RADARDREAMER: GENERATIVE RADAR WORLD MODEL FROM INSTANTANEOUS DOPPLER

Anonymous authors
Paper under double-blind review

ABSTRACT

Sensor-level world models anticipate future observations for predictive perception, planning, data synthesis, and closed-loop simulation, yet their development has largely centered on visual, occupancy, and geometric spaces. Unlike passive visual observations, a single radar measurement already encodes both spatial structure and instantaneous radial motion, making native radar particularly informative for future observation generation. The challenge is twofold: RA and DA retain spatial and Doppler evidence only as shared-azimuth marginals, obscuring their underlying structure–motion association. Meanwhile, future echoes entangle predictable motion evolution with stochastic variations in visibility, reflectivity, multipath, and measurement noise. Our key insight is to exploit instantaneous Doppler as structured dynamics evidence for stochastic future generation. We present **RadarDreamer**, a conditional rectified-flow model that maps Gaussian radar videos to five coherent future RA+DA pairs. Exact multiband states preserve measurement coefficients, azimuth-indexed coupling exchanges spatial and Doppler evidence while retaining their axis semantics, and radar-aware objectives constrain sparse innovation, temporal evolution, response structure, and projection consistency. On ColoRadar, spatial–Doppler coupling reduces CRPS by 2.16% and improves expected-sample innovation F1 by 10.98% over an architecture-matched generator, while a single RA+DA observation reduces joint log-MAE by 5.86% relative to five-frame RA history. Together with four-domain rollouts, these results demonstrate improved motion fidelity and measurement realism, establishing conditional native-radar generation as a predictive sensing primitive for radar-centric world models.

1 INTRODUCTION

A useful world model should not only anticipate how a scene evolves, but also predict what an embodied agent is likely to sense as that evolution unfolds (Hafner et al., 2020; Wang et al., 2024; Hu et al., 2023). Such sensor-level prediction can provide future evidence for downstream perception and planning, synthesize temporally coherent training observations, and enable closed-loop simulation before real measurements are acquired (Li et al., 2024; Shao et al., 2025). Recent driving world models have therefore expanded beyond image prediction toward occupancy, point clouds, and latent geometric representations (Diehl et al., 2025; Wang et al., 2025b; Song et al., 2026). Radar, however, defines a qualitatively different prediction space. Unlike passive appearance observations, a radar measurement jointly reflects scene structure, sensing geometry, material-dependent response, clutter, and instantaneous radial motion (Zhang et al., 2024; Abu-Alrub & Rawashdeh, 2023; Yeo et al., 2016). In particular, Doppler exposes motion before it becomes evident through substantial spatial displacement, making native radar measurements a potentially informative substrate for predictive sensing (Wang et al., 2025a; Chae et al., 2025). Motivated by this property, we study *native radar future-observation generation*: given one current radar observation, predict a distribution over five future measurements.

Realizing this capability is non-trivial because radar couples partially observed dynamics with stochastic measurement formation. First, **motion is observable, but its spatial association is ambiguous**. Range–azimuth (RA) localizes reflected energy, whereas Doppler–azimuth (DA) reveals the corresponding distribution of instantaneous relative-radial motion. The two projections share the azimuth axis, yet marginalizing the underlying range–Doppler–azimuth volume removes the

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.

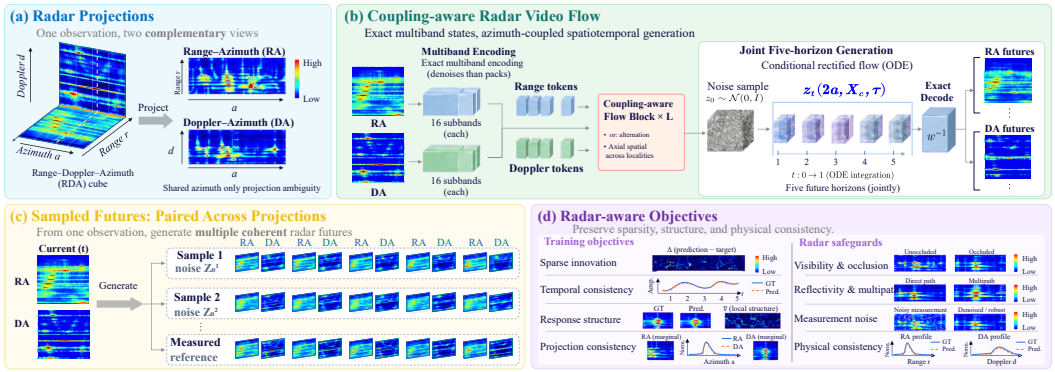

Figure 1: **RadarDreamer generates multiple coherent native-radar futures from one observation.** **(a)** Current RA and DA share azimuth while retaining distinct range and Doppler axes. **(b)** Exact multiband states condition an azimuth-coupled video flow. Independent Gaussian initial states produce paired five-frame futures. **(c)** Radar-aware objectives preserve sparse response and temporal structure. **(d)** Two complete samples from the same RaDelft condition evolve toward plausible future RA/DA measurements.

correspondence between where a reflector lies in range and how it moves in Doppler. Consequently, the same projected observation may support multiple plausible structure–motion associations, leading to different future spatial evolutions. Second, **future radar responses are not uniquely determined by motion**. Even under the same scene dynamics, future measurements can vary with visibility, aspect-dependent reflectivity, multipath, intermittent weak returns, and receiver noise. A deterministic predictor can therefore explain dominant static structure while averaging away sparse but consequential future changes. Native radar prediction must instead preserve motion-consistent evolution while representing the irreducible uncertainty of the sensing process.

Our key insight is that Doppler exposes motion tendency before displacement becomes visible, and its marginal association naturally defines a condition for stochastic future generation. The model preserves the distinct range and Doppler axes, exchanges evidence through their shared azimuth coordinate, and represents the admissible association and sensing variation through multiple coherent futures. Projection ambiguity therefore becomes a design principle: observed radial evidence shapes the conditional distribution, while latent states instantiate its plausible evolutions.

Building on this principle, we present **RadarDreamer**, a conditional radar world model that generates complete five-frame RA+DA sequences (Fig. 1). Exact multiband coding retains radar coefficients, azimuth-indexed cross-attention couples spatial and Doppler evidence, and joint-horizon rectified flow maps Gaussian video states to temporally coherent futures. Radar-aware objectives organize learning around sparse innovation, temporal evolution, response structure, and cross-projection mass. Experiments demonstrate improved distributional scores, stronger future-innovation fidelity, structured use of instantaneous Doppler, and coherent generation across four radar domains.

Our contributions are threefold:

- We present **RadarDreamer**, to our knowledge the first generative world model for native radar future observations, forecasting a distribution over multi-step RA+DA measurements directly in the radar sensing space rather than semantic, geometric, or deterministic surrogate representations.
- We introduce a spatial–Doppler coupled stochastic evolution framework that exploits instantaneous Doppler as structured dynamics evidence. Azimuth-indexed interaction associates spatial echoes with radial-motion cues while preserving the distinct range and Doppler semantics, enabling motion-consistent joint generation across future horizons.
- We establish the benefit of Doppler-conditioned distributional prediction through controlled architecture-matched comparisons, spatial-history baselines, horizon-wise analysis, and four-domain evaluation. The results demonstrate improved distributional fidelity, sparse-motion evolution, and radar measurement realism.

2 Related Work

Radar representations and Doppler-aware perception. Radar learning operates on representations with substantially different levels of abstraction, including thresholded detections, projected spectral maps, and dense range–Doppler–azimuth tensors (Kramer et al., 2022; Paek et al., 2022; Yao et al., 2025). Detection-level point sets provide an efficient interface for object-centric perception, but sparsification and thresholding suppress weak, diffuse, and spatially extended responses (Zhang & Gao, 2025). Dense radar tensors preserve the full coupling among range, azimuth, and Doppler, yet their high dimensionality makes long-horizon modeling computationally demanding. Projected heatmaps offer an intermediate representation: range–azimuth (RA) preserves dense spatial support and response intensity, whereas Doppler–azimuth (DA) retains the distribution of instantaneous relative-radial motion at each azimuth. Beyond representation design, Doppler has been extensively exploited for ego-motion estimation, scene flow, moving-object reasoning, and motion-aware temporal aggregation (Lu et al., 2020; Ding et al., 2023; Haitman & Bialer, 2025). These methods demonstrate that radial velocity provides informative dynamics cues that cannot be recovered reliably from sparse spatial displacement alone. However, they primarily use Doppler to estimate the current state or improve a deterministic downstream prediction; they do not study how instantaneous Doppler constrains a distribution over future native radar measurements.

Predictive driving models and future observations. Predictive modeling for autonomous systems has progressed from explicit state propagation to generative world modeling. Dynamic occupancy grids, Doppler-aware radar polygons, and learned bird's-eye-view predictors propagate occupied regions or motion states through time (Hoermann et al., 2018; Gao et al., 2022; Shi et al., 2024). More recent driving world models forecast occupancy fields, point clouds, and latent geometric representations, enabling scalable prediction of scene structure and agent evolution (Diehl et al., 2025; Wang et al., 2025b; Song et al., 2026). In parallel, recurrent and diffusion-based video generators model temporally coherent appearance and support stochastic future sampling in the image domain (Shi et al., 2015; Ho et al., 2022; Voleti et al., 2022). These directions establish the value of predicting future scene states or visual observations, but neither directly captures the formation of native radar measurements. Semantic occupancy and point clouds abstract away response amplitude, weak-return intermittency, clutter, and multipath; video models, by contrast, are designed around passive appearance formation rather than active sensing and Doppler-resolved motion. Consequently, forecasting the physical scene does not uniquely specify the radar measurement that will be observed from that scene. Native radar future generation must predict both motion-consistent scene evolution and the sensor-dependent variability through which that evolution is measured.

Radar simulation, generation, and stochastic dynamics. Radar synthesis has been studied through physics-based simulation, data-driven generation, and neural scene reconstruction. Existing simulators and learned generators synthesize radar waveforms, projected heatmaps, detection-level points, or dense spectral tensors (Schüßler et al., 2021; Fidelis et al., 2023; Wang et al., 2026; Shen et al., 2026). Physics-oriented simulators can explicitly approximate propagation and sensor formation, but often require detailed scene geometry, material parameters, calibration, or costly signal-level computation. Learned generators reduce this burden by approximating the radar response distribution from paired data or structured conditions; however, most formulations focus on producing an individual radar observation, translating between sensing domains, or matching a static measurement distribution. Neural radar representations such as DART and RF4D further support scene reconstruction and novel-view rendering (Huang et al., 2024; Zhang et al., 2026), but their central objective is view-consistent reconstruction rather than stochastic temporal rollout from a current radar observation. In probabilistic generation, diffusion models and flow matching provide general mechanisms for learning complex conditional distributions (Ho et al., 2020; Lipman et al., 2022). Directly applying these mechanisms to radar, however, does not by itself account for the distinct semantics of range and Doppler, the partial association induced by RA/DA projection, or the need to maintain coherent evolution across multiple future horizons.

3 Method

Figure 1 summarizes the pipeline. A synchronized RA/DA observation is normalized and encoded into separate, exactly invertible multiband packets. A dual-stream vector field exchanges evidence

only along shared azimuth and jointly transports a Gaussian five-frame video. Exact decoding returns paired RA/DA sequences. The training-time endpoint estimate supplies radar-aware supervision, while independent Gaussian initial states produce distinct futures at inference. We first formalize the target and then specify these three computational stages.

3.1 Problem Formulation

Let $S_t \in \mathbb{R}_{\geq 0}^{R\times D\times A}$ denote a processed range–Doppler–azimuth power cube. Its range–azimuth (RA) and Doppler–azimuth (DA) observations are the marginals

$$X_t^{\mathrm{RA}}(r,a) = \sum_d S_t(r,d,a), \qquad X_t^{\mathrm{DA}}(d,a) = \sum_r S_t(r,d,a). \tag{1}$$

Their first axes represent different physical variables, while azimuth is shared. In linear power, both projections preserve the same azimuth-wise mass:

$$\sum_r X_t^{\mathrm{RA}}(r,a) = \sum_d X_t^{\mathrm{DA}}(d,a) = \sum_{r,d} S_t(r,d,a). \tag{2}$$

We robustly normalize log-power using training-set medians μ_m and interquartile ranges q_m,

$$\widetilde{X}^m = \mathrm{clip}\left(\frac{\log(1+X^m)-\mu_m}{q_m}, -8, 8\right), \qquad m \in \{\mathrm{RA}, \mathrm{DA}\}. \tag{3}$$

Given one paired observation, RadarDreamer learns

$$p_\theta\left(X_{t+1:t+5}^{\mathrm{RA,DA}} \mid X_t^{\mathrm{RA}}, X_t^{\mathrm{DA}}\right), \tag{4}$$

so that each sample represents one coherent five-frame RA+DA future.

