

Angular Watermarking: Sharp Leakage Bounds and Public Subspace Validation

Bin He

Southeast University
50898115@qq.com

Abstract. Fresh payloads need not hide persistent watermark geometry. For a Gaussian latent model with refreshed fair angular bits and isotropic covariance, we derive the exact fourth-order signature, the boundary of distributional hiding, and an attainable obstruction imposed by a uniform decoding margin. A fixed-probe learner has sufficient observation count $O(p^3 \log p / c^2)$ at fixed rank, accuracy and confidence, where c is the angular signal strength. Our central statistical result is a sharp, dimension-independent centered fourth-tensor covariance bound. It yields public finite-sample lower bounds on a frozen candidate’s carrier overlap, using independent validation directions, a carrier-dimension bound and scale information, without the secret key or payload. Exact partial-index spectra expose the tradeoff between geometric compression and validation variance. On eight fixed candidates, known- and unknown-scale lower bounds average 0.173 and 0.057, while oracle overlap averages 0.600. Population identities characterize feedback signal; a paired, budgeted MIPP comparison evaluates empirical training with independent Gaussian threshold calibration and a shared public spectral initialization control. All rank failures, zero-threshold sensitivities and negative results are reported. The guarantees concern latent subspaces under the stated iid model and exact arithmetic. They do not establish convergence of same-data adaptive training, ordered bit recovery or a neural watermark attack.

Keywords: Latent watermarking · Statistical leakage · Angular robustness · Subspace validation · Non-Gaussian component analysis

1 Introduction

Refreshing a watermark’s payload can remove a code-dependent statistic while leaving its secret carrier identifiable. For geometric latent marks, this raises two distinct questions: whether the observable distribution retains a carrier signature, and whether an observer without the key can validate a learned carrier. A large empirical overlap answers the second question only if one already knows the hidden subspace. We study both questions in an explicit angular model whose covariance is exactly isotropic.

Each output encodes fresh fair bits in fixed, disjoint four-dimensional blocks of a Gaussian latent complement. Every individual coordinate pair remains

Gaussian, yet relative-angle dependence leaves a calculable fourth harmonic. We recover its carrier from public unit directions and relate its strength to the decoder’s worst-case angular margin. At a boundary parameter the full law is clean even though noiseless decoding remains possible. Thus the security conclusion depends on the mechanism and the perturbation metric, rather than on payload refreshing alone.

Threat and observation model. The observer receives exact iid unit directions in a known public p -dimensional latent complement; the fixed carrier frame and refreshed payload bits are secret. The goal is to recover its subspace U or certify a frozen candidate’s overlap, not decode ordered bits. Recovery uses the public carrier dimension m ; validation permits a public upper bound on m and either known c or only $c^2 \leq 1$. Validation data are independent of the candidate’s selection. Noisy image inversion and content anisotropy lie outside this observation model.

Contributions. First, we identify the complete fourth-order signature, the exact hiding boundary and an attainable total-variation obstruction for any encoder using the same angle-sign decoder. A degree-six moment comparison and exact probe moments improve the fixed-rank sufficient observation count from quartic to cubic dimension dependence, up to a logarithm. A separate probe-norm event makes the sample budget fixed before observation. These counts remain conservative and are not necessary or minimax rates. For the noiseless hard-support family, a computationally unrestricted support estimator is consistent at $n = p^2$ for fixed carrier rank and interior angular parameter; a universal cubic necessary count would contradict this result (Theorem 14).

Second, a sharp centered covariance bound yields explicit public lower confidence bounds for an independently frozen candidate’s carrier overlap. Exact harmonic spectra determine the variance traces when one through four tensor indices are projected. One-index compression measures overlap without geometric loss; four-index compression supports low-dimensional streaming at the cost of certifying an energy surrogate. The empirical validation reports both losses, the random-frame reference, and the number of bounds that exceed it. The exact traces refine finite-sample constants without changing the asymptotic recovery order.

Third, an exact working-subspace decomposition identifies when feedback retains the full carrier signal. A rank-dependent noise formula explains what projection removes, while counterexamples show why average overlap or rank alone cannot ensure progress. The experiments separate fixed-probe recovery, adaptive cubic refinement and a budgeted non-polynomial MIPP comparison. They retain insufficient-rank fits, weak regimes and the failure of a noise-reducing variant at a near-compute budget. Both cubic and MIPP procedures belong to NGCA; the novelty is the model-specific geometry and validation, not adaptive projection pursuit itself.

Scope and reading guide. Sections 2–5.4 contain the angular model, guarantees and empirical tests. Appendices A–F give complete angular proofs and every angular experimental outcome; the global dependency map opens Appendix A. The separately compiled technical report *Radial Watermarking: Numerical Witnesses and Complete Experimental Records* retains the complementary radial case and failed image pilot. It supplies no premise for this paper.

Relation to earlier work. Fourth-cumulant dimension reduction and non-Gaussian component analysis (NGCA) supply the general extraction principle [14,13]. Iterative NGCA adaptation, including IMAK [17], also predates this work. Log-density-gradient NGCA [18] gives another relevant estimator family. Section 2 states the operator correspondence and the limit of directly transferring all-direction guarantees from [16]. Our novelty concerns the angular law, constants, decoder boundary and public validation. Section 5.4 compares the adaptive cubic procedure with a budgeted MIPP reimplementation of [13], using identical public observations and the exact radial lift of Appendix F. This is one specified comparison, not a claim of superiority to tuned or full-budget NGCA. Principal Hessian directions [3] and watermark-subspace estimation [10,4] provide further precedents. U-statistic decompositions and block/computation tradeoffs are established tools [19,20,21]; our certificate specializes them using exact angular covariance and overlap identities.

Latent schemes have different observation laws: Tree-Ring and RingID use Fourier-ring geometry [5,11]; Gaussian Shading uses quantile encoding [6]; LAW uses angular encoding [12]; and Shallow Diffuse uses neural-predictor structure [8]. In the pinned Gaussian Shading ChaCha reference path, per-image creation refreshes the bits, key and nonce [7]; the non-ChaCha branch also redraws its bits and mask. Reusing its Python object does not establish cross-image ciphertext reuse; the repeated-codeword regimes in the radial report are defined models, not an attribution to that path. Our zero-residual calculation covers LAW’s ideal antipodal primitive under fresh fair bits, not its full scheme, repeated payload or image channel. Pseudorandom-code watermarks [1,2] address a different cryptographic construction. Neither our radial model nor its public DCT inverse is claimed to instantiate these deployed systems.

In the specified endpoint family, a positive uniform angular margin leaves a recoverable fourth-order carrier signature despite fair bit refreshing. The four-index public certificate lower-bounds overlap through retained energy; its geometric loss persists with more validation data.

2 Angular margin, fourth-order leakage and recovery

We first isolate the effect of refreshing a geometric payload. The key is a fixed unknown orthogonal frame in a public p -dimensional complement. Its carrier comprises b_0 disjoint four-dimensional blocks, $m = 4b_0 < p$. Before that rotation, one block is

$$Z_j = (R_j \cos \Theta_j, R_j \sin \Theta_j, S_j T_j \cos(\Theta_j + \pi/2 + D_j), S_j T_j \sin(\Theta_j + \pi/2 + D_j)).$$

The two radii are independent standard Rayleigh draws; the reference phase is uniform; D_j is uniform on $[-\pi a, \pi a]$; and the payload S_j is a fair sign. All these variables, the blocks and the standard-Gaussian bulk are independent. Every output refreshes all signs independently, while the frame and pairing stay fixed. We study $0 \leq a \leq 1/2$, with $D_j = 0$ at zero. An angular contraction with wrapped error $\dot{e} = -4k_{\text{ang}}t(1-t)e$ has scale $a = \exp(-2k_{\text{ang}}/3)$; integrate $t(1-t)$ over $[0, 1]$. Independent Gaussian initial phases give a uniform wrapped initial error on $[-\pi, \pi]$; contraction therefore gives the stated uniform D_j . Relative-angle decoder boundaries are $0, \pi$ in this convention. The wrapped-noise calculation later subtracts $\pi/2$ from the relative angle; its shifted ϕ is a local coordinate, not a different decoder. The observable is the unit direction $q = Z/\|Z\|$ in public coordinates. A public orthogonal projection and normalisation suffice; no radial weight, key or payload enters the learner.

2.1 What survives fair refreshing

Theorem 1. *Put $c = c(a) = \sin(2\pi a)/(2\pi a)$, with $c(0) = 1$. Then $\mathbb{E}Z = 0$, $\mathbb{E}ZZ^T = I_p$, and $\mathbb{E}qq^T = I_p/p$. In each carrier block the only nonzero fourth cumulants, including index permutations, are*

$$\mathcal{T}_{1133} = \mathcal{T}_{2244} = \mathcal{T}_{1234} = -c/2, \quad \mathcal{T}_{1144} = \mathcal{T}_{2233} = c/2.$$

For any symmetric A , the public contraction

$$C(A) = p(p+2)\mathbb{E}[(q^T A q)qq^T] - \text{tr}(A)I_p - 2A \quad (1)$$

is supported on U , and $C(I_p) = 0$. For a GOE probe with off-diagonal variance one and diagonal variance two, $\mathbb{E}_A C(A)^2 = 6c^2 P_U$. If $c \neq 0$, one such probe has range U almost surely.

Proof. Uniform phases make each pair Gaussian and fair signs cancel cross-pair covariances. The endpoint preserves every original pair radius, so its total length is unchanged. Its direction depends only on the original direction and fresh bits, independently of that length. Thus $L = \|Z\|$ remains independent of q and $\mathbb{E}L^4 = p(p+2)$. This justifies the spherical correction in (1). Integrating the relative phase uses $\mathbb{E}\cos(2D) = c$ and gives the displayed tensor. Independent blocks and bulk give all other zeros; summing $\mathcal{T}_{ijkk}$ gives zero. The GOE covariance identity contracts it to $2\sum_{s,k,l} \mathcal{T}_{iskl}\mathcal{T}_{jskl} = 6c^2(P_U)_{ij}$. In one block $A_0 = \text{diag}(1, 0, 1, 0)$ gives $C(A_0) = (c/2)\text{diag}(-1, 1, -1, 1)$. Thus the determinant polynomial is nonzero, and its zero set has Gaussian measure zero in every block.

The support principle is a concrete NGCA instance, not a new general extraction method. For the operator $\beta(h) = \mathbb{E}[Zh(Z) - \nabla h(Z)]$ of [13], choosing $h_j(z) = (z^T A z)z_j$ gives $\beta(h_j) = C(A)e_j$ under isotropy. Fourth-cumulant dimension reduction is also explicit in [14]. The model-specific content here is the tensor, its vanished identity contraction, the sphere conversion, and the relation

to the decoding boundary. The direct moment proof also covers the singular endpoint $a = 0$. Table 1 separates these claims from general NGCA. In particular, the all-direction guarantee of [16, Theorem 9] requires non-Gaussian moments in every carrier direction. Our individual pairs are exactly Gaussian, so that theorem does not directly imply our full-carrier bound. Its one-direction result remains relevant; no general sample-complexity superiority over NGCA is claimed.

Table 1. Scope of the angular contribution relative to NGCA. The rates refer to different assumptions and estimators.

Method	Setting and established ingredient	Relation to this work
NGCA [13,14]	Non-Gaussian carrier plus Gaussian noise; Stein or cumulant contractions	Supplies the extraction principle; does not supply this angle-margin tensor or its explicit comparison constants.
Rewighted PCA [16]	Polynomial one-direction guarantee; its full-carrier theorem assumes moment identifiability in every carrier direction	Exact Gaussian pair marginals prevent direct use of that full-carrier theorem. No competing minimax claim follows.
This angular model	Unit directions, fixed disjoint blocks, refreshed fair signs	Exact boundary, degree-six comparison and explicit $p^3 \log p / c^2$ sufficient count for a specified fixed-probe learner.

2.2 An explicit finite-sample guarantee

For $\ell = 1, \dots, L_0$, subtract $\text{tr}(A_\ell)I_p/p$ from each independent GOE probe. This does not change $C(A_\ell)$. With $S_n = n^{-1} \sum_i q_i q_i^T$ and $F_n(A) = n^{-1} \sum_i (q_i^T A q_i) q_i q_i^T$, use

$$\hat{C}(A) = p(p+2)F_n(A) - p \text{tr}(AS_n)I_p - 2p(AS_n + S_n A) + 2A. \quad (2)$$

This is the unbiased control-variate estimator for trace-zero probes; Eq. (19) gives the general symmetric-probe formula. The learner selects the top m eigenvectors of $\hat{G} = L_0^{-1} \sum_\ell \hat{C}(A_\ell)^2$. All probes share the same observations; they are not independent data sets. Let $B = L_0^{-1/2} [C(A_1); \dots; C(A_{L_0})]$ stack contractions vertically, with $\hat{B}$ formed from their estimates.

For the realized trace-zero probes define the public quantities $a_2 = L_0^{-1} \sum_\ell \|A_\ell\|_{\text{op}}^2$ and $f_2 = L_0^{-1} \sum_\ell \|A_\ell\|_F^2$, and put

$$J_p = \frac{2p(p^2 + p + 4)}{(p+2)(p+4)}, \quad N_p = \frac{4(p^3 - p^2 + 10p - 8)}{(p+2)(p+4)}.$$

Theorem 2. Assume the angular law, $0 \leq a < 1/2$, iid observations and independent GOE probes. Let $\delta_A, \delta_q > 0$, $\delta_A + \delta_q < 1$, and choose L_0 so that g below is positive. Define

$$\begin{aligned} u &= \log(m/\delta_A), & g &= 6 - \sqrt{72u/L_0} - 4u/L_0, \\ V &= (1+c)(J_p f_2 + N_p a_2), & R &= (p^2 + 7p + 2 + 4c)\sqrt{a_2}, \\ x &= \log(p(L_0 + 1)/\delta_q), & \tau &= \sqrt{2Vx/n} + 2Rx/(3n). \end{aligned}$$

With joint probability at least $1 - \delta_A - \delta_q$, the following holds in exact arithmetic: if $\tau < c\sqrt{g}$, the top- m frame of $\hat{G}$ satisfies

$$o = \frac{1}{m} \text{tr}(\hat{Q}^T P_U \hat{Q}) \geq \left[1 - \left(\frac{\tau}{c\sqrt{g} - \tau} \right)^2 \right]_+. \quad (3)$$

The count n may be selected from the independent probes before observing data; the probability is over both probes and observations.

Proof. Wick contraction of the six Gaussian block modes gives $\mathbb{E}_A C(A)^4 = 72c^4 P_U$. One-sided matrix Bernstein applied to $6c^2 I_m - C(A)^2$ yields $\lambda_{\min}(G|_U) \geq c^2 g$ except with probability δ_A . For the one-observation summand H_A of (2), exact clean spherical integration gives $\mathbb{E}_{Q_0} H_A^2 = J_p \text{tr}(A^2) I_p + N_p A^2$. The key comparison is $\mathbb{E}_P f \leq (1+c)\mathbb{E}_{Q_0} f$ for every nonnegative polynomial of degree at most six on the sphere: averaging reference phases and fair signs leaves at most one block harmonic, whose amplitude cannot exceed its constant term. Apply this to $x^T H_A^2 x$, reduced to degree six using $\|q\| = 1$. For the stacked centered summand both rectangular variance operators are bounded by V , even with shared observations; its norm is at most R . Rectangular Bernstein [9] gives $\|\hat{B} - B\|_{\text{op}} \leq \tau$ except with conditional probability δ_q . Projecting the right singular-vector equation onto $U^\perp$ and using $\sigma_m(\hat{B}) \geq c\sqrt{g} - \tau$ proves (3). Appendix C gives all identities.

For planning, $c \leq 1$ can replace $1+c$ by 2 in V and $4c$ by 4 in R . It cannot remove the signal $c\sqrt{g}$ from the denominator: a nonvacuous uniform recovery count still requires a positive signal floor. This differs from the unknown-scale validation branch below.

Corollary 1 (A deterministic observation budget). Under the assumptions of Theorem 2, fix a target $\alpha_* \in (0, 1)$ and positive $\delta_A, \delta_q, \delta_{\text{norm}}$ with sum less than one. Fix L_0 so that the theorem's $g > 0$, and put

$$\begin{aligned} h &= 2\sqrt{p} + 2\sqrt{\log(2L_0/\delta_{\text{norm}})}, \\ \bar{V} &= (1+c)(pJ_p + N_p)h^2, & \bar{R} &= (p^2 + 7p + 2 + 4c)h. \end{aligned}$$

Let $x = \log(p(L_0 + 1)/\delta_q)$ and

$$\rho = \frac{\sqrt{1 - o_\star}}{1 + \sqrt{1 - o_\star}}, \quad b = \rho c \sqrt{g}, \quad n_\star = \left\lceil \left(\frac{\sqrt{2\bar{V}x} + \sqrt{2\bar{V}x + 8b\bar{R}x/3}}{2b} \right)^2 \right\rceil.$$

For each fixed unknown frame and each integer $n \geq n_\star$ fixed before drawing either probes or observations, the estimator satisfies $o \geq o_\star$ with joint probability at least $1 - \delta_A - \delta_q - \delta_{\text{norm}}$, in exact arithmetic. For fixed rank, target, failure budgets and a fixed L_0 with $g > 0$, $n_\star = O(p^3 \log p / c^2)$, uniformly for $0 < c \leq 1$.

Proof. For a trace-centered GOE probe, Gaussian comparison and concentration [15, Lemma 5.33 and Proposition 5.34] give $\Pr\{\|A\|_{\text{op}} > 2\sqrt{p} + s\} \leq 2e^{-s^2/4}$. Indeed, write $A = (G + G^T)/\sqrt{2} - \sqrt{2}\text{tr}(G)I_p/p$ and compare $X_v = \sqrt{2}\langle G, vv^T - I_p/p \rangle_F$ with $Y_v = 2z^T v$, $z \sim N(0, I_p)$. Their increments obey $\mathbb{E}(X_v - X_w)^2 = 4(1 - (v^T w)^2) \leq 4\|v - w\|^2$, while $G \mapsto \lambda_{\max}(A)$ is $\sqrt{2}$ -Lipschitz. Thus $\mathbb{E}\lambda_{\max}(A) \leq 2\sqrt{p}$; concentration for A and $-A$ proves the claim. A union bound gives $a_2 \leq h^2$ and $f_2 \leq ph^2$ outside failure δ_{norm} . On this event $V \leq \bar{V}$ and $R \leq \bar{R}$, and the displayed quadratic root ensures $\tau \leq b < c\sqrt{g}$. The overlap bound is then at least $o_\star$. Combine this event with the probe-gap and conditional observation events. Finally, $\bar{V} = O(p^3)$, $\bar{R} = O(p^{5/2})$, and $n_\star = O(\bar{V}x/b^2 + \bar{R}x/b + 1)$; since $0 < c \leq 1$, the second term is bounded by the stated order.

On the usual GOE norm scale, $V = O(p^3)$ and $R = O(p^{5/2})$. For fixed positive g and target loss, the leading sufficient count is $O(p^3 \log(p(L_0 + 1)/\delta_q)/c^2)$, improving the Frobenius–Markov $O(p^4/(mc^2))$ bound at fixed rank (Appendix B). At $m = 16, L_0 = 64, a = 1/4, \delta_A = \delta_q = 0.025$ and target overlap $1/2$, the Frobenius–Markov counts at $p = 120, 240, 480$ are respectively 4.63×10^{10} , 7.47×10^{11} and 1.20×10^{13} . Using independently drawn public probe norms gives approximately 7.43×10^8 , 6.15×10^9 and 5.13×10^{10} , reductions of 62, 121 and 234 times. These are sufficient counts, not necessary or minimax rates, and remain conservative. Ordinary floating-point evaluation of the formula is not a numerical certificate for an experimental frame. Near hiding, $c \sim 2\gamma/\pi$ retains the γ^{-2} dependence. Corollary 1 allocates an additional failure probability δ_{norm} to the GOE norm event and replaces a_2, f_2 by public deterministic bounds. It gives the same fixed-rank order at total failure $\delta_A + \delta_q + \delta_{\text{norm}}$. The displayed numerical counts instead use realized probes and the original two-event budget.

2.3 Exact hiding and the price of an angular margin

Theorem 3. *At $a = 1/2$, $Z \sim N(0, I_p)$ for every fixed frame, while the noiseless sign-of-relative-angle decoder recovers each embedded fair bit almost surely. Its uniform worst-case angular margin is zero. More generally, for any encoder using that decoder on b_0 fixed disjoint blocks, if all bits remain correct under every*

relative-angle perturbation smaller than γ with probability at least $1 - \varepsilon$, then, for $0 \leq \gamma \leq \pi/2$,

$$\text{TV}(P, Q_0) \geq [1 - \varepsilon - (1 - 2\gamma/\pi)^{b_0}]_+. \quad (4)$$

Here Q_0 is the clean Gaussian or unit-direction law. The endpoint family attains equality for $\varepsilon = 0$, $a = 1/2 - \gamma/\pi$.

Proof. At $a = 1/2$ the two fair-sign half-circles partition the full relative-angle circle, giving the clean joint law. Their interiors still determine the bit. For the general bound, let R_γ be the event that every relative angle is at least γ from the sign boundaries. Robust correctness implies $P(R_\gamma) \geq 1 - \varepsilon$, whereas independence of clean relative angles gives $Q_0(R_\gamma) = (1 - 2\gamma/\pi)^{b_0}$, unchanged by global normalisation. Apply the event definition of total variation. For the explicit family, $P = Q_0(\cdot \mid R_\gamma)$ when $a > 0$, so TV equals $1 - Q_0(R_\gamma)$; the singular $a = 0$ limit has TV one.

Thus this family joins exact hiding at zero uniform margin to constructive subspace leakage at every positive margin, with an explicit sufficient recovery bound. The TV witness knows the pairing and is not itself a blind test. The constructive learner recovers U , not the ordered pairs or an authenticated message. The perturbation metric is relative angle; no pixel or inversion-error guarantee follows from it.

3 Public validation of a learned carrier

Subspace overlap is normally evaluated using the secret carrier. We instead give a lower confidence bound computed from a candidate frame and new public directions. The candidate may come from the adaptive procedure below or any other training method. Only its independence from the validation sample is required; no convergence theorem for that training method is assumed.

3.1 A sharp centered covariance bound

Let Π_4 be the orthogonal projection onto symmetric trace-free fourth tensors in ambient dimension p , and define

$$H(q) = p(p+2)\Pi_4(q^{\otimes 4}), \quad \kappa_p = \frac{24p(p+2)}{(p+4)(p+6)}, \quad C_a = (1 + |c|)^2.$$

Then $\mathbb{E}H = \mathcal{T}$ and $\|\mathcal{T}\|_F^2 = 3c^2m$. All tensor covariances below act on the trace-free tensor space.

Theorem 4 (Sharp angular covariance). *For every fixed key in the exact angular model,*

$$\text{Cov}(H(q)) \preceq C_a \kappa_p I. \quad (5)$$

If there are at least two carrier blocks, its operator norm is exactly $C_a \kappa_p$. Under the uniform spherical law the covariance is $\kappa_p I$.

Proof (Proof outline). Decompose a fourth Gaussian Hermite polynomial by its degrees in the independent blocks. Conditional on the two block radii, first aggregate equal Fourier frequencies. The degree-two and degree-three phase groups have at most two distinct frequencies and norm bound $1 + |c|$; the degree-four groups have at most three and bound $1 + 2|c|$. After centering, the only coupling between different block-degree patterns is the same-pair $(3, 1) - (1, 3)$ term, of normalized norm $c^2/2$. The resulting centered fourth-chaos variance is at most $(1 + |c|)^2$ times its Gaussian variance, without a factor depending on the block count. The positive homogeneous, norm-preserving endpoint keeps the original χ_p radius independent of q . Radial conversion multiplies the variance by $\kappa_p/24$ and subtracts a nonnegative squared-mean term. The Gaussian chaos isometry proves (5). For sharpness in our range $c \geq 0$, use $(x_2x_3 - x_1x_4)/\sqrt{2}$ on each of two independent carrier blocks. Each factor has mean zero, Gaussian second moment one and angular second moment $1 + c$. Their product is harmonic, has tensor norm squared $1/24$, and attains the bound after radial conversion. Complete frequency grouping, cross-block contractions and the equality calculation are in Appendix D.

Centering is essential: $\mathbb{E}[H \otimes H]$ also contains $\mathcal{T} \otimes \mathcal{T}$, whose norm grows with m . Sharpness of the uniform covariance constant does not mean every candidate projection attains the worst-case variance.

3.2 An explicit finite-sample certificate

Fix $B \in \mathbb{R}^{p \times r}$ with orthonormal columns before drawing $n \geq 2$ iid validation directions, and set $P = BB^T$ and $o = \text{tr}(PP_U)/m$.

Public validation protocol. (1) Freeze B and a deterministic public bound $\bar{m} \geq m$ before validation. (2) Allocate a failure probability δ to each reported statement; for a family, sum allocations within its budget. (3) Read new iid directions and accumulate U_W in (7), including cross-batch pairs. (4) Use the exact dimension-only trace and known c , or $C_a = 4$ when only $c^2 \leq 1$ is available; return the clipped bound in Theorem 5 with denominator $\bar{m}$ (return zero at known $c = 0$). No secret key, payload or oracle score enters this procedure. The iid law and candidate independence are assumptions; arithmetic errors require a separate enclosure if a numerical certificate is desired.

