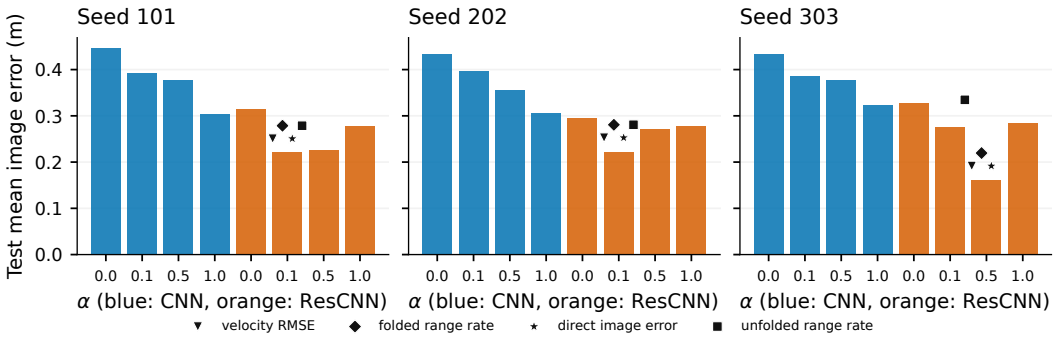


Figure 10: Sealed-test mean image error of all eight candidates at 11,520 updates per seed, with the model chosen by each validation rule marked. Velocity RMSE, folded range rate, and direct image error select the same candidate in every seed.

H DISJOINT-SUPPORT HOLDOUT EVALUATION

Protocol. This is an evaluation-only stress test of the frozen fixed-update bank of Appendix G (both architectures, $\alpha \in \{0, 0.1, 0.5, 1\}$, three seeds, both checkpoints), written down before any outcome was evaluated. Three holdout domains have supports disjoint from the training support (speed 0–0.7 m/s, direction -150 to $150°$, $x \in [-0.6, 0.6]$ m): *speed* 0.75–1.05 m/s, *direction* 151–180°, and *position* $x \in [0.625, 0.75]$ m, with every other factor unchanged (31 Hz, original geometry, same y range, SNR, reflectivity). Each domain uses its own reserved scene stream, indices 200–711 per seed (512 scenes per seed, 1,536 per domain); indices 0–199 belong to the older evaluation reported below. The sealed final-test stream of Appendix G is re-evaluated by the same script as the in-distribution reference; its 48,000 candidate rows reproduce the frozen bank rows exactly, and oracle-motion (true velocity) and static (zero velocity) imaging are added on every scene. Primary statistic per domain and architecture, at 11,520 updates: the sign of the velocity-RMSE change and the paired mean image-error difference of $\alpha = 0.5$ minus $\alpha = 0$ (4,000 within-seed resamples, conditional on the trained models). We call the outcome a *reversal* when velocity RMSE rises and the image interval lies below zero. No claim of general out-of-distribution robustness is made from three strips.

Findings (Table 6, Figure 11). *Speed holdout.* Both architectures extrapolate poorly: velocity RMSE is 0.51 m/s at $\alpha = 0$ and rises to 0.85 (CNN) and 0.78 m/s (ResCNN) at $\alpha = 0.5$, in every seed, with alias-like fractions of 72% and 53%. Mean image error nevertheless falls from 0.812 to

Table 6: Frozen bank on disjoint-support holdouts at 11,520 updates (512 scenes per seed and domain; the in-distribution rows are the 3,000 sealed scenes). Vel. is vector velocity RMSE (m/s); folded is the alias-aware range-rate RMSE (m/s); image is mean / median global-peak error (m); fail is the fraction above 0.2 m; alias-like is the fraction with nonzero folded index; the last column lists per-seed mean image error (101 / 202 / 303). Oracle motion and static SAR are reference imagers on the same scenes.

