Supplementary Material for "Branch-Aware Phase-to-Bearing Diagnostics in an Idealized Helical DAS Phase Array"

Ning Hu, Yuanyuan Cao, Chengtao Huang, and Kunjie Shi

TABLE SI
UNCERTAINTY SCOPE. NONE OF THESE SUMMARIES IS A TOTAL MEASUREMENT-UNCERTAINTY BUDGET.

Source	Variation	Summary	Not covered
IQ/noise	300 paired seeds/condition	Bootstrap/Wilson[a]	Array/profile/model
Initialization	3 locked checkpoints	Points and mean±SD	Model-family population
Profile	One D1-P/D2-J; D1-G: $1 + 50$[b]	Fixed-profile sensitivity	General profile distribution
Threshold/extractor	3×3 thresholds; extractor fixed	Sensitivity / untested	Transfer/extraction model

[a] 0/300 has a 95% Wilson upper limit of 1.26%. [b] Radius/pitch fixed; jitter resampled.

I. Purpose and Evidence Scope

This supplement reports diagnostics that are necessary for reproducibility but would obscure the evaluation-centered main narrative if reproduced in full. All E7 and E8 values below come from primary full-run assets. Statistical inference is restricted to a fixed condition and fixed checkpoint. The three learned initializations are shown separately or summarized as three locked points; they are not treated as a model-family population.

II. Experimental Units in the Primary Suites

Table SII states the replication unit and count used by each archived full suite. Within a tested condition, the same record seeds are paired across methods and locked checkpoints. Checkpoints are distinct trained instances, not additional noise replicates.

The primary transition plots summarize Monte Carlo means (with archived sample standard deviations where reported) only at grid points fixed in the archived full-run configuration. They do not estimate a continuous transition boundary. E7 adds paired-bootstrap intervals for its fixed-condition regression estimands; the D1-G profile percentiles instead describe the sampled profile distribution and are not confidence intervals.

III. Exact Itoh Increment

For sensor i, let $A_{\mathrm{eff},i} = A|a_i|$, $\omega = 2\pi f/f_s$, and $\phi_i[n] = A_{\mathrm{eff},i}\sin(\omega n+\psi_i)$. The finite-record noiseless adjacent increment is exactly

$$\max_n |\phi_i[n+1]-\phi_i[n]| = 2A_{\mathrm{eff},i}\left|\sin\left(\frac{\pi f}{f_s}\right)\right| \times \max_n \left|\cos\left(\omega(n+1/2)+\psi_i\right)\right|. \tag{1}$$

The final factor is at most one and equals one only when a sample falls at a zero crossing of the underlying sinusoid, so the main-text condition $2A_{\mathrm{eff}}^{\max}|\sin(\pi f/f_s)| < \pi$ is sufficient for every recorded sensor and every start phase, and is exact in the worst case. The small-frequency reduction $fA_{\mathrm{eff}} < f_s/2$ follows from $\sin(\pi f/f_s) \approx \pi f/f_s$. Both are local temporal conditions: they constrain neither spatial gradients nor noise-induced branch errors.

IV. Complete Metric and Tangent Definitions

Let $\phi_w^{(0)} = \mathcal{W}(\phi)$ denote noiseless wrapped truth and let $\tilde{\phi}_w = \angle z$ denote the noisy principal-phase input. The exact identity $\phi = \phi_w^{(0)} + 2\pi k^{(0)}$ applies only to the former. The evaluation uses the noise-relative reference and predicted branch labels

$$k_{\mathrm{true}} = \mathrm{round}\,\frac{\phi-\tilde{\phi}_w}{2\pi}, \qquad k_{\mathrm{pred}} = \mathrm{round}\,\frac{\hat{\phi}-\tilde{\phi}_w}{2\pi}. \tag{2}$$

Thus "true" identifies a label computed from simulated continuous truth and the same noisy algorithm input; it is not the noiseless count $k^{(0)}$.

For samplewise raw error $\epsilon_{it} = \hat{\phi}_{it} - \phi_{it}$, raw slip-energy is

$$r_{\text{slip-energy}} = \frac{\sum_{i,t}\epsilon_{it}^2\,\mathbb{I}[k_{\mathrm{pred},it} \neq k_{\mathrm{true},it}]}{\sum_{i,t}\epsilon_{it}^2}. \tag{3}$$

The single global integer alignment is

$$m^* = \arg\min_{m\in\mathbb{Z}} \|\hat{\Phi} + 2\pi m - \Phi\|_F^2, \qquad \hat{\Phi}_{\mathrm{GA}} = \hat{\Phi} + 2\pi m^*. \tag{4}$$

With $\epsilon_{it}^{\mathrm{GA}} = \hat{\phi}_{it}^{\mathrm{GA}} - \phi_{it}$ and $k_{\mathrm{pred},it}^{\mathrm{GA}} = \mathrm{round}[(\hat{\phi}_{it}^{\mathrm{GA}} - \tilde{\phi}_{w,it})/(2\pi)]$, the aligned ratio is

$$r_{\text{slip-energy}}^{\mathrm{GA}} = \frac{\sum_{i,t}(\epsilon_{it}^{\mathrm{GA}})^2\,\mathbb{I}[k_{\mathrm{pred},it}^{\mathrm{GA}} \neq k_{\mathrm{true},it}]}{\sum_{i,t}(\epsilon_{it}^{\mathrm{GA}})^2}. \tag{5}$$

The implementation returns zero when either denominator is zero. In that case, the value means that no error energy exists to allocate, rather than an empirically estimated zero fraction; perfect recovery is identified separately by zero absolute/global-aligned RMSE. The archived difference is $r_{\text{slip-energy}} - r_{\text{slip-energy}}^{\mathrm{GA}}$; the aligned ratio does not replace the raw ratio. In the archived full runs, none of the 78,000 E7 or 43,200 E8 records has a zero raw or aligned error-energy denominator, so this convention does not affect the reported aggregates.