The distributional formulation follows from projected association. For one azimuth, let $u \in \mathbb{R}_+^R$ and $v \in \mathbb{R}_+^D$ be the observed RA and DA columns. Their compatible range–Doppler couplings form

$$\mathcal{C}(u,v) = \{Q \geq 0 : Q\mathbf{1}_D = u,\ Q^\top \mathbf{1}_R = v\}. \tag{5}$$

Proposition 1 (Projected-Doppler ambiguity)**.** *If $R, D \geq 2$ and u, v are strictly positive with equal total mass, $\mathcal{C}(u,v)$ has dimension $(R-1)(D-1)$. If radial transports T_d differ across Doppler bins, two couplings can share the same RA/DA marginals yet induce different future RA observations.*

Appendix A.1 gives the constructive proof. RadarDreamer therefore conditions evolution on the observed marginals and represents the induced family of spatial–Doppler associations through its stochastic latent state.

3.2 Exact Multiband Spatial–Doppler Conditioning

Exact multiband radar state. A future generator must retain weak and high-frequency radar responses. We therefore apply a two-level orthonormal Haar packet transform $\mathcal{W}$ separately to RA and DA. Each 128×128 map becomes 16 signed subbands on a 32×32 grid, with

$$\mathcal{W}^{-1}(\mathcal{W}(X)) = X, \qquad \|\mathcal{W}(X)\|_2 = \|X\|_2. \tag{6}$$

For normalized targets $Y_{1:5}^m = (\widetilde{X}_{t+h}^m)_{h=1}^5$, the future state is

$$Z_1 = \mathrm{Std}_{\mathrm{train}}\left([\mathcal{W}(Y_{1:5}^{\mathrm{RA}}), \mathcal{W}(Y_{1:5}^{\mathrm{DA}})]\right) \in \mathbb{R}^{5\times 32\times 32\times 32}. \tag{7}$$

The transform preserves every coefficient, and RA/DA coefficients remain in separate channel groups throughout the generator.

Shared-azimuth association. The two streams combine noisy future packets with the current RA/DA condition, flow time, position, and horizon embeddings. At each horizon, states $H_R, H_D \in \mathbb{R}^{C\times 32\times A'}$ exchange evidence only within their common azimuth a:

$$A_{R\leftarrow D}^a = \mathrm{softmax}\left(\frac{Q_R(H_R^a)K_D(H_D^a)^\top}{\sqrt{C}}\right), \qquad \bar{H}_R^a = H_R^a + g_R A_{R\leftarrow D}^a V_D(H_D^a), \tag{8}$$

$$A_{D\leftarrow R}^a = \mathrm{softmax}\left(\frac{Q_D(H_D^a)K_R(H_R^a)^\top}{\sqrt{C}}\right), \qquad \bar{H}_D^a = H_D^a + g_D A_{D\leftarrow R}^a V_R(H_R^a). \tag{9}$$

Channel-wise gates g_R, g_D modulate the exchanged evidence. Azimuth indexes physically admissible interaction, whereas attention learns a soft association between distinct range and Doppler token sets. Since H_R and H_D depend on the sampled flow state, the association can change across plausible futures.

This exchange respects the two axes algebraically. For features in token-row form and independent permutations P_R, P_D,

$$\mathcal{A}(P_R H_R, P_D H_D) = P_R \mathcal{A}(H_R, H_D), \tag{10}$$

where $\mathcal{A}$ denotes the RA update in Eq. equation 9. The exchange is therefore equivariant to independent range and Doppler reindexing and preserves their axis semantics (Appendix A.12). Axis-aware filtering, axial attention, and five-horizon attention then organize spatial and temporal evolution within each draw.

3.3 Stochastic Future Evolution

For every future Z_1, we sample a Gaussian base video $Z_0 \sim \mathcal{N}(0, I)$ and time $\tau \sim \mathcal{U}(0, 1)$. Rectified flow defines the interpolant and target velocity

$$Z_\tau = (1-\tau)Z_0 + \tau Z_1, \qquad U_\tau = Z_1 - Z_0, \tag{11}$$

and trains the dual-stream network with

$$v_\theta(Z_\tau, X_t^{\mathrm{RA}}, X_t^{\mathrm{DA}}, \tau), \qquad \mathcal{L}_{\mathrm{FM}} = \mathbb{E}\|v_\theta - U_\tau\|_2^2. \tag{12}$$

At inference, each initial video is transported by

$$\frac{dZ_\tau}{d\tau} = v_\theta(Z_\tau, X_t, \tau), \qquad Z_0 \sim \mathcal{N}(0, I), \tag{13}$$

and decoded with $\mathcal{W}^{-1}$. Different initial states yield different complete radar futures while preserving a common observed condition.

The connection between projection ambiguity and distributional evaluation is explicit. Suppose two compatible couplings induce distinct futures y_+, y_- with equal conditional probability, giving $P_c = (\delta_{y_+} + \delta_{y_-})/2$.

Corollary 1 (Energy-score advantage under two-mode ambiguity)**.** *For the lower-is-better population energy score* S_E*,*

$$\inf_{\hat{y}} \mathbb{E}_{Y\sim P_c} \mathsf{S}_E(\delta_{\hat{y}}, Y) - \mathbb{E}_{Y\sim P_c} \mathsf{S}_E(P_c, Y) = \tfrac{1}{4}\|y_+ - y_-\|_2 > 0. \tag{14}$$

Appendix A.12 proves the result and its finite-ensemble form, quantifying the distributional advantage of representing the admissible future support.

3.4 Radar-Aware Training Objectives

Flow matching learns the future distribution globally, while radar responses concentrate information in sparse, structured regions. We expose a differentiable clean-endpoint estimate

$$\widehat{Z}_1 = Z_\tau + (1-\tau)v_\theta(Z_\tau, X_t, \tau), \qquad \widehat{Y} = \mathcal{W}^{-1}(\mathrm{Destd}(\widehat{Z}_1)), \tag{15}$$

and use a soft signal-and-innovation weight w derived from the current and target observations. The radar-aware objective is

$$\begin{aligned}\mathcal{L}_{\mathrm{rad}} = \|\widehat{Y} - Y\|_1 + \lambda_{\mathrm{inv}} \frac{\sum w|\widehat{Y} - Y|}{\sum w + \epsilon} + \lambda_t \|\nabla_t \widehat{Y} - \nabla_t Y\|_1 \\ + \lambda_s \|\nabla_s \widehat{Y} - \nabla_s Y\|_1 + \lambda_m \mathcal{L}_{\mathrm{mass}}(\widehat{Y}),\end{aligned} \tag{16}$$

where the five terms preserve full-frame response, sparse innovation, temporal change, spatial structure, and paired-projection consistency. After undoing log normalization, let $m_{h,a}^{\mathrm{RA}}$ and $m_{h,a}^{\mathrm{DA}}$ be the range- and Doppler-integrated linear powers. We define

$$\mathcal{L}_{\mathrm{mass}} = \frac{1}{5A} \sum_{h,a} \frac{2|m_{h,a}^{\mathrm{RA}} - m_{h,a}^{\mathrm{DA}}|}{m_{h,a}^{\mathrm{RA}} + m_{h,a}^{\mathrm{DA}} + \epsilon}. \tag{17}$$

This term has a direct geometric interpretation.

Table 1: **Conditional generation on ColoRadar.** Matched models use $K = 8$ samples and 16 ODE steps; sample and best-of-8 F1 measure typical draws and sampled support.

Method	Distribution ↓ CRPS	Distribution ↓ Energy	Innovation F1 ↑ Sample	Innovation F1 ↑ Best-of-8
Independent	0.30878	0.50237	0.16372	0.23890
RadarDreamer	**0.30212**	**0.48943**	**0.18169**	**0.25859**
Gain	▲ 2.16%	▲ 2.58%	▲ 1.80 pp	▲ 1.97 pp

Table 3: **Instantaneous Doppler improves future RA innovation.** Metrics compare one current RA+DA observation with five-frame RA history.

Method	Energy F1 ↑	Change-wtd. MAE ↓	Innov. F1 ↑
5-frame RA	0.9018	0.6907	0.3179
RA+DA	**0.9052** ▲ 0.34 pp	**0.6775** ▲ 1.91%	**0.3638** ▲ 4.59 pp

Lemma 1 (Distance to compatible projections). *For $u \in \mathbb{R}^R_{\geq 0}$ and $v \in \mathbb{R}^D_{\geq 0}$,*

$$\inf_{\substack{\bar{u},\bar{v}\geq 0 \\ \mathbf{1}^\top \bar{u} = \mathbf{1}^\top \bar{v}}} \left(\|u - \bar{u}\|_1 + \|v - \bar{v}\|_1 \right) = |\mathbf{1}^\top u - \mathbf{1}^\top v|. \tag{18}$$

Hence Eq. equation 17 is the normalized distance to a pair of marginals compatible with some nonnegative joint. The full objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{FM}} + \lambda_{\mathrm{rad}} \mathcal{L}_{\mathrm{rad}}. \tag{19}$$

The lemma and the decoded endpoint-error bounds are proved in Appendix A.12.

4 Experiments

We study conditional generation quality, the predictive value of instantaneous Doppler, and the physical structure of sampled radar futures.

Data. ColoRadar single-chip (Kramer et al., 2022) is divided by sequence into 28/4/8 train/validation/test runs (39,349/5,755/11,527 frames). Each 128^3 cube yields paired 128×128 RA/DA projections and timestamp-valid five-frame targets. At the nominal 10 Hz rate, $h = 1, \ldots, 5$ spans approximately 0.1-0.5 s. Generative models are compared on the validation protocol. Point baselines use 11,400 anchors from eight sequence-disjoint test runs. K-Radar (Paek et al., 2022), RaDelft (Roldan et al., 2024), and ColoRadar cascade extend the analysis across radar configurations.

Models and training. All matched comparisons share resolution, targets, and optimization budgets. RadarDreamer draws $K = 8$ complete futures with a 16-step Euler solver. The generative comparison varies spatial–Doppler coupling. Point baselines comprise persistence, current RA, three/five-frame RA, and dense or innovation-balanced RA+DA predictors. Architecture and optimization details appear in Appendix A.10.

Metrics. CRPS and multivariate energy score evaluate conditional distributions. Expected sample quality, $K^{-1} \sum_k m(Y^{(k)}, Y)$, measures the typical generated future, while best-of-K characterizes sampled support. Log-MAE, high-energy F1, and bandpass correlation measure response realism. Change-weighted MAE and innovation metrics measure future observation dynamics. Definitions appear in Appendices A.3–A.4.

4.1 Comparison with Baselines

Distributional generation. Table 1 shows that adding spatial-Doppler coupling reduces CRPS by 2.16% and energy score by 2.58%. Expected innovation F1 rises from 0.16372 to 0.18169, a 10.98% relative gain, while best-of-8 improves by 1.97 points. The improvement in both expected and oracle sample metrics shows that spatial–Doppler coupling strengthens typical futures and broadens the useful support of the generated distribution.

Predictive information in Doppler. The point baselines isolate input information independently of stochastic modeling. Relative to five-frame RA, one current RA+DA observation reduces joint

Table 2: **Point-forecast baselines on 11,400 test anchors from eight unseen sequences.** Blue columns measure response realism and green columns measure future innovation. Parentheses in the final row give favorable change relative to five-frame RA. "pp" denotes percentage points.

		Measurement realism			Innovation fidelity		
Method	Input	Log-MAE ↓	Energy F1 ↑	Band corr. ↑	Change-wtd. MAE ↓	Innov. cos. ↑	Innov. F1 ↑
Persistence	RA+DA	0.4261	0.8096	0.8605	1.1453	0.0000	0.0000
RA only	1 RA	0.3895	0.8215	0.9050	0.9818	0.4787	0.2574
3-frame RA	3 RA	0.3382	0.8375	0.9106	0.8285	0.5639	0.2845
5-frame RA	5 RA	0.3370	0.8399	0.9103	0.8183	0.5718	0.2852
RA+DA dense	1 RA+DA	**0.3136**	**0.8482**	**0.9202**	0.7685	0.5981	0.2629
RA+DA balanced	1 RA+DA	0.3195	0.8439	0.9123	0.7565	**0.6157**	**0.3258**
RA+DA calibrated	1 RA+DA	0.3173 ▲ 5.86%	0.8458 ▲ 0.59 pp	0.9157 ▲ 0.55 pp	**0.7562** ▲ 7.59%	0.6132 ▲ 4.14 pp	0.3129 ▲ 2.77 pp

log-MAE by 5.86% and change-weighted MAE by 7.59%, while increasing innovation F1 by 2.77 points (Table 2). Sequence-bootstrap intervals for all five reported favorable differences exclude zero (Appendix A.5). Thus instantaneous Doppler can compensate for substantial spatial history in joint future-observation prediction.

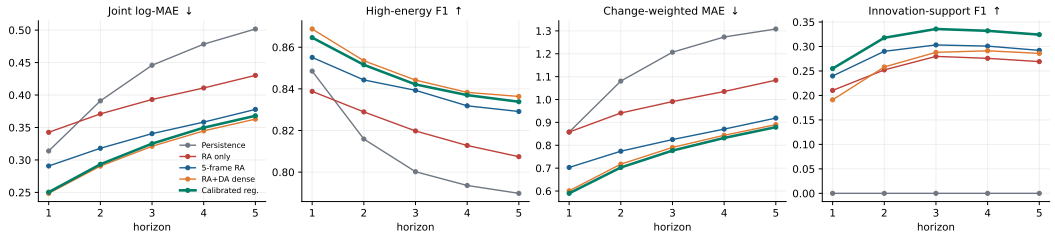

Figure 2: **Future-observation quality across horizons.** Metrics are averaged over eight unseen sequences. Instantaneous RA+DA sustains lower joint and change-weighted error across the forecast range, while innovation-balanced learning strengthens sparse response evolution.