Domain	Model	α	Vel.	Folded	Image	Median	Fail (%)	Alias (%)	Image per seed
in-distribution	CNN	0	0.138	0.215	0.438	0.367	76	0.2	0.447 / 0.433 / 0.433
in-distribution	CNN	0.5	0.332	0.198	0.370	0.307	70	13.5	0.377 / 0.356 / 0.377
in-distribution	ResCNN	0	0.096	0.160	0.313	0.257	61	0.1	0.316 / 0.296 / 0.327
in-distribution	ResCNN	0.5	0.216	0.115	0.220	0.179	45	5.2	0.227 / 0.270 / 0.162
in-distribution	oracle motion	–	0	0	0.027	0.024	0	–	–
in-distribution	static SAR	–	0.404	0.486	0.803	0.781	90	–	–
speed 0.75–1.05	CNN	0	0.511	0.453	0.812	0.795	95	33.3	0.800 / 0.838 / 0.798
speed 0.75–1.05	CNN	0.5	0.845	0.278	0.529	0.475	85	72.1	0.521 / 0.529 / 0.537
speed 0.75–1.05	ResCNN	0	0.508	0.385	0.670	0.636	88	30.7	0.673 / 0.648 / 0.690
speed 0.75–1.05	ResCNN	0.5	0.781	0.348	0.558	0.508	84	52.9	0.589 / 0.485 / 0.599
speed 0.75–1.05	oracle motion	–	0	0	0.027	0.023	0	–	–
speed 0.75–1.05	static SAR	–	0.902	0.435	0.699	0.610	93	–	–
direction 151–180°	CNN	0	0.239	0.329	0.514	0.492	88	0.6	0.496 / 0.559 / 0.489
direction 151–180°	CNN	0.5	0.214	0.363	0.486	0.475	84	0.3	0.484 / 0.524 / 0.450
direction 151–180°	ResCNN	0	0.185	0.327	0.412	0.385	78	0.5	0.443 / 0.409 / 0.385
direction 151–180°	ResCNN	0.5	0.152	0.274	0.323	0.283	65	0.0	0.345 / 0.418 / 0.206
direction 151–180°	oracle motion	–	0	0	0.032	0.029	0	–	–
direction 151–180°	static SAR	–	0.399	0.139	0.468	0.452	84	–	–
x 0.625–0.75	CNN	0	0.194	0.302	0.657	0.633	89	1.0	0.670 / 0.664 / 0.637
x 0.625–0.75	CNN	0.5	0.336	0.287	0.614	0.590	92	14.3	0.591 / 0.625 / 0.625
x 0.625–0.75	ResCNN	0	0.149	0.239	0.508	0.476	86	0.4	0.523 / 0.478 / 0.522
x 0.625–0.75	ResCNN	0.5	0.257	0.195	0.417	0.403	80	7.0	0.429 / 0.553 / 0.269
x 0.625–0.75	oracle motion	–	0	0	0.025	0.023	0	–	–
x 0.625–0.75	static SAR	–	0.399	0.468	0.878	0.817	88	–	–

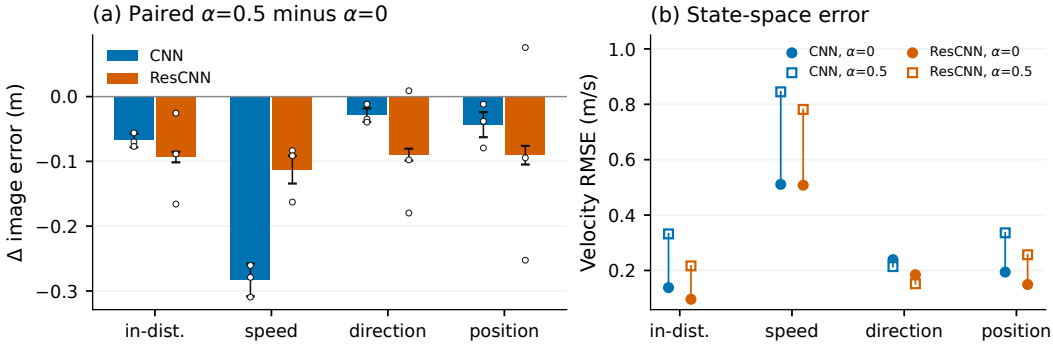


Figure 11: Holdout evaluation of the frozen bank at 11,520 updates. (a) Paired mean image-error difference of $\alpha = 0.5$ minus $\alpha = 0$ per domain and architecture, 95% intervals from 4,000 within-seed resamples; white markers are seeds. (b) Velocity RMSE of the same pairs (filled: $\alpha = 0$; open: $\alpha = 0.5$). The direction strip is the one domain in which coherence regularisation lowers both errors.

0.529 m (CNN, paired -0.283 m $[-0.308, -0.258]$) and from 0.670 to 0.558 m (ResCNN, -0.112 m $[-0.134, -0.091]$), with image gains in every seed. The reversal is therefore larger outside the speed support than inside it, consistent with the true bistatic rate of these scenes lying close to or beyond $P/2$ where the network must choose a branch. *Position holdout.* Velocity RMSE rises from 0.194 to 0.336 m/s (CNN) and 0.149 to 0.257 m/s (ResCNN) in every seed, while mean image error falls by -0.043 m $[-0.063, -0.024]$ and -0.090 m $[-0.105, -0.076]$; the ResCNN image gain is not seed-uniform (-0.094, $+0.076$, -0.253 m). *Direction holdout.* No reversal: $\alpha = 0.5$ lowers velocity RMSE (0.239 to 0.214 and 0.185 to 0.152 m/s) and image error (-0.029 m $[-0.039, -0.018]$; -0.089 m $[-0.099, -0.080]$), with alias-like fractions below 1%. In this strip the coherence term acts as a plain regulariser and both metrics agree.