For target-frequency snapshot k, spatial consistency is

$$\gamma_{f,k} = \frac{|(\mathbf{x}_{f,k}^{\mathrm{true}})^H\hat{\mathbf{x}}_{f,k}|}{\|\mathbf{x}_{f,k}^{\mathrm{true}}\|_2\|\hat{\mathbf{x}}_{f,k}\|_2}. \tag{6}$$

TABLE SII
INDEPENDENT UNIT AND ARCHIVED REPLICATION COUNT. "PER POINT" MEANS PER FULLY SPECIFIED TESTED CONDITION; NO INTERPOLATION BETWEEN GRID POINTS IS INCLUDED IN THE COUNT.

Suite	Independent unit	Count
V0 sanity	IQ/noise record seed	50 per SNR point
V1 CRB check	Scalar complex-IQ draw	100,000 per SNR point
E1/E2	IQ/noise record seed	500 per SNR/frequency point
E3	IQ/noise record seed	300 per amplitude point
E4-R	IQ/noise record seed	500 per azimuth–SNR point
E4-H	IQ/noise record seed	300 per amplitude–azimuth–SNR point
F1 controls	IQ/noise record seed	300 per amplitude and checkpoint
E7/E8	IQ/noise/drift record seed	300 per condition and instance
D1-G realization sensitivity	Profile seed	50 profiles; no noise Monte Carlo

It is stored as undefined when the denominator is at or below the fixed 10^{-12} numerical floor; record-level summaries never replace undefined values by zero. All 1,560,000 E7 snapshot coefficients (78,000 records times 20 snapshots) are defined under this rule.

For the nominal steering $\mathbf{a}$ and true snapshot matrix $\mathbf{X}^{\text{true}}$, the least-squares snapshot coefficients and the DOA-sensitive tangent basis are

$$\alpha = \frac{\mathbf{a}^H \mathbf{X}^{\text{true}}}{\mathbf{a}^H \mathbf{a}}, \quad \mathbf{g} = \left(\mathbf{I} - \frac{\mathbf{a}\mathbf{a}^H}{\mathbf{a}^H \mathbf{a}} \right) \frac{\partial \mathbf{a}}{\partial \theta}, \quad \mathbf{B} = \mathbf{g}\alpha. \tag{7}$$

The derivative is per radian. The signed prediction is $\text{Re}\langle \mathbf{B}, \mathbf{E} \rangle_F / \|\mathbf{B}\|_F^2$. A condition is marked geometrically degenerate and no prediction is computed when either $\|\mathbf{g}\|_2^2$ or $\|\mathbf{B}\|_F^2$ is at or below 10^{-12}.

Table SIII reports the corresponding empirical metrics for the representative E7 rows used later in the supplement. The values are derived from the archived full records, not from an additional run.

These rows show why the metrics are not interchangeable. The matched hard M3 row combines high slip-energy with low spatial correlation, whereas the easy D1-G M3 row has zero slip-energy and near-unit correlation despite the approximately 13° nominal-manifold bias reported below. The D1-G M4 row has substantial phase/spatial damage even though signed recovery and nominal DOA increments largely cancel. The easy D2-J M5 row retains high correlation while nominal mismatch dominates its total DOA error. Raw and aligned ratios are nearly equal in these examples; the aligned version is retained as a guard against a global integer offset, not as evidence that such alignment is always material.

V. FROZEN MISMATCH PROFILES

Table SV reproduces the frozen magnitudes used by E7/E8. A dash means zero or absent. The values are fixed archived simulator stress levels, not measured hardware tolerances.

Radius, pitch, and sound speed use the positive offsets in the table. Position jitter is a zero-mean Gaussian draw in three Cartesian coordinates, centered across sensors; tangent jitter is a Gaussian tangent-plane perturbation followed by unit-vector renormalization. With perturbed position $\widetilde{\mathbf{r}}_i$, tangent $\widetilde{\mathbf{t}}_i$, sound speed $\widetilde{c}$, lognormal response gain G_i, and response phase δ_i, the actual steering entry used to generate truth is

$$\begin{aligned} \widetilde{a}_i &= G_i (\widetilde{\mathbf{t}}_i^T \mathbf{u})^2 \exp\left[-j \frac{2\pi f}{\widetilde{c}} \widetilde{\mathbf{r}}_i^T \mathbf{u} + j\delta_i \right], \\ \phi_{it} &= A |\widetilde{a}_i| \sin(2\pi f t + \arg \widetilde{a}_i). \end{aligned} \tag{8}$$

Response and carrier gains use $\exp(sZ - s^2/2)$ with $Z \sim \mathcal{N}(0,1)$; response phase uses sZ. Thus response calibration is part of the continuous truth, whereas carrier fading acts only in observation.

For D2-J, let $\mathbf{C}_{ij} = \rho^{|i-j|}$ and $\mathbf{L}\mathbf{L}^T = \mathbf{C} + 10^{-12}\mathbf{I}$. The same spatial factor $\mathbf{L}$ is applied separately to I-noise, Q-noise, and drift innovations. Writing F_i for the fixed carrier gain after any dropout, $\eta = 0.08$ for IQ gain imbalance, $q = 3°$ for quadrature error, and $s_d = 0.08$ rad for the terminal drift scale, the implemented order is

$$\mathbf{d}_t = \mathbf{d}_{t-1} + \frac{s_d}{\sqrt{N}} \mathbf{L} \boldsymbol{\xi}_t, \qquad \boldsymbol{\xi}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{9}$$

$$I_{it}^{(0)} = F_i (1 + \eta) \cos(\phi_{it} + d_{it}), \tag{10}$$

$$Q_{it}^{(0)} = F_i (1 - \eta) \sin(\phi_{it} + d_{it} + q), \tag{11}$$

$$\mathbf{z}_t = \mathbf{I}_t^{(0)} + j\mathbf{Q}_t^{(0)} + \frac{\sigma}{\sqrt{2}} \mathbf{L}(\boldsymbol{\epsilon}_{I,t} + j\boldsymbol{\epsilon}_{Q,t}), qquad \tilde{\boldsymbol{\phi}}_{w,t} = \angle \mathbf{z}_t. \tag{12}$$

The dropout channels are selected without replacement and retain gain 0.05. The simulated continuous truth in Eq. (8) excludes drift, IQ imbalance, fading, and additive noise; those terms enter only through Eq. (12).