Long-horizon behavior. Figure 2 shows a consistent horizon-wise advantage from instantaneous Doppler. RA+DA improves change-weighted error over persistence at every horizon, averaging 33.32%, compared with 26.96% for five-frame RA. This persistent advantage shows that Doppler remains informative throughout the complete forecast interval rather than benefiting only the first predicted frame.

4.2 Ablation Studies

Spatial–Doppler coupling. The matched comparison in Table 1 isolates the bidirectional shared-azimuth exchange in Eq. equation 9. The independent-stream variant retains the same generator while removing only this exchange, so the differences isolate the coupling mechanism. Coupling reduces CRPS and energy score by 2.16% and 2.58%, while improving expected-sample and best-of-eight innovation F1 by 1.80 and 1.97 points. The concurrent gains in proper distribution scores and sparse innovation localization show that the exchange does more than broaden sample variation. It produces futures that better match the conditional distribution and its motion-sensitive support. Shared azimuth is therefore an effective physical interface for coupling RA and DA without conflating their distinct range and Doppler axes.

Output-specific effect. Table 3 isolates future RA. Relative to five-frame RA history, the single-frame RA+DA condition reduces change-weighted MAE by 1.91% and improves energy F1 and innovation F1 by 0.34 and 4.59 points. Because every metric excludes future DA, these gains cannot be explained by direct DA-to-DA prediction. They instead show that instantaneous radial-motion structure improves the predicted evolution of the spatial RA measurement itself. This result supports the central premise of RadarDreamer: Doppler acts as an early state variable for future spatial

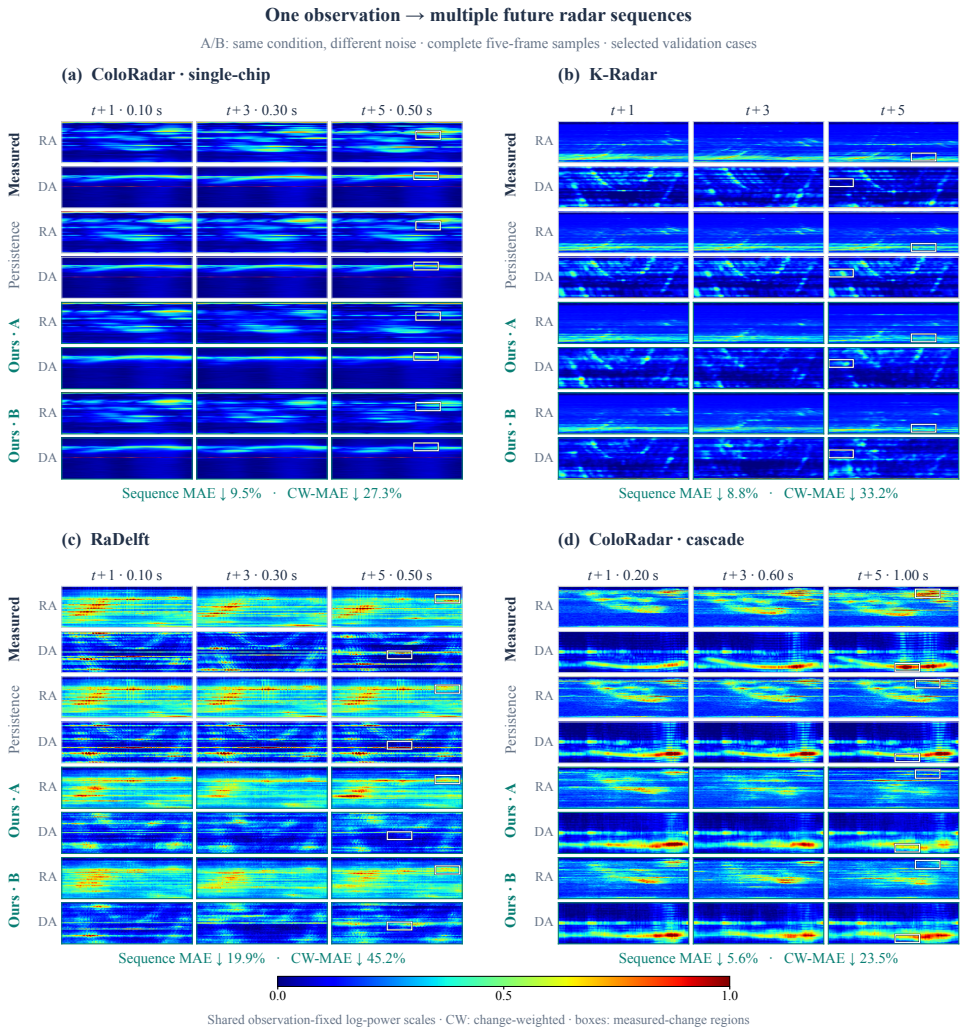

Figure 3: **Conditional future generation across four radar domains.** Measured futures, persistence, and two complete RadarDreamer draws are shown at $h = 1, 3, 5$. A/B share the observation and model but use different Gaussian initial states. Complete sequences and aggregate statistics appear in Appendix A.11.

response generation before equivalent displacement is fully expressed in RA history. The complete metric decomposition appears in Appendix A.5.

4.3 Predictive Contribution of Doppler Structure

To identify how Doppler enters the forecast, we transform its temporal alignment and local DA correspondence while preserving measured values or their histogram. Table 4 shows that a one-frame shift increases joint error by 2.94%, while RA-similar exchange and histogram-preserving block shuffle increase it by 136.36% and 118.85%. These responses isolate temporally aligned, locally organized Doppler evidence as predictive structure beyond the global intensity distribution.

Together with the coupled-flow ablation, these results establish structured instantaneous Doppler as an effective conditioning variable for the future-measurement distribution.

4.4 Analysis of Stochasticity and Physical Consistency

Coherent stochastic evolution. In Fig. 3, **A and B are independent samples from the same model and current RA+DA condition**. Each draw generates all five horizons jointly, preserving major

Table 4: **Sensitivity to structured DA interventions** on 5,701 validation anchors. Δ Joint is relative to aligned measured DA.

DA intervention	Joint MAE ↓	RA MAE ↓	Δ Joint (%)
Aligned measured DA	**0.3396**	**0.3486**	—
Shifted DA ($t-1$)	0.3496	0.3495	+2.94[a]
RA-matched DA swap	0.8027	0.5083	+136.36[b]
8×8 block shuffle	0.7432	0.4334	+118.85[c]

Sequence-bootstrap 95% intervals: [a][2.47, 3.31], [b][121.45, 155.01], [c][95.60, 136.58]. The swap uses cross-sequence nearest-RA matches (median cosine 0.9878); block shuffle exactly preserves the DA histogram.

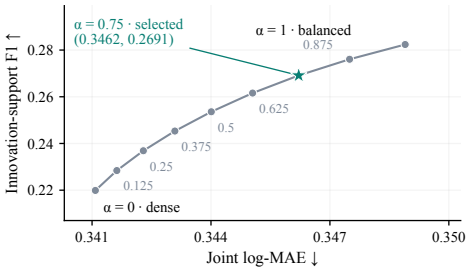

Figure 4: **Response–innovation trade-off.** At $\alpha = 0.75$, innovation F1 improves by 4.92 points over the dense endpoint ($\alpha = 0$).

RA ridges and DA bands while expressing distinct weak-response evolution. Across the displayed single-chip/K-Radar/RaDelft/cascade sequences, expected-sample joint-MAE improves over persistence by 9.5/8.8/19.9/5.6%. Complete sequences and aggregate statistics appear in Appendix A.11. The consistent advantage across these acquisition domains indicates that RadarDreamer captures transferable radar-future structure while retaining sample-dependent evolution.

Distribution quality and consistency. RadarDreamer combines improved proper scores with log-domain sample deviation 0.42113 and expected spatial-TV ratio 0.91832, preserving structured variability at the native measurement scale. Projection consistency is evaluated in linear power through the shared azimuth-wise mass of paired RA/DA columns, directly matching Eq. equation 2. Appendix A.4 reports the complete distributional and physical-consistency evaluation. Together, these measurements show that the model improves distributional accuracy while retaining radar-specific spatial statistics and projection geometry.

Association-conditioned trajectories. The quantitative and qualitative results jointly characterize the role of spatial–Doppler coupling. Table 1 shows that the coupled model improves both the expected draw and the support represented by multiple draws. Figure 3 further shows that sample variation is expressed through radar responses—ridge intensity, local support, and Doppler branches—while the dominant scene structure remains anchored to the observation. Each latent draw defines a joint element of $p_\theta(X^{\mathrm{RA,DA}}_{t+1:t+5} \mid X^{\mathrm{RA}}_t, X^{\mathrm{DA}}_t)$, coupling both projections and all horizons. The shared-azimuth association therefore acts as a sample-dependent conditioning field, generating spatial ridges, radial-motion bands, and weak-return evolution as one radar trajectory.

Sensor-level world-model behavior. The four radar domains differ in acquisition configuration and response statistics, yet the generated sequences share the same qualitative signature: stable scene-conditioned structure together with evolving sparse returns. This cross-domain behavior supports the central formulation of radar forecasting as conditional observation generation. RadarDreamer thus models the future in the native coordinates consumed by radar perception, retaining motion-sensitive measurement structure throughout the generated trajectory.

Response–innovation geometry. Figure 4 traces a smooth operating trajectory between dense response accuracy and sparse innovation localization. The selected point improves change-weighted MAE from 0.7971 to 0.7864 and raises innovation F1 by 4.92 points relative to the dense endpoint. This geometry supports the radar-aware objective design and shows that explicit innovation modeling provides a favorable operating point for predictive sensing.

5 Conclusion

RadarDreamer models native-radar futures as conditional distributions. Exact multiband states preserve measurement structure, shared-azimuth coupling injects instantaneous Doppler into joint-horizon rectified flow, and radar-aware objectives guide sparse response evolution. Coupling improves proper scores and sample innovation, while one RA+DA frame outperforms five-frame RA history on joint metrics. RadarDreamer thereby establishes native-radar generation as a predictive sensing primitive.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing, figure refinement, and limited assistance with code debugging and LaTeX formatting. We did not use generative AI to generate experimental data or results. All AI-assisted content was reviewed, revised, and verified by the authors, who take full responsibility for the final content of this work.

REFERENCES

Nader J Abu-Alrub and Nathir A Rawashdeh. Radar odometry for autonomous ground vehicles: A survey of methods and datasets. *IEEE Transactions on Intelligent Vehicles*, 9(3):4275–4291, 2023.

Yujeong Chae, Heejun Park, Hyeonseong Kim, and Kuk-Jin Yoon. Doppler-aware lidar-radar fusion for weather-robust 3d detection. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27197–27208. IEEE, 2025.

Christopher Diehl, Quinlan Sykora, Ben Agro, Thomas Gilles, Sergio Casas, and Raquel Urtasun. DIO: Decomposable implicit 4d occupancy-flow world model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27456–27466, 2025.

Fangqiang Ding, Andras Palffy, Dariu M. Gavrila, and Chris Xiaoxuan Lu. Hidden gems: 4d radar scene flow learning using cross-modal supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9340–9349, 2023.

Eduardo C Fidelis, Fabio Reway, Herick Ribeiro, Pietro L Campos, Werner Huber, Christian Icking, Lester A Faria, and Torsten Schön. Generation of realistic synthetic raw radar data for automated driving applications using generative adversarial networks. *arXiv preprint arXiv:2308.02632*, 2023.

Xiangyu Gao, Sihao Ding, and Harshavardhan Dasari Reddy. Deformable radar polygon: A lightweight and predictable occupancy representation for short-range collision avoidance. *arXiv preprint arXiv:2203.01442*, 2022.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. 2020. URL https://arxiv.org/pdf/1912.01603.pdf.

Yuval Haitman and Oded Bialer. Doppdrive: Doppler-driven temporal aggregation for improved radar object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 26085–26094, 2025.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J. Fleet. Video diffusion models. In *Advances in Neural Information Processing Systems*, 2022.

Stefan Hoermann, Martin Bach, and Klaus Dietmayer. Dynamic occupancy grid prediction for urban autonomous driving: A deep learning approach with fully automatic labeling. In *IEEE International Conference on Robotics and Automation*, pp. 2056–2063, 2018.

Anthony Hu, Lloyd Russell, Hudson Yeo, Zak Murez, George Fedoseev, Alex Kendall, Jamie Shotton, and Gianluca Corrado. Gaia-1: A generative world model for autonomous driving. *arXiv preprint arXiv:2309.17080*, 2023.

Tianshu Huang, John Miller, Akarsh Prabhakara, Tao Jin, Tarana Laroia, Zico Kolter, and Anthony Rowe. DART: Implicit doppler tomography for radar novel view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24118–24129, 2024.

Andrew Kramer, Kyle Harlow, Christopher Williams, and Christoffer Heckman. Coloradar: The direct 3d millimeter wave radar dataset. *The International Journal of Robotics Research*, 41(4): 351–360, 2022.