Secondary statistics. The outcomes are not specific to one checkpoint or weight: at 3,840 updates the speed and position reversals of the CNN persist (-0.347 and -0.063 m), while for the ResCNN at 3,840 updates the position-holdout image difference changes sign ($+0.080$ m $[+0.064, +0.095]$), as it does at $\alpha = 1$ ($+0.100$ m); $\alpha = 0.1$ improves both metrics in every domain except the speed strip, where it still reverses (-0.060 and -0.038 m). Boundary peaks are frequent outside the support (40% of state-only CNN peaks in the speed strip lie on the outer row or column, against 12% at $\alpha = 0.5$), so part of the holdout image error is a field-of-view effect of the kind quantified in Appendix G; the oracle imager stays at 0.025–0.032 m in every domain, so the failures are estimator failures, not imaging-operator failures.

Table 7: Older validation-selected models (Appendix C protocol; 2k, 8k and 32k training scenes) on the same three holdout domains, holdout-stream indices 0–199 per seed (600 scenes). Columns: state-only and coherence velocity RMSE (m/s); paired coherence-minus-state-only mean image difference (m) with a 95% interval; per-seed differences; alias-like fractions (state-only / coherence).

Train	Domain	Vel. state	Vel. coh.	Δ image [95%]	Per seed	Alias (%)
2k	speed 0.75–1.05	0.524	0.831	-0.433 [-0.471, -0.394]	-0.489 / -0.381 / -0.430	40.0 / 72.0
2k	direction 151–180°	0.200	0.192	-0.009 [-0.026, +0.008]	+0.097 / +0.016 / -0.140	0.2 / 0.2
2k	x 0.625–0.75	0.214	0.354	-0.075 [-0.109, -0.042]	-0.120 / -0.094 / -0.012	1.3 / 17.5
8k	speed 0.75–1.05	0.481	0.693	-0.271 [-0.311, -0.231]	-0.384 / -0.234 / -0.195	32.8 / 48.2
8k	direction 151–180°	0.215	0.192	-0.055 [-0.068, -0.041]	-0.095 / -0.035 / -0.034	0.2 / 0.3
8k	x 0.625–0.75	0.180	0.245	-0.099 [-0.129, -0.070]	-0.154 / -0.029 / -0.113	0.8 / 6.0
32k	speed 0.75–1.05	0.715	0.647	-0.058 [-0.094, -0.022]	-0.013 / -0.014 / -0.148	29.5 / 31.2
32k	direction 151–180°	0.137	0.117	-0.018 [-0.029, -0.008]	-0.025 / +0.013 / -0.043	0.3 / 0.0
32k	x 0.625–0.75	0.074	0.078	-0.049 [-0.067, -0.032]	-0.047 / -0.062 / -0.038	0.2 / 0.3

Older protocol and data scale. Table 7 reports the 2k/8k/32k models trained under the original validation-selected protocol (Appendix C) on the same strips. The speed-strip reversal is large at 2k and 8k (-0.433 and -0.271 m, every seed) and disappears at 32k, where the coherence model has *lower* velocity RMSE (0.647 versus 0.715 m/s) and a smaller image gain (-0.058 m); the position-strip reversal is present at 2k and 8k and reduces to a near-tie in velocity at 32k (0.078 versus 0.074 m/s); the direction strip never reverses. These agree in direction with the controlled bank at 2k.

What this does and does not show. The reversal is not confined to the training support: with the same frozen models it appears, and grows, in the speed strip and remains in the position strip, in both architectures. It is not universal: the direction strip and, for the ResCNN, a lower weight or the earlier checkpoint remove it, and in one ResCNN configuration the sign of the image difference flips. Absolute accuracy outside the support is poor for every candidate (failure rates 65–95%), so these are boundary tests of the ranking question, not evidence that either model localises reliably out of distribution.