The fixed profile seed is 20260714 and is combined with a stable profile-name substream. It fixes geometry, response calibration, fading, and dropout for the primary E7/E8 profile realization. E7 uses sample seeds 520000–520299 and E8 uses 530000–530299 for observation noise and drift. Truth is generated by the perturbed profile, while recovery and reported nominal MUSIC retain the nominal model. The D1-G sensitivity instead uses profile seeds 20260715–20260764: radius and pitch remain fixed at +2% and +5%, while only position and tangent jitter are redrawn.

VI. TARGET-FREQUENCY EXTRACTION AND MUSIC RECONSTRUCTION

All E7 and E8 DOA quantities use one frozen narrowband reconstruction. A record of length N is partitioned into $K = 20$

TABLE SIII
REPRESENTATIVE E7 METRIC ROLES. SLIP-ENERGY ENTRIES AND SNAPSHOT CORRELATION ARE MEDIANS OVER 300 RECORDS; $\Pr(m^* \neq 0)$ IS THE EMPIRICAL RECORD FRACTION. FOR EACH RECORD, $\bar{\gamma}_f$ IS THE MEAN OVER ITS 20 DEFINED TARGET-FREQUENCY SNAPSHOTS.

Condition	Instance	Raw r_{SE}	Aligned $r_{\mathrm{SE}}^{\mathrm{GA}}$	$\Pr(m^* \neq 0)$	Median $\bar{\gamma}_f$
D0, $A = 3$, 30°	M3	0	0	0	0.999953
D0, $A = 10$, 30°	M3	0.999958	0.999957	0.240	0.583408
D0, $A = 10$, 30°	M5 seed 17	0	0	0	0.999996
D1-G, $A = 3$, 30°	M3	0	0	0	0.999953
D1-G, $A = 10$, 30°	M4 seed 17	0.991783	0.991783	0	0.646118
D2-J, $A = 3$, 30°	M5 seed 17	0.082764	0.082764	0	0.998598
D2-J, $A = 10$, 30°	M3	0.999596	0.999482	0.373	0.108540

TABLE SIV
E8 HARD-CONDITION OUTCOMES (D0, $A = 10$ RAD, 20 DB, 200 HZ, 30°, $N = 10000$), FROM THE LOCKED E8 RECORDS. LEARNED ROWS ARE MEAN±SD OVER THREE LOCKED CHECKPOINTS; FAILURE RATES USE THE PRIMARY THRESHOLDS ($\rho_{\mathrm{br}} > 10^{-3}$; $|e_\theta| > 1°$), AND A ZERO RATE OVER 300 RECORDS HAS A WILSON 95% UPPER LIMIT OF 1.26%.

Instance	DOA MAE (°)	P95 $|e_\theta|$ (°)	Phase fail	DOA fail
M1	88.242	88.325	1.000	1.000
M2	100.779	103.857	1.000	1.000
M3	2.077	2.228	1.000	1.000
M4	1.852 ± 0.789	2.414 ± 1.667	1.000	1.000
Cont-2D	0.185 ± 0.028	0.193 ± 0.029	0.000	0.000
M5	0.0011 ± 0.0002	0.0028 ± 0.0004	0.000	0.000

nonoverlapping equal segments ($L = N/K$ for every tested length). In segment k, each sensor sequence is multiplied by the symmetric length-L Hann window and transformed with $N_{\mathrm{FFT}} = \max(L, 64)$, without DFT-amplitude normalization. The retained bin is $b = \mathrm{round}(f N_{\mathrm{FFT}}/f_s)$. Stacking the complex bin values gives $\mathbf{X}_f \in \mathbb{C}^{60\times 20}$, and the covariance is

$$\widehat{\mathbf{R}} = \frac{1}{20}\mathbf{X}_f\mathbf{X}_f^H. \tag{13}$$

No temporal demeaning or diagonal loading is applied.

The number of sources is fixed to one. If $\widehat{\mathbf{R}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$ has ascending eigenvalues, the first 59 eigenvectors form $\mathbf{U}_n$ and $\mathbf{P}_n = \mathbf{U}_n\mathbf{U}_n^H$. The scanned quantity is the normalized denominator

$$D_N(\theta) = \frac{\mathbf{a}(\theta)^H\mathbf{P}_n\mathbf{a}(\theta)}{\mathbf{a}(\theta)^H\mathbf{a}(\theta)}. \tag{14}$$

The nominal scan contains 721 uniform azimuths from −90° to 90° at fixed zero elevation. After selecting the grid minimum, a parabola is fitted to the three adjacent values of D_N, not to its reciprocal. The refined offset is clipped to one grid step; an endpoint, nonpositive curvature, or nonuniform neighborhood returns the grid point. Actual-manifold oracle MUSIC uses the identical extractor, covariance, scan, and refinement, changing only the steering vector.

VII. E7 ALGEBRA AND NUMERICAL CLOSURE

For the notation of the main paper, the two exact vector identities are

$$\mathbf{E}_{\mathrm{tot}} = \mathbf{E}_{\mathrm{rec},\perp} + \mathbf{E}_{\mathrm{nom}}, \tag{15}$$

$$\mathbf{E}_{\mathrm{nom}} = \mathbf{E}_{\mathrm{mis}} + \mathbf{E}_{\mathrm{oracle}}. \tag{16}$$

For either identity $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$, the stored scalar decomposition is

$$\|\mathbf{E}\|_F^2 = \|\mathbf{E}_1\|_F^2 + \|\mathbf{E}_2\|_F^2 + 2\,\mathrm{Re}\langle\mathbf{E}_1, \mathbf{E}_2\rangle_F. \tag{17}$$

The signed normalized components divide all three right-hand terms by $\|\mathbf{E}\|_F^2$ only when the denominator is nonzero. Alignment is $\mathrm{Re}\langle\mathbf{E}_1, \mathbf{E}_2\rangle_F/(\|\mathbf{E}_1\|_F\|\mathbf{E}_2\|_F)$ only when both factors are nonzero. Undefined quantities are stored as JSON null, never NaN or infinity.