Xinhai Li, Jialin Li, Ziheng Zhang, Rui Zhang, Fan Jia, Tiancai Wang, Haoqiang Fan, Kuo-Kun Tseng, and Ruiping Wang. Robogsim: A real2sim2real robotic gaussian splatting simulator. *arXiv preprint arXiv:2411.11839*, 2024.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The eleventh international conference on learning representations*, 2022.

Chris Xiaoxuan Lu, Muhamad Risqi U Saputra, Peijun Zhao, Yasin Almalioglu, Pedro PB De Gusmao, Changhao Chen, Ke Sun, Niki Trigoni, and Andrew Markham. milliego: single-chip mmwave radar aided egomotion estimation via deep sensor fusion. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pp. 109–122, 2020.

Dong-Hee Paek, Seung-Hyun Kong, and Kevin Tirta Wijaya. K-radar: 4d radar object detection for autonomous driving in various weather conditions. *Advances in Neural Information Processing Systems*, 35:3819–3829, 2022.

Ignacio Roldan, Andras Palffy, Julian F. P. Kooij, Dariu M. Gavrila, Francesco Fioranelli, and Alexander Yarovoy. A deep automotive radar detector using the radelft dataset. *IEEE Transactions on Radar Systems*, 2:1062–1075, 2024. doi: 10.1109/TRS.2024.3485578.

Christian Schüßler, Marcel Hoffmann, Johanna Bräunig, Ingrid Ullmann, Randolf Ebelt, and Martin Vossiek. A realistic radar ray tracing simulator for large mimo-arrays in automotive environments. *IEEE Journal of Microwaves*, 1(4):962–974, 2021.

Yidi Shao, Mu Huang, Chen Change Loy, and Bo Dai. Gaussim: Foreseeing reality by gaussian simulator for elastic objects. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7841–7850. IEEE, 2025.

Dazhong Shen, Jingjing Gu, Qiang Zhou, Meng Zhao, and Ying Sun. LiDAR-to-4DRadar diffusion bridge via cross-modal alignment and translation in latent space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17110–17120, 2026.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, 2015.

Yining Shi, Kun Jiang, Ke Wang, Jiusi Li, Yunlong Wang, Mengmeng Yang, and Diange Yang. Streamingflow: Streaming occupancy forecasting with asynchronous multi-modal data streams via neural ordinary differential equation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14833–14842, 2024.

Jingyu Song, Yi Liu, and Katherine A. Skinner. CRISP: A spatiotemporal camera–radar backbone for driving via forecasting-based world-model pretraining. *arXiv preprint arXiv:2607.04541*, 2026.

Vikram Voleti, Alexia Jolicoeur-Martineau, and Chris Pal. MCVD: Masked conditional video diffusion for prediction, generation, and interpolation. In *Advances in Neural Information Processing Systems*, 2022.

Shengpeng Wang, Xin Luo, Yulong Xie, and Wei Wang. SDDiff: Boosting radar perception via spatial-doppler diffusion. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 8805–8813, 2025a. doi: 10.24963/ijcai.2025/979.

Xiaofeng Wang, Zheng Zhu, Guan Huang, Xinze Chen, Jiagang Zhu, and Jiwen Lu. Drivedreamer: Towards real-world-drive world models for autonomous driving. In *European conference on computer vision*, pp. 55–72. Springer, 2024.

Yuping Wang, Xiangyu Huang, Xiaokang Sun, Mingxuan Yan, Shuo Xing, Zhengzhong Tu, and Jiachen Li. UniOcc: A unified benchmark for occupancy forecasting and prediction in autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 25560–25570, 2025b.

Zhaoze Wang, Changxu Zhang, Tai Fei, Christopher Grimm, Yi Jin, Claas Tebruegge, Ernst Warsitz, and Markus Gardill. Synthetic FMCW radar range azimuth maps augmentation with generative diffusion model, 2026.

Shanliang Yao, Runwei Guan, Zitian Peng, Chenhang Xu, Yilu Shi, Weiping Ding, Eng Gee Lim, Yong Yue, Hyungjoon Seo, Ka Lok Man, et al. Exploring radar data representations in autonomous driving: A comprehensive review. *IEEE Transactions on Intelligent Transportation Systems*, 2025.

Hui-Shyong Yeo, Gergely Flamich, Patrick Schrempf, David Harris-Birtill, and Aaron Quigley. Radarcat: Radar categorization for input & interaction. In *Proceedings of the 29th annual symposium on user interface software and technology*, pp. 833–841, 2016.

Jiarui Zhang, Zhihao Li, Chong Wang, and Bihan Wen. RF4D: Neural radar fields for novel view synthesis in outdoor dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15387–15397, 2026.

Ruibin Zhang and Fei Gao. Sem-radiff: Diffusion-based 3d radar semantic perception in cluttered agricultural environments. *arXiv preprint arXiv:2509.02283*, 2025.

Ruibin Zhang, Donglai Xue, Yuhan Wang, Ruixu Geng, and Fei Gao. Towards dense and accurate radar perception via efficient cross-modal diffusion model. *IEEE Robotics and Automation Letters*, 9(9):7429–7436, 2024.

Appendix Contents

A.1 Theory. Projected-Doppler non-identifiability proof . . . . . p. 13
A.2 Radar physics. FMCW observation, ego confounding, and radial transport . . . . . p. 14
A.3 Metrics. Measurement and innovation definitions . . . . . p. 15
A.4 Probabilistic evaluation. Proper scores, sample quality, and calibration . . . . . p. 15
A.5 Additional metrics. Structure and temporal statistics . . . . . p. 16
A.6 Robustness. Objective trade-offs, thresholds, mass, and ego speed . . . . . p. 17
A.7 Timing. Physical horizon durations . . . . . p. 18
A.8 Ambiguity cases. Real projected-neighbor examples and full-RDA coupling . . . . . p. 18
A.9 Alternative formulations. Stochastic, support, and transport variants . . . . . p. 18
A.10 Reproducibility. Architecture, training, and hashes . . . . . p. 18
A.11 Extended sequences. Full rollouts, aggregate statistics, and boundary cases . . . . . p. 22
A.12 Design theory. Axis structure, marginal feasibility, and distributional ambiguity . . . . . p. 25
A.13 Responsible use. Deployment context and ethical considerations . . . . . p. 27

A Supplementary Material

A.1 Proof of Projected-Doppler Non-Identifiability

We prove Proposition 1 for one azimuth. The ambiguity applies independently to every azimuth column. Define the transportation polytope

$$\mathcal{C}(u,v) = \left\{ Q \in \mathbb{R}_+^{R\times D} : Q\mathbf{1}_D = u,\ Q^\top \mathbf{1}_R = v \right\}, \qquad \mathbf{1}_R^\top u = \mathbf{1}_D^\top v = M > 0. \tag{20}$$

Strict positivity of u and v gives an interior feasible point $Q_0 = uv^\top/M$. The R row-sum and D column-sum constraints contain one redundancy because both totals equal M. Their affine rank is therefore $R + D - 1$. The affine hull of $\mathcal{C}(u,v)$ consequently has dimension

$$RD - (R + D - 1) = (R-1)(D-1). \tag{21}$$

For the second statement, select (r_1, r_2, d_1, d_2) satisfying the condition in Proposition 1. Let E be zero except on this 2×2 submatrix, where

$$E_{r_1 d_1} = E_{r_2 d_2} = 1, \qquad E_{r_1 d_2} = E_{r_2 d_1} = -1. \tag{22}$$

Every row and column of E sums to zero. Because Q_0 is strictly positive, $Q_\pm = Q_0 \pm \epsilon E$ remain nonnegative and feasible for sufficiently small $\epsilon > 0$. Let the transported RA marginal be

$$\Phi(Q) = \sum_{d=1}^{D} T_d Q_{:d}. \tag{23}$$

Linearity gives

$$\Phi(Q_+) - \Phi(Q_-) = 2\epsilon\left[T_{d_1} e_{r_1} - T_{d_2} e_{r_1} - T_{d_1} e_{r_2} + T_{d_2} e_{r_2}\right] \tag{24}$$

$$= 2\epsilon (T_{d_1} - T_{d_2})(e_{r_1} - e_{r_2}) \neq 0. \tag{25}$$

Thus Q_+ and Q_- have exactly the same projected RA and DA observations but yield different transported futures. Fractional radial translations induced by distinct Doppler bins satisfy the stated condition except in degenerate cases. Hence marginal conservation, including an exact Sinkhorn solution, cannot by itself select the physically correct target association. □

A.2 RADAR-PHYSICS INTERPRETATION OF THE DESIGN

From an FMCW return to the RDA cube. For a linear chirp with carrier f_c, slope κ, wavelength $\lambda = c/f_c$, and a point scatterer at range R with relative radial velocity v_r, the first-order dechirped phase is

$$\phi(\tau, n, p) \simeq 2\pi \left(\frac{2\kappa R}{c}\tau + \frac{2v_r}{\lambda} n T_c + \frac{pd \sin\theta}{\lambda} \right) + \phi_0, \tag{26}$$

where τ is fast time, n indexes chirps separated by T_c, and p indexes antennas with spacing d. Fourier processing along these three coordinates resolves range, Doppler, and azimuth, respectively, yielding a nonnegative power cube $S_t(r, d, a)$. Equation equation 1 then follows by incoherent marginalization. Linearity of summation gives

$$\sum_r X_t^{\mathrm{RA}}(r, a) = \sum_{r,d} S_t(r, d, a) = \sum_d X_t^{\mathrm{DA}}(d, a), \tag{27}$$

which explains why the consistency constraint is physically meaningful only in linear power: in general, $\sum_i \log(1 + s_i) \neq \log(1 + \sum_i s_i)$.

Lemma 2 (Ego-motion and tangential-velocity ambiguity)**.** *Let $n(\theta) = (\cos\theta, \sin\theta)^\top$ be the line-of-sight direction and let $v_o, v_e \in \mathbb{R}^2$ denote object and ego velocity. A monostatic Doppler bin measures*

$$f_D = \frac{2}{\lambda} v_r, \qquad v_r = n(\theta)^\top (v_o - v_e). \tag{28}$$

Without v_e, the same f_D is compatible with infinitely many ego/object decompositions. Even if v_e is known, the component of v_o orthogonal to $n(\theta)$ remains unobserved from one return.

Proof. For any w and any z satisfying $n^\top z = 0$, replacing (v_o, v_e) by $(v_o + w + z, v_e + w)$ leaves $n^\top(v_o - v_e)$ unchanged. If v_e is fixed, adding any such z to v_o still leaves the observation unchanged. The null space has dimension one in the planar case. Therefore DA supplies relative-radial evidence but cannot by itself identify object velocity or trajectory. □

Lemma 3 (Radial transport depends on the missing coupling)**.** *Let ΔR be range-bin spacing and $v(d)$ the relative radial velocity represented by Doppler bin d. Under locally constant velocity and line of sight,*

$$R_{t+h} \simeq R_t + h\Delta t\, v_r, \qquad \delta_h(d) = \frac{h\Delta t\, v(d)}{\Delta R}. \tag{29}$$

If $K(r' \mid r + \delta)$ is an interpolation kernel, the idealized future RA column generated by an RDA coupling Q_a is

$$u_{t+h,a}(r') = \sum_{r,d} K(r' \mid r + \delta_h(d))\, Q_a(r, d). \tag{30}$$

Thus the future spatial marginal is not determined by the separate present marginals $u = Q\mathbf{1}$ and $v = Q^\top \mathbf{1}$.

Proof. Differentiating range gives $\dot{R} = n^\top(v_o - v_e) = v_r$, whose first-order integration yields Eq. equation 29. Substituting the velocity-conditioned displacement into a mass-preserving renderer gives Eq. equation 30. The kernel multiplies each (r, d) entry before Doppler is summed, so changing the association $Q(r, d)$ while preserving its two marginals can change the output. Proposition 1 constructs such a change explicitly. □

Why balance signed innovations. Let $\rho = N_+/N$ be the changed-support fraction. Plain pixel-mean ℓ_1 regression can be written as

$$\mathcal{L}_{\mathrm{den}} = \rho\, \bar{e}_+ + (1 - \rho)\, \bar{e}_-, \qquad \bar{e}_\pm = \frac{1}{N_\pm} \sum_{i \in \mathcal{I}_\pm} |e_i|. \tag{31}$$

For $\rho \simeq 0.01$, comparable per-pixel residuals make the changed region contribute roughly one percent of the data term. The innovation-balanced point predictor minimizes $(\bar{e}_+ + \bar{e}_-)/2$ together with the dense term, increasing the contribution of rare response changes. Signed regression additionally captures their direction and magnitude, providing a measurement-level reference for the decoded flow objective.

Lemma 4 (Convex calibration bound)**.** *For any target Y, predictions P_0, P_1, and $\alpha \in [0,1]$, the calibrated predictor satisfies*

$$\|(1-\alpha)P_0 + \alpha P_1 - Y\|_1 \le (1-\alpha)\|P_0 - Y\|_1 + \alpha\|P_1 - Y\|_1. \tag{32}$$

Proof. Apply the triangle inequality to $(1-\alpha)(P_0 - Y) + \alpha(P_1 - Y)$ and use nonnegativity of the two coefficients. The bound motivates a global validation-selected operating point, but does not guarantee improvement over the better endpoint and therefore complements the single-network result in Table 7. □

A.3 Complete Metric Definitions

All metrics are evaluated in denormalized log-power unless stated otherwise. Let P, Y, and X denote a prediction, target, and current observation for one horizon and modality, and let $\Delta P = P - X$ and $\Delta Y = Y - X$.