I Additional geometries under the controlled protocol

Protocol. The complete fixed-update bank protocol of Appendix G (both architectures, $\alpha \in \{0, 0.1, 0.5, 1\}$, seeds 101/202/303, 11,520 updates, update-indexed learning rate, no validation-based selection, checkpoints at 3,840 and 11,520 updates) was rerun in two further geometries, written down before any outcome was evaluated: *G2*, the transmitter moved to $(-4, 0)$ m with the receiver path unchanged (the geometry of the older validation-selected replication in Section 5.3), and *G3*, the transmitter at $(3, 3)$ m beside the target region and the receiver path rotated by $30°$ about the origin, a bistatic angle near $90°$ at the prior centre. Target states are drawn from the same streams as the original bank, so scene identities pair across geometries while the CSI differs. Every candidate is evaluated on the sealed final-test stream of its geometry (1,000 scenes per seed) with oracle-motion and static reference imagers; the sealed selection experiment was not repeated. The geometries differ in their sampled branch structure: the median reference gradient norm $\|\mathbf{g}_0\|$ is 1.95, 1.79, and 1.43, the bisector spacing $P/\|\mathbf{g}_0\|$ is 0.90, 0.97, and 1.22 m/s, and the fraction of test scenes whose true bistatic rate $|\mathbf{g}_0^\top \mathbf{v}|$ exceeds $P/2$ is 16.5%, 12.6%, and 3.2% for the original, G2, and G3 geometries ($P = 1.747$ m/s throughout).

Table 8: Fixed-update bank in three geometries on each geometry's sealed test (1,000 scenes per seed, 11,520 updates). Velocity is vector RMSE (m/s); image is mean global-peak error (m); alias-like is the fraction with nonzero folded index; per-seed columns list seeds 101 / 202 / 303.

Geometry	Model	α	Vel.	Image	Alias (%)	Vel. per seed	Image per seed
original	CNN	0	0.138	0.438	0.2	0.139 / 0.136 / 0.138	0.447 / 0.433 / 0.433
original	CNN	0.1	0.126	0.392	0.1	0.125 / 0.126 / 0.128	0.393 / 0.397 / 0.387
original	CNN	0.5	0.332	0.370	13.5	0.342 / 0.327 / 0.328	0.377 / 0.356 / 0.377
original	CNN	1	0.355	0.311	15.3	0.354 / 0.347 / 0.365	0.305 / 0.306 / 0.323
original	ResCNN	0	0.096	0.313	0.1	0.096 / 0.089 / 0.102	0.316 / 0.296 / 0.327
original	ResCNN	0.1	0.075	0.240	0.0	0.070 / 0.068 / 0.086	0.221 / 0.223 / 0.277
original	ResCNN	0.5	0.216	0.220	5.2	0.121 / 0.347 / 0.075	0.227 / 0.270 / 0.162
original	ResCNN	1	0.354	0.280	14.9	0.351 / 0.346 / 0.366	0.278 / 0.278 / 0.285
original	oracle motion	–	0	0.027	–	–	–
G2 shifted TX	CNN	0	0.136	0.405	0.1	0.139 / 0.132 / 0.136	0.413 / 0.399 / 0.403
G2 shifted TX	CNN	0.1	0.131	0.380	0.2	0.140 / 0.125 / 0.129	0.395 / 0.363 / 0.380
G2 shifted TX	CNN	0.5	0.145	0.378	0.9	0.165 / 0.121 / 0.144	0.383 / 0.363 / 0.389
G2 shifted TX	CNN	1	0.316	0.325	9.6	0.318 / 0.308 / 0.323	0.321 / 0.318 / 0.337
G2 shifted TX	ResCNN	0	0.096	0.299	0.0	0.099 / 0.092 / 0.099	0.307 / 0.284 / 0.306
G2 shifted TX	ResCNN	0.1	0.084	0.269	0.0	0.088 / 0.080 / 0.083	0.292 / 0.256 / 0.260
G2 shifted TX	ResCNN	0.5	0.061	0.183	0.0	0.066 / 0.063 / 0.054	0.184 / 0.195 / 0.170
G2 shifted TX	ResCNN	1	0.297	0.267	8.1	0.300 / 0.267 / 0.321	0.281 / 0.272 / 0.249
G2 shifted TX	oracle motion	–	0	0.029	–	–	–
G3 lateral TX + rotated RX	CNN	0	0.157	0.378	0.1	0.157 / 0.154 / 0.161	0.384 / 0.372 / 0.379
G3 lateral TX + rotated RX	CNN	0.1	0.155	0.358	0.2	0.149 / 0.154 / 0.161	0.355 / 0.350 / 0.368
G3 lateral TX + rotated RX	CNN	0.5	0.149	0.351	0.3	0.149 / 0.149 / 0.150	0.348 / 0.337 / 0.368
G3 lateral TX + rotated RX	CNN	1	0.144	0.323	0.3	0.142 / 0.147 / 0.142	0.324 / 0.313 / 0.332
G3 lateral TX + rotated RX	ResCNN	0	0.121	0.301	0.0	0.112 / 0.124 / 0.127	0.286 / 0.307 / 0.311
G3 lateral TX + rotated RX	ResCNN	0.1	0.127	0.309	0.0	0.122 / 0.127 / 0.133	0.299 / 0.311 / 0.318
G3 lateral TX + rotated RX	ResCNN	0.5	0.115	0.272	0.1	0.113 / 0.102 / 0.127	0.272 / 0.252 / 0.292
G3 lateral TX + rotated RX	ResCNN	1	0.105	0.251	0.1	0.105 / 0.097 / 0.112	0.250 / 0.239 / 0.263
G3 lateral TX + rotated RX	oracle motion	–	0	0.032	–	–	–