Compress all four indices:

$$H_W = (B^T)^{\otimes 4} H, \quad \theta_W = 6 \|P^{\otimes 4} \mathcal{T}\|_F^2, \quad \rho_4 = \frac{\theta_W}{18c^2 m} \quad (c \neq 0).$$

The tensor identity $\sum_{a,b,d} \mathcal{T}_{iabd} \mathcal{T}_{jabd} = 3c^2 (P_U)_{ij}$ implies

$$[4o - 3]_+ \leq \rho_4 \leq o. \tag{6}$$

Indeed, the four coordinate projections commute, $P^{\otimes 4} \preceq P_j$ and $I - P^{\otimes 4} \preceq \sum_{j=1}^4 (I - P_j)$. Taking their quadratic forms at $\mathcal{T}$ proves both inequalities. If

$Q^T P Q = oI_m$, then $\rho_4 = o^4$: retained fourth-order energy is a conservative surrogate for overlap, not an unbiased overlap estimate.

Use the complete U-statistic

$$U_W = \frac{6}{n(n-1)} \left(\left\| \sum_i H_W(q_i) \right\|_F^2 - \sum_i \|H_W(q_i)\|_F^2 \right). \quad (7)$$

Theorem 5 (Public overlap certificate). *Assume the angular law and a candidate B fixed independently of the iid validation directions. Set $\tau_r = \tau_{p,r}^{(4)} = \text{tr}[(\Pi_4 P^{\otimes 4} \Pi_4)^2]$, $A_n = 24C_a \kappa_p / n$, and $B_n = 72C_a^2 \kappa_p^2 \tau_r / [n(n-1)]$. Then $\mathbb{E}U_W = \theta_W$ and $\text{Var}(U_W) \leq A_n \theta_W + B_n$. For $0 < \delta < 1$, put $h_\delta = (1 - \delta)/\delta$ and define*

$$\ell(u) = \begin{cases} 0, & u \leq \sqrt{h_\delta B_n}, \\ u + h_\delta A_n / 2 - \sqrt{h_\delta A_n u + h_\delta^2 A_n^2 / 4 + h_\delta B_n}, & u > \sqrt{h_\delta B_n}. \end{cases}$$

For known $c \neq 0$, with probability at least $1 - \delta$,

$$o \geq \rho_4 \geq \frac{\ell(U_W)}{18c^2 m}. \quad (8)$$

If c is unknown, use $C_a = 4$ throughout to obtain $o \geq \ell(U_W)/(18m)$. This gives a conservative bound also at $c = 0$.

Proof. Put $J = \Pi_4 P^{\otimes 4} \Pi_4$. It satisfies $0 \preceq J \preceq I$ and $\text{tr}(J^2) = \tau_r \leq \binom{r+3}{4}$. The correct norm identity is $\|H_W\|^2 = \langle H, JH \rangle$; J need not be a projection. The Hoeffding decomposition of (7) and Theorem 4 bound its linear and degenerate variances by $24C_a \kappa_p \theta_W / n$ and $72C_a^2 \kappa_p^2 \text{tr}(J^2) / [n(n-1)]$, respectively. Cantelli's inequality gives $U_W \leq \theta_W + \sqrt{h_\delta (A_n \theta_W + B_n)}$ except with probability δ . Invert this increasing function of $\theta_W \geq 0$ and apply (6). For unknown scale, $\theta_W \leq 18mo$ follows from $c^2 \leq 1$.

The exact trace is given by the harmonic spectrum below. Replacing it by any deterministic upper bound, including the dimension bound $\binom{r+3}{4}$, preserves coverage with a potentially smaller lower bound. The choice depends only on the fixed dimensions and operator, not on the observed validation score. A zero endpoint preserves coverage but certifies no positive overlap. If only $\bar{m} \geq m$ is known, replace the denominator m by $\bar{m}$: $\theta_W \leq 18c^2 mo \leq 18c^2 \bar{m} o$ preserves the same coverage event (use $18\bar{m} o$ for unknown scale). This certifies overlap normalized by the true m , with a potentially smaller lower bound. If the upper bound is random and $\Pr\{\bar{m} < m\} \leq \delta_m$, allocate total failure probability at least $\delta + \delta_m$; an unqualified dimension estimate does not suffice.

3.3 Partial-index geometry and exact variance traces

For $k = 1, \dots, 4$, let A_k average the $\binom{4}{k}$ commuting tensor operators that apply P in exactly k positions and the identity in the rest; put $L_k = \Pi_4 A_k \Pi_4$. For $c \neq 0$, the kernel $6\langle H(q), L_k H(q') \rangle$ has mean $\theta_k = 18c^2 m \rho_k$, where

$$[ko - (k-1)]_+ \leq \rho_k \leq o, \quad \rho_1 = o.$$

If $Q^T P Q = o I_m$, then $\rho_k = o^k$; this is a sufficient condition, not a universal identity. Its U-statistic obeys the same mean-dependent variance and confidence formula, with τ_r replaced by $\tau_{p,r}^{(k)} = \text{tr}(L_k^2)$. The gain in retained geometry at smaller k must be balanced against a larger variance trace and feature space.

Theorem 6 (Exact partial-index spectrum). *For $1 \leq r < p$, the harmonic decomposition associated with $\mathbb{R}^p = \mathbb{R}^r \oplus \mathbb{R}^{p-r}$ diagonalizes L_k . Its branches are indexed by $d+e+2j = 4$ and have multiplicities $h_d(r)h_e(p-r)$, where $h_d(t)$ is the degree- d harmonic dimension on $\mathbb{R}^t$. Their eigenvalues are explicit weighted averages of $\binom{d+2i}{k} / \binom{4}{k}$, $0 \leq i \leq j$; the complete weights, endpoint conventions and proof are given in Theorem 13. Summing squared eigenvalues with multiplicity gives $\tau_{p,r}^{(k)}$ exactly.*

For $p = 240, r = 16$, the exact $k = 4$ trace is $209990519484/54479161 \approx 3854.510892$, compared with the dimension bound 3876. The one-index trace also follows from Lemma 4; its inclusion in this unified spectrum is not separate improvement over that result. These trace reductions do not change the asymptotic order or establish a faster recovery algorithm.

Order-two U-statistic decompositions and block/computation tradeoffs are established tools [19,20,21]. The contribution here is the exact model constant, public compressed statistic and its quantitative link to carrier overlap. Equation (7) is evaluated from accumulated second and fourth moments of $B^T q$, including every cross-batch pair. It needs $O(npr + nr^4)$ operations and $O(pr + r^4 + \beta(p + r^2) + N_b)$ storage, where β is the maximum batch size and N_b is the number of saved scalar batch partials. This includes the candidate frame and batch quadratic features. At $r = 16$, symmetric quadratic features have width 136 and their Gram has only 136^2 entries. All trace corrections use ambient p . Since $\kappa_p < 24$, the validation variance avoids a cubic ambient dimension factor at fixed r ; computation still depends on p .

3.4 Validation of all eight primary outputs

All eight adaptive outputs of Section 5 were fixed before validation. Each received 65,536 fresh directions, for a total of 524,288. Allocating $\delta = 0.05/16$ to each known- and unknown-scale statement gives simultaneous coverage at least 95% by a union bound, conditional on the frozen outputs. Both trace bounds are evaluated on the same saved public statistics. The exact-trace calculation changes only a deterministic constant; it adds no observations, output selection or optional sample-size increase. The NGCA training panels use separate seeds. The panel-mean certificate losses are visible separately:

$$\underbrace{0.599854}_{\text{oracle overlap } o} \xrightarrow{\text{geometry}} \underbrace{0.301215}_{\text{oracle energy } \rho_4} \xrightarrow{\text{confidence}} \underbrace{0.173227}_{\text{public known-scale LCB}}.$$

Oracle scores are never validation inputs, and mean bounds do not predict future training success. No bound exceeds 0.5; five unknown-scale bounds and one

Table 2. Numerical evaluations of exact-arithmetic overlap bounds for all eight fixed outputs ($p = 240, r = m = 16$), using the exact trace. The sixteen statements retain the same 5% total failure budget. LCBs are rounded down to six decimals. The random-frame expectation is $r/p = 1/15$; a smaller lower bound does not imply a worse-than-random candidate.

Scale information	Minimum	Mean	Positive	Above r/p
Known $c = 2/\pi$	0.036199	0.173227	8/8	7/8
Unknown c , only $c^2 \leq 1$	0.003466	0.056994	8/8	3/8

known-scale bound lie below r/p . A lower bound below that expectation does not imply that the candidate itself is worse than a random frame. Removing compression with $k = 1$ increases the trace: at $c = 2/\pi$ and the same dimensions, sample count and failure allocation, any frozen candidate has probability at most 0.011613 of obtaining a positive one-index bound (Corollary 2). This is a theoretical power bound, not eight executed $k = 1$ measurements. Here the confidence penalty uses $\sqrt{(1 - \delta)/\delta} = \sqrt{319}$. Appendix D.9 gives a complementary sample-budget calculation conditional on a stated candidate-quality floor. A separate single-output validation gives a known-scale individual 95% bound of 0.219432 using another 65,536 directions. It uses an independent validation draw, so its difference from this familywise panel is not solely a confidence-budget or trace adjustment. Appendix D.12 reports every row and external oracle scores, which were never inputs to the validation API. Coverage is a theorem for the exact iid angular model and exact arithmetic. The implementation passes independent float64 tensor and streaming checks; it does not supply a directed-rounding enclosure or certify a neural image channel. The frontend certificate-evaluator uses the exact rational trace and records its constants and decimal precision. It reproduces all sixteen bounds; neither decimal evaluation nor the preceding float64 feature accumulation is enclosed.

4 Cubic features and subspace feedback

Cubic contractions $b(v) = \mathcal{T}(\cdot, v, v, v)$ offer a streamed alternative to dense quadratic probes. The population map below explains which carrier directions a working subspace can retain. We then specify the empirical update and test it with controlled feature budgets.

4.1 Population signal and its limits

Proposition 1 (Cubic population signal in a working subspace). *Let $Q \in \mathbb{R}^{p \times m}$ be the ordered carrier basis of the angular construction, and let $\tilde{T}$ be its stated block tensor. For any fixed rank- r orthogonal projector P_W , $r \geq 1$, put $M = Q^T P_W Q$ and $S_W = \mathbb{E}_{v \sim N(0, P_W)} b(v) b(v)^T$. Define the mode-one flattening*

$F_{i,(a,b,d)} = \tilde{T}_{iabd}$ and $(D_M)_{ia} = \sum_{b,d} \tilde{T}_{iabd} M_{bd}$. Then $S_W = P_U S_W P_U$ and

$$Q^T S_W Q = 6F(M^{\otimes 3})F^T + 9D_M M D_M^T. \quad (9)$$

In particular, if $\lambda = \lambda_{\min}(M)$, then

$$S_W \succeq 18c^2 \lambda^3 P_U. \quad (10)$$

Thus $c \neq 0$ and $\lambda > 0$ suffice for $\text{range}(S_W) = U$. If W contains U , then $S_W = 18c^2 P_U$. For a uniformly distributed unit vector in W , the covariance is instead $S_W/[r(r+2)(r+4)]$.

Proof. Support of $\mathcal{T}$ gives $b(v) \in U$, and $Q^T v \sim N(0, M)$. Of the fifteen pairings of six Gaussian coordinates, six pair every coordinate of the first cubic with one of the second; they give the first term in (9). The remaining nine pair two coordinates within each cubic and the two remaining coordinates across cubics, giving $9D_M M D_M^T \succeq 0$. The displayed block coefficients give $FF^T = 3c^2 I_m$. Since $M^{\otimes 3} \succeq \lambda^3 I_{m^3}$, (10) follows. For $W \supseteq U$, $M = I_m$ and the zero tensor traces imply $D_M = 0$. Finally write $v = R\theta$ for Gaussian v in W ; R and the uniform unit vector θ are independent, $\mathbb{E}R^6 = r(r+2)(r+4)$, and b is homogeneous of degree three.

This identity holds pointwise for every projector, so it also describes the population geometry of a realized learned frame. It does not assert conditional unbiasedness of estimates computed from the same observations that selected that frame. Mean coverage $\text{tr}(M)/m$ alone need not give a positive λ ; moreover, $\lambda > 0$ is sufficient, not necessary. The implementation uses unnormalized Gaussian features and hence S_W without the spherical divisor.

The lower bound is attained when $M = \lambda I_m$, so its constant and cubic exponent are sharp for this population class. This does not establish an optimal observation-count exponent. The exact full-signal criterion is $\text{rank}(F(L^{\otimes 3})) = m$ for a factorization $M = LL^T$; Appendix C.5 proves it and gives two limits of simpler rules. Working spaces with the same mean coverage can have full or zero signal, and a rank-two working space can retain all four directions of one carrier block. Thus neither mean coverage nor working rank alone provides a bootstrap guarantee for empirical feedback.

For an independent uniform unit direction v in the full ambient space, the same proposition, with $W = \mathbb{R}^p$, gives the exact scale

$$\mathbb{E}\|b(v)\|^2 = \frac{18c^2 m}{p(p+2)(p+4)}.$$

This suggests that weak initial population signal can matter in high dimension. It is a cubic population identity, not the empirical signal-to-noise ratio of a non-polynomial MIPP update. The calibrated panel below controls the initial public span for MIPP; it does not turn this cubic identity into a nonlinear MIPP error bound.

Conditional on a fixed working frame, if $S_W \succeq 0$, $\text{range}(S_W) = U$, $g = \lambda_m(S_W) > 0$, and a symmetric estimate satisfies $\|\hat{S} - S_W\|_{\text{op}} \leq \eta < g$, its rank- m leading projection obeys

$$\|\hat{P} - P_U\|_F \leq \frac{\sqrt{2m}\eta}{g - \eta}.$$

Proposition 9 proves this and gives a model-internal counterexample if full signal is replaced by support containment alone. Splitting observations can justify conditioning on the working frame; it does not create $g > 0$ or an error rate for η . A separate held-out acceptance rule in Proposition 8 certifies retained carrier mass, without asserting a universal training contraction.

4.2 Streamed score and empirical update

For each fixed v independent of the observations, a spherical harmonic summand with mean $b(v)$ is

$$h_v(q) = p(p+2)(v^T q)^3 q - 3d_p\{(v^T q)^2 v + \|v\|^2(v^T q)q\} + \frac{3p}{p+4}\|v\|^2 v,$$

$$d_p = p(p+2)/(p+4).$$

It is a trace-free fourth-tensor projection; substituting isotropic second moments verifies unbiasedness directly. For K vectors fixed independently of the observations, let $\bar{h}_v = n^{-1} \sum_i h_v(q_i)$. The diagonal-free matrix

$$\hat{G}_U = \frac{1}{K} \sum_v \left\{ \frac{n}{n-1} \bar{h}_v \bar{h}_v^T - \frac{1}{n(n-1)} \sum_i h_v(q_i) h_v(q_i)^T \right\}$$

is an order-two U-statistic with expectation $K^{-1} \sum_v b(v) b(v)^T$. Writing $h_v(q) = a_v(q)q + d_v(q)v$ allows the diagonal term to be accumulated using weighted second moments and matrix products, without forming a fourth-order tensor. With feature blocks, its cost is $O(np(K+p) + p^2 K)$ and storage is $O(pK + p^2 + \text{batch} \cdot (p + \text{feature\_batch}))$, including the observation batch.

Adaptive cubic algorithm (empirical). Input: the same n public directions, public rank $m = 16$, and random seeds. (1) Use 24 GOE probes to estimate a rank-64 initial frame. (2) For target ranks 64, 32, 16, draw 1024 Gaussian vectors in the current span, accumulate the diagonal-free matrix above on the same observations, and retain its leading target-rank frame. (3) Return the final rank-16 frame; use fresh data for any certificate. This makes four sample passes. Stage seeds are the recorded feature seed plus 0, 1, 2; the two NGCA panels reserve disjoint stage-seed ranges across groups.

The vectors in the three refinement passes depend on earlier reads of the same data. Fixed-vector unbiasedness and independent-probe concentration do not apply conditionally to these outputs. We evaluate them empirically and retain weak regimes and unsuccessful variants.

4.3 Exact rank-dependent noise and a tested alternative

The centered covariance theorem also quantifies the noise of the cubic score. For a fixed rank- r projector P_W , independent $v \sim N(0, P_W)$, and a uniform spherical q_0 , define

$$L_{p,r} = \frac{6p(p+1)(r+2)(r+4)}{(p+4)^2(p+6)}.$$

Exact spherical moments give

$$\mathbb{E}_{q_0,v} h_v h_v^T = L_{p,r} [(p+2)rI + 3(p-2)P_W]. \quad (11)$$

Because $h_v - b(v)$ is a fixed linear image of $H - \mathcal{T}$, its corresponding angular-model covariance, averaged over v , is bounded by C_a times (11). Appendix E gives the full derivation and fixed-probe cross-covariance. This is a variance proxy, not a standalone sample complexity theorem: at approximately random coverage $M \simeq (r/p)I_m$, the six-pairing signal also shrinks as $(r/p)^3$.

For $v = Bg$, $g \sim N(0, I_r)$, the third Gaussian-chaos projection $h_{Bg}^{(3)} = h_{Bg} - \frac{1}{2}\Delta_g h_{Bg}$ removes the first chaos. Its population signal is $6FM^{\otimes 3}F^T$, with the same image as the full six-plus-nine expression. Orthogonality reduces the average centered noise covariance, but also removes signal and does not give a uniform signal-to-noise or overlap advantage. At $p = 240$, working ranks 16, 32 and 64 lose respectively 14.07%, 7.26% and 3.26% of their $W^\perp$ null noise.

Two separately frozen eight-seed pilots at $p = 120$ use an independent one-step update. At the same $K = 1024$, mean overlap improvements are 0.01815 and 0.01993. In the second pilot, however, the original $K = 1536$ estimator and optimized third-chaos $K = 1024$ estimator take about 2.082 and 2.040 seconds per update and reach 0.50584 and 0.47122. The paired difference is -0.03462 , negative in all eight cases. We therefore retain the original algorithm. Appendix E.6 preserves both complete panels, timing scope and approximate uncertainty; these pilots do not replace the main-scale confirmation below.

5 Controlled evaluation

5.1 Confirmation and dimension sensitivity

At $p = 120, m = 16$, the 24-probe confirmation panel uses three fresh data/key/probe triples. At $n = 1048576$, default strength $k_{\text{ang}} = 8$ gives mean overlap 0.9655, versus unmarked 0.1320 and random $16/120 = 0.1333$. All 48 records (twelve scenarios, four nested counts) remain in angular-confirmation; Appendix A gives details.

The separate fixed-rank dimension panel was hash-bound before new seeds. All 81 rows (27 scenarios, three nested counts) and frames are retained. At $n = 262144$, increasing p from 60 to 240 drops mean 45° -margin overlap from 0.9611 to 0.0815 (Table 3). This is a budget-dependent limit of the baseline, not a necessary dimension law. Appendix B gives ranges and second-moment controls. The three-seed refinement panel has a separate protocol binding: its

corresponding baseline mean is 0.0799, using different seeds from the 0.0815 here.

Table 3. Fixed-rank dimension panel: mean fourth-order overlap at $n = 262144$, three triples per row. Margins are 45° , 18° , and zero.

p	$a = 1/4$	$a = 2/5$	$a = 1/2$	Random m/p
60	0.9611	0.5324	0.2789	0.2667
120	0.4177	0.1740	0.1431	0.1333
240	0.0815	0.0697	0.0681	0.0667

5.2 Equal total feature budgets

The three-seed refinement panel gave 0.6810 mean overlap (0.5709–0.7623), versus 0.2004 for 48 probes. To distinguish feature budget from feedback, a second prospective panel fixes eight new seed groups at $p = 240$, $a = 1/4$, $n = 2^{20}$. Its two controls use 3072 fixed cubic features, either in the ambient space or in the same learned initial 64-dimensional frame. The latter does not update that frame. All three cubic methods remove diagonal products; only the ambient vectors are independent of the observations. Their feature counts match, not their wall times. Table 4 reports the complete primary comparison. The paired adaptive-minus-frozen-initial difference has mean 0.2716 and range 0.1691–0.3648 (8/8 positive). This comparison evaluates the complete feedback schedule; it does not isolate rank reduction from every other adaptive choice. The paired adaptive-minus-frozen difference has sample SD 0.0756 (descriptive standard error 0.0267). All cubic variants use shared feature-seed blocks within each group; changing the frame changes the resulting vectors. These are paired, correlated comparisons. Ambient cubic features reach only 0.1249, below both 24- and 48-probe baselines (0.1453 and 0.1925). The gain therefore concerns the tested feedback procedure; more cubic features alone do not improve every comparator, and a finite-sample variance explanation has not been proved.

The fixed-probe theorem and fixed-vector unbiasedness do not certify either data-dependent cubic estimator. These timings exclude shared generation and charge initialization consistently. Initialization runs first; subsequent method order rotates by seed, so initialization timing can still carry a fixed-order effect. The 24/48-probe designs are nested within seed; methods are paired, seed groups independent. The low-budget and 480-dimensional cases remain weak (0.0784 and 0.0372), while the exact-hiding control remains near chance (0.0648 versus 0.0667). Appendix C reports all 60 refinement-panel rows, the equal-feature-budget records, working-rank diagnostics, negative results and public-input replay.

Table 4. Frozen eight-seed confirmation at $p = 240, a = 1/4, n = 2^{20}$. Each cubic method uses 3072 total features. Brackets are minima and maxima, not confidence intervals. Times exclude data generation and include initialization where used. They are separate from the two-thread NGCA measurements in Table 5.

Method	Passes	Overlap mean [min, max]	Seconds
24 GOE probes	1	0.1453 [0.1359, 0.1517]	21.40
48 GOE probes	1	0.1925 [0.1797, 0.2040]	41.08
Fixed ambient cubic	1	0.1249 [0.1089, 0.1355]	29.96
Frozen initial cubic	2	0.3282 [0.2972, 0.3887]	51.33
Adaptive cubic	4	0.5999 [0.4804, 0.7131]	52.00

5.3 A paired intervention on the middle rank

On the eight already observed seed groups, schedules 64–32–16 and 64–64–16 have mean overlaps 0.5999 and 0.4357. Their first update and second-update operator agree; final Gaussian coefficients are coupled on the common 32-space. The paired difference is positive in all eight groups, with mean 0.1641 and sample SD 0.0534. This is a post-hoc diagnostic under a specified coupling, not a fresh confirmation or a universal schedule. It does not identify a signal-to-noise mechanism; greater mean overlap need not increase the weakest signal direction (Table 8). Appendix C.8 retains the full protocol, Table 13, every pair and all sixteen final frames.

5.4 Calibrated non-polynomial NGCA controls

We compare adaptive cubic with a budgeted MIPP reimplementation of [13]: four nonlinear families, 256 function/start pairs, empirical whitening and ten direction updates. This is not the original 4,000-function budget. Eight fresh groups use $p = 240, m = 16, a = 1/4$ and $n = 2^{20}$; all three learners receive the same public directions within each group. MIPP uses an independent replayed χ_p lift (Appendix F).

Following the reference’s Gaussian-calibration recommendation, four independent null datasets of the same dimensions and size run the complete MIPP pipeline. A separate pooled empirical 95th percentile fixes each initialization arm’s threshold before two null holdouts and the eight angular test groups. The ambient arm starts in the full whitened space. The spectral arm starts in the whitened covector image of the same public rank-64 frame used by adaptive cubic, computed from identical 24 GOE probes; later MIPP updates are unrestricted. Both arms share their initial Gaussian draws and lift seeds. This controls initial public information without an oracle. It introduces no familywise or exact false-positive guarantee: functions are dependent and heterogeneous, with only four calibration datasets and two null holdouts.

Mean overlaps are 0.6968 for adaptive cubic (8/8 frames), 0.2608 for calibrated ambient MIPP (4/8), and 0.4159 for calibrated spectral MIPP (8/8).

Table 5. Gaussian-calibrated paired panel, $p = 240$, $n = 2^{20}$ per group. Primary MIPP thresholds are calibrated before testing. Success counts include every rank failure; overlap summaries use returned frames only. Zero is a prespecified sensitivity. Times include preprocessing; threshold variants share a fit.

Method	Threshold	Rank-16	Overlap mean (SD)	Seconds
Adaptive cubic	—	8/8	0.6968 (0.0760)	47.4
MIPP ambient	1.0456	4/8	0.2608 (0.0982)	57.6
MIPP spectral64	1.0404	8/8	0.4159 (0.1335)	78.1
MIPP ambient	0	8/8	0.2053 (0.0895)	57.6
MIPP spectral64	0	8/8	0.4074 (0.1323)	78.1

The ambient mean is conditional on success, so this is not an eight-pair comparison. At zero threshold all eight pairs exist; the spectral-minus-ambient difference averages 0.2021. Equal angular training data do not imply equal computation or search effort.

Zero threshold is a prespecified sensitivity, not a chosen winning cutoff. The fixed-threshold panel, including every rank failure and $T = 0$ ablation, remains in Appendix F.1. The methods match n angular training directions per group; MIPP also uses four Gaussian calibration datasets, whose costs are reported separately. Computation is not matched: adaptive cubic uses four passes, ambient MIPP twelve and spectral MIPP thirteen, including its separately recomputed initializer. MIPP evaluates $2816n$ scalar functions and the same number of derivatives. Times include all fitting and replay costs; threshold variants share a fit. The cosine family has zero population Stein moment by central symmetry and remains in the budget. Oracle-scored development favored the cubic schedule; these controls do not equalize that tuning effort or establish superiority to tuned full-budget MIPP, IMAK or other NGCA families. Complete moments, all calibration and test outcomes, costs and analysis rules are in Appendix F.2.