Measurement realism. Log-MAE is $|P - Y|$ averaged over pixels. For high-energy F1, the target 99th percentile $q_{.99}(Y)$ defines positives in both the target and prediction: $Y \ge q_{.99}(Y)$ and $P \ge q_{.99}(Y)$. Bandpass correlation is the Pearson correlation between

$$B(Z) = \mathrm{AvgPool}_{3\times3}(Z) - \mathrm{AvgPool}_{9\times9}(Z) \tag{33}$$

for prediction and target. Additional diagnostics are absolute log-ratio errors for total linear power and spatial total variation, Euclidean strongest-peak error in bins, and the mean correlation between row- and column-integrated linear-power profiles.

Innovation fidelity. We compute soft signal-and-change weights in normalized coordinates,

$$w = \sigma\left(\frac{Y - 1.0}{0.5}\right)\sigma\left(\frac{|Y - X| - 0.25}{0.1}\right). \tag{34}$$

Change-weighted MAE is $\sum w|P - Y| / \sum w$ in log-power. Innovation cosine is

$$\mathrm{Cos}_\Delta = \frac{\sum w \Delta P \Delta Y}{\sqrt{\sum w \Delta P^2}\sqrt{\sum w \Delta Y^2}}. \tag{35}$$

Innovation-support F1 applies the same target-threshold protocol as high-energy F1 to $|\Delta P|$ and $|\Delta Y|$. We also measure absolute log-ratio errors between predicted and target frame-to-frame change and between their horizon separations, plus correlation between the five-horizon innovation-magnitude curves. These statistics characterize measurement evolution. Object-level velocity and trajectory evaluation additionally requires tracked targets.

Aggregation and uncertainty. We first average anchors for every sequence, horizon, and modality, then average the resulting cells. This prevents long sequences from dominating the benchmark. Paired confidence intervals resample the eight test sequences with replacement 10,000 times while retaining all horizons within each selected sequence.

A.4 Generative Evaluation Protocol

Let $Y^{(1)}, \ldots, Y^{(K)} \sim p_\theta(\cdot \mid X_t)$ be independent complete five-frame futures and let $Y^\star$ be the observed future. For any scalar coefficient, the finite-ensemble CRPS estimator is

$$\widehat{\mathrm{CRPS}} = \frac{1}{K}\sum_{k=1}^{K}|Y^{(k)} - Y^\star| - \frac{1}{2K^2}\sum_{k,j=1}^{K}|Y^{(k)} - Y^{(j)}|. \tag{36}$$

For a vectorized RA+DA sequence with n entries, the multivariate energy score replaces absolute values by RMS distances $\|\cdot\|_2/\sqrt{n}$. Both scores reward accuracy and distributional spread, but neither alone guarantees spatially coherent radar returns.

We therefore distinguish three quantities for every nonlinear radar metric m:

$$M_{\text{sample}} = \frac{1}{K}\sum_{k=1}^{K} m(Y^{(k)}, Y^\star), \quad M_{\text{mean}} = m\left(\frac{1}{K}\sum_{k} Y^{(k)}, Y^\star\right), \\ M_{\text{oracle}} = \operatorname*{best}_{k}\, m(Y^{(k)}, Y^\star). \tag{37}$$

M_{sample} is the expected quality of a deployed draw and is the primary structural statistic. M_{mean} measures the ensemble point estimator. It can blur mutually exclusive sharp futures and never substitutes for M_{sample}. M_{oracle} measures whether the sampled support contains a compatible future and is explicitly non-deployable.

For nominal level $q \in \{0.5, 0.8\}$, coverage is the fraction of target coefficients inside the empirical central-q sample interval. Diversity is the mean pairwise sample distance,

$$D_K = \frac{2}{K(K-1)}\sum_{k<j} \|Y^{(k)} - Y^{(j)}\|_1, \tag{38}$$

and spread–skill compares ensemble standard deviation with ensemble-mean RMSE. We additionally compute per-sample high-energy F1, bandpass correlation, total-variation ratio, innovation F1, and the linear-power projection error in Eq. equation 39. Evaluation uses $K = 8$ fixed sampling seeds and 16 ODE evaluations. K/NFE curves use the CRPS-selected checkpoint.

Model-selection criteria. The validation protocol jointly considers proper-score improvement, nonzero diversity, oracle support gain, expected-sample response and innovation F1, a spatial-TV ratio in $[0.80, 1.20]$, interval coverage, projection-mass error, sensitivity to structured DA interventions, and localization of variation around radar responses. These criteria are reported separately because no scalar summary captures accuracy, diversity, calibration, and measurement structure simultaneously.

RadarDreamer has log-domain sample standard deviation 0.42113, expected spatial-TV ratio 0.91832, and empirical 80% interval coverage 0.59291, compared with 0.59322 for the independent-stream model. These statistics complement the proper scores reported in the main paper.

A.5 Additional Test Metrics

Sequence-level uncertainty of the information controls. For calibrated RA+DA versus five-frame RA, the sequence-bootstrap 95% intervals for favorable differences in joint log-MAE, high-energy F1, change-weighted MAE, innovation cosine, and innovation-support F1 are respectively $[0.0116, 0.0290]$, $[0.0027, 0.0096]$, $[0.0438, 0.0825]$, $[0.0245, 0.0611]$, and $[0.0238, 0.0314]$. These intervals quantify the sequence-level variation of the test-set point-forecast comparison.

Table 5: Complete future-RA decomposition for instantaneous Doppler conditioning. The main paper reports the three response-evolution metrics most directly associated with sparse future innovation.

Method	Log-MAE ↓	Energy F1 ↑	Change-wtd. MAE ↓	Innov. cos. ↑	Innov. F1 ↑
5-frame RA	**0.3198**	0.9018	0.6907	**0.7220**	0.3179
RA+DA dense	0.3264	0.9047	0.6967	0.6975	0.3200
RA+DA calibrated	0.3286	**0.9052**	**0.6775**	0.7062	**0.3638**

Table 6: Additional test-set metrics spanning aggregate response structure and future innovation.

Method	Power err. ↓	TV err. ↓	Peak bins ↓	Axis corr. ↑	Change err. ↓	Horizon err. ↓
Persistence	**0.0645**	**0.0392**	**4.668**	**0.9563**	14.6490	14.8593
RA only	0.2506	0.4032	16.656	0.9009	1.5029	1.5942
3-frame RA	0.1752	0.3781	17.277	0.9422	1.5148	1.5324
5-frame RA	0.1778	0.3758	15.134	0.9427	1.5209	1.5445
RA+DA dense	0.1465	0.3649	**7.960**	**0.9545**	1.2946	1.0268
RA+DA calibrated regression	**0.1420**	**0.3514**	8.514	0.9536	**1.2207**	**0.9481**

Table 6 complements the main metrics with power, variation, peak, and temporal statistics. Persistence is strongest on several static frame statistics because most radar energy changes slowly, yet its change

and horizon errors are more than an order of magnitude worse. Among the learned point predictors, calibrated regression best balances power, variation, and temporal evolution. The dense branch retains a slightly more accurate strongest peak and axis profile.

A.6 Robustness and Physical Consistency

Figure 4 in the main paper motivates the joint realism–innovation evaluation. Table 7 additionally includes the balanced single network, separating the effect of its training objective from output interpolation.

Single-network control and projection consistency. For predicted denormalized log-power $P^{\mathrm{RA}}, P^{\mathrm{DA}}$, let $Z^m = \max\{\exp(P^m) - 1, 0\}$ and define

$$E_{\mathrm{mass}} = \frac{1}{A} \sum_a \frac{2\left|\sum_r Z^{\mathrm{RA}}(r,a) - \sum_d Z^{\mathrm{DA}}(d,a)\right|}{\sum_r Z^{\mathrm{RA}}(r,a) + \sum_d Z^{\mathrm{DA}}(d,a) + \epsilon}. \tag{39}$$

Equation equation 39 is evaluated only after undoing the two modality-specific normalizations. The deterministic calibration row rescales both maps at each azimuth toward the geometric-mean mass $g(a) = \sqrt{m_{\mathrm{RA}}(a) m_{\mathrm{DA}}(a)}$:

$$\bar{Z}^{\mathrm{RA}}(:,a) = \frac{g(a)}{m_{\mathrm{RA}}(a)} Z^{\mathrm{RA}}(:,a), \qquad \bar{Z}^{\mathrm{DA}}(:,a) = \frac{g(a)}{m_{\mathrm{DA}}(a)} Z^{\mathrm{DA}}(:,a). \tag{40}$$

This target-free operation guarantees the shared mass identity and is used only to quantify its accuracy trade-off.

Table 7: Test-set robustness of the point-forecast baselines. AP varies the prediction threshold for target-defined top-1% innovation support. Mass error is the symmetric relative discrepancy in Eq. equation 39.

Method	Log-MAE ↓	Change-wtd. MAE ↓	Innov. F1 ↑	Innov. AP ↑	Mass error ↓
Ground truth	–	–	–	–	**0.0080%**
5-frame RA	0.3370	0.8183	0.2852	0.1685	22.73%
RA+DA dense	**0.3136**	0.7685	0.2629	0.1764	13.60%
RA+DA balanced (single)	0.3195	0.7565	**0.3258**	**0.1901**	15.32%
RA+DA calibrated regression	0.3173	**0.7562**	0.3129	0.1880	14.87%
Calibrated regression + exact mass projection	0.3328	0.7658	0.3048	0.1851	1.08×10^{-4}**%**

The single balanced network exceeds the selected blend on innovation F1 and AP, while the blend retains lower full-frame and slightly lower change-weighted error. Thus the training objective itself improves localization. Conversely, exact mass projection removes nearly all RA–DA inconsistency but costs 4.88% relative log-MAE, motivating the differentiable consistency term in Eq. equation 17 rather than post-hoc projection.

Support-threshold robustness. Table 8 varies the target support fraction while retaining the same predictions. Calibrated regression improves over 5-frame RA at every tested fraction, so the localization result is not specific to 1%. Average precision in Table 7 additionally removes a fixed prediction threshold, although its target support remains top-1% by definition.

Table 8: Joint innovation-support F1 under target support fractions from 0.5% to 5%.

Method	0.5%	1%	2%	5%
5-frame RA	0.2322	0.2852	0.3349	0.4009
RA+DA calibrated regression	**0.2615**	**0.3129**	**0.3599**	**0.4158**
Gain (pp)	▲ 2.93	▲ 2.78	▲ 2.49	▲ 1.49

Ego-speed stratification. The Native128 metadata provide radar-frame ego velocity. We stratify the 11,400 anchors by current ego speed with fixed predictions and thresholds. Table 9 shows lower reconstruction error in every stratum and stronger innovation metrics at moderate and high ego speed. Near-stationary scenes favor spatial history on innovation cosine and F1, revealing a speed-dependent balance between spatial and Doppler evidence.

Table 9: Anchor-weighted test metrics stratified by current ego speed and averaged over five horizons and both modalities.

Ego speed	Method	Anchors	Log-MAE ↓	Change-wtd. MAE ↓	Innov. cos. ↑	Innov. F1 ↑
≤ 0.2 m/s	5-frame RA	1,482	0.1752	0.3886	**0.4413**	**0.1372**
	RA+DA calibrated regression		**0.1692**	**0.3750**	0.3788	0.0988
0.2–2.0 m/s	5-frame RA	5,787	0.3535	0.9012	0.6064	0.3025
	RA+DA calibrated regression		**0.3389**	**0.8444**	**0.6456**	**0.3393**
> 2.0 m/s	5-frame RA	4,131	0.3167	0.7992	0.5184	0.3418
	RA+DA calibrated regression		**0.2698**	**0.6768**	**0.6325**	**0.3785**

A.7 Physical Horizon Duration

ColoRadar timestamps are Unix seconds and the single-chip stream is nominally 10 Hz. Table 10 reports actual offsets over the exact 11,400 final-test anchors. Scheduling jitter explains the finite intervals. Evaluation remains indexed by contiguous valid frames rather than resampled timestamps.

Table 10: Physical time after the current observation for each forecast horizon.

h	Median (s)	Mean (s)	2.5–97.5 percentile (s)
1	0.1000	0.1000	0.0538–0.1462
2	0.2000	0.2000	0.1540–0.2460
3	0.3000	0.3000	0.2540–0.3460
4	0.4000	0.4000	0.3544–0.4455
5	0.5000	0.5000	0.4540–0.5460

A.8 Real Projection-Ambiguity Cases

We retrieve cross-sequence validation neighbors using downsampled, standardized current RA+DA descriptors and then select, within the top 0.5% of projection similarity, pairs whose $h = 5$ futures diverge most. To expose information retained only by the cube, define the per-azimuth normalized RDA joint $q_a(r, d)$, its marginals q_a^r, q_a^d, and the standardized association residual

$$A_a(r, d) = \frac{q_a(r, d) - q_a^r(r) q_a^d(d)}{\sqrt{q_a^r(r) q_a^d(d) + \epsilon}}. \tag{41}$$

Unlike RA and DA, A_a retains range–Doppler coupling. Figure 5 gives two target-independent, disjoint-sequence cases. Their current projection cosines are 0.981 and 0.979, but their future cosines fall to 0.828 and 0.836. The corresponding full-RDA association-residual cosines are only 0.614 and 0.452. These examples instantiate the ambiguity behind Proposition 1 and empirically support stochastic association modeling.