The full audit covers 78,000 E7 records. The maximum signed MUSIC increment closure and linear-prediction closure are both 7.1054×10^{-15} degrees. The maximum vector identity residual norm is 1.8133×10^{-12}. No record is geometrically degenerate and no linear bearing prediction is null. These checks are numerical quality gates, not evidence that the first-order approximation is accurate.

The nonlinear MUSIC increments are defined independently of the tangent linearization. Let $\mathcal{M}_N(\cdot)$ and $\mathcal{M}_A(\cdot)$ denote the frozen nominal- and actual-manifold MUSIC maps, $\mathbf{X}^{\mathrm{true}}$ the truth snapshot matrix, $\widehat{\mathbf{X}}$ the recovered snapshot matrix, and θ_0 the generating azimuth. Then

$$\begin{aligned}
\theta_{\mathrm{true,N}} &= \mathcal{M}_N(\mathbf{X}^{\mathrm{true}}),\\
\theta_{\mathrm{est,N}} &= \mathcal{M}_N(\widehat{\mathbf{X}}),\\
\theta_{\mathrm{true,A}} &= \mathcal{M}_A(\mathbf{X}^{\mathrm{true}}),\\
\Delta\theta_{\mathrm{rec}}^{\mathrm{MUSIC}} &= \theta_{\mathrm{est,N}} - \theta_{\mathrm{true,N}},\\
\Delta\theta_{\mathrm{nom}}^{\mathrm{MUSIC}} &= \theta_{\mathrm{true,N}} - \theta_0,\\
\Delta\theta_{\mathrm{tot}}^{\mathrm{MUSIC}} &= \theta_{\mathrm{est,N}} - \theta_0,\\
e_{\mathrm{oracle}}^{\mathrm{MUSIC}} &= \theta_{\mathrm{true,A}} - \theta_0.
\end{aligned} \tag{18}$$

Consequently, $\Delta\theta_{\mathrm{tot}}^{\mathrm{MUSIC}} = \Delta\theta_{\mathrm{rec}}^{\mathrm{MUSIC}} + \Delta\theta_{\mathrm{nom}}^{\mathrm{MUSIC}}$ exactly up to floating-point arithmetic. The oracle error is reported as a simulator/extractor check and is not inserted as a third term in this nonlinear identity.

VIII. E7 INFERENTIAL BOUNDARY

Within a fixed profile, amplitude, azimuth, and checkpoint, the nominal component is deterministic. Its regression predictor has zero Monte Carlo variance; slope, intercept, and R^2 are consequently unidentified and its bootstrap interval degenerates. The manuscript therefore reports the nominal term through signed point comparisons across fixed tested conditions. Recovery and total terms retain fixed-condition/checkpoint slopes, intercepts, R^2, MAE, sign agreement, and paired-bootstrap 95% intervals.

TABLE SV

Frozen mismatch-profile magnitudes. "Std." denotes the generating scale before realization; gain and fading scales are log-domain.

Component	D1-G	D1-P	D2-J	Units / role
Radius / pitch offset	+2%/+5%	–	+5%/+10%	Fixed systematic offsets
Position / tangent jitter std.	1 mm / 0.5°	–	3 mm / 1.5°	Sensor-level geometry scales
Sound-speed offset	–	+1.5%	+4%	Fixed relative offset from 1500 m/s
Response gain / phase std.	–	0.03 / 0.02 rad	0.08 / 0.08 rad	Sensor-level fixed calibration draw
Spatial noise correlation ρ	–	–	0.50	Covariance $\rho^{\lvert i-j \rvert}$ for I, Q, and drift innovations
Terminal drift scale	–	–	0.08 rad	Cumulative spatially correlated random walk
Carrier-fading log std.	–	–	0.15	Fixed sensor-level lognormal gain
Dropout fraction / retained gain	–	–	5% / 0.05	Three of 60 sensors selected without replacement
IQ gain imbalance / quadrature error	–	–	0.08 / 3°	I/Q scale and phase nonideality

TABLE SVI

Representative E7 condition means. Signed triples are recovery/nominal/total errors in degrees.

Condition	Instance	Phase RMSE	Branch-mismatch density	MUSIC triple	Linear triple
D0, $A = 3$, 30°	M3	0.071	0	−0.001/−0.000/−0.001	−0.001/−0.000/−0.001
D0, $A = 10$, 30°	M3	7.559	0.5863	2.100/−0.000/2.100	−0.978/−0.000/−0.978
D0, $A = 10$, 30°	M5 seed 17	0.071	0	−0.000/−0.000/−0.000	−0.000/−0.000/−0.000
D1-G, $A = 3$, 30°	M3	0.071	0	−0.001/13.279/13.278	−0.001/13.122/13.121
D1-G, $A = 10$, 30°	M4 seed 17	3.623	0.0987	−12.584/13.279/0.694	−12.995/13.122/0.127
D2-J, $A = 3$, 30°	M5 seed 17	0.315	0.0007	0.011/31.960/31.971	0.891/26.621/27.512
D2-J, $A = 10$, 30°	M3	8.051	0.6550	−94.506/31.960/−62.546	−32.208/26.621/−5.587

For each eligible group, ordinary least squares with an intercept regresses the observed signed MUSIC component on its corresponding first-order prediction. Slope and intercept are undefined when predictor variance is at or below floating-point epsilon; R^2 is additionally undefined when observed variance is zero. MAE is the mean paired absolute discrepancy. Sign agreement excludes only a pair in which both values are exactly zero; a pair with exactly one zero is counted as disagreement. Each 95% interval is the 2.5th–97.5th percentile of 2,000 resamples of the 300 paired record indices with replacement. Resampling is confined to one condition and one method/checkpoint. A base seed of 20260717 and a deterministic SHA-256-derived group token define independent reproducible bootstrap substreams; no records are pooled across groups.