6 Limitations and conclusion

The angular model links uniform decoding margin to fourth-order leakage and an attainable hiding obstruction. Fixed-probe bounds give conservative sufficient recovery counts. Sharp centered covariance instead permits public validation of any candidate frozen before new data. Four-index compression certifies energy below overlap; more data cannot remove that geometric loss. Candidate-specific covariance bounds may improve the uniform penalty.

Same-data adaptive training has no end-to-end convergence proof and remains weak in some larger or lower-budget regimes. The exact rank-noise formula does not make its third-chaos variant win the near-compute-budget test. The calibrated NGCA panel controls threshold calibration and initial public information for a 256-function MIPP implementation. It does not cover tuned full-budget MIPP, IMAK or log-density-gradient methods, or equalize development tuning. Four null calibration datasets provide an empirical cutoff, not an

exact error guarantee. Neither the one-step pilots nor post-hoc rank intervention establishes a universal schedule. Sample splitting cannot supply the full-signal condition $\text{range}(S_W) = U$ or a positive m th eigenvalue.

The model assumes fixed pairing, fresh fair bits, a Gaussian complement and exact iid directions. It identifies subspaces; sign symmetry prevents ordered-bit identification without an anchor. A fixed-dimensional weighted-chi-square limit cannot certify a finite-sample extreme-tail ROC; the conservative alternative can have severe power cost (Proposition 17). Coverage assumes exact arithmetic; the checked floating-point frontend has no directed-rounding enclosure. The separate radial report has its own numerical contract and reports blind removal of a public linear carrier and the 24-image diffusion pilot with no detections in 48 inversions. Those records do not validate the angular learner or establish a successful neural attack.

Reproducibility. The appendices give complete proofs, protocols and outcomes. **ARTIFACTS.json** in the source archive indexes separately maintained code and records by descriptive identifiers and SHA256 digests. The validator uses public inputs; the secret carrier enters only oracle evaluation. Validation and NGCA panels have distinct seeds and protocols; pre-execution hashes bind code and inputs. Generative AI tools assisted drafting, mathematical exploration, code and checking. Claims rest on proofs and reproducible records; responsibility remains with the author.

References

1. Christ, M., Gunn, S.: Pseudorandom error-correcting codes. In: CRYPTO 2024, pp. 325–347 (2024). <https://arxiv.org/abs/2402.09370>
2. Gunn, S., Zhao, X., Song, D.: An undetectable watermark for generative image models. In: ICLR (2025). <https://openreview.net/forum?id=j1hBFm7T2J>
3. Li, K.-C.: On principal Hessian directions for data visualization and dimension reduction: Another application of Stein’s lemma. J. Amer. Statist. Assoc. 87(420), 1025–1039 (1992).
4. Bas, P., Westfeld, A.: Two key estimation techniques for the Broken-Arrows watermarking scheme. In: ACM Workshop on Multimedia and Security, pp. 1–8 (2009). <https://doi.org/10.1145/1597817.1597819>
5. Wen, Y., Kirchenbauer, J., Geiping, J., Goldstein, T.: Tree-Ring Watermarks: Fingerprints for Diffusion Images that are Invisible and Robust. In: NeurIPS (2023). <https://arxiv.org/abs/2305.20030>
6. Yang, Z., Zeng, K., Chen, K., Fang, H., Zhang, W., Yu, N.: Gaussian Shading: Provable Performance-Lossless Image Watermarking for Diffusion Models. In: CVPR, pp. 12162–12171 (2024). <https://arxiv.org/abs/2404.04956>
7. Gaussian Shading authors: Reference implementation, commit 09c678fadc7545acf7be12647ddf2a5e66f6a9dc. `run_gaussian_shading.py`, lines 59–65; `watermark.py`, lines 34–56 and 131–143. Inspected 13 September 2026. <https://github.com/bsmhmmf/Gaussian-Shading/tree/09c678fadc7545acf7be12647ddf2a5e66f6a9dc>

8. Li, W., Zhang, H., Qu, Q.: Shallow Diffuse: Robust and Invisible Watermarking through Low-Dimensional Subspaces in Diffusion Models. In: NeurIPS (2025). <https://arxiv.org/abs/2410.21088>
9. Tropp, J. A.: User-Friendly Tail Bounds for Sums of Random Matrices. *Foundations of Computational Mathematics* 12, 389–434 (2012).
10. Doërr, G., Dugelay, J.-L.: Danger of low-dimensional watermarking subspaces. In: ICASSP, vol. 3, pp. 93–96 (2004).
11. Ci, H., Yang, P., Song, Y., Shou, M. Z.: RingID: Rethinking Tree-Ring Watermarking for Enhanced Multi-Key Identification. In: ECCV (2024). <https://arxiv.org/abs/2404.14055>
12. Zheng, Y., An, H., Hua, G., Lin, Z., Fang, Y.: Correlation-Aware and Gaussianity-Preserving Robust Latent Angular Watermarking for Diffusion Models. arXiv:2607.22386v1 (2026), Appendix A, Eq. (44). <https://arxiv.org/abs/2607.22386v1>
13. Blanchard, G., Kawanabe, M., Sugiyama, M., Spokoiny, V., Müller, K.-R.: In Search of Non-Gaussian Components of a High-Dimensional Distribution. *J. Mach. Learn. Res.* 7, 247–282 (2006). <https://www.jmlr.org/papers/v7/blanchard06a.html>
14. Kawanabe, M.: Linear Dimension Reduction Based on the Fourth-Order Cumulant Tensor. In: ICANN 2005, LNCS 3697, pp. 151–156 (2005). https://doi.org/10.1007/11550907_25
15. Vershynin, R.: Introduction to the non-asymptotic analysis of random matrices. In: *Compressed Sensing: Theory and Applications*, pp. 210–268 (2012). Preprint v7, Lemma 5.33 and Proposition 5.34. <https://arxiv.org/abs/1011.3027v7>
16. Tan, Y. S., Vershynin, R.: Polynomial Time and Sample Complexity for Non-Gaussian Component Analysis: Spectral Methods. In: COLT, PMLR 75, pp. 498–534 (2018). <https://proceedings.mlr.press/v75/tan18a.html>
17. Kawanabe, M., Sugiyama, M., Blanchard, G., Müller, K.-R.: A new algorithm of non-Gaussian component analysis with radial kernel functions. *Annals of the Institute of Statistical Mathematics* 59, 57–75 (2007). <https://doi.org/10.1007/s10463-006-0098-9>
18. Sasaki, H., Niu, G., Sugiyama, M.: Non-Gaussian Component Analysis with Log-Density Gradient Estimation. In: AISTATS, PMLR 51, pp. 1177–1185 (2016). <https://proceedings.mlr.press/v51/sasaki16.html>
19. Joly, E., Lugosi, G.: Robust estimation of U-statistics. *Stochastic Processes and their Applications* 126, 3760–3773 (2016). <https://doi.org/10.1016/j.spa.2016.04.021>
20. Minsker, S., Wei, X.: Moment inequalities for matrix-valued U-statistics of order 2. arXiv:1801.05921v4 (2019). <https://arxiv.org/abs/1801.05921>
21. Zaremba, W., Gretton, A., Blaschko, M.: B-tests: Low Variance Kernel Two-Sample Tests. In: NeurIPS (2013). <https://arxiv.org/abs/1307.1954>
22. Goldberg, P. W., Jerrum, M. R.: Bounding the Vapnik–Chervonenkis dimension of concept classes parameterized by real numbers. *Machine Learning* 18, 131–148 (1995). <https://www.cs.ox.ac.uk/people/paul.goldberg/papers/GJ-ML95.pdf>
23. Franks, W. C., Moitra, A.: Rigorous Guarantees for Tyler’s M-Estimator via Quantum Expansion. In: COLT, PMLR 125, pp. 1601–1632 (2020). <https://proceedings.mlr.press/v125/franks20a.html>
24. Rudelson, M., Vershynin, R.: Hanson–Wright inequality and sub-gaussian concentration. arXiv:1306.2872 (2013). <https://arxiv.org/abs/1306.2872>
25. Giné, E., Latała, R., Zinn, J.: Exponential and moment inequalities for U-statistics. arXiv:math/0003228 (2000). <https://arxiv.org/abs/math/0003228>

A Refreshed angular payloads: fourth-order identification and an exact boundary

Global dependency map. The rows below separate theorem prerequisites from optional analyses and empirical records. No angular proof depends on the separate radial report.

Claim or use	Proof route and scope
Angular signature and hiding	Endpoint $\rightarrow$ signature (Theorems 1, 7); hiding and margin (Theorem 3). The margin uses the specified angle-sign decoder.
Fixed-probe recovery	Fourth moments $\rightarrow$ Appendix B $\rightarrow$ sharp probe/observation bounds in Appendix C; requires $c > 0$.
Public overlap validation	Centered covariance (Theorem 4) $\rightarrow$ U-statistic variance $\rightarrow$ Cantelli inversion (Theorem 5); frozen candidate and independent data.
Exact compression constants	Harmonic branches (Theorem 13) $\rightarrow$ exact traces; improves the preceding certificate constants.
Feedback and experiments	Population decomposition (Appendix C.5) and noise (Appendix E); experiments do not prove adaptive contraction.
Optional scope analyses	Section D.13: support learning, preprocessing, bit symmetry, tails and rank selection have their own assumptions; they are not premises of the main certificate.
NGCA evidence	Appendix F: exact public lift, fixed-threshold panel and calibration/initialization controls; every outcome retained.

The endpoint law below is a standalone definition of the angular model. It asks whether payload refreshing hides the carrier or only its second moment. The calculation also covers the ideal antipodal encoding primitive in LAW [12], Eq. (44), when bits are independently refreshed, by taking the zero-residual limit below. It is not an experiment on LAW, LAW-M or a diffusion generator. Fourth-cumulant subspace extraction is an established statistical idea, encompassed by non-Gaussian component analysis [13]. Our additional results are the exact angular coefficient, the failure of isotropic fourth-order contraction, recovery from normalised observations, and an attainable robustness–distribution boundary for the specified decoder.

A.1 An exact endpoint, without numerical integration

Work in the public complement, of dimension p , in units where the clean variance is one. There are b_0 disjoint four-dimensional carrier blocks, so $m = 4b_0 < p$. Before a fixed unknown orthogonal change of coordinates, each block is

$$Z_j = (R_j \cos \Theta_j, R_j \sin \Theta_j, S_j T_j \cos(\Theta_j + \pi/2 + D_j), S_j T_j \sin(\Theta_j + \pi/2 + D_j)). \quad (12)$$

Here R_j, T_j are independent standard Rayleigh radii, Θ_j is uniform on the circle, S_j is an independent fair sign, and $D_j \sim \text{Unif}[-\pi a, \pi a]$, independently across blocks. The remaining $p - m$ coordinates are independent standard Gaussians. The sign is the refreshed payload; the orthogonal frame, including the pairing, stays fixed. At $a = 0$, interpret $D_j = 0$.

For the angular ODE $\dot{e} = -4k_{\text{ang}}t(1-t)e$ with a constant reference phase, the wrapped relative-phase error keeps its sign except for an initial branch boundary of probability zero. Integration gives

$$a = \exp(-2k_{\text{ang}}/3). \quad (13)$$

Initial independent Gaussian phases give the stated uniform error. The common radial factor $\sqrt{1/2}$ and all mode dynamics cancel after projecting onto the public complement and normalising there. Thus the attacker can use

$$q = \frac{T_0^T X}{\|T_0^T X\|},$$

where T_0 is any public orthonormal basis of the mode complement and X is either the original endpoint or its globally normalised version. No endpoint norm, payload, secret pairing or mode-energy weight is supplied. Direct standard-Gaussian latents are the special case with no mode space.

A.2 Second-order blindness and a nonzero fourth harmonic

Theorem 7. For (12), $\mathbb{E}Z = 0$, $\mathbb{E}ZZ^T = I_p$, and $\mathbb{E}qq^T = I_p/p$. Put

$$c(a) = \begin{cases} \sin(2\pi a)/(2\pi a), & a > 0, \\ 1, & a = 0. \end{cases}$$

In one block numbered 1, 2, 3, 4, the nonzero fourth cumulants, together with their index permutations, are

$$\begin{aligned} \mathcal{T}_{1133} = \mathcal{T}_{2244} = \mathcal{T}_{1234} &= -c(a)/2, \\ \mathcal{T}_{1144} = \mathcal{T}_{2233} &= c(a)/2. \end{aligned} \quad (14)$$

All cumulants involving Gaussian bulk coordinates or more than one independent block vanish. In particular, $c(a) > 0$ for $0 \leq a < 1/2$, although the population second moment has no spectral contrast.

Proof. Uniform reference phases make every individual pair exactly standard Gaussian. Averaging the fair sign removes cross-pair second moments. The embedding is homogeneous and preserves every radius, so the original Gaussian length $L = \|Z\|$ remains independent of $q = Z/L$, with $L^2 \sim \chi_p^2$. Consequently $\mathbb{E}qq^T = I_p/p$.

For the fourth moment, an odd total power in either pair vanishes after averaging S_j and the phase. In a mixed even term, $\mathbb{E}R_j^2T_j^2 = 4$ and $\mathbb{E}\cos(2D_j) = c(a)$, $\mathbb{E}\sin(2D_j) = 0$. For example,

$$\mathbb{E}Z_1^2Z_3^2 = 1 - c(a)/2, \quad \mathbb{E}Z_1^2Z_4^2 = 1 + c(a)/2, \quad \mathbb{E}Z_1Z_2Z_3Z_4 = -c(a)/2.$$

Subtracting the Gaussian pairings gives (14). The other entries follow by exchange of the two axes. Independence across blocks and Gaussianity of the bulk give the remaining zeros.

The bit sign disappears from the fourth moments because it is squared. Refreshing it therefore erases the signed covariance, but not this dependence. Moreover,

$$\sum_k \mathcal{T}_{ijkk} = 0. \quad (15)$$

Thus contracting the fourth moment with the identity, as in a radial fourth-order statistic, also fails. A non-isotropic quadratic probe is needed; replacing the original public weight merely by the observed squared length does not repair the blind spot.

A.3 A key-blind population identity and a consistent learner

For a symmetric public probe A , define

$$C(A) = p(p+2)\mathbb{E}[(q^T A q)qq^T] - \text{tr}(A)I_p - 2A. \quad (16)$$

Since $\mathbb{E}L^4 = p(p+2)$, this is exactly the contraction of the fourth cumulant tensor of the standardised complement. The factor is essential: applying a Gaussian cumulant formula directly to unit vectors would create a nonzero null statistic.

Theorem 8. *Let P_U denote the fixed m -dimensional carrier projector and suppose $c(a) \neq 0$. Then $C(A) = P_U C(A) P_U$ for every A . If A is an independent Gaussian orthogonal ensemble (GOE) matrix with off-diagonal variance one and diagonal variance two, then*

$$\mathbb{E}_A[C(A)^2] = 6c(a)^2 P_U. \quad (17)$$

For one such probe, $C(A)$ has range U with probability one. Consequently, for any positive fixed number L_0 of independent probes, $G = L_0^{-1} \sum_{\ell=1}^{L_0} C(A_\ell)^2$ has range U almost surely. Sample averages of (16) consistently recover that subspace as the number of independent observations tends to infinity.

Proof. Only within-block cumulants survive, so each contraction is supported on U . In one block, choose $A_0 = \text{diag}(1, 0, 1, 0)$. Direct contraction gives $C(A_0) = \frac{c(a)}{2} \text{diag}(-1, 1, -1, 1)$, which is nonsingular. The block determinant is therefore a nonzero polynomial in the entries of a general symmetric probe. Its zero set has measure zero. An independent GOE probe, expressed in the secret frame, has a density; a finite union of these blockwise null sets still has probability zero.

The GOE identity $\mathbb{E}A_{kl}A_{rs} = \delta_{kr}\delta_{ls} + \delta_{ks}\delta_{lr}$ gives

$$\mathbb{E}_A[C(A)^2]_{ij} = 2 \sum_{s,k,l} \mathcal{T}_{iskl} \mathcal{T}_{jskl}.$$

For each coordinate of a carrier block, the sum contains twelve squared entries of magnitude $|c(a)|/2$; the off-diagonal sums vanish. It is therefore $6c(a)^2 I_4$ on each block and zero on the bulk. Orthogonal equivariance gives (17) in the public coordinates.

For fixed probes, each summand based on q is bounded. The strong law gives convergence of every estimated contraction, hence of $\hat{G}$. The positive eigengap of G then gives convergence of its top m eigenspace. More explicitly, if λ is the smallest positive eigenvalue of G and $e = \|\hat{G} - G\|_{\text{op}} < \lambda/2$, a top m frame $\hat{Q}$ satisfies

$$\frac{1}{m} \text{tr}(\hat{Q}^T P_U \hat{Q}) \geq 1 - \left(\frac{e}{\lambda - e} \right)^2. \quad (18)$$

Indeed, projecting its eigenequation onto $U^\perp$ leaves only the perturbation, while its smallest selected eigenvalue is at least $\lambda - e$. The squared operator bound controls the mean squared sine of the principal angles. This is a deterministic perturbation statement, not a numerical finite-sample confidence certificate.

In the implementation, lower-order control variates reduce sampling noise. With $S_n = n^{-1} \sum_i q_i q_i^T$ and $F_n(A) = n^{-1} \sum_i (q_i^T A q_i) q_i q_i^T$, we use the unbiased contraction

$$\begin{aligned} \hat{C}(A) &= p(p+2)F_n(A) - p \text{tr}(AS_n)I_p - 2p(AS_n + S_n A) \\ &\quad - p \text{tr}(A)S_n + \text{tr}(A)I_p + 2A. \end{aligned} \quad (19)$$

Its expectation is (16). Subtracting the identity component from a GOE probe does not change its population contraction, by (15). The reported learner fixes 24 such probes independently of both key and samples, accumulates them in batches, and returns the top 16 directions of their mean square. It never receives the key or the payload. The true frame is used only to score its output.

A.4 An exact hiding point and an attainable robustness bound

Theorem 9. *At $a = 1/2$, equivalently $k_{\text{ang}} = \frac{3}{2} \log 2$, the entire vector Z has law $N(0, I_p)$ for every fixed secret frame, despite zero bit error under the noiseless sign-of-relative-angle decoder. Thus q is uniform on the sphere and no number of such independently refreshed public samples can identify the secret frame from their distribution. This claim requires independent fair payload bits and does not cover a repeated fixed message.*

Proof. Given $S_j = +1$, the relative phase is uniform on one half-circle; given $S_j = -1$, it is uniform on the other. The equally weighted mixture is uniform on the whole circle and independent of the reference phase. The original independent

Rayleigh radii are unchanged, so the two pairs are jointly independent standard Gaussians. Independence across blocks completes the proof. The half-circle also determines the bit except on boundaries of probability zero. Its distance to a decision boundary, however, has essential infimum zero.

For $0 \leq a < 1/2$, that decoder has a positive worst-case relative-angle margin

$$\gamma = \pi(1/2 - a), \quad c(a) = \frac{\sin(2\gamma)}{\pi - 2\gamma}, \quad (20)$$

with the continuous limit $c = 1$ at $\gamma = \pi/2$. Hence every positive margin in this family produces a nonzero fourth-order recovery signal.

Theorem 10. *Consider any encoder with b_0 fixed, disjoint reference/encoding pair blocks and the same sign-of-relative-angle decoder. Let Q_0 be the clean standard-Gaussian law, or its unit-direction law. Suppose that with probability at least $1 - \varepsilon$ under the marked law P , all decoded bits are correct under every relative-angle perturbation of size less than γ , where $0 \leq \gamma \leq \pi/2$. Then*

$$\text{TV}(P, Q_0) \geq [1 - \varepsilon - (1 - 2\gamma/\pi)^{b_0}]_+. \quad (21)$$

This bound makes no assumption about the feedback law. For $\varepsilon = 0$, the endpoint family (12) attains equality when $a = 1/2 - \gamma/\pi$ and bits are independent and fair.

Proof. Let R_γ be the event that every relative angle is at distance at least γ from both sign-decision boundaries. Robust correct decoding implies this event, so $P(R_\gamma) \geq 1 - \varepsilon$. Under Q_0 , the relative angles are independent and uniform, including after global normalisation. Thus $Q_0(R_\gamma) = (1 - 2\gamma/\pi)^{b_0}$. Applying the definition of total variation to this event proves the lower bound. The endpoint family is exactly Q_0 conditioned on R_γ : its relative angles are uniform on the retained arcs and all other variables keep their clean laws. The total variation of a distribution conditioned on an event of probability r is $1 - r$, proving equality. At $a = 0$ use the limiting law supported on the exact target angles, with total variation one.

The total-variation witness uses the fixed secret pairing. It establishes a distributional obstruction, not an efficient key-blind test or a deployed query bound. The fourth-order learner supplies the separate constructive result for the explicit endpoint family. The perturbation is in relative angle, not pixel distance or neural inversion error. At the exact hiding point, average robustness to a small perturbation can still be nonzero; what disappears is a positive uniform worst-case margin.

A.5 Confirmation and reproducibility

An exploratory key and sample seed were used to check feasibility, then a separate specification fixed three new data seeds (201–203), three new key seeds

(84011–84013), three probe seeds (92011–92013), four regimes, 24 probes and four sample sizes before those runs. Every one of the 48 checkpoint rows is retained: twelve scenarios with four nested sample sizes; regimes share random streams within a replicate. This is a held-out confirmation design, not a preregistration of the whole investigation. Its SHA-256 and full parameters are in the artifact. No estimator setting was selected using the confirmation overlaps.

Table 6. Fourth-order recovery from normalised public directions, $p = 120, m = 16$. Means across three fresh data/key/probe triples; brackets give the observed range, not a confidence interval. All bits are refreshed. The random overlap is $16/120 = 0.1333$.

Regime	n	Fourth order, mean [range]	Second order
Default angular	65536	0.2728 [0.2547, 0.2865]	0.1372
Default angular	262144	0.7161 [0.7016, 0.7246]	0.1282
Default angular	1048576	0.9655 [0.9647, 0.9666]	0.1379
$a = 1/4$	1048576	0.8725 [0.8650, 0.8773]	0.1378
$a = 1/2$ (exact null)	1048576	0.1270 [0.1172, 0.1360]	0.1362
Unmarked	1048576	0.1320 [0.1295, 0.1362]	0.1327

All rows use the exact continuous endpoint in the public complement, normalised to unit length. The second-order comparator uses the uncentered sample second moment and the 16 eigenvalues furthest from its median; the population mean is known to be zero. The radial report’s leakage-regime tables use a sample-centered covariance with a different normalisation; they therefore concern a different observable. A half-angle endpoint ($a = 1/2$) and unmarked Gaussian directions are two matched negative-control regimes. Three replicates describe sensitivity to key, data and probes; the displayed ranges are not confidence intervals.

Independent phase/radius quadrature checks all 256 entries of the block fourth tensor at six contraction parameters, to residual below 9×10^{-16} in ordinary floating point. A spherical fourth-moment design checks the null normalisation, and direct tensor contraction checks the factor 6 and the zero identity contraction. On 64 inputs, comparison with the angular RK4 implementation at 240, 480 and 960 steps gives maximum public-direction errors 9.12×10^{-9} , 5.77×10^{-10} and 3.63×10^{-11} . These are regression and algebra checks, not directed numerical enclosures. The analytic endpoint derivation, rather than those discrepancies, connects the experiment to the continuous law.

An additional interface check runs the learner in a temporary directory containing only its code and a public observation matrix. Its arguments include the rank and independent probe seed, but no producer seed, key or payload. External scoring reproduces the corresponding confirmation overlap to 4×10^{-16} . This checks the data interface, not operating-system isolation; the research reproduction bundle intentionally records synthetic producer and evaluation seeds.

The angular model exhibits a population mechanism beyond radial targets and admits a demonstrated blind learner. The main text gives a conservative analytic finite-sample guarantee; it does not certify the numerical confirmation frames or supply image-only recovery, payload authentication, or a verified neural inverse. Existing non-Gaussian component methods remain the statistical precedent; the scope added here is the exact angular analysis and its matching counterexample.

B Finite-probe and finite-sample angular bounds

The following Frobenius–Markov bound provides a baseline comparison.

Theorem 11. *Assume the stated angular law, $c \neq 0$, iid observations and independent GOE probes. For $\delta_A, \delta_q \in (0, 1)$ with $\delta_A + \delta_q < 1$, put*

$$\rho = 1 - \sqrt{6/L_0} - \sqrt{2\log(2b_0/\delta_A)/L_0} > 0, \quad (22)$$

$$K_p = 2p^4 - 4p^3 + 14p^2 - 20p + 8. \quad (23)$$

In exact arithmetic the returned frame satisfies, with probability at least $1 - \delta_A - \delta_q$ over probes and observations,

$$\frac{1}{m} \text{tr}(\widehat{Q}^T P_U \widehat{Q}) \geq \left[1 - \frac{2(K_p - 6mc^2)}{3mn\delta_q c^2 \rho^2} \right]_+. \quad (24)$$

The population finite-probe matrix obeys $\lambda_{\min}(G|_U) \geq 6c^2\rho^2$ except with probability δ_A .