A.9 Analysis of Alternative Formulations

Table 11 reveals complementary trade-offs among support localization, distributional scores, response structure, and marginal reconstruction. Their joint pattern motivates coupling-aware generation together with separate realism and innovation metrics.

A.10 Implementation and Reproducibility Details

The point-forecast baselines use channel widths $(32, 64, 128, 256)$, two residual blocks per level, GroupNorm, SiLU activations, strided 3×3 downsampling, bilinear upsampling, and a 1×1 ten-channel output head. Each learned model has 4,744,074 parameters. RadarDreamer uses width 256, eight coupling/flow blocks, eight attention heads, an exact two-level Haar packet, and 28,043,040 trainable parameters. Its future-code mean and standard deviation are estimated from 8,192 training anchors across all five horizons. Coefficient standard deviations range from 0.008361 to 6.696867. We optimize with AdamW, cosine decay, gradient clipping at 1.0, global batch 16 through accumulation, and bfloat16 mixed precision for ten epochs. Evaluation uses $K = 8$ samples and 16 explicit-Euler

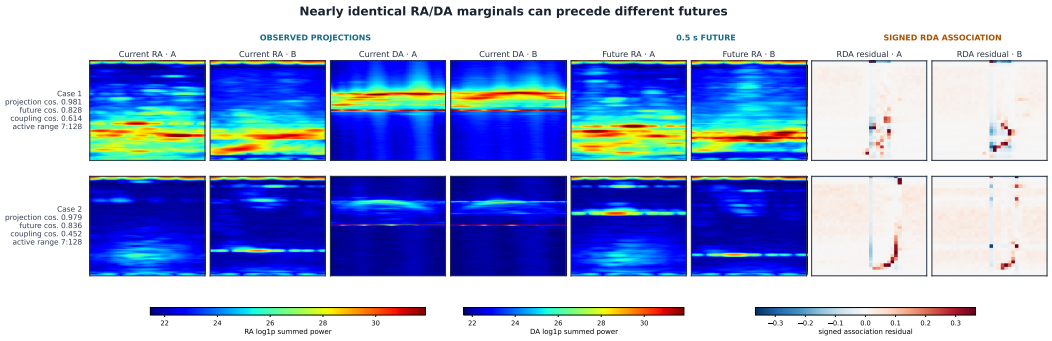

Figure 5: **Real validation cases for projection ambiguity.** Each row compares samples from different sequences. Current RA/DA projections are close, whereas their 0.5-s future RA observations separate. Leading and trailing all-zero range rows are removed. Every RA panel shares one robust log-power scale and every DA panel shares another, both rendered with the jet color map. The final columns use a sign-preserving diverging scale for the full-RDA association residual in Eq. equation 41 at the azimuth of maximum coupling difference.

Table 11: Validation results for alternative stochastic, support, and transport formulations. Metrics are reported jointly because a better distribution score or marginal constraint can coexist with degraded innovation fidelity.

Variant	Intended mechanism	Positive evidence	Observed trade-off
Support BCE/Dice	threshold-crossing localization	innovation F1 $+5.20$ pp, cosine $+4.44$ pp	log-MAE 0.4939, energy F1 0.6785, change-wtd. MAE 0.9609
Latent rectified flow	stochastic residual	CRPS -20.06%, energy score -21.89%	change-wtd. gain -9.75%, TV ratio 0.748, F1 0.809
Multiband pixel flow	frequency-balanced stochastic target	CRPS -18.12%, energy score -20.19%	change-wtd. gain -6.58%, power ratio 0.862, F1 0.774
Maximum-entropy transport	parameter-free RA–DA coupling	98.7% linear power retained	RA MAE $+191.98\%$, wrong range allocation
Learned power transport	learned $P(d \mid r, a)$ with protected rendering	structure/power recovered, all horizons beat persistence	change-wtd. MAE 0.8936, three of seven criteria satisfied
Sinkhorn association	exact marginal-constrained joint	marginal errors 1.38×10^{-8}/0.00357	change-wtd. MAE 0.8939, temporal-change error 0.3384

ODE steps. The quantitative generative comparison reports training seed 20260921. The A/B visualizations use two sampling seeds from the same trained checkpoint.

The test evaluator records each checkpoint hash, the training-set normalization hash, ordered-anchor identity hash, and all sequence–horizon metrics in a non-overwritable JSON artifact. The coefficient $\alpha = 0.75$ is selected on validation before test evaluation. The detailed single-chip example in Fig. 6 is a sequence-balanced fixed validation anchor recorded by the data module. Its selection uses only sequence identity and anchor position and is independent of model output. The RA crop is the smallest contiguous range interval containing every nonzero row in the current and target maps. RA and DA use separate 2nd–99.7th percentile limits computed from observations and then fixed across all draws.

Four-domain generation protocol. Every domain uses the same 28.043M-parameter generator and objective. The three added domains reset their optimizer and domain-specific training statistics, initialize weights from the single-chip checkpoint, and fine-tune on their own training splits. The

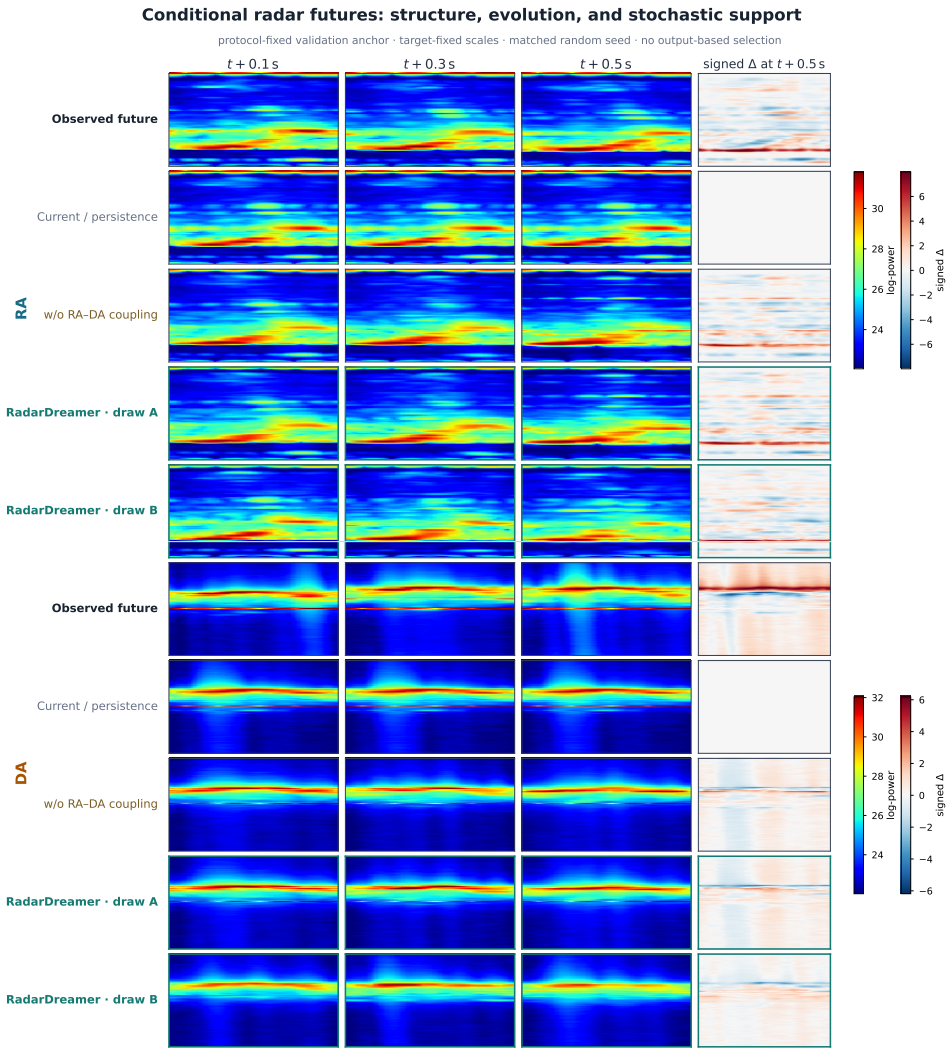

Figure 6: **Detailed single-chip temporal and innovation view.** Rows compare the measured future, persistence, a matched generator without RA–DA coupling, and two independent RadarDreamer draws on the protocol-fixed validation anchor. The right column shows signed innovation from the current observation at 0.5 s. Persistence has zero innovation. The generative rows depart from the input while retaining dominant RA ridges and the DA manifold. Full and ablated generators use the same noise seed where applicable. Crops and robust scales are fixed from observations.

five-epoch transfer schedule uses learning rate 5×10^{-5} and effective batch 16. Checkpoints are selected by validation CRPS. Figure 3 uses epoch 9 for single-chip, epoch 3 for K-Radar, and epoch 0 for RaDelft and cascade, forming a qualitative transfer study across the four acquisition domains. Appendix A.11 reports complete sequences and aggregate statistics.

Multi-domain gallery protocol. Figure 7 is built from fixed validation manifests of ColoRadar single-chip, K-Radar, RaDelft, and ColoRadar cascade. For each domain, the script sorts sequences by group identifier, selects the middle sequence and its middle valid anchor, and only then reads radar arrays. No intensity, motion, target, or model-output statistic enters selection. It displays the current frame and offsets $h \in \{1, 3, 5\}$ from the same anchor. The RA crop removes only leading and trailing rows that are identically zero across those four frames. Separate RA and DA limits are the 2nd and 99.7th percentiles within each domain and modality, followed by a common $[0, 1]$ jet mapping. This normalization makes morphology legible across sensors whose native log-power scales differ and precludes cross-panel amplitude comparison.

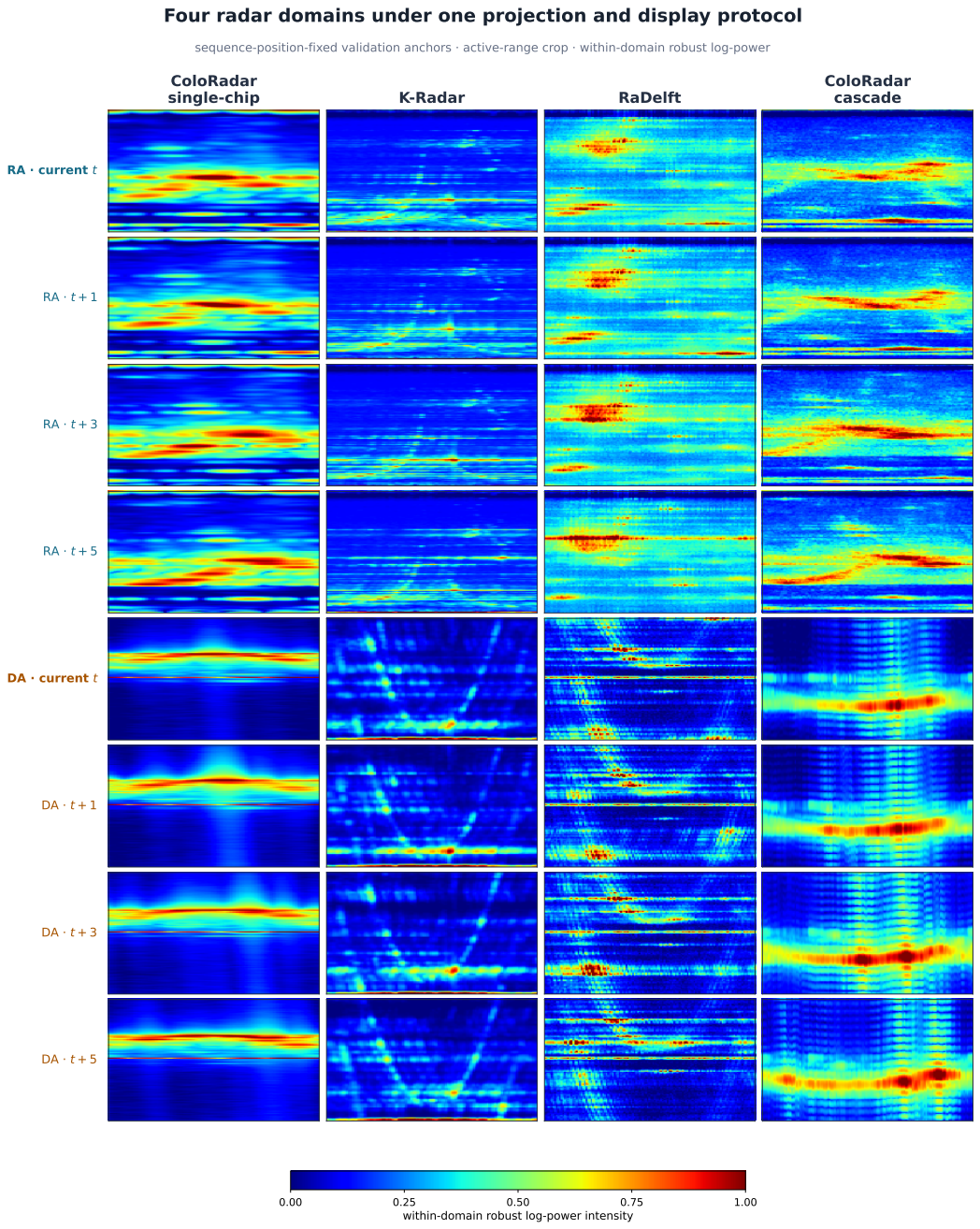

Figure 7: **Four radar domains under one projection and display protocol.** Columns show ColoRadar single-chip, K-Radar, RaDelft, and ColoRadar cascade. Rows show paired RA and DA observations at the current frame and offsets $h \in \{1, 3, 5\}$. Each panel is a protocol-fixed validation observation. For display only, identically zero leading and trailing range rows are removed, and RA/DA use per-domain robust log-power limits mapped through the same jet color scale. The gallery documents hardware- and dataset-dependent response structure across four acquisition configurations.