Among 150 D0/D1-G/D1-P condition–checkpoint groups with no observed record-level exceedance of the fixed branch-mismatch threshold (threshold-clean), the median recovery and total prediction MAEs are 7.93×10^{-4} and 1.19×10^{-3} degrees; median sign agreement is 1.0 and median R^2 is 0.99995. Among 50 non-D2 groups containing phase failures, median recovery and total MAEs are approximately 1.70°, 29 groups exceed 1° MAE, and ten have negative slope. Among 60 D2-J groups, median recovery MAE is 1.93°, recovery sign agreement is 0.048, total MAE is 4.40°, and total R^2 is approximately 0.40 over the 50 groups with defined R^2 (the remaining ten are undefined and excluded rather than zero-filled). These groups define three descriptive regimes rather than post hoc acceptance thresholds.

Table SVIII gives auditable examples from the three regimes. Each interval resamples the 300 paired Monte Carlo seeds within exactly one condition and one locked instance. The complete 260 condition–instance rows, including all three learned checkpoints separately, are retained in `E7_derived_summary.json` and `E7_condition_method_summary.csv`; no cross-condition bootstrap is used.

TABLE SVII

E7 descriptive no-exceedance medians by condition. The strata have zero observed record-level exceedances of the primary branch-mismatch threshold; they are not proofs of zero failure probability.

Statistic	D0	D1-P	D1-G
Groups	60	45	45
Recovery-prediction MAE (°)	5.95×10^{-5}	6.01×10^{-4}	0.0982
Total-prediction MAE (°)	7.30×10^{-5}	1.19×10^{-3}	0.291
Recovery sign agreement	1.000	0.997	0.867
Total R^2	0.999998	0.999902	0.828

IX. D1-G Geometry-Realization Distribution

The main E7/E8 matrices keep the archived D1-G profile seed fixed. To test whether its approximately 13° nominal bias is realization-specific, an additional deterministic sensitivity run used 50 consecutive, independent profile seeds (20260715–20260764), excluding the original seed 20260714. The +2% radius and +5% pitch offsets remained fixed systematic stress levels; each seed resampled the sensor-level 1-mm position and 0.5° tangent perturbations implemented by the frozen mismatch generator. No learned method or noise Monte Carlo was run. For each profile, noiseless true phase was passed through the same 20-segment extractor and 721-point nominal MUSIC scan at the three fixed azimuths 0°, 30°, and 60°. The primary profile statistic was the median absolute nominal bias across those azimuths.

TABLE SVIII
REPRESENTATIVE FIXED-CONDITION/CHECKPOINT E7 REGRESSION RESULTS. EACH ENTRY IS ESTIMATE [PAIRED-BOOTSTRAP 95% CI]; MAE IS IN DEGREES. R AND T DENOTE RECOVERY AND TOTAL COMPONENTS.

Condition	Instance	Component	Slope	Intercept	R^2	MAE	Sign agreement
D0, $A = 3$, 30°	M3	R	1.00005 [0.99991, 1.00019]	-3.34×10^{-7} [-1.68×10^{-6}, 1.07×10^{-6}]	0.9999983 [0.9999979, 0.9999987]	9.47×10^{-6} [8.63×10^{-6}, 1.03×10^{-5}]	1.000 [1.000, 1.000]
		T	1.00005 [0.99991, 1.00020]	-3.99×10^{-5} [-4.12×10^{-5}, -3.86×10^{-5}]	0.9999983 [0.9999979, 0.9999987]	4.00×10^{-5} [3.86×10^{-5}, 4.13×10^{-5}]	1.000 [1.000, 1.000]
D1-G, $A = 10$, 30°	M4 seed 17	R	5.253 [5.112, 5.417]	55.677 [53.843, 57.804]	0.9935 [0.9922, 0.9955]	0.661 [0.584, 0.746]	1.000 [1.000, 1.000]
		T	5.253 [5.109, 5.416]	0.028 [0.006, 0.050]	0.9935 [0.9922, 0.9955]	0.714 [0.635, 0.803]	1.000 [1.000, 1.000]
D2-J, $A = 10$, 30°	M3	R	−0.552 [−1.006, −0.060]	−112.29 [−126.90, −96.38]	0.0217 [0.0005, 0.0739]	62.298 [62.195, 62.404]	1.000 [1.000, 1.000]
		T	−0.552 [−1.034, −0.078]	−65.63 [−68.30, −62.95]	0.0217 [0.0005, 0.0789]	56.959 [56.855, 57.057]	1.000 [1.000, 1.000]

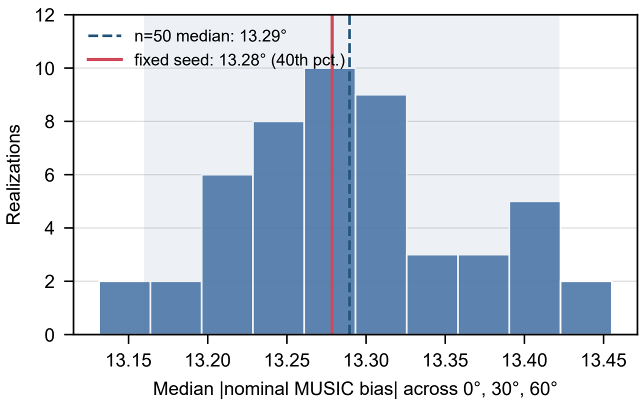

Fig. S1. Empirical D1-G geometry-realization distribution ($n = 50$). Each bar counts the three-azimuth median absolute nominal MUSIC bias for one independently resampled position/tangent perturbation profile. The systematic radius and pitch offsets remain fixed. Shading marks the empirical 2.5th–97.5th percentiles, not a confidence interval; the red line is the original fixed profile.