This appendix proves the identities used in Theorem 11. Let $D = \text{diag}(1, -1)$, $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Use the following six symmetric 4×4 matrices, displayed in 2×2 blocks:

$$\begin{aligned} B_1 &= \text{diag}(D, 0), & B_2 &= \text{diag}(X, 0), & B_3 &= \text{diag}(0, D), & B_4 &= \text{diag}(0, X), \\ B_5 &= \sqrt{2} \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix}, & B_6 &= \sqrt{2} \begin{pmatrix} 0 & J \\ -J & 0 \end{pmatrix}. \end{aligned}$$

For a standard GOE block write

$$\begin{aligned} u &= (A_{44} - A_{33})/2, & v &= -A_{34}, & w &= (A_{22} - A_{11})/2, & t &= -A_{12}, \\ r &= -(A_{13} + A_{24}), & s &= A_{14} - A_{23}. \end{aligned}$$

These variables are independent centered Gaussians with variances 1, 1, 1, 1, 2, 2. The exact contraction is

$$C(A)|_{U_j} = c \begin{pmatrix} u & v & r & s \\ v & -u & -s & r \\ r & -s & w & t \\ s & r & t & -w \end{pmatrix} = c \sum_{r=1}^6 g_r B_r,$$

with independent standard normal g_r . Scalar trace subtraction has no effect on these six coordinates. Direct squaring gives $\sum_r B_r^2 = 6I_4$.

Let H_j have the L_0 independent six-mode coefficient vectors as rows. The bound in [15, Corollary 5.35] implies $s_{\min}(H_j)/\sqrt{L_0} \geq \rho$ with failure at most δ_A/b_0 , for ρ in (22). If it holds, then for every x in that block

$$\frac{1}{L_0} \sum_{\ell} \|C(A_{\ell})x\|^2 \geq c^2 \rho^2 \sum_r \|B_r x\|^2 = 6c^2 \rho^2 \|x\|^2.$$

Summing blocks and taking a union bound proves the finite-probe gap. The argument requires $\rho > 0$; squaring a negative value would be invalid. The population mean identity alone would not prove this finite-probe bound.

Exact integrated variance. For trace-zero A , the one-observation summand is

$$H_A(q) = p(p+2)(q^T A q) q q^T - p(q^T A q) I_p - 2p(A q q^T + q q^T A) + 2A.$$

It has expectation $C(A)$. Orthogonal invariance of trace-centered GOE and equivariance of H_A show that $\mathbb{E}_A \|H_A(q)\|_F^2$ is constant for every unit q , even though the marked direction law is not uniform. Evaluate at $q = e_1$. The entries are

$$\begin{aligned} H_{11} &= (p-1)(p-2)A_{11}, & H_{jj} &= -pA_{11} + 2A_{jj} \quad (j > 1), \\ H_{1j} &= -2(p-1)A_{1j}, & H_{jk} &= 2A_{jk} \quad (j, k > 1, j \neq k). \end{aligned}$$

For trace-centered GOE, $\mathbb{E} A_{ii}^2 = 2(1-1/p)$, $\mathbb{E} A_{ii} A_{jj} = -2/p$ for $i \neq j$, and $\mathbb{E} A_{ij}^2 = 1$. Therefore the diagonal, first-row off-diagonal and remaining off-diagonal contributions sum to

$$\begin{aligned} \mathbb{E}_A \|H_A(q)\|_F^2 &= \frac{2(p-1)^3(p-2)^2}{p} + (p-1)\{2p(p-1) + 16 - 8/p\} \\ &\quad + 8(p-1)^3 + 4(p-1)(p-2) = K_p. \end{aligned}$$

For fixed A , independence of observations gives variance divided by n . Integrating over A and using $\mathbb{E}_A \|C(A)\|_F^2 = 6mc^2$ gives

$$\mathbb{E} \|\hat{B} - B\|_F^2 = \frac{1}{L_0} \sum_{\ell} \mathbb{E} \|\hat{C}(A_{\ell}) - C(A_{\ell})\|_F^2 = \frac{K_p - 6mc^2}{n}.$$

Cross-probe independence of the errors is unnecessary for this sum of squared norms. Increasing the number of probes therefore improves the gap bound but does not make this integrated error energy decrease as $1/L_0$ when all probes use the same observations.

From operator error to mean overlap. Write $\hat{P} = \hat{Q}\hat{Q}^T$ and $\Delta = \hat{B} - B$. Since $\hat{P}$ is a best rank- m right projection for $\hat{B}$ in Frobenius norm and $B(I - P_U) = 0$,

$$\|B(I - \hat{P})\|_F \leq \|\hat{B}(I - \hat{P})\|_F + \|\Delta(I - \hat{P})\|_F \leq 2\|\Delta\|_F.$$

Also $B^T B \succeq \lambda P_U$, hence the square of the left side is at least $\lambda \text{tr}(P_U(I - \hat{P})) = m\lambda(1 - o)$. Markov bounds $\|\Delta\|_F^2$ by $(K_p - 6mc^2)/(n\delta_q)$ except with probability δ_q . Combining events proves (24), including when its positive part is zero. This bound avoids replacing a realized random gap by its expectation.

Meaning and verification. At fixed failure probabilities and loss target, the bound is polynomial in ambient dimension and diverges as c^{-2} near exact hiding. It gives a sufficient count, with no matching lower bound. Its conservative Markov and Frobenius steps leave a large gap to measured performance; it cannot validate any particular 24-probe frame. The verification code checks K_p using exact rational GOE basis integration for $p = 2, \dots, 10$, independently checks the six-mode covariance and square identity, and exercises the variational inequality on 100 numerical matrices. These checks accompany, and do not replace, the proof.

Prospective dimension panel. Before executing the dimension panel, dimension-protocol bound the sampler, learner, runner, parameters and environment summary. This is a local prospective freeze after exploratory development, without an external timestamp authority. The exploratory study’s specification bound parameters but omitted implementation hashes; the absence of tuning in that study remains an author declaration rather than a verified pre-execution code commitment.

The dimension panel uses dimensions 60, 120 and 240; rank 16; $a = 1/4, 2/5, 1/2$; 24 probes; three data/key/probe triples; and nested counts 16384, 65536 and 262144. Within each dimension/replicate, regimes reuse the same Gaussian and bit streams. Seed identifiers recur across dimensions, so these are paired sensitivity comparisons, not 27 independent studies. All 81 rows and every checkpoint frame/eigenspectrum are retained in dimension-results. None were dropped or used to tune the algorithm. The table reports empirical ranges, not confidence intervals.

Table 7. Complete dimension-panel summary at $n = 262144$; three-run minima and maxima are empirical ranges. All 81 checkpoint rows remain in the data artifact.

p	a	Fourth mean [min, max]	Second mean	m/p
60	0.25	0.9611 [0.9587, 0.9638]	0.2665	0.2667
60	0.4	0.5324 [0.5139, 0.5605]	0.2704	0.2667
60	0.5	0.2789 [0.2768, 0.2820]	0.2688	0.2667
120	0.25	0.4177 [0.4057, 0.4378]	0.1380	0.1333
120	0.4	0.1740 [0.1692, 0.1769]	0.1354	0.1333
120	0.5	0.1431 [0.1297, 0.1498]	0.1339	0.1333
240	0.25	0.0815 [0.0780, 0.0840]	0.0653	0.0667
240	0.4	0.0697 [0.0649, 0.0723]	0.0652	0.0667
240	0.5	0.0681 [0.0643, 0.0715]	0.0646	0.0667

C Sharper concentration and adaptive cubic sketches

C.1 A dimension-free comparison for degree-six observables

Lemma 1. *Let P be the angular unit-direction law with $0 \leq a \leq 1/2$ and $c = \sin(2\pi a)/(2\pi a)$, with the continuous value at zero. Let Q_0 be uniform on the sphere. If a real polynomial f has degree at most six and is nonnegative on the sphere, then*

$$(1 - c)\mathbb{E}_{Q_0}f \leq \mathbb{E}_P f \leq (1 + c)\mathbb{E}_{Q_0}f.$$

The constants do not grow with the number of carrier blocks, and the statement includes the singular endpoint $a = 0$.

Proof. Use the secret coordinates only for this proof. Condition on all pair radii and on the bulk coordinates. Their joint law, and the total length, are unchanged by angular embedding. Average f over each uniform reference phase and over the fair sign of each encoding pair. This preserves nonnegativity. A surviving nonconstant relative-angle Fourier term in a block must have even frequency at least two, and must balance that frequency between its reference and encoding pair. It therefore requires degree at least four. Two nonconstant blocks, or relative frequency four, would require degree at least eight. The averaged degree-six polynomial consequently has the form

$$F(\phi_1, \dots, \phi_{b_0}) = F_0 + \sum_j (u_j \cos 2\phi_j + v_j \sin 2\phi_j).$$

Since the angles can be minimized independently, nonnegativity gives $\sum_j \sqrt{u_j^2 + v_j^2} \leq F_0$. Under the clean law its mean is F_0 . Under the marked law, $\mathbb{E} \cos 2\phi_j = -c$ and $\mathbb{E} \sin 2\phi_j = 0$, so its mean is $F_0 - c \sum_j u_j$, between $(1 - c)F_0$ and $(1 + c)F_0$. Integrating the conditioning variables proves the result. This argument uses a finite Fourier expansion, not a bound on the full density ratio, which would diverge at the singular endpoint.

C.2 Exact probe fluctuations and a stronger gap

Write the six modes in Appendix B with independent standard Gaussian coefficients. Wick contraction gives

$$\begin{aligned} \mathbb{E}_A C(A)^2 &= 6c^2 P_U, \\ \mathbb{E}_A C(A)^4 &= c^4 \sum_{r,s} (B_r^2 B_s^2 + B_r B_s B_r B_s + B_r B_s^2 B_r) = 72c^4 P_U. \end{aligned}$$

The identities hold on every block and hence on U . Restrict to U and put $X_\ell = 6c^2 I_m - C(A_\ell)^2$. Then

$$\mathbb{E} X_\ell = 0, \quad \lambda_{\max}(X_\ell) \leq 6c^2, \quad \mathbb{E} X_\ell^2 = 36c^4 I_m.$$

The one-sided hypothesis of matrix Bernstein [9, Theorem 6.1] requires an upper eigenvalue bound, not an absolute norm bound. Thus the unbounded negative eigenvalues of X_ℓ cause no problem and no Gaussian probes are truncated. With $u = \log(m/\delta_A)$, the theorem gives

$$\lambda_{\min}(G|_U) \geq c^2 g, \quad g = 6 - \sqrt{72u/L_0} - 4u/L_0$$

except with probability δ_A , whenever $g > 0$. For $m = 16$, $\delta_A = 0.025$, the factors g at 24 and 64 probes are approximately 0.5203 and 2.9000. The six-dimensional smallest-singular-value argument gives no positive bound at 24 probes and about 0.4350 at 64. This improves the proof for the same probes, without changing their realized population matrix.

C.3 Operator variance instead of total error energy

For a trace-zero symmetric probe the original control-variate summand is

$$H_A(q) = p(p+2)(q^T A q) q q^T - p(q^T A q) I_p - 2p(A q q^T + q q^T A) + 2A.$$

It is the summand used by the existing fixed-probe implementation. Although H_A has degree four, using $(q q^T)^2 = q q^T$ and $\|q\| = 1$ reduces every entry of $H_A(q)^2$ to a polynomial of degree at most six on the sphere. For each fixed x , $x^T H_A(q)^2 x$ is nonnegative. Lemma 1 therefore implies

$$\mathbb{E}_P H_A(q)^2 \preceq (1+c) \mathbb{E}_{Q_0} H_A(q)^2.$$

Direct spherical moments give the exact clean identity

$$\begin{aligned} \mathbb{E}_{Q_0} H_A(q)^2 &= J_p \operatorname{tr}(A^2) I_p + N_p A^2, \\ J_p &= \frac{2p(p^2 + p + 4)}{(p+2)(p+4)}, \quad N_p = \frac{4(p^3 - p^2 + 10p - 8)}{(p+2)(p+4)}. \end{aligned} \tag{25}$$

For completeness, write $s = q^T A q$, $Q = q q^T$, $M = A Q + Q A$ and $r = q^T A^2 q$. For $H = \alpha s Q + \beta s I + \eta M + \zeta A$, expand its square with $Q^2 = Q$, $Q M + M Q = 2s Q + M$ and $M^2 = s M + A Q A + r Q$. Thus, with $d_1 = \alpha^2 + 2\alpha\beta + 2\alpha\eta$ and $d_2 = \eta^2 + \alpha\eta + \alpha\zeta + 2\beta\eta$,

$$H^2 = d_1 s^2 Q + \beta^2 s^2 I + d_2 s M + \eta^2 (A Q A + r Q) + \zeta^2 A^2 + 2\beta\zeta s A + \eta\zeta (A M + M A).$$

Use $\mathbb{E}_{Q_0} s A = 0$, $\mathbb{E}_{Q_0} (A M + M A) = 4A^2/p$, and

$$\begin{aligned} \mathbb{E}_{Q_0} s^2 &= 2 \operatorname{tr}(A^2)/(p(p+2)), \\ \mathbb{E}_{Q_0} s^2 Q &= (2 \operatorname{tr}(A^2) I + 8A^2)/(p(p+2)(p+4)), \\ \mathbb{E}_{Q_0} s M &= 4A^2/(p(p+2)), \quad \mathbb{E}_{Q_0} r Q = (\operatorname{tr}(A^2) I + 2A^2)/(p(p+2)), \\ \mathbb{E}_{Q_0} A Q A &= A^2/p. \end{aligned}$$

Substitute $(\alpha, \beta, \eta, \zeta) = (p(p+2), -p, -2p, 2)$ to get (25). Exact rational integration of unreduced polynomial entries, including degree-eight terms, independently checks it.

Condition on the independent probes and set $a_2 = L_0^{-1} \sum_{\ell} \|A_{\ell}\|_{\text{op}}^2$ and $f_2 = L_0^{-1} \sum_{\ell} \|A_{\ell}\|_F^2$. Centering a symmetric summand subtracts $C(A)^2$ from its second moment. Hence its variance norm is at most $(1+c)(J_p \text{tr}(A^2) + N_p \|A\|_{\text{op}}^2)$. For the stacked centered summand $Z = L_0^{-1/2} [H_{A_1} - C(A_1); \dots; H_{A_{L_0}} - C(A_{L_0})]$, both rectangular variance norms are bounded by

$$V = (1+c)(J_p f_2 + N_p a_2).$$

For $\mathbb{E} Z^T Z$ this follows by summing the individual bounds. For $\mathbb{E} Z Z^T$, apply the triangle inequality in L^2 to $\sum_{\ell} (H_{A_{\ell}} - C(A_{\ell})) x_{\ell}$, then Cauchy–Schwarz. This does not assume independent probe errors on the shared observations.

The block contraction map has Frobenius operator norm at most $2c$; its four-dimensional input block has Frobenius norm at most $2\|A\|_{\text{op}}$. Thus $\|C(A)\|_{\text{op}} \leq 4c\|A\|_{\text{op}}$. Indeed, the displayed block coordinates give squared Frobenius norm $c^2(2u^2 + 2v^2 + 2w^2 + 2t^2 + 4r^2 + 4s^2)$, at most four times c^2 times the input block’s squared Frobenius norm. The elementary bound on H_A then gives

$$\|Z\|_{\text{op}} \leq R = (p^2 + 7p + 2 + 4c)\sqrt{a_2}.$$

Rectangular matrix Bernstein [9, Theorem 1.6], in dimension $pL_0 \times p$, yields $\|\widehat{B} - B\|_{\text{op}} \leq \tau$ with conditional failure at most δ_q , where

$$x = \log(p(L_0 + 1)/\delta_q), \quad \tau = \sqrt{2Vx/n} + 2Rx/(3n).$$

The sample size may be selected from the independently drawn probes before seeing any observations. Combining the conditional data event with the unconditional probe event gives total failure at most $\delta_A + \delta_q$. If $\tau < c\sqrt{g}$, the smallest retained singular value is at least $c\sqrt{g} - \tau$. Projecting the right singular-vector equation onto $U^{\perp}$ gives

$$o \geq \left[1 - \left(\frac{\tau}{c\sqrt{g} - \tau} \right)^2 \right]_+.$$

For GOE probes on the usual scale $a_2 = O(p)$ and $f_2 = O(p^2)$, the variance term is $O(p^3)$ and $R = O(p^{5/2})$. At fixed positive g , the leading sufficient count is $O(p^3 \log(p(L_0 + 1)/\delta_q)/c^2)$. This is a sufficient bound, not a minimax rate. The observation law, independent probe design and exact-arithmetic scope remain essential.

C.4 A budget fixed before drawing the probes

The norm and gap events need not be independent. No probes are clipped, screened or redrawn. The corollary concerns each fixed unknown frame, not a

single event uniform over all frames, and does not allow data-dependent stopping. Its deterministic norm allowance is deliberately conservative. The numerical examples in Section 2 use realized probe norms with failure budget $\delta_A + \delta_q$; they are not instances of this three-event deterministic count.

Which part of the count can improve? Put $b = \rho c \sqrt{g}$, $n_V = 2Vx/b^2$ and $\kappa = Rb/(3V)$. Before rounding to an integer, inverting the displayed τ gives

$$n_\tau = n_V \left(\frac{1 + \sqrt{1 + 4\kappa}}{2} \right)^2, \quad n_V(1 + \kappa) \leq n_\tau \leq n_V(1 + 2\kappa).$$

For the realized-probe examples at $p = 120, 240, 480$, $a = 1/4$, even formally deleting the entire R term lowers these sufficient counts by only 1.664%, 1.166% and 0.827%. This is a diagnostic limit holding V, g, x, c and the target fixed, not a realizable zero-norm bound or an estimation lower bound. In the same examples $J_p f_2$ accounts for 94.0%–98.4% of V . It arises from the exact clean second moment, so its large share alone does not make it removable slack. Further gains require justified variance, spectral or estimator analysis; these calculations do not establish an optimal dimension exponent.

C.5 Exact working-subspace support and offline geometry

Exact population support. Write $S_U = Q^T S_W Q$. If $s = \text{rank}(M) > 0$, factor $M = LL^T$ with $L \in \mathbb{R}^{m \times s}$ of full column rank, and set $A = F(L^{\otimes 3})$ and $J_{(a,b,d),e} = \delta_{ae}\delta_{bd}$. Then $D_M L = AJ$ and $J^T J = sI_s$, so the covariance identity becomes

$$S_U = A(6I_{s^3} + 9JJ^T)A^T, \quad 6AA^T \preceq S_U \preceq (6 + 9s)AA^T.$$

Thus $\text{range}(S_U) = \text{range}(A)$. Full carrier signal holds exactly when A has row rank m . The internal-pairing term can change eigenvalues but cannot enlarge this range. Symmetry gives $\text{rank}(S_U) \leq \min\{m, \binom{s+2}{3}\}$. If $s = 0$, then $S_U = 0$. This is an algebraic support criterion, with no finite-sample or adaptive guarantee.

Two examples, both with $c = 1$, clarify the limits of simpler criteria. First let $p = 8$, $m = r = 4$, $U = \text{span}(e_1, \dots, e_4)$, and

$$W_1 = \text{span}\{(e_i + e_{i+4})/\sqrt{2} : 1 \leq i \leq 4\}, \quad W_2 = \text{span}(e_1, e_2, e_5, e_6).$$

Their compressions are $M_1 = I_4/2$ and $M_2 = \text{diag}(1, 1, 0, 0)$, with the same mean coverage $1/2$. Nevertheless $S_U(M_1) = 9I_4/4$, whereas $b(v) = 0$ for every $v \in W_2$ and hence $S_U(M_2) = 0$. More generally, coverage $\omega = \text{tr}(M)/m$ implies only $\lambda_{\min}(M) \geq \max\{0, m\omega - (m - 1)\}$ without further structure.

Second, a working rank below m need not lose any carrier direction. For one four-dimensional block, take $W = \text{span}(e_1 + e_3, e_2 + 2e_4)$, so $r = 2 < m = 4$. In carrier coordinates a Gaussian feature is $v = (x, y, x, 2y)$, with independent $x \sim N(0, 1/2)$ and $y \sim N(0, 1/5)$. Direct substitution gives

$$b(v) = (-3x^3/2, -9x^2y/2 - 6y^3, -3x^3/2 - 9xy^2/2, -3y^3)^T.$$

Its coefficient matrix in (x^3, x^2y, xy^2, y^3) has determinant $729/8$. Thus every nonzero output linear combination is a nonzero polynomial, whose squared expectation under this nondegenerate Gaussian is positive. The covariance therefore has rank four although $\lambda_{\min}(M) = 0$. All these statements concern the population map; neither mean coverage nor a positive population gap certifies the same-data empirical update.

Geometry of saved trajectories. We evaluate Proposition 1 on all 127 saved records in the two refinement panels. Repeated final/stage records are retained, so these are not 127 independent trials. The diagnostic uses the ordered synthetic secret carrier and the full Gaussian covariance formula, including both its six- and nine-pairing terms. It is an offline model calculation in ordinary floating point, not a public certificate or an input to the learner. At exact hiding $c = 0$, the actual population covariance is zero regardless of a counterfactual unit- c geometry score.

Table 8. Offline geometry of the eight saved primary trajectories. $\omega = \text{tr}(M)/16$, $\lambda = \lambda_{\min}(M)$ and $\eta = \lambda_{\min}(Q^T S_W Q)/(18c^2)$. Entries are means; full ranges are in the machine-readable record. The key is used only for this diagnostic.

Frame	Rank	ω	λ	η
Initial	64	0.42408	0.17100	0.04509
After stage 1	64	0.48343	0.20861	0.05872
After stage 2	32	0.40167	0.08008	0.01440
Final	16	0.59985	0.00532	0.00712

Table 8 shows that average coverage and weakest population direction need not improve together. From the initial to final frame, mean coverage rises from 0.4241 to 0.5999, while mean $\lambda_{\min}(M)$ falls from 0.1710 to 0.00532; its smallest final value is 9.03×10^{-5} . These values do not assert signal loss or explain a finite-sample noise mechanism. Geometry of the final rank-16 frame describes a potential *next* feature draw; the recorded final update instead uses the preceding rank-32 frame. The exact population formula, eight rational covariance cases, fifteen numerical cases, four support factorizations and four independent 15-pairing archive checks are supplied in geometry-checks. The full diagnostic is geometry-diagnostics.

C.6 Verification and prospective confirmation

The six-mode second and fourth moments are checked by exact integer Wick contractions. The clean operator variance is independently integrated with exact rational spherical moments at three dimensions for both the original and harmonic summands. A two-block Fourier grid checks the degree-six support

and nonnegativity argument. Separate direct per-observation calculations check the streamed diagonal subtraction and cubic moments.

Exploration used data seed 501 only. The final rank schedule, feature count, six source/parameter files and environment were then bound in refinement-protocol before data seeds 601–603 were run. The four conditions compare a low observation budget, the primary 240-dimensional case, its exact-hiding control, and transfer to 480 dimensions. Both baselines and all three refinement stages are retained. The same seed triple is reused across conditions, and the low budget is a prefix; the records are not independent replicates. Timings are descriptive single-machine measurements. Baseline times exclude shared synthetic data generation; refinement times include initialization plus all three passes and their synthetic replay overhead. Thus these timings conservatively charge data regeneration to refinement and are not pure kernel benchmarks. The saved runner’s timing-scope string omits this asymmetry; the reported measurements therefore require the timing interpretation above.

Table 9. Frozen confirmation: mean rank-16 overlap over three held-out seed groups. Refinement uses four passes over the same n observations; baselines use one.

p	a	n	24 probes	48 probes	Refined	m/p
240	0.25	262144	0.0799	0.0845	0.0784	0.0667
240	0.25	1048576	0.1486	0.2004	0.6810	0.0667
240	0.5	1048576	0.0716	0.0704	0.0648	0.0667
480	0.25	1048576	0.0382	0.0382	0.0372	0.0333

C.7 Controls with the same total cubic-feature budget

The equal-feature-budget comparison fixes $p = 240, m = 16, a = 1/4, n = 2^{20}$ and eight seed groups (701–708). The schedule, all implementation dependencies and analysis scope were bound before their first run in equal-feature-protocol. This is a local prospective freeze after method selection, not externally time-stamped preregistration. Each of the two fixed controls pools three blocks of 1024 Gaussian vectors into one diagonal-free matrix. The ambient control draws these in $\mathbb{R}^p$ independently of observations. The frozen-initial control instead multiplies all three blocks by the same learned initial 64-frame. It therefore reuses data for initialization, while omitting later feedback. The adaptive method retains the unchanged 64–32–16 schedule.

All cubic methods use 3072 total vectors and the same diagonal removal. They have different numbers of data passes and do not have equal time budgets. All methods read the identical public float64 array; reported times exclude generation and saving/scoring, and include probe/vector construction, accumulation and eigensolves. Methods requiring initialization are charged its separately

Table 10. All refinement stages. Ranges are three-run minima and maxima, not confidence intervals. Seconds include cumulative fit and adaptive replay overhead as described above.