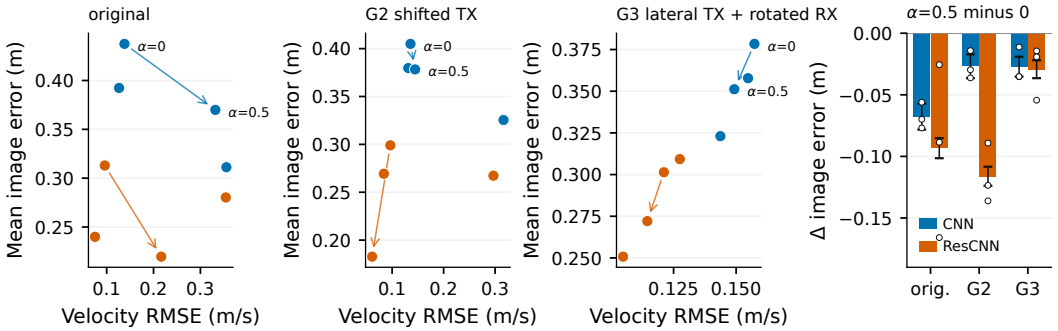


Figure 12: Velocity RMSE against mean image error for every candidate at 11,520 updates in each geometry (blue CNN, orange ResCNN; arrows join $\alpha = 0$ to $\alpha = 0.5$ for each architecture, labelled for the CNN), and the paired $\alpha = 0.5$ minus $\alpha = 0$ image difference per geometry and architecture (95% intervals from 4,000 within-seed resamples; white markers are seeds).

Findings (Table 8, Figure 12). *Original geometry.* The reversal of Appendix G: $\alpha = 0.5$ raises velocity RMSE (0.138 to 0.332 and 0.096 to 0.216 m/s) and lowers image error (-0.068 and -0.093 m), with alias-like fractions of 13.5% and 5.2%. *G2.* For the CNN the velocity penalty is small and not seed-uniform (0.136 to 0.145 m/s) and the image gain is -0.027 m $[-0.037, -0.017]$; for the ResCNN, $\alpha = 0.5$ *lowers* velocity RMSE (0.096 to 0.061 m/s) and image error (-0.116 m $[-0.124, -0.108]$) in every seed. Alias-like fractions are below 1%. Only $\alpha = 1$ produces alias-like selection (9.6% and 8.1%) and a velocity penalty. *G3.* No reversal for either architecture: $\alpha = 0.5$ lowers both velocity RMSE (0.157 to 0.149; 0.121 to 0.115 m/s) and image error (-0.027 and -0.029 m), and alias-like fractions stay at or below 0.3% for every weight.

Interpretation and older protocol. Across the three geometries the reversal is present, marginal, and absent in the same order as the fraction of scenes whose true rate approaches the first sampled

branch (16.5%, 12.6%, 3.2%) and as the alias-like fractions at $\alpha = 0.5$. This is consistent with the mechanism of Section 5.3: the coherence term changes the ranking when the prior reaches a distant near-equivalent branch, and acts as an ordinary regulariser that improves both metrics when it does not. The older validation-selected replication in G2 (velocity RMSE 0.153 versus 0.174 m/s, image -0.127 m, alias-like 3.0%; Section 5.3) showed a larger reversal than the controlled bank in the same geometry, so training protocol and geometry interact. The controlled tests therefore bound the claim: the ranking reversal is a property of sampled geometries in which the prior reaches an approximate branch, not of the coherence objective in general, and it is not a universal property of the testbed.