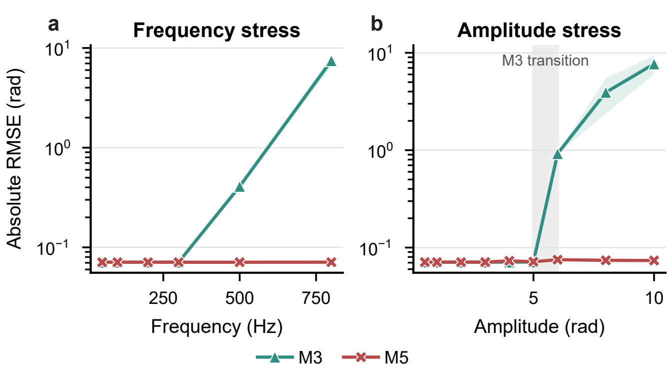

Fig. S2. Observed frequency and amplitude transitions at 20 dB. The shaded 5–6 rad interval is defined by tested grid points for the implemented M3 instance; neither panel estimates a general recovery boundary.

Across the 50 profiles, this statistic has median 13.2896°, sample SD 0.0728°, range 13.1314°–13.4552°, and empirical 2.5th–97.5th percentiles 13.1599°–13.4222°. The percentile range describes the sampled realization distribution; it is not a confidence interval. The original profile has 13.2786°: 20 of the 50 resampled values are smaller, placing it at the 40th empirical percentile. Its three azimuth-level biases also reproduce the archived oracle values within 1.5×10^{-13} degrees.

X. E8 SHORT-RECORD EXTRACTOR SEMANTICS

The $N = 100$ E8 condition is a locked-pipeline stress rather than an isolated record-length intervention. With 20 segments, its segment length is five and the symmetric Hann window is $[0, 0.5, 1, 0.5, 0]$; zero-padding to 64 points gives the retained bin $f_b = 187.5$ Hz for the 200-Hz target. For a real sinusoid with sensor phase ψ_i, its retained coefficient has the form

$$X_i(f_b) = \frac{A_i}{2j}\left[e^{j\psi_i}W(f_b - f_0) - e^{-j\psi_i}W(f_b + f_0)\right], \quad (19)$$

where $W(\nu) = \sum_{n=0}^{4} w_n e^{-j2\pi\nu n/f_s}$. Under the implemented values, $|W(f_b + f_0)|/|W(f_b - f_0)| \approx 0.673$. The retained snapshot can therefore mix the nominal steering term with its conjugate; zero-padding does not remove this contribution. The E8 result is accordingly specific to the fixed segment count, real-signal Hann extraction, target-bin rule, padding, and MUSIC scan. Exact-frequency or analytic-signal extraction and alternative segment counts were not tested, so E8 does not identify a general causal effect of record length alone.

XI. E8 THRESHOLD SENSITIVITY

The primary thresholds are branch-mismatch density 10^{-3} and absolute DOA error 1°. Table SIX shows why sensitivity and checkpoint identity matter. At the short matched hard condition, M5 seeds 17 and 29 label every Monte Carlo record as a phase failure at 10^{-3} while no record is a DOA failure even at 0.5°. Raising the slip threshold to 10^{-2} removes that label. Under D1-G, all three checkpoints are DOA failures at every reported threshold despite little phase failure. Under D2-J, every checkpoint is a DOA failure and the phase label itself depends on the slip threshold for seed 43.

For a fixed checkpoint with $n = 300$, an observed failure rate of zero has Wilson 95% interval approximately $[0, 0.0126]$, and a rate of one has interval approximately $[0.9874, 1]$. Accordingly, the text uses "zero observed failures" rather than "failure probability is zero."

XII. E7/E8 PHASE–TASK DISCORDANCE CENSUS

The census groups records only by the fixed `condition_id×method_instance_id` key. Every group contains 300 records. A phase-clean/task-fail cell has zero stored phase failures and 300 stored DOA failures; a phase-fail/task-success cell has the reverse counts. The stored flags were rechecked record by record against slip density $> 10^{-3}$ and absolute DOA error $> 1°$ before aggregation.

Both directions therefore recur in the locked matrices: branch-clean recovery can coexist with systematic task failure under nominal-manifold mismatch, and branch failure can

TABLE SIX
M5 checkpoint-specific E8 sensitivity. Entries are empirical failure rates over 300 paired Monte Carlo seeds.

Condition	Checkpoint	DOA MAE	DOA> 0.5°	DOA> 1°	DOA> 2°	Slip> 10^{-4}	Slip> 10^{-3}	Slip> 10^{-2}
D0, $A = 10$, $N = 100$	seed 17	0.0462	0	0	0	1	1	0
	seed 29	0.0462	0	0	0	1	1	0
	seed 43	0.0602	0	0	0	0.997	0	0
D0, $A = 10$, $N = 10000$	seed 17	0.0010	0	0	0	0	0	0
	seed 29	0.0013	0	0	0	0	0	0
	seed 43	0.0010	0	0	0	0	0	0
D1-G, $A = 10$, $N = 2000$	seed 17	13.2777	1	1	1	0	0	0
	seed 29	13.2798	1	1	1	0	0	0
	seed 43	13.2807	1	1	1	0.063	0	0
D2-J, $A = 10$, $N = 2000$	seed 17	31.9780	1	1	1	1	1	0
	seed 29	31.9150	1	1	1	1	1	0
	seed 43	30.6506	1	1	1	1	1	1

TABLE SX
Descriptive census of unanimous phase/task-discordant locked cells. Counts are design cells, not independent population samples or occurrence probabilities.

Suite	Records	Cells	Phase-clean/task-fail	Phase-fail/task-success	Profile breakdown
E7	78,000	260	53	3	45 D1-G + 8 D2-J; 3 D0 opposite
E8	43,200	144	17	2	16 D1-G + 1 D2-J; 2 D0 opposite

coexist with task success in selected matched cells. This is a descriptive audit of designed conditions, not a cross-cell inferential test. The complete discordant-cell list appears in Table **??**; all 404 cells, including mixed outcomes, are archived in `E7_E8_phase_task_census_cells.csv`. The companion JSON records the thresholds, group key, derivation command, and SHA-256 hashes of both full-record inputs.