Condition	Method	Overlap mean [min, max]	Seconds
$p = 240, a = 1/4, n = 2^{18}$	24 probes	0.0799 [0.0756, 0.0864]	4.67
	48 probes	0.0845 [0.0827, 0.0868]	9.24
	Pass 2	0.0790 [0.0753, 0.0842]	7.27
	Pass 3	0.0743 [0.0708, 0.0763]	9.80
	Pass 4	0.0784 [0.0708, 0.0840]	12.32
$p = 240, a = 1/4, n = 2^{20}$	24 probes	0.1486 [0.1448, 0.1548]	18.88
	48 probes	0.2004 [0.1816, 0.2127]	37.81
	Pass 2	0.1922 [0.1651, 0.2189]	29.21
	Pass 3	0.3393 [0.2559, 0.4001]	39.69
	Pass 4	0.6810 [0.5709, 0.7623]	50.19
$p = 240, a = 1/2, n = 2^{20}$	24 probes	0.0716 [0.0651, 0.0779]	19.72
	48 probes	0.0704 [0.0662, 0.0784]	39.46
	Pass 2	0.0698 [0.0666, 0.0749]	30.53
	Pass 3	0.0657 [0.0581, 0.0726]	41.18
	Pass 4	0.0648 [0.0612, 0.0683]	52.11
$p = 480, a = 1/4, n = 2^{20}$	24 probes	0.0382 [0.0343, 0.0420]	84.41
	48 probes	0.0382 [0.0376, 0.0392]	168.35
	Pass 2	0.0345 [0.0297, 0.0401]	104.35
	Pass 3	0.0333 [0.0305, 0.0375]	124.78
	Pass 4	0.0372 [0.0344, 0.0402]	145.74

measured cost, even when that computation is shared by the runner. Subsequent method order rotates by seed. The larger 96-probe comparison is restricted prospectively to the first three groups; it is not an eight-seed benchmark. Raw records include every scheduled result.

Table 11. All paired final overlaps in the equal-feature-budget panel. The 96-probe comparison was prespecified for seeds 701–703 only; dashes mean not scheduled.

Seed	24	48	Fixed ambient	Frozen initial	Adaptive	96
701	0.1477	0.2005	0.1355	0.3082	0.5497	0.3016
702	0.1404	0.2012	0.1285	0.3483	0.7131	0.2986
703	0.1359	0.1797	0.1336	0.3271	0.6540	0.2671
704	0.1485	0.1851	0.1089	0.2972	0.5118	–
705	0.1455	0.2040	0.1272	0.3135	0.6391	–
706	0.1431	0.1907	0.1304	0.3113	0.4804	–
707	0.1517	0.1863	0.1185	0.3887	0.5810	–
708	0.1493	0.1924	0.1167	0.3318	0.6698	–

Table 12. Adaptive working-subspace coverage $\text{tr}(P_{\text{stage}}P_U)/16$, using every retained column. Working ranks vary, and successive frames need not be nested.

Seed	Initial 64	Stage 1: 64	Stage 2: 32	Stage 3: 16
701	0.4135	0.4556	0.3763	0.5497
702	0.4295	0.5179	0.4593	0.7131
703	0.4079	0.4841	0.4132	0.6540
704	0.4164	0.4582	0.3702	0.5118
705	0.4237	0.4800	0.4099	0.6391
706	0.4238	0.4660	0.3439	0.4804
707	0.4477	0.5146	0.4337	0.5810
708	0.4301	0.4909	0.4069	0.6698

All 60 refinement-panel frames were also rescored at their entire working rank. Their initial 64-frame coverage means are 0.3056 (low budget), 0.4266 (primary), 0.2719 (hiding), and 0.1476 (480 dimensions), versus random 64/ p . These are post-hoc diagnostics, not evidence of a proved bootstrap threshold. Coverage of 64-, 32- and 16-dimensional frames is not directly comparable as a monotone sequence, and updates may leave the preceding frame.

A separate public-input CLI process reproduces refinement-panel seed 602 at overlap 0.7099290 with the same projector as the archive. Its inputs contain only the public direction array and independent probe/feature seeds; neither key nor production seed is passed to the learner. This checks the trusted implementa-

tion’s input interface, not operating-system isolation. The complete records and hashes remain in public-input-replay.

C.8 Protocol for the paired middle-rank intervention

Table 13. Post-hoc intervention on the eight existing primary seeds. Each schedule has three blocks of 1024 Gaussian features and the same observations. The common 32-dimensional coefficients of the last block are coupled; the added directions use a separate frozen stream. This is a mechanism diagnostic, not a new confirmation panel.

Schedule	Mean overlap	Sample SD	Minimum–maximum
64–32–16	0.5999	0.0820	0.4804–0.7131
64–64–16	0.4357	0.0579	0.3344–0.5086
Paired difference (32 minus 64)	0.1641	0.0534	0.0752–0.2279

This post-hoc diagnostic uses all eight previously observed seed groups 701–708, $p = 240$, $m = 16$, $a = 1/4$, $n = 2^{20}$, and 1024 features per stage. Its specification, analysis script and implementation dependencies were hash-bound in rank-protocol before execution. This is a local pre-run record, not prospective confirmation on unseen seeds. The primary statistic is the continuous paired difference; no significance or equivalence cutoff, extra seeds or best-branch selection was introduced after observing these results.

The original public observations are regenerated from the frozen recipe. Seed 701 freshly computes the baseline and first update, verifying both against archived projectors; the other seven reuse the archived first-update frame. Every second-update operator is recomputed and its top 32-space checked against the archive. Let Q_{32} be that common orthonormal frame, and let E_{32} complete it to the top 64-space of the same operator. With independent matrices $G_a, G_b \in \mathbb{R}^{32 \times 1024}$ of iid standard Gaussian entries, the last feature blocks are

$$V_{32} = Q_{32}G_a, \quad V_{64} = Q_{32}G_a + E_{32}G_b.$$

The common stream is the original last-stage feature stream. The extra stream uses the separately frozen seed offset 10^6 . Merely using the same RNG seed for arrays with different row counts would align the wrong eigenvector rows. The displayed coupling gives the correct Gaussian law on each space, without promising reduced variance of the comparison. It is not a claim of bitwise identity with a separately seeded 64–64–16 implementation. Both branches retain 16 final directions.

The saved second-stage spectra also permit a numerical cutoff diagnostic. Across the eight groups, the smallest gaps at ranks 32 and 64 are 0.00958 and 0.01095, respectively; their smallest ratios to the operator norm are 0.00298 and 0.00353. These post-hoc checks use the saved eigenvalues and do not certify stability under sampling or arithmetic perturbations. All values are in rank-spectra.

Each logical schedule uses 3072 features and four passes over the same observations. Reused prefix work is distinguished from fresh work; recorded durations are operational measurements, with no new end-to-end speed ranking. Branch order alternates by seed. Each standard final branch is checked against its archived projector and overlap, and an independent projector calculation rescored all sixteen final frames with zero recorded score discrepancy. The raw rows, shared spaces, final frames, freeze and execution records are supplied in rank-records and rank-protocol-records; rank-summary regenerates the tables. These controls assess this rank intervention, while same-data feedback remains outside the fixed-probe theorem.

Table 14. Every paired final overlap in the rank intervention.

	Seed 64–32–16	64–64–16	Difference
701	0.549692	0.415575	0.134117
702	0.713119	0.508557	0.204562
703	0.653989	0.426075	0.227914
704	0.511780	0.392863	0.118917
705	0.639092	0.441293	0.197800
706	0.480369	0.334396	0.145973
707	0.580972	0.505758	0.075214
708	0.669815	0.461230	0.208585

D Public validation: proofs and implementation

This appendix proves the statements used in Theorems 4 and 5. All probabilities refer to the exact angular model of Theorem 1, conditional on a fixed key. The parameter range of the main results is $0 \leq a \leq 1/2$. Write

$$c = \frac{\sin(2\pi a)}{2\pi a}, \quad c(0) = 1, \quad m = 4b_0 < p, \quad P_U = QQ^T,$$

where Q is an orthonormal carrier frame. A candidate $B \in \mathbb{R}^{p \times r}$ has orthonormal columns and defines $P = BB^T$ and overlap $o(P) = \text{tr}(PP_U)/m$. The candidate can be the output of any training procedure, including one that reuses its training sample adaptively. It must be fixed before the validation directions are observed. Conditional on the training history, these are fresh iid observations from the same fixed-key model. Integrating conditional coverage over the training history gives unconditional coverage.

Reading routes and proof dependencies. The covariance and confidence arguments are organized below by their role. For the direct one-index certificate, read Sections D.1–D.4 in order. For the compressed streaming certificate, add

Sections D.5–D.6. The general partial-index construction in Section D.11 reuses the same covariance and confidence argument with a different compression operator; it supplies the exact spectrum and trace of the squared operator, $\text{tr}(L_k^2)$, for all four choices of index count. Each inversion uses a choice of k fixed before validation. Reporting several choices together requires a prespecified joint confidence budget as in Section D.10; choosing k after inspecting the data is not covered by an individual fixed- k bound. The repeated inversion statement makes the general compression result self-contained.

Task	Proof or evidence entry
Feature and sharp variance	Sections D.1–D.2, page 40: radial transfer, centered covariance and the sharpness witness.
Direct overlap inference	Sections D.3–D.4, page 43: unbiased kernel, variance, one-sided inversion and incomplete statistics.
Streaming and its cost	Sections D.5–D.7, page 47: retained energy versus overlap, closed kernel and computational cost.
Power and sample planning	Sections D.8–D.9, page 50: power bounds and sample budgets conditional on a candidate-quality premise.
Coverage and exact spectra	Sections D.10–D.11, page 52: joint coverage, numerical scope, partial-index spectra and stage acceptance.
All validation outcomes	Section D.12, page 59: every saved candidate, both scale branches and reproducibility limits.
Recovery and testing limits	Section D.13, page 61: full-signal conditions, support estimation, whitening, ordered bits, finite-sample tails and rank selection.

The final group states separate extensions and applicability limits; its support-learning and preprocessing results are not premises of the public-certificate theorem. Coverage, positive-certificate power and candidate quality are distinct claims throughout these reading routes.

D.1 The trace-free feature and radial transfer

Let $\mathcal{H}_4$ be the real Hilbert space of symmetric trace-free fourth-order tensors on $\mathbb{R}^p$, with Frobenius inner product, and let Π_4 be its orthogonal projection. Put $A = p(p+2)$, $d = p+4$, and $H(q) = A\Pi_4(q^{\otimes 4})$. For unit q ,

$$H(q) = A \left[q^{\otimes 4} - \frac{\text{Sym}_6(q^{\otimes 2} \otimes I_p)}{d} + \frac{\text{Sym}_3(I_p \otimes I_p)}{(p+2)d} \right]. \quad (26)$$

The symbols $\text{Sym}_6, \text{Sym}_3$ denote sums over the six and three distinct placements, not averages. The cubic feature is $h_v(q) = H(q)(\cdot, v, v, v)$. The block tensor of Theorem 1 gives

$$\mathbb{E}H(q) = \mathcal{T}, \quad \sum_{\alpha, \beta, \gamma} \mathcal{T}_{i\alpha\beta\gamma} \mathcal{T}_{j\alpha\beta\gamma} = 3c^2(P_U)_{ij}, \quad \|\mathcal{T}\|_F^2 = 3c^2m. \quad (27)$$

Here $\mathcal{T} \in \mathcal{H}_4$. These are identities of the specified block law, not consequences of isotropy alone.

Lemma 2 (Independence of the total radius). *The marked vector admits $Z = Lq$, with L independent of q , $L^2 \sim \chi_p^2$, $\mathbb{E}L^4 = A$, and $\mathbb{E}L^8 = p(p+2)(p+4)(p+6)$.*

Proof. Write the original Gaussian input as $G = R\Omega$, where $R^2 \sim \chi_p^2$ and R is independent of the uniform direction Ω and fresh payload randomness ε . The angular endpoint map E_a retains every pair radius and the Gaussian bulk. Its changes of angle depend only on angles and ε , so it preserves the norm and is positively homogeneous:

$$E_a(R\Omega, \varepsilon) = RE_a(\Omega, \varepsilon).$$

Thus $Z = RE_a(\Omega, \varepsilon)$, $L = R$, and $q = E_a(\Omega, \varepsilon)$. This proves independence including the ratios of the different block radii; independence of angles alone would not suffice. The moments follow from the chi-square law.

D.2 A centered fourth-order covariance comparison

Define

$$\kappa_p = \frac{24p(p+2)}{(p+4)(p+6)} < 24, \quad C = (1 + |c|)^2 \leq 4, \quad \Sigma = \text{Cov}(H(q)). \quad (28)$$

Lemma 3 (Local Hermite blocks). *For one marked block, let G_k be the raw second-moment matrix in an orthonormal basis, under the standard Gaussian measure, of the k th homogeneous Gaussian Hermite chaos. Then*

$$G_1 = I, \quad G_2 \preceq (1 + |c|)I, \quad G_3 \preceq (1 + |c|)I, \quad G_4 \preceq (1 + 2|c|)I.$$

The first-, second-, and third-chaos means vanish. The cross-moment operator between the first and third chaoses has norm $|c|/\sqrt{2}$.

Proof. Use polar coordinates $(R \cos \theta, R \sin \theta, T \cos \phi, T \sin \phi)$. Conditional on the two radii, a polynomial in Gaussian Hermite chaos k has Fourier frequencies (u, v) satisfying $|u| + |v| \leq k$ and $u + v \equiv k \pmod{2}$. First combine all terms with the same frequency (u, v) into one coefficient, which may depend on the radii. Different ordinary monomials need not have different Fourier frequencies. Under independent uniform phases, the conditional squared norm is the sum of squared absolute values of these combined coefficients.

In the marked law, $\phi = \theta + \pi/2 + D + \pi J$, where J is a fair bit and D is uniform on $[-\pi a, \pi a]$. Two frequencies interact only when their totals $u + v$ agree and their second-frequency difference is even. Each interacting group has at most two members for $k = 2, 3$, and at most three for $k = 4$. Its Hermitian Gram matrix has diagonal entries one. Off-diagonal entries have, up to a unit-modulus phase, magnitude $|c|$ for a frequency difference of two and magnitude

$$|c_4| = \left| \frac{\sin(4\pi a)}{4\pi a} \right| = |c \cos(2\pi a)| \leq |c|$$

for a difference of four, using continuous values at $a = 0$. The row-sum bound gives largest eigenvalue at most $1 + |c|$ or $1 + 2|c|$ for a two- or three-member group, respectively. The middle row of a three-member group can contain two entries of magnitude $|c|$; a bound $1 + |c| + |c_4|$ does not follow. Integrating the unchanged independent Rayleigh radii gives the asserted raw Gram bounds in Gaussian-normalized Hermite bases.

Isotropy gives $G_1 = I$ and the zero second-chaos mean. Central symmetry of the block gives zero odd-chaos means. For a symmetric cubic coefficient tensor D , Gaussian normalization is $\mathbb{E}_0 \langle D, :X^{\otimes 3}: \rangle^2 = 3! \|D\|_F^2$. The cross moment of $:X^{\otimes 3}:$ and X is the $3|1$ flattening of the fourth cumulant, since Hermite subtraction cancels the Gaussian pairings. Its norm is $\sqrt{3}|c|$ by (27). Dividing by $\sqrt{3!}$ gives $|c|/\sqrt{2}$. The corresponding Gaussian-bulk cross moment is zero.

Proposition 2 (Centered covariance and its sharp constant). *In the angular model,*

$$\Sigma \preceq C \kappa_p I_{\mathcal{H}_4}. \quad (29)$$

For a uniform-sphere direction, $\Sigma = \kappa_p I_{\mathcal{H}_4}$. If the carrier has at least two blocks, then

$$\|\Sigma\|_{\text{op}} = \kappa_p (1 + |c|)^2. \quad (30)$$

Proof. First consider the unnormalised vector Z . Decompose Gaussian fourth Hermite chaos by the local degrees on independent blocks, including the Gaussian bulk. The types are

$$(4), \quad (3, 1) + (1, 3), \quad (2, 2), \quad (2, 1, 1), \quad (1, 1, 1, 1).$$

Center each single-block fourth-chaos component within that block. Other types already have zero mean. Distinct supports or degree types are orthogonal under the marked law except for $(3, 1)$ and $(1, 3)$ on the same pair of blocks. Indeed, a block appearing in only one support contributes a zero local mean, unless it is a single fourth-chaos block, which has just been centered. On matching supports, a first- versus second-chaos pairing vanishes by central symmetry. Precisely the third- versus first-chaos coupling remains.

Relative to their Gaussian norms, the five covariance operators are bounded respectively by

$$1 + 2|c|, \quad 1 + |c| + \frac{c^2}{2}, \quad (1 + |c|)^2, \quad 1 + |c|, \quad 1. \quad (31)$$

For the first, centering subtracts a positive semidefinite matrix from the raw Gram matrix. For the second, each diagonal block is at most $(1 + |c|)I$, and its off-diagonal block has norm at most $(|c|/\sqrt{2})^2 = c^2/2$ by independence. The other bounds are tensor products of the local Gram bounds and first-chaos identities. Every quantity in (31) is at most C .

If $D \in \mathcal{H}_4$, then $F_D(Z) = \langle D, Z^{\otimes 4} \rangle$ lies entirely in Gaussian fourth chaos: lower-order Hermite corrections contract a trace of D . The Gaussian isometry therefore yields

$$\text{Var } F_D(Z) \leq 24C \|D\|_F^2. \quad (32)$$

Set $f_D(q) = \langle D, H(q) \rangle$. Lemma 2 gives $F_D(Z) = L^4 f_D(q)/A$ and $\mathbb{E}F_D = \mathbb{E}f_D$. With $\eta = A^2/\mathbb{E}L^8 = \kappa_p/24$, the exact centered relation is

$$\text{Var } f_D = \eta \text{Var } F_D - (1 - \eta)(\mathbb{E}F_D)^2 \leq C\kappa_p \|D\|_F^2. \quad (33)$$

This proves (29). For a Gaussian vector the mean is zero and the fourth-chaos isometry gives equality with $C = 1$, proving the uniform-sphere assertion.

For sharpness, on one carrier block take $f_2(x) = (x_2x_3 - x_1x_4)/\sqrt{2}$. Here $c \geq 0$ in the main parameter range. Its moments are

$$\mathbb{E}f_2 = 0, \quad \mathbb{E}_0f_2^2 = 1, \quad \mathbb{E}f_2^2 = \frac{(1 + c/2) + (1 + c/2) - 2(-c/2)}{2} = 1 + c.$$

Multiply this polynomial on two different blocks: $F(Z) = f_2(Z_1)f_2(Z_2)$. Independence gives $\mathbb{E}F = 0$, $\text{Var}_0 F = 1$, and $\text{Var } F = (1 + c)^2$. Every monomial uses four distinct coordinates, so F is harmonic and homogeneous of degree four. Write $F(Z) = \langle D, Z^{\otimes 4} \rangle$ with $D \in \mathcal{H}_4$. The Gaussian isometry gives $24\|D\|_F^2 = 1$. Its mean correction in (33) is zero, so the Rayleigh quotient is $\kappa_p(1 + c)^2$, proving equality.

As an algebraic extension of this block-law calculation, at a parameter with $c < 0$ the choice $(x_1x_3 + x_2x_4)/\sqrt{2}$ has second moment $1 - c$ and proves the same equality with $|c|$. This does not extend the parameter range of the main training or angular-margin theorems. No single-block sharpness assertion is needed, and not every candidate direction attains the worst-case variance.

Centering is essential: $\mathbb{E}[H \otimes H] = \Sigma + \mathcal{T} \otimes \mathcal{T}$ contains a mean direction of squared norm $3c^2m$. For any fixed linear map L_0 , however,

$$\text{Cov}(L_0H(q)) \preceq C \text{Cov}(L_0H(q_0)),$$

where q_0 is uniform on the sphere. The map must be independent of the directions used in this assertion.

D.3 The direct unbiased overlap kernel

For independent $v \sim N(0, I_p)$, define $K_P(q, q') = \mathbb{E}_v[h_v(q)^T P h_v(q')]$. For unit q, q' , put $s = q^T q'$, $x = q^T P q$, $y = q'^T P q'$, and $t = q^T P q'$. Exact Gaussian integration gives

$$\begin{aligned} K_P(q, q') = 6A^2 \left[t \left(s^3 - \frac{3ps}{d^2} \right) - \frac{3(x+y)}{d} \left(s^2 - \frac{1}{d} \right) \right. \\ \left. + \frac{3r}{d^2} \left(s^2 - \frac{1}{p+2} \right) \right]. \end{aligned} \quad (34)$$

Of the fifteen sixth-moment Gaussian pairings, nine contract two indices within the same trace-free tensor and vanish. The remaining six give

$$K_P(q, q') = 6 \sum_{i,j,\alpha,\beta,\gamma} P_{ij} H(q)_{i\alpha\beta\gamma} H(q')_{j\alpha\beta\gamma}. \quad (35)$$

Substitute (26) and collect the coefficients of $t, x + y, r$ to obtain (34). Independence and (27) imply

$$\theta_P := \mathbb{E}K_P(q, q') = 18c^2 \text{tr}(PP_U). \quad (36)$$

In particular,

$$K_I(q, q') = 6A^2 \left[s^4 - \frac{6s^2}{d} + \frac{3}{(p+2)d} \right], \quad \theta_I = 18c^2 m, \quad (37)$$

$$K_P(q, q) = \frac{6p^2(p+2)^2(p+1)}{d^2} \left[(p-2)q^T P q + \frac{3r}{p+2} \right]. \quad (38)$$

This positive diagonal must be removed from an unbiased pair statistic.

Let P_j act on tensor index j . On $\mathcal{H}_4$ define

$$\mathcal{A}_P = \Pi_4 \left(\frac{1}{4} \sum_{j=1}^4 P_j \right) \Pi_4.$$

It is a positive contraction and, by symmetry, $K_P(q, q') = 6\langle H(q), \mathcal{A}_P H(q') \rangle$.

Lemma 4 (Exact trace for the direct kernel). *Put $D_4(p) = p(p-1)(p+1)(p+6)/24 = \dim \mathcal{H}_4$. Then $\text{tr} \mathcal{A}_P = D_4(p)r/p$ and*

$$\begin{aligned} t_1(p) &= \left(\frac{p+1}{p+4} \right)^2 + (p-1) \left(\frac{3(p+2)}{4(p+4)} \right)^2 \\ &\quad + \frac{(p-2)(p+1)}{2} \left(\frac{p+3}{2(p+4)} \right)^2 + \frac{(p-1)(p-2)(p+3)}{96}, \\ \tau_p(r) &:= \text{tr}(\mathcal{A}_P^2) = \frac{r(p-r)}{p-1} t_1(p) + \frac{r(r-1)}{p(p-1)} D_4(p). \end{aligned} \quad (39)$$

For independent uniform-sphere directions,

$$\mathbb{E}_0 K_P = 0, \quad \mathbb{E}_0 K_P^2 = 36\kappa_p^2 \tau_p(r), \quad \mathbb{E}_0 K_I^2 = 36\kappa_p^2 D_4(p). \quad (40)$$

Proof. For $P = e_1 e_1^T$, represent tensors by homogeneous harmonic quartics. The operator acting on such a polynomial f is

$$f \mapsto \frac{x_1 \partial_1 f}{4} - \frac{\|x\|^2 \partial_1^2 f}{4(p+4)}.$$

Indeed, $\partial_1^2 f$ is a harmonic quadratic, $\Delta(x_1 \partial_1 f/4) = \partial_1^2 f/2$, and $\Delta(\|x\|^2 \partial_1^2 f) = 2(p+4)\partial_1^2 f$. Thus the correction is a trace term and the result is harmonic.

Under rotations fixing e_1 , separate transverse harmonic degree $k = 0, \dots, 4$. For a transverse harmonic Y_k , its harmonic extension has the form

$$Y_k(x_\perp) \sum_{j=0}^{\lfloor (4-k)/2 \rfloor} a_j x_1^{4-k-2j} \|x_\perp\|^{2j}, \quad a_0 = 1,$$

with the remaining coefficients determined by harmonicity. Comparing the leading coefficient gives eigenvalue

$$\alpha_k = \frac{4-k}{4} - \frac{(4-k)(3-k)}{4(p+4)} = \frac{(4-k)(p+1+k)}{4(p+4)}.$$

For $k = 0, 1, 2, 3$, the multiplicities are 1, $p-1$, $(p-2)(p+1)/2$, and $(p-1)(p-2)(p+3)/6$; $\alpha_4 = 0$. Their squared sum gives $t_1(p)$. For general symmetric P , orthogonal equivariance makes $\text{tr}(\mathcal{A}_P^2)$ an invariant quadratic form $\alpha(\text{tr } P)^2 + \beta \text{tr}(P^2)$. The rank-one value $t_1(p)$ and identity value $D_4(p)$ determine its coefficients and give (39). Linear invariance similarly gives the trace formula. Finally, the uniform-sphere law has $\mathbb{E}H = 0$ and $\text{Cov } H = \kappa_p I$; independence proves (40).

For fixed r , $\tau_p(r) = rp^3/96 + O_r(p^2)$ and $D_4(p) \sim p^4/24$. Use the exact formula when r also grows.