J CROSS-SIMULATOR DIGITAL-TWIN VALIDATION

Motivation. The analytic simulator deliberately isolates the sampled-physics mechanism: a point target, a unit-modulus phase kernel, no propagation loss, and no static environment. To test whether the observed ranking mismatch depends on that forward model, we regenerated the training, monitoring, and sealed-test scene streams with an independent ray-traced RF digital twin (SensingSP with Blender's `bpy` ray casting) and retrained the same models under the same fixed-update protocol. This is a cross-simulator validation, not measured-hardware validation.

Setup. Acquisition parameters are matched (5.32 GHz, 80 MHz, 32 subcarriers, 1 s, 31 Hz or 100 Hz, TX at $(-1.5, 0)$ m, the same receiver path, the same target prior); every twin scene reuses the analytic scene's state, reflectivity phase, and SNR draw, so scenes pair across simulators. The twin differs in propagation: the target is a 0.02 m icosphere whose facets are individually visibility-tested by ray casting; path amplitudes follow spreading loss, facet area, and RCS factors; and, in the wall conditions, four wall planes with 1 m facets add single-scatter multipath calibrated to a clutter-to-target power near 0 dB (median 0.3 dB). The direct TX–RX path is excluded, as in the analytic model. The CSI is synthesised coherently from the ray-traced path list with the same negative-propagation-phase convention, $H_{nk} = \sum_p a_p \exp(-j2\pi f_k d_p/c)$; the receiver positions, target centre distances, and delay slopes reproduce the analytic geometry to 10^{-7} m and 0.3%, and the twin target-only CSI has coherence 0.998 with the analytic steering vector of the same state. A 0.05 m sphere gave 0.67 (wavelength-scale glint) and was rejected before any training. Figure 13 shows one stored scene of the wall condition. Conditions at 31 Hz (three seeds each): ray-traced phases with their spreading-loss amplitudes ("loss"), the same with walls, and amplitude-flattened variants of the no-wall dataset in which the target magnitude is replaced while its phase is kept: an early variant with fresh noise, and *same-noise* variants in which the exact stored noise realisation $h - h_{\text{target}} - h_{\text{clutter}}$ is re-added (reconstruction residual $< 10^{-20}$), flattening either the whole magnitude, only its slow-time variation (per subcarrier), or only its frequency variation (per snapshot). One 100-Hz wall condition (one seed) was run. Training and imaging are identical to Appendix G: CNN, $\alpha \in \{0, 0.1, 0.5\}$, identical initialisation and minibatch order per seed, 11,520 updates, analytic imaging operator with the known receiver path, 300 sealed scenes per seed.

Result. With ray-traced phases and a flat target amplitude the reversal reproduces (Table 9, Figure 14). The same-noise control isolates the amplitude: it changes only the target magnitude while keeping the exact phase history, states, clutter, and noise realisation, and $\alpha = 0.5$ raises velocity RMSE in every seed while lowering mean image error (paired -0.069 m $[-0.089, -0.047]$, image gain in every seed, 12.0% alias-like estimates); the earlier fresh-noise variant agrees, with range-rate errors clustering at 1.68 m/s against the predicted $P = 1.75$ m/s. Flattening only the slow-time amplitude variation restores the reversal as well (-0.070 m $[-0.090, -0.050]$), whereas flattening only the frequency variation does not (both metrics improve, no alias-like estimate). Analytic-trained models evaluated on twin CSI reverse in the same way (image -0.097 m). With the ray-traced spreading-loss amplitudes retained, the reversal is absent: both metrics improve and no estimate is alias-like, even though the slow-time modulation of the target echo is small (median coefficient of variation 3.6%, Figure 15) and leaves the first-branch coherence unchanged (0.716 twin versus 0.715 analytic). With walls near 0 dB there is neither a reversal nor a resolved image gain ($+0.018$ m $[-0.003, +0.039]$). At 100 Hz with walls, $\alpha = 0.5$ improves both metrics and no estimate is alias-like, consistent with the analytic 100-Hz result.