Table 12: **Aggregate statistics for qualitative-example selection.** Every row includes all 32 validation anchors and both sampling seeds. P/O denotes persistence/expected-sample error. Active RA excludes only all-zero boundary range rows. DA retains its full grid. Raw RA shows the uncropped result. A win requires lower full-frame and change-weighted MAE in both modalities.

Domain	Active RA MAE P/O	Full DA MAE P/O	Raw RA MAE P/O	CRPS P/O	Wins	Case gains (%)† MAE / CW
Single-chip	0.5338/0.5363	0.3998/0.4169	0.5046/0.5095	0.4668/0.3748	6/32	9.5 / 27.3
K-Radar	0.3034/0.3334	0.4734/0.4744	0.3034/0.3334	0.3884/0.3230	5/32	8.8 / 33.2
RaDelft	0.1738/0.1657	0.2020/0.1962	0.1642/0.3192	0.1879/0.1438	12/32	19.9 / 45.2
Cascade	0.8499/0.8755	0.7107/0.7828	0.8034/0.8288	0.7803/0.6694	2/32	5.6 / 23.5

† Joint MAE / continuously change-weighted MAE decrease on the selected whole sequence, averaged over both draws.

A.11 Multi-domain Future Sequences and Selection Statistics

What the qualitative comparison establishes. Each sample jointly generates five future RA+DA pairs from the current observation, without intermediate measured futures or frame-wise draw switching. Figures 8–11 expose the condition and every horizon. They illustrate stochastic sensor evolution, not action-conditioned planning. The single-chip and RaDelft examples span approximately 0.5 s and cascade spans 1.00015 s. K-Radar remains frame-indexed because its cached timestamps are unavailable. The accompanying animation uses slowed playback for inspection.

Example selection and aggregate statistics. We use 32 uniformly spaced indices per 128-anchor validation subset and the CRPS-selected checkpoint. Sampling seeds 20260921/20260922 generate complete sequences using 16 ODE steps in float32. The displayed anchor maximizes the joint gain across RA/DA and full-frame/change-weighted MAE by averaging both draws and all horizons. Seeds and individual frames are fixed. Table 12 includes every candidate, and Fig. 12 complements the selected examples with additional cases from each domain. The numbers of anchors improving all four criteria are 6/32, 5/32, 12/32, and 2/32 for single-chip, K-Radar, RaDelft, and cascade, respectively.

Across the full subset, empirical two-sample CRPS is below the point-mass persistence score in all four domains, while expected-sample MAE remains mixed. This difference illustrates why stochastic support and per-sample accuracy are reported separately. These subset statistics complement the $K = 8$ single-chip evaluation.

For modality m, the continuous change weight is $w^m_{h,u} = |Y^m_{h,u} - X^m_u|$ in log-power units. The displayed change-weighted error is

$$E^m_{\mathrm{CW}} = \frac{1}{KH} \sum_{k=1}^{K} \sum_{h=1}^{H} \frac{\sum_{u \in \mathcal{V}_m} w^m_{h,u} \, |\hat{Y}^{m,(k)}_{h,u} - Y^m_{h,u}|}{\max(\sum_{u \in \mathcal{V}_m} w^m_{h,u}, \varepsilon)}, \qquad K = 2, \quad H = 5. \tag{42}$$

This visualization diagnostic uses continuous measured change without a top-k mask. It is distinct from the threshold-weighted headline metric. Full-frame MAE uses uniform weights on the same support. Relative reductions compare the mean of the two modality errors with persistence rather than best-of-two or ensemble-mean errors.

Display support and the normalization floor. All methods share per-domain/modality 2nd–99.7th percentile jet limits from the current and measured futures. The RA support $\mathcal{V}_{\mathrm{RA}}$ removes only all-zero boundary range rows. DA keeps every bin. RaDelft has a decoded normalization floor near 2.97 on seven zero-padding rows. Table 12 therefore reports both raw-grid and display-support RA errors. This display diagnostic does not change training or headline evaluation. Shared boxes mark the 24×24 terminal-frame region with greatest mean measured absolute innovation. Crops, boxes, and color limits never use predicted errors.

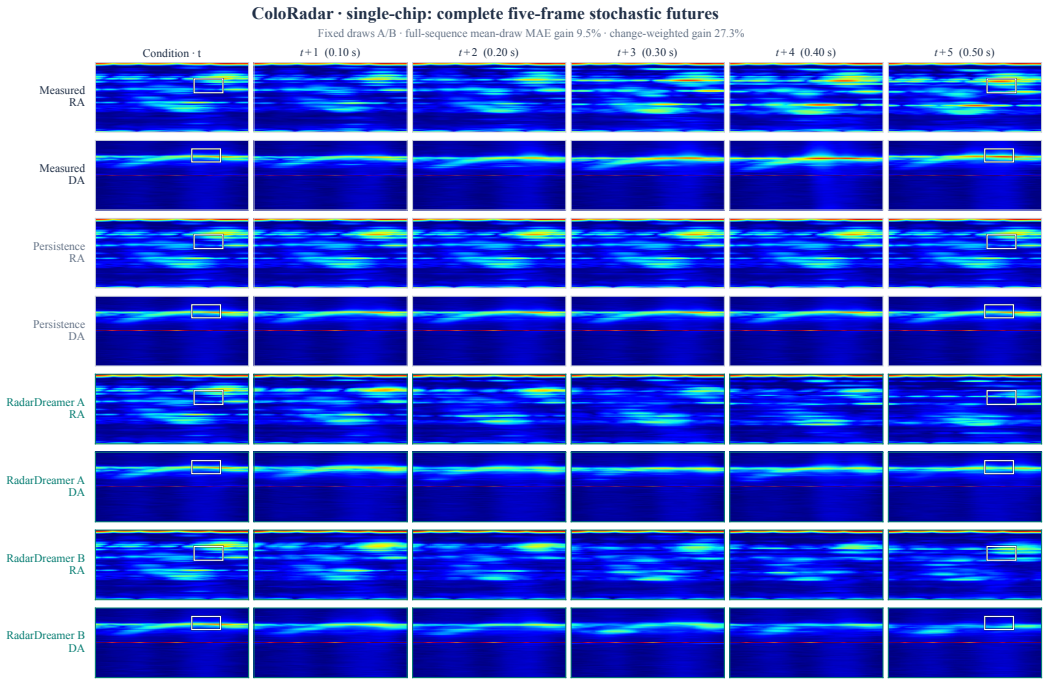

Figure 8: **Single-chip: complete generated sequences.** The same current observation conditions two independently sampled five-frame futures. RA/DA rows compare the measured future, persistence, and fixed draws A/B. Every frame of each draw is retained. The selected-case gains average both draws over all five horizons. Boxes and jet scales are shared across methods.

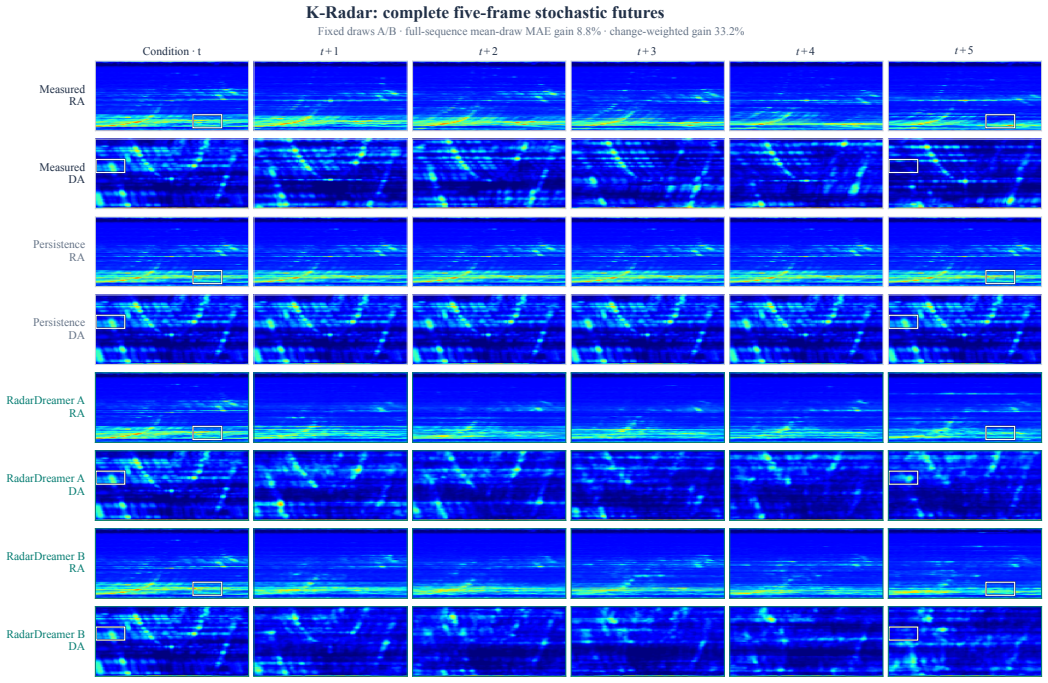

Figure 9: **K-Radar: complete generated sequences.** The selected transfer checkpoint generates distinct future RA ridges and DA branches from one condition. Persistence repeats the input. Horizons remain frame-indexed because the cached sequence has no physical timestamps. The full-sequence selection rule and all-candidate results are given in Table 12.

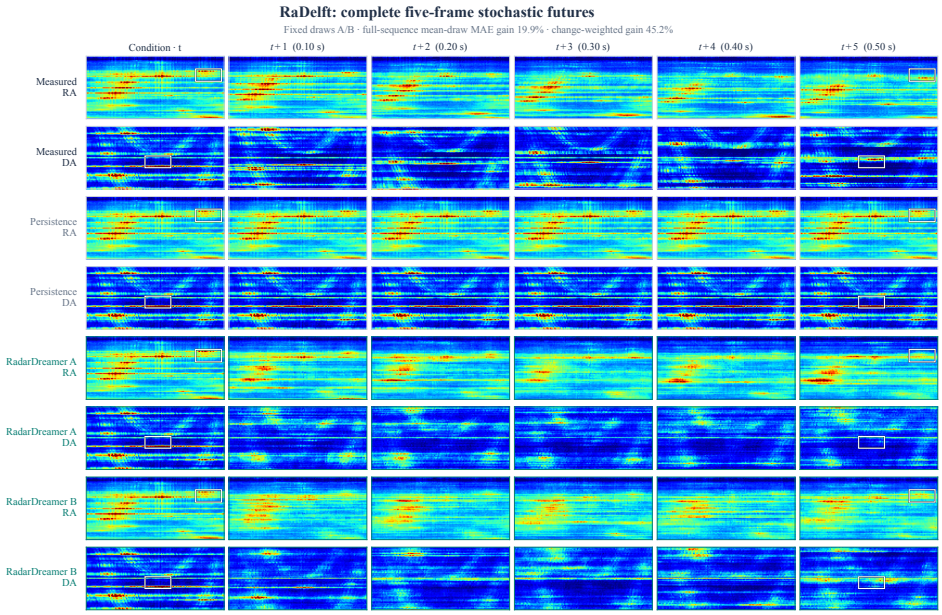

Figure 10: **RaDelft: complete generated sequences.** Measured changes in response strength are compared with a static copy and two complete generated futures over approximately 0.5 s. Zero-padding range rows are removed uniformly from every method. The associated raw-grid normalization-floor error is disclosed separately, so the displayed-support improvement cannot be mistaken for an uncropped error reduction.