D.4 Orthogonal decomposition and finite-sample inversion

Let $\mathcal{L}$ be any fixed self-adjoint operator on $\mathcal{H}_4$ with $0 \preceq \mathcal{L} \preceq I$, and set

$$K_{\mathcal{L}}(q, q') = 6\langle H(q), \mathcal{L}H(q') \rangle, \quad \theta = 6\langle \mathcal{T}, \mathcal{L}\mathcal{T} \rangle \geq 0.$$

With $X = H - \mathcal{T}$, the exact Hoeffding decomposition is

$$\begin{aligned} K_{\mathcal{L}}(q, q') - \theta &= \psi(q) + \psi(q') + \psi_2(q, q'), \\ \psi(q) &= 6\langle X(q), \mathcal{L}\mathcal{T} \rangle, \quad \psi_2(q, q') = 6\langle X(q), \mathcal{L}X(q') \rangle. \end{aligned} \tag{41}$$

The last term has conditional mean zero given either argument; the three terms are orthogonal. This standard U-statistic structure is described in [19, Section 2.1]; the feature and constants here are specific to the angular model. Writing $v_1 = \mathbb{E}\psi^2$, $v_2 = \mathbb{E}\psi_2^2$, one obtains

$$v_1 = 36\langle \mathcal{L}\mathcal{T}, \Sigma\mathcal{L}\mathcal{T} \rangle \leq 6C\kappa_p\theta, \tag{42}$$

$$v_2 = 36 \text{tr}(\mathcal{L}\Sigma\mathcal{L}\Sigma) \leq 36C^2\kappa_p^2 \text{tr}(\mathcal{L}^2). \tag{43}$$

The first bound uses $\mathcal{L}^2 \preceq \mathcal{L}$; the second follows from $\text{tr}(\mathcal{L}\Sigma\mathcal{L}\Sigma) = \|\Sigma^{1/2}\mathcal{L}\Sigma^{1/2}\|_F^2 \leq \|\Sigma\|_{\text{op}}^2 \|\mathcal{L}\|_F^2$. Commutation is not assumed. Also $\mathbb{E}K_{\mathcal{L}}^2 = \theta^2 + 2v_1 + v_2$.

For n directions the full U-statistic satisfies

$$U_n = \binom{n}{2}^{-1} \sum_{i < j} K_{\mathcal{L}}(q_i, q_j), \quad \text{Var } U_n = \frac{4v_1}{n} + \frac{2v_2}{n(n-1)}.$$

Alternatively split $n = Nb$ directions into N predetermined disjoint blocks of length $b \geq 2$ and average their full U-statistics. Then

$$\begin{aligned} \text{Var } \bar{U}_b &= \frac{4v_1}{n} + \frac{2v_2}{n(b-1)} \leq A_n\theta + B_{n,b}, \\ A_n &= \frac{24C\kappa_p}{n}, \quad B_{n,b} = \frac{72C^2\kappa_p^2\tau}{n(b-1)}, \end{aligned} \tag{44}$$

where $\tau \geq \text{tr}(\mathcal{L}^2)$ is any public bound. The full statistic is the case $b = n$, with $B_n = B_{n,n}$; these are the variance coefficients used in Theorem 5. Throughout, $h_\delta = (1 - \delta)/\delta$ denotes the confidence ratio, whereas k denotes the number of compressed tensor indices.

Proposition 3 (Computable one-sided bound). *Let $\delta \in (0, 1)$, $h_\delta = (1 - \delta)/\delta$, and observe $u = \bar{U}_b$. Put $A = A_n$, $B = B_{n,b}$ and*

$$\ell_\delta(u) = \begin{cases} 0, & u \leq \sqrt{h_\delta B}, \\ u + \frac{h_\delta A}{2} - \sqrt{h_\delta A u + \frac{h_\delta^2 A^2}{4} + h_\delta B}, & u > \sqrt{h_\delta B}. \end{cases} \quad (45)$$

Then $\Pr\{\theta \geq \ell_\delta(\bar{U}_b)\} \geq 1 - \delta$.

Proof. Cantelli's inequality and (44) give

$$\Pr\{\bar{U}_b - \theta \geq t\} \leq \frac{A\theta + B}{A\theta + B + t^2}.$$

The function $g(\theta) = \theta + \sqrt{h_\delta(A\theta + B)}$ is increasing on $[0, \infty)$. Above $g(0)$ its inverse is the positive branch of (45); otherwise use $\theta \geq 0$. The event $\bar{U}_b > g(\theta)$ has probability at most δ . Zero-variance limiting cases follow directly.

For its positive branch, the numerically preferable rationalized form is

$$\ell_\delta(u) = \frac{u^2 - h_\delta B}{u + h_\delta A/2 + \sqrt{h_\delta A u + h_\delta^2 A^2/4 + h_\delta B}}. \quad (46)$$

It removes the subtraction of the square-root term in the original formula, but its numerator can still suffer cancellation near $u^2 = h_\delta B$. It is not a directed numerical enclosure or a uniform numerical-stability guarantee. A valid upper bound for θ is

$$u_+ + \frac{h_\delta A}{2} + \sqrt{h_\delta A u_+ + \frac{h_\delta^2 A^2}{4} + h_\delta B}, \quad u_+ = \max(u, 0). \quad (47)$$

If θ exceeds this quantity, then $\theta - u > \sqrt{h_\delta(A\theta + B)}$; apply the other one-sided Cantelli inequality.

For the direct kernel, take $\tau = \tau_p(r)$. Known $c \neq 0$ permits division of $\ell_\delta(u)$ by $18c^2 m$. A simpler, usually wider, bound uses $\theta_P \leq 18c^2 \min(r, m)$ and

$$V_{n,b} = \frac{432C\kappa_p c^2 \min(r, m)}{n} + \frac{72C^2 \kappa_p^2 \tau_p(r)}{n(b-1)}. \quad (48)$$

It is $[u - \sqrt{h_\delta V_{n,b}}]_+ / (18c^2 m)$. For unknown c , set $C = 4$, and assign failure probability $\delta/2$ to a lower bound on θ_P and an upper bound on θ_I . Their ratio bounds $o(P) = \theta_P/\theta_I$ for $c \neq 0$, jointly with probability at least $1 - \delta$; the statistics need not be independent. Return zero if the computed denominator is zero. At $c = 0$ the numerator lower bound is zero outside its failure event, so this convention does not assert a spurious positive overlap.

Shared observations and incomplete statistics. For predetermined edge weights with $\sum_{i < j} w_{ij} = 1$, let $d_i = \sum_{j \neq i} w_{ij}$. The decomposition gives exactly

$$\text{Var} \left(\sum_{i < j} w_{ij} K_{\mathcal{L}}(q_i, q_j) \right) = v_1 \sum_i d_i^2 + v_2 \sum_{i < j} w_{ij}^2. \quad (49)$$

Distinct canonical terms have zero covariance even when they share one observation, by their conditional mean-zero property. Thus $\binom{n}{2}$ pairs are not $\binom{n}{2}$ independent observations. For $n/2$ disjoint equally weighted pairs, the degenerate term is $2v_2/n$, instead of $2v_2/[n(n-1)]$ for full U or $2v_2/[n(b-1)]$ for blocks. If M pairs are sampled independently with replacement from all distinct pairs after observing the data, their average has variance

$$\text{Var} U_{\text{inc}} = \text{Var} U_n + \frac{2v_1 + v_2 - \text{Var} U_n}{M}. \quad (50)$$

Conditionally its mean is U_n . The expected conditional variance of one sampled pair is $2v_1 + v_2 - \text{Var} U_n$, proving the formula. Increasing M cannot create additional independent directions.

D.5 Four-index compression and overlap

Set

$$H_W = (B^T)^{\otimes 4} H, \quad \mathcal{J}_P = \Pi_4 P^{\otimes 4} \Pi_4, \quad \theta_W = 6 \|P^{\otimes 4} \mathcal{T}\|_F^2.$$

This positive contraction factors through symmetric fourth tensors on $\mathbb{R}^r$, so

$$\text{tr}(\mathcal{J}_P^2) \leq \text{rank } \mathcal{J}_P \leq \binom{r+3}{4}. \quad (51)$$

In general $\mathcal{J}_P$ is not a projection. The correct relations are

$$\|H_W\|_F^2 = \langle H, \mathcal{J}_P H \rangle, \quad \|\mathcal{J}_P H\|_F^2 = \langle H, \mathcal{J}_P^2 H \rangle \leq \langle H, \mathcal{J}_P H \rangle.$$

These two norms must not be identified.

Proposition 4 (Retained fourth energy). *For $c \neq 0$, let $\rho_4 = \theta_W/(18c^2m)$. Then*

$$[4o(P) - 3]_+ \leq \rho_4 \leq o(P). \quad (52)$$

If $Q^T P Q = oI_m$, then $\rho_4 = o^4$.

Proof. The four P_j commute and are orthogonal projections, so

$$P^{\otimes 4} \preceq P_j, \quad I - P^{\otimes 4} \preceq \sum_{j=1}^4 (I - P_j).$$

By (27), $\langle \mathcal{T}, P_j \mathcal{T} \rangle = 3c^2 \text{tr}(P P_U) = o(P) \|\mathcal{T}\|_F^2$. Taking quadratic forms and dividing by $\|\mathcal{T}\|_F^2$ proves (52). For the special case, write $\mathcal{T}_U$ in carrier coordinates:

$$\|P^{\otimes 4} \mathcal{T}\|_F^2 = \langle \mathcal{T}_U, (Q^T P Q)^{\otimes 4} \mathcal{T}_U \rangle = o^4 \|\mathcal{T}\|_F^2.$$

Use Proposition 3 with $\mathcal{L} = \mathcal{J}_P$ and the exact $\tau = \tau_{p,r}^{(4)}$ from Theorem 6. The simpler bound $\tau = \binom{r+3}{4}$ remains valid and defines the dimension-bound baseline; the exact-trace comparison changes only that deterministic constant. For known $c \neq 0$, $\ell_\delta(U_W)/(18c^2m)$ is an overlap lower bound. The statistic estimates retained fourth energy, not overlap itself. If only $c^2 \leq 1$ is known, use $C = 4$ and the public normalization $\ell_\delta(U_W)/(18m)$. Its population limit is $c^2\rho_4$; unknown scale causes a real loss. All overlap lower bounds may be clipped to $[0, 1]$.

D.6 Closed compressed kernel and streaming sufficient statistics

Use only candidate coordinates $z = B^T q$, and define $t = z^T z'$, $x = \|z\|^2$, $y = \|z'\|^2$, $\eta_1 = 1/(p+4)$, and $\eta_2 = 1/[(p+2)(p+4)]$. The compressed kernel is

$$\begin{aligned} k_W(q, q') &= 6\langle H_W(q), H_W(q') \rangle \\ &= 6A^2 \{ t^4 - 6\eta_1(x+y)t^2 + 3\eta_2(x^2 + y^2) \\ &\quad + 6\eta_1^2[(r+4)t^2 + xy] - 6\eta_1\eta_2(r+2)(x+y) + 3\eta_2^2 r(r+2) \}. \end{aligned} \quad (53)$$

Indeed, set $S_6(z) = \text{Sym}_6(z^{\otimes 2} \otimes I_r)$ and $S_3 = \text{Sym}_3(I_r \otimes I_r)$. The required contractions are

$$\begin{aligned} \langle z^{\otimes 4}, S_6(z') \rangle &= 6xt^2, & \langle z^{\otimes 4}, S_3 \rangle &= 3x^2, \\ \langle S_6(z), S_6(z') \rangle &= 6[(r+4)t^2 + xy], & \langle S_6(z), S_3 \rangle &= 6(r+2)x, \\ \langle S_3, S_3 \rangle &= 3r(r+2). \end{aligned}$$

Substitute $H_W = A[z^{\otimes 4} - \eta_1 S_6(z) + \eta_2 S_3]$. The trace coefficients retain ambient dimension p . Using an r -dimensional trace-free projection instead changes the mean and invalidates the certificate. When $r = p$ and $x = y = 1$, the result is (37).

Define the sufficient accumulations

$$S = \sum_i H_W(q_i), \quad D_{\text{diag}} = \sum_i \|H_W(q_i)\|_F^2, \quad U_W = \frac{6(\|S\|_F^2 - D_{\text{diag}})}{n(n-1)}. \quad (54)$$

Expanding the square proves that this is the exact full U-statistic, including all pairs from different computational batches. Batching this accumulation does not replace it by the statistical block design in (44). Specifically, accumulate

$$\begin{aligned} M_2 &= \sum_i z_i z_i^T, & M_4 &= \sum_i z_i^{\otimes 4}, \\ S &= A[M_4 - \eta_1 \text{Sym}_6(M_2 \otimes I_r) + n\eta_2 S_3]. \end{aligned} \quad (55)$$

For $w = \|z\|^2$, the diagonal is

$$\begin{aligned} \frac{\|H_W(q)\|_F^2}{A^2} &= w^4 - \frac{12w^3}{d} + \left[\frac{6}{(p+2)d} + \frac{6(r+5)}{d^2} \right] w^2 \\ &\quad - \frac{12(r+2)}{(p+2)d^2} w + \frac{3r(r+2)}{(p+2)^2 d^2}. \end{aligned} \quad (56)$$

Hence its sum requires only scalar accumulations.

For matrix multiplication, form the packed symmetric quadratic feature

$$y_{jk}(z) = \sqrt{2 - \mathbf{1}\{j = k\}} z_j z_k, \quad 1 \leq j \leq k \leq r,$$

and accumulate $\sum_i y(z_i) y(z_i)^T$. An entry of M_4 is obtained by sorting the indices within each of its two pairs, taking the corresponding Gram entry, and dividing by its two known square-root weights. At $r = 16$, the quadratic width is 136, its Gram matrix is 136 by 136, and a dense r^4 float64 tensor occupies 0.5 MiB. A packed symmetric fourth tensor has $\binom{r+3}{4} = 3876$ entries.

The matrix-multiplication implementation costs $O(npr + nr^4)$ operations. With computational batch size b_{batch} its workspace is

$$O(pr + r^4 + b_{\text{batch}}(p + r^2) + N_{\text{batches}}).$$

This includes the stored frame, batch quadratic features, and the executed implementation's list of one diagonal partial sum per batch. The last list is combined by an accurate scalar summation routine. A constant-space online diagonal accumulation is algebraically possible, but was not substituted into the frozen implementation used for the reported results. No full collection of directions or Gaussian probes needs to be stored.

The public API receives the candidate frame, public directions, dimensions, a confidence budget, and the stated scale information. It receives no producer key, secret pairing, payload, true carrier frame, true overlap, or external score. Synthetic generation and external scoring may use those quantities separately. This is separation of data interfaces, not a claim of operating-system isolation. For full U, the two confidence-error terms have orders

$$\sqrt{\frac{C\kappa_p\theta_W}{n\delta}} \quad \text{and} \quad \frac{C\kappa_p\sqrt{\binom{r+3}{4}}}{n\sqrt{\delta}}.$$

At fixed rank the variance has no ambient p^3 factor. For $r = m$, the normalized degenerate term has order $m/(nc^2\sqrt{\delta})$ at known scale. Projection still costs $O(npr)$, high ranks incur the r^4 accumulation cost, and $\rho_4 \leq o(P)$ remains a loss of the estimand.

D.7 Cost boundary of direct overlap validation

Direct pairwise evaluation of (34) costs $O(npr + n^2(p+r))$ for full U or $O(npr + nb(p+r))$ for equal statistical blocks. Gram tiles control the pairwise memory. These are implementation costs, not lower bounds: a full ambient fourth-order feature representation exists but has dimension $O(p^4)$. Four-index compression changes the estimand to obtain smaller sufficient statistics.

Table 15 records a variance-design calculation at $p = 240$, $m = r = 16$, $a = 1/4$, $n = 2^{20}$, and total failure probability 0.05. Then $c = 2/\pi$, $C = 2.6785242793$, and $\tau_{240}(16) = 2860630.4918$. The hypothetical observed numerator is set to

Table 15. Direct-kernel design at hypothetical overlap 0.9. The known-scale margin is in overlap units; values are rounded.

Block length b	Number of pairs	Known- c margin	Known- c lower bound	Unknown- c lower bound
512	267,911,168	1.44027	0	0
1,024	536,346,624	1.01798	0	0
16,384	8,589,410,304	0.25481	0.64519	0.07361
1,048,576	549,755,289,600	0.03525	0.86475	0.55938

its mean at overlap 0.9; the denominator, where used, is also set to its mean. Known-scale columns use (48). The unknown-scale column uses $C = 4$, the mean-dependent lower and upper bounds, and failure probability 0.025 for each. These are design values, not validation outcomes or empirical coverage rates. None of these large direct pairwise sums was executed.

A favorable full-U degenerate variance therefore does not make short statistical blocks informative at these parameters. At unknown scale, the identity-kernel denominator also has a null second moment of order p^4 . The executed validation instead uses the exact low-dimensional full statistic (54), avoiding a large direct double sum.

D.8 The power cost of removing geometric compression

The $k = 1$ kernel directly estimates $18c^2mo$, but its larger variance trace can prevent informative certification at the current sample count. The following calculation uses the theoretical variance bound; it does not require, or reconstruct, empirical one-index statistics from the archived four-index panel.

Corollary 2 (A one-index positive-certificate power bound). *Set $p = 240$, $r = m = 16$, $a = 1/4$, $c = 2/\pi$, $n = 65,536$ and $\delta = 0.05/16$. For any candidate fixed independently of validation, the known-scale one-index certificate of Proposition 7 satisfies*

$$\Pr\{\ell(U_1) > 0\} \leq 0.011612247954.$$

Proof. Write $h_\delta = (1 - \delta)/\delta = 319$ to distinguish the confidence ratio from the index count. The exact trace is $\tau_{240,16}^{(1)} = 174498460/61$. With A_n, B_n as in Theorem 5, using this trace, set $t = \sqrt{h_\delta B_n}$ and $\theta_{\max} = 18c^2m$. Then $t > 243.2900555$ and $\theta_{\max} < 116.7220036 < t$. For every candidate, $0 \leq \theta_1 \leq \theta_{\max}$ and $\text{Var}(U_1) \leq V_{\max} := A_n\theta_{\max} + B_n$. Cantelli and monotonicity in the variance and the positive mean gap give

$$\Pr\{U_1 > t\} \leq \frac{V_{\max}}{V_{\max} + (t - \theta_{\max})^2}.$$

The event on the left is exactly $\{\ell(U_1) > 0\}$. For the displayed decimal bound we enclose π by the alternating rational sums in $\pi =$

$16 \arctan(1/5) - 4 \arctan(1/239)$, bracket square roots by integer arithmetic, and round the resulting rational upper bound upward. The reproducible calculation is certificate-power. This enclosure concerns the scalar theoretical constants, not upstream float64 observations.

Substituting the largest population mean into the inversion gives zero, but that weaker observation alone would not imply zero empirical bounds. No eight-candidate $k = 1$ measurement was executed. This result also differs from Proposition 19, which bounds the alternative power of a specified four-index test at null level 10^{-6} by less than 0.06%. That latter number is a power ceiling, not a false-positive rate or a single-image decoder guarantee.

D.9 Conditional planning for a positive one-index certificate

This diagnoses the specified Cantelli certificate under the exact iid model and exact arithmetic; it adds no validation run, convergence theorem or rate.

Corollary 3 (A conditional positivity budget). *Fix a rank- r candidate and $n \geq 2$ before independent validation. Suppose $c > 0$ is known and its true overlap satisfies $o \geq o_0 > 0$. Use A_n, B_n from (71) with $\tau = \tau_{p,r}^{(1)}$. Put $h_\delta = (1 - \delta)/\delta$, $t_n = \sqrt{h_\delta B_n}$ and $\theta_0 = 18c^2 m o_0$. For the inversion in Proposition 7,*

$$\Pr\{\ell_\delta(U_1) > 0\} \geq G_n(o_0) := \begin{cases} \frac{(\theta_0 - t_n)^2}{A_n \theta_0 + B_n + (\theta_0 - t_n)^2}, & \theta_0 > t_n, \\ 0, & \theta_0 \leq t_n. \end{cases} \quad (57)$$

Proof. Here $\theta = 18c^2 m o \geq \theta_0$ and $\ell_\delta(U_1) > 0$ iff $U_1 > t_n$. For $\theta_0 > t_n$, lower-tail Cantelli and Proposition 6 give the ratio with θ in place of θ_0 . It increases in $\theta > t$, since

$$\frac{d}{d\theta} \frac{(\theta - t)^2}{A\theta + B + (\theta - t)^2} = \frac{(\theta - t)[A(\theta + t) + 2B]}{[A\theta + B + (\theta - t)^2]^2} > 0 \quad (\theta > t).$$

Thus substituting θ_0 is conservative; otherwise zero is a trivial lower bound, not a zero-power conclusion.

Artifact certificate-planning verifies the table using rational π enclosures and integer square roots. Neither panel means nor the random-frame expectation establishes o_0 . No entry measures empirical power or the cost of computing the complete one-index statistic.

Let E mean issuance by a fixed rule using only earlier information. If an earlier certificate gives $\Pr\{o < o_0\} \leq \delta_{\text{old}}$, choose n on E so $G_n(o_0) \geq 1 - \eta$ before independent new validation. Then $\Pr\{E \cap (\{o < o_0\} \cup \{\ell_\delta(U_1) \leq 0\})\} \leq \delta_{\text{old}} + \eta$ by conditioning on the earlier information and a union bound. This is unconditional, not a guarantee conditional on issuance. Requiring new numerical coverage as well adds its separate δ ; observing the new data cannot select or extend n .

Table 16. Conditional budgets at $p = 240$, $r = m = 16$, $c = 2/\pi$, $\delta = 1/320$, $\tau_{240,16}^{(1)} = 174498460/61$. Threshold intervals round outward and probability lower bounds down.

n	Enclosure of t_n	$G_n(1/15)$ lower	$G_n(1/2)$ lower
2^{16}	[243.290055, 243.290056]	0.000000	0.000000
2^{17}	[121.644563, 121.644564]	0.000000	0.000000
2^{18}	[60.822165, 60.822166]	0.000000	0.000000
2^{19}	[30.411053, 30.411054]	0.000000	0.996091
2^{20}	[15.205519, 15.205520]	0.000000	0.999566
2^{21}	[7.602758, 7.602759]	0.146048	0.999913
2^{22}	[3.801378, 3.801379]	0.996974	0.999977

D.10 Coverage, multiple outputs, and numerical scope

For a predetermined finite collection of frames and scale assumptions, assign failure probabilities δ_j before validation. A union bound gives joint coverage at least $1 - \sum_j \delta_j$, without independence between claims. For example, eight previously trained frames, each validated with known-scale and unknown-scale bounds at $\delta_j = 0.05/16$, have joint coverage at least 95% under the stated model. These archived validations used fresh data for eight existing outputs, not eight additional training trials. The exact-trace evaluation draws no additional observations. The frames, sample counts, synthetic benchmark seeds, confidence allocation, and code hashes can be frozen before validation. External overlap or retained-energy scores do not enter the public API.

The finite-sample statements assume exact iid observations and exact arithmetic. The implementation uses ordinary float64 and has been compared with independent dense tensor contractions and direct small-dimensional pair sums. These algebraic cross-checks do not supply a directed-rounding enclosure of the final real-valued statistic. The mathematical coverage guarantee and the numerical limitations of the implementation are therefore distinct. No normal approximation, estimated standard error, or independent-pair assumption is used in the proof.

This result certifies an output using new data. It does not prove convergence of training that reuses a sample or make its intermediate empirical updates unbiased. Changing frames, thresholds, sample counts, or confidence allocations after inspecting validation requires new independent data, a prespecified simultaneous guarantee, or separately justified sequential inference.

D.11 Partial-index certificates and their exact harmonic spectrum

The direct overlap kernel and the four-index compressed kernel of Appendix D are endpoints of a common family. This appendix constructs the intermediate two- and three-index kernels, computes the spectrum of all four operators, and supplies their public confidence bounds. Throughout the statistical statements, the key is fixed, $m = 4b_0 < p$, $0 < a < 1/2$, and $c = \sin(2\pi a)/(2\pi a) > 0$. Let Q

be the ordered carrier frame, $P_U = QQ^T$, and let $P = BB^T$ be an orthogonal candidate projector of rank r , frozen before the iid validation sample. Write $o = \text{tr}(PP_U)/m$. The operator and spectral identities below do not depend on a . The endpoint results in Appendix D retain their stated parameter ranges.

We use its trace-free feature $H(q) = p(p+2)\Pi_4(q^{\otimes 4})$ on $\mathcal{H}_4$, and its proved identities

$$\mathbb{E}H = \mathcal{T}, \quad \mathcal{T}_{(1)}\mathcal{T}_{(1)}^T = 3c^2P_U, \quad \|\mathcal{T}\|_F^2 = 3c^2m, \quad \Sigma := \text{Cov}(H) \preceq C\kappa_p I, \quad (58)$$

where $C = (1+c)^2$ and $\kappa_p = 24p(p+2)/[(p+4)(p+6)]$. The flattening identity is (27); the full covariance proof, including sharpness, is Proposition 2.

Construction and the geometry of retained energy Let P_j apply P to tensor index j . For $1 \leq k \leq 4$, define

$$A_k = \left(\binom{4}{k}\right)^{-1} \sum_{\substack{S \subseteq \{1,2,3,4\} \\ |S|=k}} \prod_{j \in S} P_j, \quad L_k = \Pi_4 A_k \Pi_4|_{\mathcal{H}_4}. \quad (59)$$

For independent directions q, q' , put

$$K_k(q, q') = 6\langle H(q), L_k H(q') \rangle, \quad U_k = \binom{n}{2}^{-1} \sum_{i < j} K_k(q_i, q_j), \quad n \geq 2. \quad (60)$$

Theorem 12 (Partial-index energy and general overlap bounds). *Each L_k is a positive contraction. With*

$$\rho_k = \frac{\langle \mathcal{T}, A_k \mathcal{T} \rangle}{\|\mathcal{T}\|_F^2}, \quad \theta_k = \mathbb{E}U_k = 18c^2 m \rho_k, \quad (61)$$

one has

$$[ko - (k-1)]_+ \leq \rho_k \leq o, \quad \rho_1 = o. \quad (62)$$

The condition $Q^T P Q = oI_m$ is sufficient for $\rho_k = o^k$. For $o > 0$ this condition itself implies $r \geq m$. In particular, it gives a two-index kernel with mean $18c^2 mo^2$.