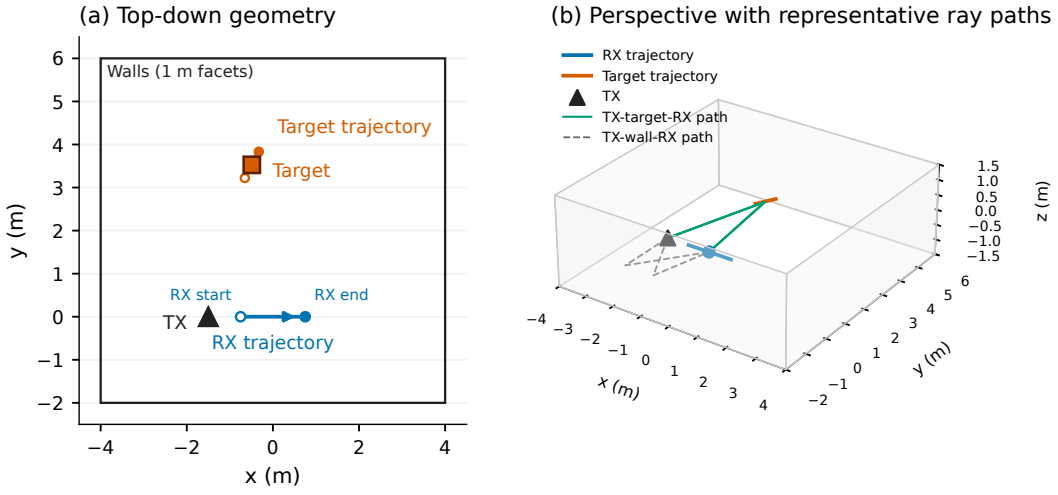


Figure 13: Representative SensingSP/Blender digital-twin geometry used for the cross-simulator validation (stored wall-condition scene, seed 303, index 66). (a) Top-down view: fixed TX, the 1.5 m receiver trajectory, the target trajectory over the 1 s acquisition with its reference state, and the four 1 m-facet walls. (b) Perspective view with three TX–target–RX paths and the two strongest TX–wall–RX paths at the centre snapshot, drawn from the saved ray-tracing output.

Table 9: Digital-twin conditions at 11,520 updates (300 sealed scenes per seed). Vel. is vector velocity RMSE (m/s); folded is the alias-aware range-rate RMSE (m/s); image is mean / median global-peak error (m); fail is the fraction above 0.2 m; alias-like is the fraction with nonzero folded index; the last column lists per-seed mean image error. Each block ends with the paired $\alpha = 0.5$ minus $\alpha = 0$ image difference (4,000 within-seed resamples) and the oracle-motion image error.

Forward model	α	Vel.	Folded	Image	Median	Fail (%)	Alias (%)	Image per seed
ray-traced phases, flat amplitude (fresh noise)	0	0.140	0.219	0.450	0.378	77	0.0	0.459 / 0.451 / 0.440
ray-traced phases, flat amplitude (fresh noise)	0.5	0.312	0.192	0.370	0.296	68	11.7	0.380 / 0.352 / 0.378
paired α=0.5 minus 0: -0.080 m [-0.101, -0.060], per seed -0.079, -0.099, -0.061; oracle 0.031 m								
same-noise amplitude-flat	0	0.140	0.221	0.438	0.364	76	0.0	0.451 / 0.417 / 0.445
same-noise amplitude-flat	0.5	0.315	0.194	0.369	0.302	68	12.0	0.365 / 0.355 / 0.388
paired α=0.5 minus 0: -0.069 m [-0.089, -0.047], per seed -0.086, -0.063, -0.057; oracle 0.031 m								
same-noise, slow-time-only flat	0	0.138	0.214	0.441	0.376	77	0.1	0.454 / 0.440 / 0.429
same-noise, slow-time-only flat	0.5	0.304	0.199	0.371	0.313	70	10.8	0.355 / 0.364 / 0.393
paired α=0.5 minus 0: -0.070 m [-0.090, -0.050], per seed -0.099, -0.077, -0.035; oracle 0.031 m								
same-noise, frequency-only flat	0	0.073	0.082	0.202	0.147	35	0.0	0.190 / 0.221 / 0.195
same-noise, frequency-only flat	0.5	0.056	0.062	0.144	0.110	20	0.0	0.136 / 0.147 / 0.149
paired α=0.5 minus 0: -0.058 m [-0.068, -0.049], per seed -0.054, -0.074, -0.046; oracle 0.031 m								
ray-traced, spreading loss	0	0.077	0.081	0.203	0.154	36	0.0	0.187 / 0.218 / 0.204
ray-traced, spreading loss	0.5	0.058	0.062	0.148	0.118	20	0.0	0.143 / 0.147 / 0.154
paired α=0.5 minus 0: -0.054 m [-0.063, -0.046], per seed -0.044, -0.070, -0.049; oracle 0.031 m								
ray-traced, spreading loss + walls	0	0.150	0.184	0.403	0.326	76	0.2	0.425 / 0.360 / 0.425
ray-traced, spreading loss + walls	0.5	0.163	0.217	0.421	0.327	71	1.1	0.436 / 0.378 / 0.449
paired α=0.5 minus 0: +0.018 m [-0.003, +0.039], per seed +0.011, +0.019, +0.024; oracle 0.066 m								
ray-traced + walls, 100 Hz	0	0.109	0.141	0.321	0.280	65	0.0	0.321
ray-traced + walls, 100 Hz	0.5	0.100	0.124	0.268	0.204	51	0.0	0.268
paired α=0.5 minus 0: -0.053 m [-0.074, -0.030], per seed -0.053; oracle 0.047 m								