XIII. Classical Method Implementation Details

M0 returns the noisy principal phase $\tilde{\phi}_w$ without modification. M1 operates independently on each sensor IQ sequence: it computes $\angle(z_{t-1}^* z_t)$, cumulatively sums these adjacent increments, and anchors the first output sample at zero; it uses no smoothing. M2 is likewise sensorwise and uses a scalar random-walk Kalman recursion with $x_0 = 0$, $P_0 = 1$, process variance $Q = 0.01$, and measurement variance $R = 0.1$. Its innovation is wrapped as $\mathcal{W}(\tilde{\phi}_{w,t} - x_{t|t-1})$ before the ordinary scalar update.

M3 calls `skimage.restoration.unwrap_phase` (scikit-image version $\geq$ 0.22 in the frozen requirements) once on the complete $60 \times N$ noisy principal-phase array, using the library defaults. Boundary wrap-around is disabled, so sensors 0 and 59 are not treated as neighbors. No method is retuned by condition, profile, amplitude, azimuth, or record length.

XIV. Implemented Learned Instances and Training

Table SXI records the implemented architectures rather than proposing a generic model family. M4 operates on each sensor sequence independently. The two 2-D instances share the same U-Net-style backbone and differ only in output representation and loss.

The seven 2-D input channels are $\cos\tilde{\phi}_w$, $\sin\tilde{\phi}_w$, wrapped temporal and spatial differences, normalized IQ-amplitude reliability, the wrapped Kalman guide, and its scaled coarse-count guide. The sensor axis is reflection-padded from 60 to 64 only for the pooling path and cropped back to 60 at output. Windows have length 256 and stride 128. If $N < 256$, the existing time record is right-reflection-padded to 256; no future observations are acquired. For longer records, the start-index list includes a final right-aligned window whenever the regular stride would miss the endpoint. Continuous-2D phase outputs are averaged over overlapping windows. M5 class probabilities are averaged first and the final class is then selected by argmax. Outputs are cropped to the original record length.

For M4, the locked objective is

$$\mathcal{L}_{\mathrm{M4}} = \mathrm{MSE}(\hat{\phi}, \phi) + 0.1\ \mathrm{MSE}\big(\mathrm{clip}(\Delta_t \hat{\phi}, -\pi, \pi) - \Delta_t \hat{\phi}, 0\big)\ . \tag{20}$$

For wrap-count output, let $\tilde{k} = \sum_c c\ \mathrm{softmax}(l)_c$ and $\tilde{\phi} = \tilde{\phi}_w + 2\pi\tilde{k}$. Its objective is

$$\begin{aligned}\mathcal{L}_{\mathrm{wrap}} = \mathcal{L}_{\mathrm{focal}}^{\gamma=2} &+ 0.25\ \mathrm{SL1}(\tilde{\phi}, \phi) \\ &+ 0.05\big[\mathrm{SL1}(\Delta_t \tilde{k}, \Delta_t k) + \mathrm{SL1}(\Delta_s \tilde{k}, \Delta_s k)\big]\ .\end{aligned} \tag{21}$$

The continuous-2D control replaces the focal/count terms with $\mathrm{SL1}(\hat{\phi}, \phi)$ plus 0.1 times the mean temporal and spatial phase-gradient SL1 losses.

All learned instances use the same 40,000 training seeds (0–39,999) and 5,000 validation seeds (100,000–104,999). The 2-D runs use AdamW, learning rate 3×10^{-4}, weight decay 10^{-4}, batch size 16, at most 80 epochs, cosine scheduling, mixed precision during training, and patience 12. M4 uses the same sample sets and stopping rule with AdamW, learning rate 10^{-3}, batch size 16, and per-sensor sequence flattening. A checkpoint minimizes validation globally aligned RMSE; a numerical tie is broken by boundary F1. Test data do not select checkpoints.

Each seed deterministically specifies one 60×256 training field. The source-count draw is single-component with probability 0.8 and dual-component with probability 0.2. One of three SNR strata, $[0, 10]$, $[10, 20]$, or $[20, 40]$ dB, and one of three amplitude strata, $[0.5, 2]$, $[2, 5]$, or $[5, 10]$ rad, are

TABLE SXI
IMPLEMENTATION-ESSENTIAL ARCHITECTURE SPECIFICATION. GN DENOTES EIGHT-GROUP NORMALIZATION; CONVOLUTIONAL BLOCKS USE SILU ACTIVATION AND A RESIDUAL PROJECTION.

Instance	Input	Encoder/context/decoder	Output	Parameters
M4 fair	$[\cos\tilde{\phi}_w, \sin\tilde{\phi}_w]$ per sensor	2-layer LSTM, hidden 64; linear head	continuous phase	50,753
Continuous-2D	seven-channel 60×256 field	32/64/128/256 GN blocks; 1, 2, 4, 8 dilated context; decoder	one channel	2,670,273
M5 wrap-2D	same seven-channel field	same backbone	logits for $k \in \{-2, -1, 0, 1, 2\}$	2,670,405

selected deterministically from the seed; the value is uniform within the selected stratum. Frequency is uniform on 50–800 Hz, azimuth on 0°–90°, and elevation on −30°–30°. In a dual-component field, the second frequency is independently uniform on 50–800 Hz, its mixture weight is uniform on 0.3–0.7, and initial phases are independent uniform draws on $[-\pi, \pi]$; both components share the sampled direction. The training generator contains no D1-G, D1-P, or D2-J mismatch. M4 reshapes the same field examples into sensor sequences, whereas the 2-D methods retain the joint field; their seed-level data distributions are therefore shared even though their input organizations differ.

Class weights use the first 1,000 training seeds. Across 15,360,000 labels, the supports for $[-2, -1, 0, 1, 2]$ are $[851, 751124, 13854913, 752192, 920]$, with frequencies $[5.54 \times 10^{-5}, 0.04890, 0.90201, 0.04897, 5.99 \times 10^{-5}]$. The clipped inverse-square-root weights are $[2.455, 0.5, 0.5, 0.5, 2.361]$. Every implemented class has nonzero training support, but the outer classes are rare and the finite alphabet remains an explicit extrapolation limit.