Proof. The P_j commute and are orthogonal projections, so every product in (59) is a positive contraction. Averaging and orthogonal compression preserve this property. Independence and $\Pi_4 \mathcal{T} = \mathcal{T}$ give (61). For $i \in S$,

$$\prod_{j \in S} P_j \preceq P_i, \quad I - \prod_{j \in S} P_j \preceq \sum_{j \in S} (I - P_j).$$

By symmetry of $\mathcal{T}$ and (58), each index satisfies

$$\langle \mathcal{T}, P_j \mathcal{T} \rangle = 3c^2 \text{tr}(PP_U) = o\|\mathcal{T}\|_F^2.$$

Taking quadratic forms in the preceding inequalities proves the upper bound and the lower bound $ko - (k - 1)$; positivity supplies the positive part. For $k = 1$ the average is exactly o . For the final assertion, write $\mathcal{T} = Q^{\otimes 4} \mathcal{T}_U$. In the quadratic form of each product, every selected index then contributes $Q^T P Q = o I_m$, and every other index contributes I_m . The resulting value is $o^k \|\mathcal{T}\|_F^2$.

The sufficient condition involves the true carrier geometry. A mean overlap, or the fact that a candidate was drawn at random, does not establish it for a realized candidate. No necessity assertion is made. For $k = 1$, the construction is exactly $\mathcal{A}_P$ and K_P in Appendix D.3, including its direct mean and closed pair formula (34). For $k = 4$, it is exactly $\mathcal{J}_P$ and the compressed kernel of Appendix D.5. The intermediate choices change the retained energy and its variance together.

One explicit feature representation is

$$F_k(q) = ((B^T)^{\otimes k} \otimes I_p^{\otimes (4-k)}) H(q).$$

Symmetry makes every choice of k positions have the same inner product, hence $K_k = 6 \langle F_k(q), F_k(q') \rangle$ and

$$U_k = \frac{6}{n(n-1)} \left(\left\| \sum_i F_k(q_i) \right\|_F^2 - \sum_i \|F_k(q_i)\|_F^2 \right). \quad (63)$$

This follows by expanding the square and removing its diagonal. It includes all pairs, including pairs in different computational batches. The single-observation feature is quartic; a quadratic feature with zero mean does not have this kernel mean.

Proposition 5 (A dimension bound before harmonic compression). *For $1 \leq r < p$, set*

$$d_j = \binom{r+j-1}{j} \binom{p-r+3-j}{4-j}, \quad D_k = \sum_{j=k}^4 d_j, \quad \bar{\tau}_k = \sum_{j=k}^4 \left(\frac{\binom{j}{k}}{\binom{4}{k}} \right)^2 d_j.$$

Then

$$\text{rank}(L_k) \leq D_k, \quad \text{tr}(L_k^2) \leq \bar{\tau}_k \leq D_k. \quad (64)$$

Proof. Decompose $\text{Sym}^4(\mathbb{R}^p)$ according to the number j of indices in $\text{range}(P)$. The j th summand has dimension d_j and A_k acts on it by $\binom{j}{k} / \binom{4}{k}$, with value zero when $j < k$. This proves the rank and squared Frobenius norm formulas for A_k . Multiplication on either side by an orthogonal projection cannot increase rank or Frobenius norm, proving (64).

The feature $A_k^{1/2} H$ can therefore be represented with D_k symmetric coefficients, giving an $O(D_k)$ storage upper bound for the accumulated feature. Direct storage of F_k instead uses $O(r^k p^{4-k})$ coefficients. These are representation bounds, not memory or running-time lower bounds for all implementations. The

direct $k = 1$ pair formula requires only candidate projections and inner products, and does not require the D_1 coefficient array. For the symmetric representation, r Householder reflections can implicitly extend B to orthogonal coordinates, with $O(pr^2)$ preprocessing and $O(npr)$ operations to transform n observations. Explicit quartic coefficient accumulation has an additional dimension cost. These counts describe a possible representation, not a measured algorithmic improvement. A pairwise kernel implementation can exchange that storage for pairwise computation. Thus an improved mean under special geometry alone does not establish an improved confidence bound at a fixed computational budget.

Exact spectrum under orthogonal splitting For $t \geq 2$ and $d \geq 0$, let

$$h_d(t) = \binom{t+d-1}{d} - \binom{t+d-3}{d-2}$$

be the dimension of harmonic homogeneous polynomials of degree d in t variables; a binomial coefficient with negative lower index or with upper index smaller than the nonnegative lower index is zero. In particular, $h_4(p) = p(p-1)(p+1)(p+6)/24 = \dim \mathcal{H}_4$.

Theorem 13 (Complete partial-index spectrum). *Suppose first that $\mathbb{R}^p = \mathbb{R}^r \oplus \mathbb{R}^s$, $s = p-r$, with $r, s \geq 2$, and P projects onto the first summand. For each nonnegative triple (d, e, j) with $d+e+2j = 4$, put $A = d+r/2$ and $B = e+s/2$. The operator L_k has an eigenvalue*

$$\lambda_{d,e,j}^{(k)} = \sum_{i=0}^j w_i \frac{\binom{d+2i}{k}}{\binom{4}{k}} \quad \text{with multiplicity } h_d(r)h_e(s), \quad (65)$$

where the weights sum to one and are

$$\begin{array}{c|c} j & (w_0, \dots, w_j) \\ \hline 0 & (1) \\ 1 & \left(\frac{A}{A+B}, \frac{B}{A+B} \right) \\ 2 & \frac{(A(A+1), 2(A+1)(B+1), B(B+1))}{(A+B+1)(A+B+2)}. \end{array} \quad (66)$$

The last case has $d = e = 0$. These nine branches, counting coincident eigenvalues with their multiplicities, exhaust the spectrum. Consequently the exact public squared trace is

$$\tau_{p,r}^{(k)} := \text{tr}(L_k^2) = \sum_{d+e+2j=4} h_d(r)h_e(s) (\lambda_{d,e,j}^{(k)})^2. \quad (67)$$

For $r = 1$ or $s = 1$, use $h_0(1) = h_1(1) = 1$ and $h_d(1) = 0$ for $d \geq 2$, deleting zero-multiplicity branches; the same formulas apply. For $r = 0$, $L_k = 0$ and $\tau_{p,0}^{(k)} = 0$. For $r = p$, $L_k = I_{\mathcal{H}_4}$ and $\tau_{p,p}^{(k)} = h_4(p)$.

Proof. Associate a symmetric tensor T with $f_T(x, y) = \langle T, (x, y)^{\otimes 4} \rangle$. Trace-freeness is equivalent to $\Delta f_T = 0$. For fixed harmonic homogeneous polynomials $H_d(x)$ and $G_e(y)$, consider the radial sector

$$f(x, y) = H_d(x)G_e(y) \sum_{i=0}^j a_i \|x\|^{2i} \|y\|^{2(j-i)}. \quad (68)$$

The radial Laplacian identity is

$$\Delta_x (\|x\|^{2i} H_d(x)) = 4i(d + r/2 + i - 1) \|x\|^{2i-2} H_d(x).$$

Combining it with the analogous identity in y , and equating coefficients of the radial terms in Δf , gives

$$a_0 = 1, \quad \frac{a_{i+1}}{a_i} = -\frac{(j-i)(B+j-i-1)}{(i+1)(A+i)} \quad (0 \leq i < j). \quad (69)$$

All denominators are positive, so each fixed $H_d G_e$ sector has exactly one harmonic radial combination.

For completeness, decompose the homogeneous polynomials in x and y separately into harmonic polynomials times powers of their squared radii. This gives all sectors in (68); within each, the displayed recurrence is exactly the condition for the total polynomial to be harmonic. The resulting dimension is $\sum_{d+e+2j=4} h_d(r)h_e(s)$. Indeed, the generating identity

$$\sum_{d \geq 0} h_d(t) z^d = \frac{1 - z^2}{(1 - z)^t}$$

shows that multiplying the two harmonic series and then $1/(1 - z^2)$ gives $(1 - z^2)/(1 - z)^p$, whose degree-four coefficient is $h_4(p)$. This also verifies completeness.

Use the Fischer inner product $\langle f, g \rangle_{\text{Fis}} = f(\partial)g|_0$. On polynomials of total degree four, $\langle f_T, f_{T'} \rangle_{\text{Fis}} = 4! \langle T, T' \rangle_F$, so it defines the same orthogonal harmonic projection. Distinct i terms in (68) have different x degrees and are orthogonal. Multiplication by $\|x\|^2$ is adjoint to Δ_x . Iterating the radial identity therefore gives

$$\|\|x\|^{2i} H_d\|_{\text{Fis}}^2 = 4^i i! (A)_i \|H_d\|_{\text{Fis}}^2,$$

where $(A)_i = A(A+1) \cdots (A+i-1)$ and $(A)_0 = 1$. After removing the common factor $\|H_d\|_{\text{Fis}}^2 \|G_e\|_{\text{Fis}}^2$, the squared norm of the i th term is

$$v_i = a_i^2 4^j i! (A)_i (j-i)! (B)_{j-i}. \quad (70)$$

The normalized quantities $v_i / \sum_{\ell} v_{\ell}$ are the weights. Substitution of (69) gives exactly (66) for $j = 0, 1, 2$.

The i th radial term has $d + 2i$ indices in $\mathbb{R}^r$. Thus A_k multiplies it by $\binom{d+2i}{k} / \binom{4}{k}$. Both A_k and harmonic projection preserve the fixed $H_d G_e$ radial

Table 17. Exact rational squared traces from (67). The decimal column is rounded for display. The $k = 1$ values are the pre-existing direct-kernel trace (39).

(p, r)	k	$\tau_{p,r}^{(k)}$	Decimal value
(240, 16)	1	174498460/61	2860630.491803
	2	7874095086092/54479161	144534.073975
	3	830626374708/54479161	15246.680739
	4	209990519484/54479161	3854.510892
(2048, 16)	1	251640555868/171	1471582198.058480
	2	84181250910586988/10240441875	8220470.555680
	3	1100538791492648/10240441875	107469.853833
	4	39664780720148/10240441875	3873.346600

sector: the former preserves its bidegrees, and the latter is obtained by subtracting radial multiples of Laplacians. Its harmonic part is the one-dimensional radial combination already found. The compressed operator is therefore scalar on that combination, with scalar equal to its Rayleigh quotient. The orthogonal norm weights in (70) give (65). Choosing bases for H_d and G_e gives the stated multiplicity and (67). The same radial identities are valid in one variable; its only nonzero harmonic degrees are zero and one. The zero- and full-rank endpoints follow directly from (59).

For $k = 1$, Theorem 13 recovers $\tau_{p,r}^{(1)} = \tau_p(r)$ of Lemma 4; that exact trace and the direct mean already belong to the baseline construction. For $p = 240, r = 16$, replacing $\bar{\tau}_k$ by the exact trace reduces the $k = 2, 3, 4$ constants by approximately 0.16435%, 0.34849%, and 0.55441%, respectively. These reductions concern the degenerate variance constant. They change neither the feature dimension nor the order of a stated computational cost.

Variance and public confidence inversion

Proposition 6 (Complete variance for a frozen positive kernel). *Let L be any fixed self-adjoint operator on $\mathcal{H}_4$ with $0 \preceq L \preceq I$, let $K(q, q') = 6\langle H(q), LH(q') \rangle$, and let U be its complete order-two U -statistic. Put*

$$\theta = 6\langle \mathcal{T}, L\mathcal{T} \rangle, \quad v_1 = 36\langle L\mathcal{T}, \Sigma L\mathcal{T} \rangle, \quad v_2 = 36 \operatorname{tr}(L\Sigma L\Sigma).$$

For every public $\tau \geq \operatorname{tr}(L^2)$,

$$\begin{aligned} \operatorname{Var}(U) &= \frac{4v_1}{n} + \frac{2v_2}{n(n-1)} \leq A_n\theta + B_n, \\ A_n &= \frac{24C\kappa_p}{n}, \quad B_n = \frac{72C^2\kappa_p^2\tau}{n(n-1)}. \end{aligned} \tag{71}$$

In particular one may use $\tau = \tau_{p,r}^{(k)}$ for $L = L_k$.

Proof. This is the general inference calculation of Appendix D.4, recalled to apply it to the full partial-index family. With $X = H - \mathcal{T}$,

$$\begin{aligned} K - \theta &= \psi(q) + \psi(q') + \psi_2(q, q'), \\ \psi(q) &= 6\langle X(q), L\mathcal{T} \rangle, \quad \psi_2(q, q') = 6\langle X(q), LX(q') \rangle. \end{aligned}$$

The last term has conditional mean zero given either argument. Consequently $U - \theta = (2/n) \sum_i \psi(q_i) + 2 \sum_{i < j} \psi_2(q_i, q_j) / [n(n-1)]$ is an orthogonal decomposition. Distinct canonical pair terms are uncorrelated even when they share one observation, by conditioning on that observation. Counting terms gives the variance equality. The covariance bound (58) and $L^2 \preceq L$ give $v_1 \leq 6C\kappa_p\theta$. Without any commutation assumption,

$$v_2 = 36\|\Sigma^{1/2}L\Sigma^{1/2}\|_F^2 \leq 36\|\Sigma\|_{\text{op}}^2\|L\|_F^2 \leq 36C^2\kappa_p^2\tau.$$

This proves the inequality.

Proposition 7 (Public overlap bounds for every partial index). *For $\delta \in (0, 1)$, write $h_\delta = (1 - \delta)/\delta$, and define*

$$\ell_\delta(u) = \begin{cases} 0, & u \leq \sqrt{h_\delta B_n}, \\ u + h_\delta A_n/2 - \sqrt{h_\delta A_n u + h_\delta^2 A_n^2/4 + h_\delta B_n}, & u > \sqrt{h_\delta B_n}. \end{cases} \quad (72)$$

Then $\Pr\{\theta \geq \ell_\delta(U)\} \geq 1 - \delta$. For U_k , a valid overlap lower bound at known $c > 0$ is $\ell_\delta(U_k)/(18c^2m)$. If only $c^2 \leq 1$ is known, use $C = 4$ in (71) and $\ell_\delta(U_k)/(18m)$. Each bound can be clipped to $[0, 1]$.

Proof. Cantelli's inequality gives, with probability at least $1 - \delta$,

$$U \leq \theta + \sqrt{h_\delta(A_n\theta + B_n)}.$$

The right side increases on $\theta \geq 0$, and its inverse above its value at zero is the positive branch of (72). Below that value, $\theta \geq 0$ suffices. For known scale, apply $\theta_k = 18c^2m\rho_k \leq 18c^2mo$. For unknown scale, $\theta_k \leq 18mo$ and $C \leq 4$ give the claim.

In general this inference does not justify a square, cubic, or fourth root of the normalized lower bound. Such a power inversion requires an additional justified condition ensuring $\rho_k = o^k$, such as the sufficient geometry in Theorem 12. It cannot be inferred from the same unknown overlap being certified. For multiple predetermined claims, budgets summing to δ_{total} give simultaneous coverage by a union bound; the statistics may share observations.

All these guarantees are for exact observations and exact arithmetic. A finite-precision implementation inherits them if it encloses the entire real-valued statistic and evaluates the confidence endpoints outwards, using valid upper bounds for the variance coefficients. For example, an enclosure $\underline{u} \leq U \leq \bar{u}$ permits the lower endpoint to use $\ell_\delta(\underline{u})$, since ℓ_δ is nondecreasing. Rounding only the trace upwards does not enclose feature accumulation, normalization, subtraction of the diagonal, or evaluation of the inverse. The rationalized positive branch (46) can improve numerical stability but by itself supplies no such enclosure.

A verifiable rule for accepting a proposed stage For the same variance coefficients, an upper confidence endpoint is

$$R_\delta(u) = u_+ + \frac{h_\delta A_n}{2} + \sqrt{h_\delta A_n u_+ + \frac{h_\delta^2 A_n^2}{4} + h_\delta B_n}, \quad u_+ = \max(u, 0). \quad (73)$$

Indeed, if $\theta > R_\delta(u)$, then $\theta - u > \sqrt{h_\delta(A_n\theta + B_n)}$; the opposite one-sided Cantelli inequality bounds this event by δ . Together (72) and (73) give intervals with the sum of the two one-sided failure budgets.

Proposition 8 (Conditional acceptance by public validation). *At stage s , condition on the complete previous history and freeze both P_{old} and P_{new} . Use new independent validation directions and the direct $k = 1$ kernel to form simultaneous intervals $[L_{\text{old}}, R_{\text{old}}]$ and $[L_{\text{new}}, R_{\text{new}}]$ for their respective θ_1 values, with conditional total failure probability at most δ_s . Accept the proposal precisely when $L_{\text{new}} > R_{\text{old}}$; otherwise retain the old candidate. For any finite or countable set of stages with predetermined deterministic budgets $\delta_s \in (0, 1)$ satisfying $\sum_s \delta_s \leq \delta_{\text{total}}$, with probability at least $1 - \delta_{\text{total}}$ every accepted stage satisfies*

$$\text{tr}(P_{\text{new}}P_U) > \text{tr}(P_{\text{old}}P_U). \quad (74)$$

Proof. On the stage’s simultaneous coverage event, $\theta_1(P_{\text{new}}) \geq L_{\text{new}} > R_{\text{old}} \geq \theta_1(P_{\text{old}})$. The direct identity (36) divides this inequality by the common positive factor $18c^2$. The tower property bounds each unconditional stage failure by δ_s , and the union bound proves the joint assertion.

The candidates need not have the same rank: the rule compares captured carrier dimension $\text{tr}(PP_U)$. For a nested rank reduction, $P_{\text{new}} \preceq P_{\text{old}}$, that quantity cannot increase, so the rule does not accept such a reduction on its coverage event. It does not guarantee acceptance, convergence of the proposals, an increase in the smallest principal-angle coverage, or an improvement in purity or computational cost. Using $k = 4$ in the comparison instead certifies an increase of retained fourth energy. Unknown c does not prevent comparison of a common θ_1 scale when the intervals use a public variance bound, but sample splitting alone supplies none of the missing optimization or signal conditions.

D.12 Complete held-out validation outcomes

The following table includes every archived primary adaptive output, ordered by training data seed. Each frame received a distinct new stream of 65,536 angular directions at $p = 240, m = r = 16, a = 1/4$. The candidate, code, seeds, count and sixteen-way failure allocation were frozen before generation. Each known- and unknown-scale claim uses $\delta = 0.003125$, so all sixteen cover simultaneously with probability at least 0.95 under the exact iid model. This is a local prospective freeze, not an externally timestamped registration. The panel uses 524,288 validation directions; the exact-trace evaluation adds no observations. Neither NGCA panel enters this family. The candidates’ mean true overlap is 0.5998535,

Table 18. Saved public statistics and exact-trace lower bounds. LCBs are rounded down to six decimals; other columns are rounded. External true overlap and retained energy are independently evaluated oracle diagnostics; neither is an input to the certificate.

Seed	U_W	Known LCB	Unknown LCB	True o	True ρ_4
701	28.502506	0.126910	0.038887	0.549692	0.244690
702	51.908437	0.290080	0.102689	0.713119	0.441495
703	35.576216	0.175304	0.057788	0.653989	0.303096
704	25.544336	0.106981	0.031106	0.511780	0.220602
705	39.895717	0.205292	0.069509	0.639092	0.337290
706	14.735176	0.036199	0.003466	0.480369	0.127133
707	47.748222	0.260509	0.091110	0.580972	0.412466
708	36.910997	0.184538	0.061396	0.669815	0.322949

whereas the exact-trace mean simultaneous lower bounds are approximately 0.1732270662 and 0.0569944499. Using the exact trace 209990519484/54479161 instead of the dimension bound 3876 increases the means by about 0.00010453 and 0.00007225. All frames, statistics and sixteen failure allocations are unchanged. The exact trace is determined by fixed p, r, k ; choosing this algebraic improvement does not select a candidate or statistic using validation outcomes. The 70-digit recalculation uses archived public decimal values and does not enclose upstream feature-accumulation error. No frame was excluded. No reported lower bound exceeds 0.5. The unknown-scale bounds exceed the random-frame expectation $r/p = 1/15$ for seeds 702, 705 and 707 (3/8); the known-scale bounds do so for every seed except 706 (7/8). Seed 706 has the smallest unknown-scale bound: its saved statistic exceeds the scalar zero-bound threshold by more than 1.3986560977. This rationally bracketed scalar margin does not bound errors from the preceding float64 accumulation. The public API consumes only the candidate, directions, carrier dimension and confidence parameters. Synthetic generation and external scoring use the producer metadata separately; interface separation is not an operating-system isolation proof.

Separate single-output validation. Before the eight-output panel, the first archived frame (seed 701) was selected by its index and independently validated using another 65,536 directions. Its statistic is $U_W = 29.59534007$; known-scale and unknown-scale lower bounds are 0.21943266 and 0.08490182. Each of these two statements has individual 95% coverage; we do not claim their joint coverage is 95%. The eight-output table uses its separately generated data and a different multiplicity budget. These two individual bounds use the dimension trace bound 3876 and are separate from the sixteen-statement family. It is not permissible to select the larger bound across these experiments while retaining either experiment’s original confidence claim.

Numerical and reproducibility scope. The complete U-statistic represents 2, 147, 450, 880 unordered pairs per frame via sufficient accumulations; these are

dependent pairs, not extra independent observations. The checked implementation uses float64 and keeps one diagonal partial sum per computational batch. It is not a directed-rounding enclosure. The frame hashes, implementation hash, pre-execution protocol, all public statistics and separate oracle scores are preserved in the accompanying evidence under validation-manifest, which maps each saved object to its role and SHA256 digest.

D.13 Conditions for recovery, preprocessing, and statistical testing

This appendix separates population identifiability from finite-sample estimation, computational feasibility, and ordered-bit decoding. Unless a spherical null law is specified, the observations follow the exact angular model with a fixed key, $m = 4b_0 < p$, and $0 < a < 1/2$. Thus $c = \sin(2\pi a)/(2\pi a) > 0$. The statements about support learning additionally take a and m as known fixed parameters. We use $H, \mathcal{T}, \Sigma, C, \kappa_p$ from (58). None of these restrictions changes the endpoint theorems already proved in Appendix A.

Full signal is required for a carrier perturbation bound

Proposition 9 (Conditional full-signal perturbation). *Fix a working frame W and a population matrix $S_W \succeq 0$ such that $\text{range}(S_W) = U$, where $\dim U = m$. Put $g = \lambda_m(S_W) > 0$. If a symmetric estimate satisfies $\|\hat{S} - S_W\|_{\text{op}} \leq \eta < g$, then every valid top- m spectral projector $\hat{P}$ of $\hat{S}$ satisfies*

$$\|\hat{P} - P_U\|_F \leq \frac{\sqrt{2m}\eta}{g - \eta}. \quad (75)$$

Proof. Let $\hat{Q}$ contain orthonormal selected eigenvectors and let $\hat{\Lambda}$ contain their eigenvalues. Weyl's inequality gives $\lambda_{\min}(\hat{\Lambda}) \geq g - \eta > 0$. Since $P_{U^\perp} S_W = 0$, projection of the eigenvector equation gives

$$P_{U^\perp} \hat{Q} \hat{\Lambda} = P_{U^\perp} (\hat{S} - S_W) \hat{Q}.$$

Hence $\|P_{U^\perp} \hat{Q}\|_F \leq \sqrt{m}\eta/(g - \eta)$. For two rank- m projectors, $\|\hat{P} - P_U\|_F^2 = 2\|P_{U^\perp} \hat{Q}\|_F^2$, proving the result.

For the cubic population construction, Proposition 1 gives the sufficient signal condition

$$g \geq 18c^2 \lambda_{\min}(Q^T P_W Q)^3 > 0 \quad \text{if } \lambda_{\min}(Q^T P_W Q) > 0.$$

The true minimum coverage in this expression is not a public error bound. Conditional on past training, fresh samples allow Proposition 9 to be applied to a frozen W , provided its signal and estimation-error premises are established. Sample splitting does not establish these premises. The stronger condition $\eta < g/2$ makes (75) smaller than the trivial bound $\sqrt{2m}$ and separates the estimated m th and $(m + 1)$ st eigenvalues. Only $\eta < g$ is required for (75); possible eigenvalue ties do not invalidate its proof.

Proposition 10 (Support alone does not imply full signal). *There is an exact angular model with $\text{range}(S_W) \subsetneq U$ and a positive smallest nonzero eigenvalue g , for which replacing the full-signal hypothesis of Proposition 9 by these two properties makes (75) false.*

Proof. Take $p = 9, m = 8, a = 1/4$, and two carrier blocks $U_1 = \text{span}(e_1, \dots, e_4)$ and $U_2 = \text{span}(e_5, \dots, e_8)$. Set $W = U_1$. The block cubic identity gives $S_W = gP_{U_1}$, where $g = 18c^2$ and $c = 2/\pi$. Let $\eta = g/100$ and $\hat{S} = S_W + \eta e_9 e_9^T$. Its four eigenvalues g and its eigenvalue η are positive. Every top-eight subspace therefore contains U_1 and e_9 , and contains a three-dimensional subspace of U_2 . It follows that

$$\|\hat{P} - P_U\|_F = \sqrt{2}, \quad \|\hat{S} - S_W\|_{\text{op}} = \eta, \quad \frac{\sqrt{2m\eta}}{g - \eta} = \frac{4}{99}.$$

This is a strict violation, independent of how the remaining zero-eigenvalue directions are chosen.