Interpretation. The phenomenon is reproduced by an independent ray-traced propagation model when the target echo is amplitude-flattened so that the measurement is phase-dominated, and its alias-like error modes fall at the predicted branch spacing. Because the same-noise controls change only the target magnitude, and the slow-time-only variant restores the reversal while the frequency-only variant does not, the deterministic slow-time amplitude variation of the ray-traced echo supplies an additional disambiguating cue in this digital twin; this supports, without proving, the amplitude

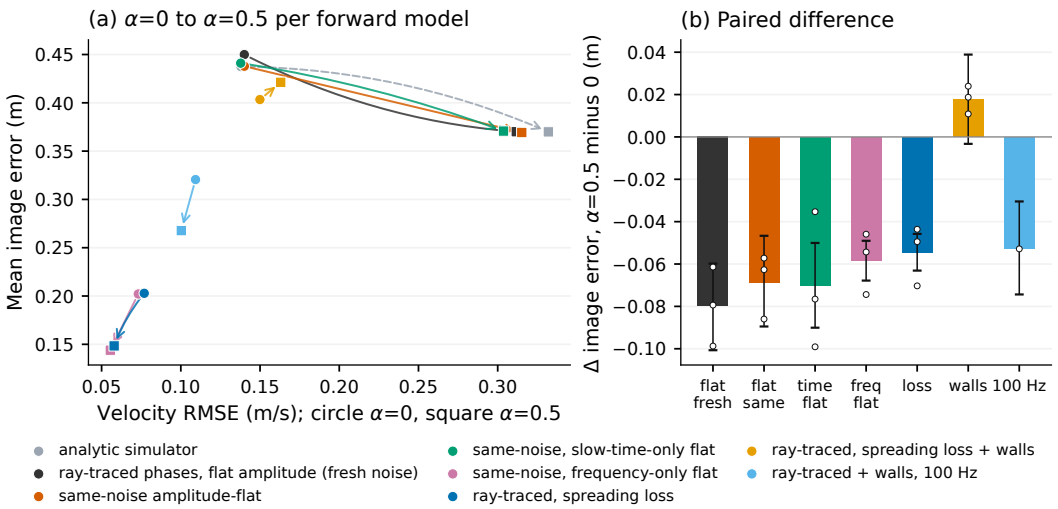


Figure 14: Digital-twin validation. (a) Velocity RMSE against mean image error for the state-only ($\alpha = 0$, circles) and coherence-regularised ($\alpha = 0.5$, squares) CNN under each forward model; the analytic bank of Appendix G is shown for reference. An arrow pointing right and down is a reversal; left and down means both metrics improve. (b) Paired coherence-minus-state-only image difference per condition (95% intervals from 4,000 within-seed resamples; white markers are seeds).

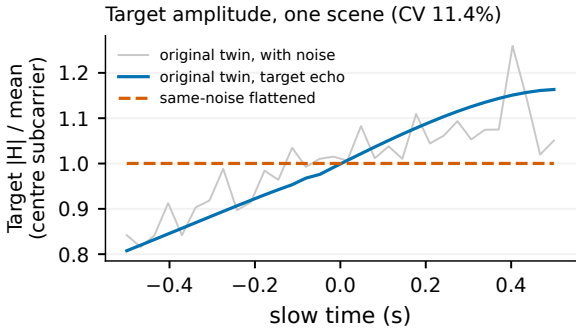


Figure 15: Target-echo amplitude over slow time at the centre subcarrier for one twin scene (no walls): the ray-traced spreading-loss modulation, the same trace with the stored noise, and the same-noise flattened version used in the control.

interpretation. Dense static multipath at 0 dB removes the coherence advantage altogether. Whether measured CSI retains such a clean amplitude cue after gain control and hardware amplitude noise is not tested here.

Limitation. This is cross-simulator validation with one ray tracer, one room, single-scatter facet multipath, an omnidirectional antenna model, and no measured data; it does not prove that multipath preserves or destroys the mechanism in general.