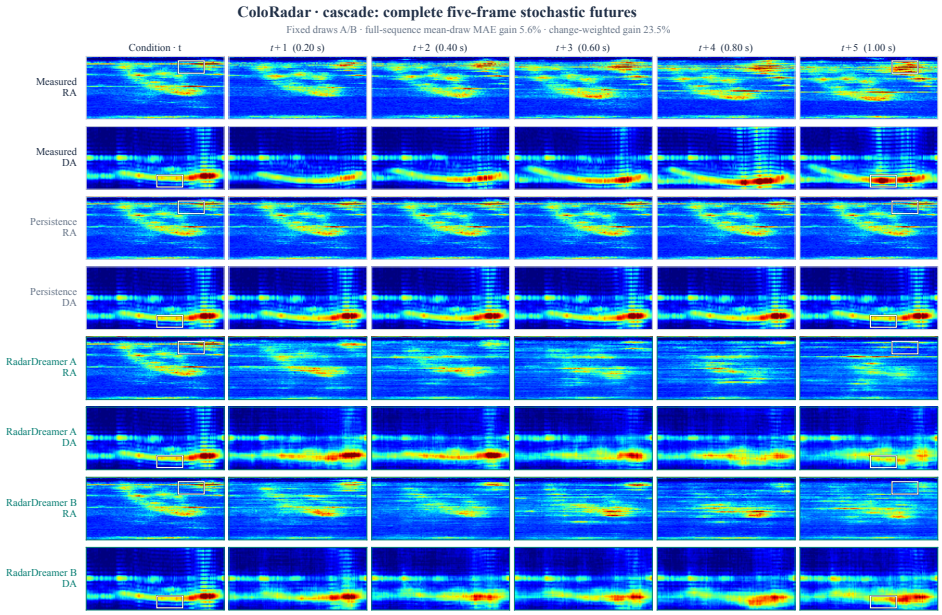

Figure 11: **ColoRadar cascade: complete generated sequences.** Five future frames span approximately 1.0 s for this anchor. The two fixed samples preserve a common Doppler response band while differing in echo placement and amplitude. Temporal evolution is visible against unchanged persistence, but weak echoes and localized response strength still deviate from the realized future.

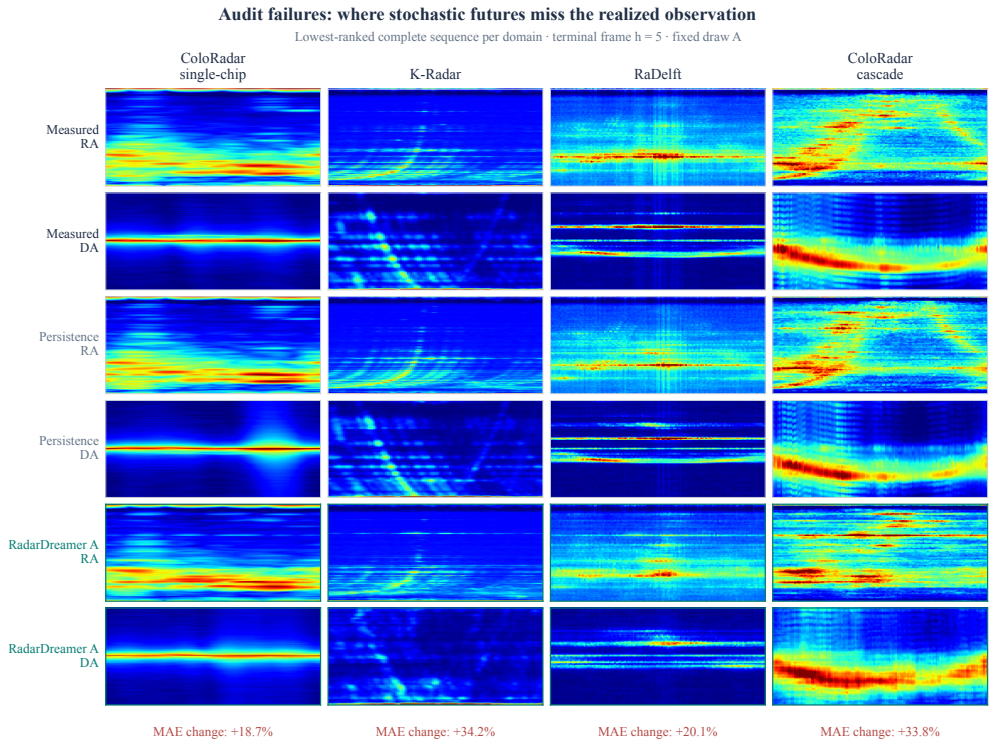

Figure 12: **Additional cases from the same candidate pool.** Each column shows a terminal frame spanning complementary response patterns. Measured, persistence, and fixed draw A share one crop and scale. Percentages summarize the complete sequence and both draws.

A.12 Structural Properties of the Generative Design

The following results interpret specific design choices. They formalize the exchange operator, compatibility of nonnegative projected measurements, and the distributional consequence of an ambiguous conditional law.

Independent-axis reindexing of cross-projection exchange. Fix a horizon and azimuth, and write feature tokens as rows, $H_R \in \mathbb{R}^{R' \times C}$ and $H_D \in \mathbb{R}^{D' \times C}$. For one attention head, define

$$\begin{aligned} B(H_R, H_D) &= \mathrm{softmax}_{\mathrm{row}}\left(\frac{Q(H_R)K(H_D)^\top}{\sqrt{d_k}}\right)V(H_D), \\ \mathcal{A}(H_R, H_D) &= H_R + B(H_R, H_D)G, \end{aligned} \tag{43}$$

where d_k is the head dimension and G is a diagonal channel gate. Token-wise normalization and affine query/key/value maps commute with row permutations. For independent permutation matrices P_R, P_D, if $L = Q(H_R)K(H_D)^\top/\sqrt{d_k}$, row-wise softmax satisfies

$$\mathrm{softmax}_{\mathrm{row}}(P_R L P_D^\top) = P_R\,\mathrm{softmax}_{\mathrm{row}}(L)P_D^\top. \tag{44}$$

Consequently,

$$B(P_R H_R, P_D H_D) = P_R\,\mathrm{softmax}_{\mathrm{row}}(L)P_D^\top P_D V(H_D) = P_R B(H_R, H_D), \tag{45}$$

$$\mathcal{A}(P_R H_R, P_D H_D) = P_R H_R + P_R B(H_R, H_D)G = P_R \mathcal{A}(H_R, H_D). \tag{46}$$

The identity survives head concatenation and a token-wise output projection, and holds symmetrically for the DA update. It therefore applies to the implemented exchange, including its channel gates and layer normalization. No identification of a range index with a Doppler index is needed. Crucially, the inputs here are already constructed features: permuting raw bins while keeping their learned positions fixed need not preserve the full model, nor should a radar model ignore physically calibrated positions. This is an algebraic property of cross-attention, not a new general attention theorem.

Proof of Lemma 1. Let $a = \mathbf{1}^\top u$ and $b = \mathbf{1}^\top v$. For any feasible corrected pair of common mass m, the triangle inequality gives

$$\|u - \bar{u}\|_1 + \|v - \bar{v}\|_1 \geq |a - m| + |b - m| \geq |a - b|. \tag{47}$$

If $a \geq b$ and $a > 0$, choose $\bar{u} = (b/a)u$ and $\bar{v} = v$. The correction is exactly $a - b$. If $b > a$, rescale v instead. If $a = b = 0$, both vectors already vanish. This proves attainability, including zero-mass columns. For common mass $m > 0$, the explicit witness

$$\bar{Q} = \frac{\bar{u}\bar{v}^\top}{m}, \qquad \bar{Q}\mathbf{1} = \bar{u}, \qquad \bar{Q}^\top \mathbf{1} = \bar{v} \tag{48}$$

is nonnegative. If $m = 0$, use $\bar{Q} = 0$. Conversely, every nonnegative matrix has equal total row and column mass. Thus the equal-mass condition is necessary and sufficient for compatibility with some nonnegative matrix. □

Writing this minimal correction as $d_1(u, v)$, the per-column penalty is $2d_1(u, v)/(a + b + \epsilon)$. It lies in $[0, 2]$ and vanishes exactly at compatibility. Without the numerical stabilizer, it is invariant to a common positive power rescaling. Averaging over horizons and azimuths gives Eq. equation 17. The witness in Eq. equation 48 is only an existence construction, not an inference module or a correct physical association. Propagation, velocity, visibility, and electromagnetic scene constraints can further restrict the feasible set. In particular, this lemma does not contradict Proposition 1: feasibility does not imply uniqueness.

Proof of Corollary 1. For a distribution P with finite first moment, let U, U' be independent draws from P, independent of the realized target y. We use the standard population energy score (Gneiting & Raftery, 2007), in lower-is-better orientation:

$$\mathsf{S}_E(P, y) = \mathbb{E}\|U - y\|_2 - \tfrac{1}{2}\mathbb{E}\|U - U'\|_2. \tag{49}$$

The evaluator uses RMS distance, $\|\cdot\|_2/\sqrt{n}$ for n scalar entries. This common positive rescaling divides all following scores and gaps by $\sqrt{n}$ without changing the comparison. Set $\Delta = \|y_+ - y_-\|_2 > 0$. Under the stipulated conditional law P_c, two independent futures differ with probability $1/2$. Therefore

$$\mathbb{E}_{Y \sim P_c}\mathsf{S}_E(P_c, Y) = \frac{\Delta}{2} - \frac{1}{2}\frac{\Delta}{2} = \frac{\Delta}{4}. \tag{50}$$

For any point forecast $\hat{y}$, its distribution $\delta_{\hat{y}}$ has zero pairwise spread, so

$$\mathbb{E}_{Y \sim P_c}\mathsf{S}_E(\delta_{\hat{y}}, Y) = \tfrac{1}{2}\big(\|\hat{y} - y_+\|_2 + \|\hat{y} - y_-\|_2\big) \geq \frac{\Delta}{2}. \tag{51}$$

Equality holds for every point on the line segment joining the two futures. Subtracting Eq. equation 50 proves the corollary. □

The physical source of the two futures is explicit. For the cycle perturbation E in Appendix A.1, let $Q_\pm = Q_0 \pm \eta E$ and define $b_h = (T_{h,d_1} - T_{h,d_2})(e_{r_1} - e_{r_2})$. Under the constant-velocity radial transport approximation, the concatenated multi-horizon RA futures obey

$$\Delta_H^2 = \|\Phi_{1:H}(Q_+) - \Phi_{1:H}(Q_-)\|_2^2 = 4\eta^2 \sum_{h=1}^{H} \|b_h\|_2^2. \tag{52}$$

The same present RA/DA observations thus permit distinct future sequences. Stacking further horizons cannot reduce their unnormalized Euclidean separation. This is a statement about the idealized transported signals, not monotonic measured per-pixel error: normalization by horizon count, changing visibility, and finite observation windows can change that comparison. Assigning the two couplings equal probability is an explicit example, not a distribution implied by marginalization alone. The score gap applies in any fixed evaluation coordinates where the two futures remain distinct.

Finite-ensemble qualification. The implemented score uses the K^2 empirical pair average, including zero diagonal terms, as in Eq. equation 36. For independent $U_1, \ldots, U_K, Y \sim P_c$, its expected energy score is

$$\mathbb{E}\widehat{\mathsf{S}}_{E,K}(P_c, Y) = \frac{\Delta}{2} - \frac{K(K-1)}{2K^2}\frac{\Delta}{2} = \frac{\Delta}{4}\left(1 + \frac{1}{K}\right). \tag{53}$$

The idealized advantage over the best point distribution is therefore $\Delta(1 - 1/K)/4$, positive for $K > 1$. This calculation neither changes the finite-ensemble estimator nor predicts a numerical gain on the dataset. It explains why comparisons keep K fixed, why one draw cannot characterize a conditional law, and why an oracle best-of-K metric is not a replacement for a distribution score. For this example, an independent sample and a point on the segment both have expected Euclidean error $\Delta/2$: a pointwise error alone does not reward the correct uncertainty.

Endpoint supervision and exact multiband decoding. Let $D = \mathrm{diag}(s_i)$ contain the positive train-only coefficient scales and let b be the coefficient center. Before output clipping, the decoder is $\mathcal{D}(z) = \mathcal{W}^{-1}(Dz+b)$. For the training interpolant, the true endpoint identity is $Z_1 = Z_\tau + (1-\tau)U_\tau$. Defining $e_v = v_\theta(Z_\tau, X_t, \tau) - U_\tau$, orthonormality gives

$$\widehat{Z}_1 - Z_1 = (1-\tau)e_v, \tag{54}$$

$$\|\mathcal{D}(\widehat{Z}_1) - Y\|_2^2 = (1-\tau)^2 \|De_v\|_2^2, \tag{55}$$

$$(1-\tau)^2 s_{\min}^2 \|e_v\|_2^2 \le \|\mathcal{D}(\widehat{Z}_1) - Y\|_2^2 \le (1-\tau)^2 s_{\max}^2 \|e_v\|_2^2. \tag{56}$$

Here Y is the clipped normalized target, not raw linear power, and $s_{\min}, s_{\max}$ are the extreme fixed coefficient scales. Projection of the decoded estimate onto the target's clipping box is nonexpansive, so the upper bound remains valid after the implementation's clipping, but the equality and lower bound need not. Log denormalization and exponentiation are not isometries.

This identity explains why decoded objectives can target radar response structure without introducing a separately regressed future anchor. The one-step endpoint is a training surrogate, not the result of solving the full sampling ODE. At $\lambda_{\mathrm{end}} = 0$, the squared-error population minimizer is $v^*(z, c, \tau) = \mathbb{E}[U_\tau \mid Z_\tau = z, c]$, which conditions on the noisy state as well as the current observation c. Its local velocity estimate is not a single future predicted from c. Nonzero endpoint weights can shift this optimum, and finite-step ODE sampling adds numerical error. Neither the identity nor the mass lemma guarantees calibrated or exactly mass-consistent generated samples. Those remain empirical checks in the main evaluation.

A.13 Responsible Use

RadarDreamer is developed for offline predictive-sensing research and generates sensor measurements rather than personal identity or protected attributes. Safety-critical deployment calls for sensor-specific calibration, closed-loop validation, and integration with downstream perception and planning.