XV. Implemented Factor Controls

The historical CE-only secondary control is reported transparently but is not used to replace the locked main instance after test inspection. At its matched $A = 3$ rad, 20-dB, 200-Hz, 30° condition, the single CE-only checkpoint has absolute RMSE 0.071046 rad, branch-mismatch density 5.33×10^{-7}, wrap-count macro-F1 0.9697, boundary F1 0.9393, and DOA RMSE 0.009680°. The corresponding three-checkpoint structured instance has 0.070893 rad, zero slips, F1 values 1.0, and 0.009661°. These easy-condition differences are practically small and the CE-only row is supporting, not a new factor claim. Its copied source is `E4_secondary_ablations_full_corrected.json`. The factor table shows that 2-D context and output representation are not isolated by a single M4–M5 comparison. It also does not establish a universal 2-D gain because the 1-D and 2-D implementations are not capacity matched.

XVI. E8 Timing and Resource Semantics

Each condition/method has 100 raw timing calls after ten warm-ups. Learned calls synchronize CUDA immediately before and after timing. At $N = 10000$, M4 has approximately 3.47 GB peak GPU memory, while continuous-2D and M5 use approximately 0.09–0.11 GB. This is an implementation comparison between a full recurrent sequence and overlapping windows, not an architecture-only or parameter-count result. CPU-only methods store GPU memory as not applicable.

All reported RTF values are below one, but RTF is defined from offline median processing time divided by N/f_s. It excludes acquisition, padding, transfer, model loading, and external decision logic. At 2 kHz, a native 256-sample window spans 0.128 s. For $N = 100$, however, the implementation uses only the available 0.05-s record and reflection-pads it internally; 0.128 s is therefore a model-window horizon, not an acquisition lower bound for that tested path.

The E8 payload records CUDA execution for learned methods, float32 inputs, one record per call, synchronization, and all raw timings; classical methods run on the host CPU. The exact GPU and CPU models, Python/PyTorch/CUDA versions, and thread settings were not captured in the E8 run manifest. Training manifests contain a separate RTX 4080 SUPER environment record, but that record is not used to infer the E8 timing device. Hence the latency values support within-run implementation tradeoffs only and are not portable hardware benchmarks.

XVII. AI-Assisted Figure Provenance

The graphical abstract was drafted with OpenAI GPT Image 2 and is the only AI-assisted image in this submission; every numbered figure in the manuscript and in this supplement is generated directly from the archived runs by the regeneration scripts. The archived graphical-abstract prompt requested a horizontal white-background scientific diagram with five left-to-right modules: helical DAS array, coherent IQ observation, phase recovery, branch-aware evaluation, and MUSIC-DOA plus engineering outputs. It specified restrained flat-vector styling, readable labels after two-column reduction, and prohibited gradients, watermarks, decorative backgrounds, and pseudo-formulas. After generation, the authors replaced malformed symbols and labels, aligned module spacing and arrows, removed method-promotional emphasis, and manually checked every scientific label against the manuscript. The full prompt is archived as `fig1_system_overview_prompt.md`; no generated pixel value enters a numerical result.

XVIII. Reproducibility Records and Availability

The asset manifest additionally records parent assets, frozen configuration hashes, checkpoint hashes, generation commands, byte counts, and evidence status. The research package also retains the E7/E8 configuration overlays, experiment and derivation scripts, locked checkpoints, record-level full-run outputs, and commands needed to regenerate the derived tables. The SHA-256 values establish the identities of those local assets; they do not by themselves provide public access. No repository URL or archival DOI has yet been assigned.

TABLE SXII
LOCKED FACTOR CONTROLS AT 20 dB AND 200 Hz. VALUES ARE ABSOLUTE PHASE RMSE IN RADIANS, REPORTED AS MEAN OVER THREE CHECKPOINT MEANS WITH THEIR SD.

Variant	Context	Output	$A = 3$	$A = 10$
M4 fair	1-D temporal	continuous	0.06264 (0.00549)	3.1507 (0.2962)
Continuous-1D control	1-D temporal	continuous	0.07036 (0.01184)	3.7249 (0.1555)
Wrap-1D control	1-D temporal	wrap count	0.07089 (0)	4.0449 (0.0940)
Continuous-2D	2-D sensor–time	continuous	0.00722 (0.00041)	0.06667 (0.00477)
M5 wrap-2D	2-D sensor–time	wrap count	0.07089 (0)	0.07364 (0.00444)

TABLE SXIII
PRIMARY FULL-RUN, REALIZATION, AND IMPLEMENTATION HASHES.

Identity	SHA-256
E7 full output	`6d93b5149212967ceb8ced5b670b9ae9` `4a205b6b451f0607672c60456bdf5934`
E8 full output	`9f9c1d5865ab2ff044ed624c7dc1ab9` `10ac54d94de8375c4ae00d6bcb06e64fd`
D1-G $n = 50$ records	`778a1c0fc4eda82f9a4343b608f880ba` `5207eae745fbc6322e4ed6b02d759147`
Experiment code	`f45cd2ee64d2b60c7a5835d6d77b93c` `f9e26c5af5ced18c6a45818cbbe8b26c7`
E7 overlay	`1413c2fc9adcfb41e65493fe28240914` `a0c97def4b43088f714a0a5a13d360de`
E8 overlay	`525404eaca50a74493085bb0dbb66bbd` `02228c8500a654000981081b4ce08e35`
Census cells	`9b841870b549528010187216a1cc02c3` `36fe072a2feca928b6e6ef6398173ffc`
Census summary	`f8513c21efd5b255f291e390324dd908` `a39a81f07dcf55c226a4e9dce87d13d2`
Census script	`820cca8a688183646ccde984b784e01c` `4c9bbac7a8bbe4bb6705f34aa9e47469`

A stable public locator should therefore be added to the final data/code-availability statement before public release.