Adding a term βP_W to an estimate requires its perturbation to be included, for example by replacing η by $\eta + |\beta|$ and checking the resulting full-signal condition. Likewise, squaring a noisy estimate introduces a separate issue: if $\hat{b} = b + \varepsilon$ with $\mathbb{E}\varepsilon = 0$, then $\mathbb{E}\hat{b}\hat{b}^T = bb^T + \mathbb{E}\varepsilon\varepsilon^T$. Freezing W does not remove this positive noise term. At the population level a full-signal S_W already has top eigenspace U ; a universal contraction factor for a noisy adaptive update does not follow from that identity.

Support learning without a computational restriction Let $F = (f_1, \dots, f_m)$ range over $\text{St}(m, p) = \{F \in \mathbb{R}^{p \times m} : F^T F = I_m\}$. Its blocks are ordered, with

$$u_{F,j}(q) = \begin{pmatrix} f_{4j-3}^T q \\ f_{4j-2}^T q \end{pmatrix}, \quad v_{F,j}(q) = \begin{pmatrix} f_{4j-1}^T q \\ f_{4j}^T q \end{pmatrix}.$$

Write P_F for the observation law and $P_F = FF^T$ for the carrier projector. This distinguishes a distribution from a matrix. With $t_a = \sin(\pi a)$, define the closed set

$$E_F = \bigcap_{j=1}^{b_0} \{q \in S^{p-1} : (u_{F,j}^T v_{F,j})^2 \leq t_a^2 \|u_{F,j}\|^2 \|v_{F,j}\|^2\}. \quad (76)$$

Let Q_0 denote uniform measure on S^{p-1} and $r_a = (2a)^{b_0}$.

Lemma 5 (Equal-volume support representation). *For every ordered frame,*

$$Q_0(E_F) = r_a, \quad P_F = Q_0(\cdot \mid E_F), \quad \frac{dP_F}{dQ_0} = r_a^{-1} \mathbf{1}_{E_F}. \quad (77)$$

For two frames with the same a, m ,

$$\text{TV}(P_F, P_G) = P_F(E_G^c) = \frac{Q_0(E_F \triangle E_G)}{2r_a}. \quad (78)$$

Furthermore, $D_{\text{KL}}(\mathbf{P}_F \| Q_0) = b_0 \log(1/(2a))$, whereas $D_{\text{KL}}(\mathbf{P}_F \| \mathbf{P}_G) = \infty$ whenever $Q_0(E_F \setminus E_G) > 0$.

Proof. Under the clean Gaussian law, each relative phase is uniform, independently across the disjoint blocks and independently of the pair radii. The inequality in (76) selects the two arcs centered at $\pi/2$ and $3\pi/2$, of total length $4\pi a$. Each block thus has event probability $2a$. Conditioning makes its relative phase the fair mixture of the two retained uniform arcs, exactly as in (12), and does not change any radius or the Gaussian bulk. Global normalization changes neither the relative angles nor these inequalities. This proves (77). Integrating the difference of the two indicator densities gives the last expression in (78). Their supports have equal measure, so their two set differences have equal measure; this gives the middle expression. The KL statements follow by integrating the log density ratio, including its infinite value on a positive-measure support difference.

The clean-versus-marked TV identity and its exact relation to the relative-angle decoding margin are already the content of Theorem 10; no additional universal distortion or channel-capacity relation is needed here. The support behavior also excludes a general local quadratic KL argument for this unsmoothed model. For example, in one block take unnormalized coordinates $u = (1, 0)$, $v = (t_a, \sqrt{1 - t_a^2})$, and a bulk coordinate equal to one. This point is on the boundary of E_F . Rotate the first axis of v toward that bulk coordinate by an arbitrarily small positive angle ε . Its new first coordinate is $t_a \cos \varepsilon + \sin \varepsilon > t_a$ for sufficiently small ε , while its second coordinate is unchanged. The new support inequality is violated strictly. A small perturbation of the original point into the interior of E_F still violates it; after normalization this gives an open spherical set in $E_F \setminus E_G$. Other blocks can be placed strictly inside their supports. Thus arbitrarily small such frame rotations have infinite $D_{\text{KL}}(\mathbf{P}_F \| \mathbf{P}_G)$.

Lemma 6 (TV controls carrier loss in the fixed angular family). *For frames with the same a, m ,*

$$\|\mathbf{P}_F - \mathbf{P}_G\|_F^2 \leq K_{m,c} \text{TV}(\mathbf{P}_F, \mathbf{P}_G), \quad K_{m,c} = \frac{8(C\kappa_p + 3c^2m)}{3c^2} \leq \frac{256}{c^2} + 8m. \quad (79)$$

Proof. If $\mathcal{T}_F \neq \mathcal{T}_G$, set $D = (\mathcal{T}_F - \mathcal{T}_G) / \|\mathcal{T}_F - \mathcal{T}_G\|_F$ and $f(q) = \langle D, H(q) \rangle$. Under either law, the covariance and mean bounds give $\mathbb{E}f^2 \leq M := C\kappa_p + 3c^2m$. For their densities p_F, p_G relative to Q_0 , Cauchy–Schwarz gives

$$\begin{aligned} \|\mathcal{T}_F - \mathcal{T}_G\|_F^2 &= \left(\int f(p_F - p_G) dQ_0 \right)^2 \\ &\leq \left(\int f^2 |p_F - p_G| dQ_0 \right) \left(\int |p_F - p_G| dQ_0 \right) \\ &\leq 4M \text{TV}(\mathbf{P}_F, \mathbf{P}_G). \end{aligned}$$

The first index of $\mathcal{T}_G$ is supported on G . Applying $I - P_G$ to that index and using the flattening identity yields

$$\|\mathcal{T}_F - \mathcal{T}_G\|_F^2 \geq 3c^2 \operatorname{tr}((I - P_G)P_F) = \frac{3}{2}c^2\|P_F - P_G\|_F^2.$$

Combining the two inequalities proves the first bound. If the tensors coincide, the same last identity already forces the projectors to coincide. Finally $C \leq 4$ and $\kappa_p < 24$ give the stated dimension-independent upper bound on $K_{m,c}$.

Theorem 14 (A measurable support estimator and its risk). *Suppose $a \in (0, 1/2)$ and $m = 4b_0$ are known. There exists a Borel measurable estimator $\hat{F} : (S^{p-1})^n \rightarrow \operatorname{St}(m, p)$ which, whenever a feasible frame exists, chooses one satisfying $q_i \in E_{\hat{F}}$ for every i . Define*

$$V = \lceil 2mp \log_2(64eb_0) \rceil, \quad \epsilon_n = \min \left\{ 1, \frac{8 + 4[V \log(2en/V) + \log(2/\delta)]}{n} \right\}. \quad (80)$$

For $n \geq V$ and $\delta \in (0, 1)$, under every true frame F ,

$$\Pr_F\{\operatorname{TV}(\mathbf{P}_F, \mathbf{P}_{\hat{F}}) > \epsilon_n\} \leq \delta, \quad (81)$$

$$\mathbb{E}_F\|P_{\hat{F}} - P_F\|_F^2 \leq K_{m,c}\epsilon_n + 2m\delta. \quad (82)$$

The assertion imposes no bound on the time required to find a feasible support.

Proof. Membership in (76) is a conjunction of b_0 polynomial inequalities in mp frame parameters and the sample coordinates. Each polynomial has degree four in the frame parameters and total degree at most eight jointly. The Goldberg–Jerrum bound [22, Theorem 2.2] applied in its total-degree formulation with $d = 8$ bounds the VC dimension by $2mp \log_2(8e \cdot 8b_0)$, and hence by V . Restricting the parameters to orthonormal frames cannot increase the VC dimension. The complement class $\{E_G^c\}$ has the same VC dimension.

We next construct a measurable choice explicitly. The Stiefel manifold and the sample space are compact. For a sample $x = (q_1, \dots, q_n)$, its feasible fiber

$$\mathcal{F}(x) = \{G \in \operatorname{St}(m, p) : q_i \in E_G \text{ for every } i\}$$

is compact, and the graph of $\mathcal{F}$ is closed, because all its defining inequalities are weak and continuous. The set of x having a nonempty fiber is consequently closed. On its complement choose a fixed default frame. For each integer $\ell \geq 1$, fix a finite cover of the Stiefel manifold by closed ambient balls of radius $2^{-\ell}$, with centers in the manifold and a fixed order. On the feasible domain, choose the first ball in the first cover that meets the fiber. Recursively choose the first ball in the ℓ th cover whose intersection with the fiber and all previously chosen balls is nonempty. For any fixed finite sequence of balls, this nonemptiness event is the projection of a compact set and is closed. The first-index choices are thus Borel measurable by induction. Their intersections with the fiber are nested

nonempty compact sets and have diameters at most $2^{1-\ell}$. They have a unique common point; their selected centers converge to it. This pointwise limit defines a Borel measurable $\hat{F}(x) \in \mathcal{F}(x)$. The true frame is feasible almost surely under P_F , so the default case has probability zero in the model.

Here is a double-sample proof with explicit constants. If $\epsilon_n = 1$, (81) is immediate, so consider $\epsilon = \epsilon_n < 1$. Suppose the selected support misses true probability more than ϵ while containing all n training points. For an independent ghost sample of size n , the number falling outside that support is binomial with success probability greater than ϵ . Chebyshev's inequality implies that it is at least $n\epsilon/2$ with probability at least $1/2$ when $n\epsilon \geq 8$. It follows that the bad training event has probability at most twice the probability that some E_G^c misses the training sample and contains at least $n\epsilon/2$ ghost points.

Condition on the pooled $2n$ observations and assign uniformly which n form the training half. For a fixed labeling with $j \geq n\epsilon/2$ error points, the probability that all j fall in the ghost half is zero if $j > n$, and otherwise is

$$\frac{\binom{n}{j}}{\binom{2n}{j}} \leq 2^{-j} \leq e^{-n\epsilon/4}.$$

Sauer's bound limits the number of labelings to $(2en/V)^V$, since $n \geq V$. Union bounding gives $2(2en/V)^V e^{-n\epsilon/4} \leq \delta$ by (80). Equal support volumes identify the missed probability with TV by (78), proving (81). Finally use Lemma 6 on the good event and the bound $\|P_{\hat{F}} - P_F\|_F^2 \leq 2m$ on its complement to obtain (82).

For fixed a, m , take $n = p^2$ and $\delta = 1/p$. Eventually $n \geq V$, and (82) yields

$$\sup_F \mathbb{E}_F \|P_{\hat{F}} - P_F\|_F^2 = O_{m,a} \left(\frac{\log p}{p} \right) \rightarrow 0. \quad (83)$$

Consequently, no universal lower bound of the form $c_0 \min\{m, p^3/(nc^2)\}$, with a positive constant c_0 independent of p, n , can hold over all estimators in this fixed hard-support model: at $n = p^2$ its right side eventually equals $c_0 m$. This does not contradict a cubic sufficient count for a specified moment estimator, and does not prove that the support feasibility problem is computationally easy. The conclusion is asymptotic. For example, at $m = 16$ the displayed V is approximately $302p$; $n = p^2$ fails $n \geq V$ at $p = 240$, and the displayed bound remains truncated at one at $p = 2048$ with $\delta = 1/p$. Noise or contamination removes the exact support premise and requires a different analysis.

Proposition 11 (A row-sparse support class). *Suppose the true frame is supported on an unknown set of at most s public coordinate rows, where $m \leq s \leq p$. Put*

$$V_s = \lceil 2sm \log_2(64eb_0) \rceil, \\ \epsilon_{n,s} = \min \left\{ 1, \frac{8 + 4[V_s \log(2en/V_s) + \log \binom{p}{s} + \log(2/\delta)]}{n} \right\}.$$

For $n \geq V_s$, a measurable feasible estimator restricted to this class satisfies (81)–(82) with ϵ_n replaced by $\epsilon_{n,s}$.

Proof. For a fixed set of s rows there are sm frame parameters, giving VC bound V_s by the preceding argument. Any smaller row support is contained in one of the $\binom{p}{s}$ size- s supports. The growth function for their union is therefore at most $\binom{p}{s}(2en/V_s)^{V_s}$. Insert this factor into the same double-sample proof. The parameter set is a finite union of compact Stiefel sets, so the preceding measurable-selection construction also applies. The TV-to-projector estimate is unchanged.

Since $\log \binom{p}{s} \leq s \log(ep/s)$, this gives an explicit statistical benefit from row sparsity, without an efficient algorithm claim. Columnwise sparsity is a different class and has a different support count; a rank- m row-sparse frame necessarily has $s \geq m$.

Tyler whitening under a fixed affine transformation

Proposition 12 (Exact affine preservation and its scale). *Suppose $Y = AZ$, where A is a fixed invertible matrix and Z has the exact angular law. The representative $S = AA^T$ solves the population Tyler equation*

$$S = p \mathbb{E} \frac{YY^T}{Y^T S^{-1} Y}. \quad (84)$$

Writing $O = S^{-1/2}A$, one has $O^T O = I_p$ and the normalized whitened direction is $\tilde{q} = Oq$. Thus its second moment is I_p/p and its fourth tensor is $O^{\otimes 4} \mathcal{T}$, supported on the rotated carrier OU . If instead $S = tAA^T$ for $t > 0$, the orthogonal matrix is $O = \sqrt{t} S^{-1/2}A$ and the same normalized conclusion holds.

Proof. Theorem 7 gives $p \mathbb{E} ZZ^T / \|Z\|^2 = I_p$. For $S = AA^T$, substitution into the right side of (84) gives $A(p \mathbb{E} ZZ^T / \|Z\|^2)A^T = AA^T$. The identity $S^{-1/2}AA^T S^{-1/2} = I_p$ makes O orthogonal. Normalization cancels the length of Z , and, for a rescaled representative, the additional factor $t^{-1/2}$. Orthogonal equivariance proves the moment and carrier claims.

This calculation uses angular isotropy and the specified fixed affine map; it does not require elliptical symmetry of Z . For nonzero finite observations, the strict condition $\#\{i : Y_i \in V\}/n < \dim(V)/p$ for every proper nonzero subspace V ensures an SPD Tyler solution unique up to scale [23, Theorem 2.1]. With that scale fixed, the finite-sample equation is $\hat{S} = (p/n) \sum_i Y_i Y_i^T / (Y_i^T \hat{S}^{-1} Y_i)$. When $n < p$ it cannot hold with $\hat{S}$ positive definite, because its right side has rank at most n .

Proposition 13 (Same-sample quadratic cancellation). *If an SPD Tyler solution exists and $\tilde{q}_i = \hat{S}^{-1/2} Y_i / \|\hat{S}^{-1/2} Y_i\|$, then*

$$\frac{1}{n} \sum_i \tilde{q}_i \tilde{q}_i^T = \frac{I_p}{p}. \quad (85)$$

For any orthonormal $B \in \mathbb{R}^{p \times r}$, even one chosen from these same fitted observations, define

$$h_i = B^T \tilde{q}_i \tilde{q}_i^T B - \frac{\|B^T \tilde{q}_i\|_F^2}{r} I_r, \quad U_2^{\text{quad}} = \binom{n}{2}^{-1} \sum_{i < j} \langle h_i, h_j \rangle_F.$$

Then $\sum_i h_i = 0$ and

$$U_2^{\text{quad}} = -\frac{\sum_i \|h_i\|_F^2}{n(n-1)} = -\frac{1-1/r}{n(n-1)} \sum_i \|B^T \tilde{q}_i\|_F^4. \quad (86)$$

It is identically zero for $r = 1$. For $r > 1$,

$$-\frac{r-1}{p(n-1)} \leq U_2^{\text{quad}} \leq -\frac{r(r-1)}{p^2(n-1)} < 0. \quad (87)$$

Proof. Multiplying the finite-sample fixed-point equation on both sides by $\hat{S}^{-1/2}$ proves (85). Sandwiching this identity by B and taking its trace gives $\sum_i h_i = 0$, pointwise for every B . Expansion of $\|\sum_i h_i\|_F^2$ proves the first equality in (86). For $z = B^T \tilde{q}_i$, $\|zz^T - \|z\|^2 I_r / r\|_F^2 = (1-1/r)\|z\|^4$. Finally put $x_i = \|B^T \tilde{q}_i\|_F^2$. One has $0 \leq x_i \leq 1$ and $\sum_i x_i = nr/p$. Hence $nr^2/p^2 \leq \sum_i x_i^2 \leq nr/p$ by Cauchy-Schwarz and $x_i^2 \leq x_i$; substitution gives the bounds.

The fitted directions are dependent through $\hat{S}$. The exact sample identity is consequently distinct from a population claim for independent future observations. Regularized or approximate Tyler fits change the equation and must retain their actual residuals.

Lemma 7 (Deterministic error on fresh whitened directions). *Freeze a fitted whitening transformation before applying it to new observations. After removing its irrelevant scalar and a fixed orthogonal factor, suppose its relative action is $I + D$, where $\|D\|_{\text{op}} \leq \eta < 1$. For a unit direction q , set $\hat{q} = (I + D)q / \|(I + D)q\|$. Then*

$$\begin{aligned} \|\hat{q} - q\| &\leq 2\eta, \\ \|\mathbb{E} \hat{q} \hat{q}^T - I_p / p\|_{\text{op}} &\leq 4\eta, \\ \|\mathbb{E} H(\hat{q}) - \mathcal{T}\|_F &\leq 8p(p+2)\eta. \end{aligned} \quad (88)$$

Expectations can be conditional on the fitting history.

Proof. For $z = (I + D)q$, the lower bound $\|z\| \geq 1 - \eta > 0$ makes normalization valid, and

$$\|z/\|z\| - q\| \leq |1 - \|z\|| + \|Dq\| \leq 2\eta.$$

The two-term outer-product expansion bounds $\|\hat{q} \hat{q}^T - qq^T\|_{\text{op}}$ by $2\|\hat{q} - q\| \leq 4\eta$. The four-term telescoping expansion of $\hat{q}^{\otimes 4} - q^{\otimes 4}$ has Frobenius norm at most $4\|\hat{q} - q\| \leq 8\eta$. Apply the contraction Π_4 , multiply by $p(p+2)$, and take expectations to obtain (88).

A sample size and confidence level for η require a separate estimation theorem for the actual observation law. A condition-number restriction on A alone supplies neither a dimension-independent rate nor exact coverage of the original angular certificate after estimated whitening.

Ordered bits require an orientation convention

Proposition 14 (An unanchored sign ambiguity). *With independently refreshed fair bits and no shared sign anchor, even knowledge of P_U does not permit a uniformly accurate ordered-bit decoder from public observations. For any binary decoder using those observations and P_U , there are two allowed keys whose full-message average bit-error probabilities sum to one. At least one is therefore at least $1/2$.*

Proof. In every block, negate both columns of its second two-dimensional frame, and negate every payload sign in that block. Their products in the generated vector are unchanged, so this transformation preserves every public observation pointwise and preserves the full carrier projector. The negated payloads remain independent and fair; thus the two keys have the same marginal observation law, including for any number of training observations. Couple the decoder's internal randomness identically under the two keys. Its predicted bit is identical, whereas every target bit is opposite. For a binary output, the two error indicators sum to one. Average over all bits, observations, and randomness.

Negating only one block proves the analogous ambiguity for that block's error rate; negating all blocks gives the stated full-message conclusion. Pilots, synchronization, a signed ordered frame, or asymmetric code constraints can alter the problem by breaking the symmetry. Their provision and their effect on the refreshed payload law must be part of a decoding model.

Proposition 15 (Likelihood ratio for wrapped Gaussian angle noise). *Suppose the signed, ordered pair frame is shared, and write ϕ for the observed relative angle after subtracting $\pi/2$. Assume independent wrapped Gaussian angle noise with $\sigma > 0$, having circular density*

$$w_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{\ell \in \mathbb{Z}} \exp\left(-\frac{(x + 2\pi\ell)^2}{2\sigma^2}\right).$$

For the angular residual uniform on $[-\pi a, \pi a]$, the two conditional densities and the fair-bit log likelihood ratio are

$$f_+(\phi) = \frac{1}{2\pi a} \int_{-\pi a}^{\pi a} w_\sigma(\phi - d) dd, \quad f_-(\phi) = f_+(\phi - \pi), \quad L(\phi) = \log \frac{f_+(\phi)}{f_-(\phi)}. \quad (89)$$

Proof. Conditional on a positive bit, the observed residual is the sum modulo 2π of the uniform residual and the independent noise. Circular convolution gives f_+ . A negative bit adds π , yielding f_- . The prior odds are one, so the log density ratio is the posterior log odds.

The expression $2 \cos \phi / \sigma^2$ instead follows from the separately specified channel $f_s(\phi) \propto \exp(s \cos \phi / \sigma^2)$. It is not the general likelihood ratio in (89). For a Huber-type clipped score $|\psi| \leq M$ and a contamination law $R = (1 - \varepsilon)P + \varepsilon Q$, one can assert

$$|\mathbb{E}_R \psi - \mathbb{E}_P \psi| = \varepsilon |\mathbb{E}_Q \psi - \mathbb{E}_P \psi| \leq 2\varepsilon M.$$

Boundedness alone gives no numerical BER: a decoding guarantee also needs the conditional channel laws, contamination size, orientation convention, and decision or coding rule. The displayed likelihood and bounded-score statements do not assume an implemented end-to-end decoder or a measured BER.

Fixed-dimensional limits and finite-sample null control

Proposition 16 (The degenerate limit and its normalization). *Fix p, r, k and a candidate independent of iid validation directions. Under Q_0 , if λ_j are all eigenvalues of L_k on $\mathcal{H}_4$, then*

$$nU_k \implies \sum_j 6\kappa_p \lambda_j (G_j^2 - 1), \quad \sqrt{n}U_k \longrightarrow 0 \quad \text{in } L^2, \quad (90)$$

where the G_j are independent standard Gaussians. For an angular alternative with fixed parameters, the nondegenerate conclusion $\sqrt{n}(U_k - \theta_k) \implies N(0, 4v_1)$ holds if $v_1 > 0$, with v_1 as in Proposition 6. Positive overlap alone does not ensure this condition for $k = 3, 4$.

Proof. Under Q_0 , H has mean zero and covariance $\kappa_p I$. Put $Z_i = L_k^{1/2} H(q_i)$. The exact square expansion gives

$$nU_k = \frac{6n}{n-1} \left(\|n^{-1/2} \sum_i Z_i\|_F^2 - \frac{1}{n} \sum_i \|Z_i\|_F^2 \right).$$

The finite-dimensional central limit theorem and law of large numbers give a centered Gaussian quadratic form with covariance $\kappa_p L_k$, proving the first claim. The exact variance in Proposition 6 gives $n\mathbb{E}_0 U_k^2 = 72\kappa_p^2 \text{tr}(L_k^2)/(n-1) \rightarrow 0$. For an alternative, its Hoeffding linear term has variance $4v_1/n$ and obeys the scalar central limit theorem. The canonical remainder multiplied by $\sqrt{n}$ converges to zero in L^2 , proving the conclusion when $v_1 > 0$.

For the last assertion, take $p = 8, m = r = 4, a = 1/4$, $U = \text{span}(e_1, e_2, e_3, e_4)$, and $\text{range}(B) = \text{span}(e_1, e_2, e_5, e_6)$. Here $o = 1/2$, but each nonzero block-tensor entry has two indices in each carrier plane. Projection of three or four indices to the candidate therefore kills every entry: $A_3 \mathcal{T} = A_4 \mathcal{T} = 0$. It follows that $L_3 \mathcal{T} = L_4 \mathcal{T} = 0$ and $v_1 = 0$.

There is also a distinct quadratic statistic. In an orthonormal basis of symmetric trace-free r by r matrices, identify $h(q) = B^T q q^T B - \|B^T q\|^2 I_r / r$ with a vector. For frozen B , isotropy gives $\mathbb{E}h = 0$ under both the exact angular law

and Q_0 . If $\Sigma_h = \text{Cov}(h)$ has eigenvalues ν_j , its independent sample statistic satisfies

$$\begin{aligned} U^{\text{quad}} &= \frac{\|\sum_i h(q_i)\|^2 - \sum_i \|h(q_i)\|^2}{n(n-1)}, & \text{Var}(U^{\text{quad}}) &= \frac{2 \text{tr}(\Sigma_h^2)}{n(n-1)}, \\ nU^{\text{quad}} &\implies \sum_j \nu_j (G_j^2 - 1), & \sqrt{n} U^{\text{quad}} &\longrightarrow 0 \quad \text{in } L^2. \end{aligned} \quad (91)$$

The proof is the same finite-dimensional square expansion, without the factor six. Under Q_0 , a trace-free coefficient matrix D has $\mathbb{E}(q^T B D B^T q)^2 = 2\|D\|_F^2/[p(p+2)]$, so $\nu_j = 2/[p(p+2)]$ for all $j = 1, \dots, r(r+1)/2 - 1$. For $r = 2$, continuity of the limiting law gives $\Pr_0(U^{\text{quad}} > 0) \rightarrow \Pr(\chi_2^2 > 2) = e^{-1}$. The vanishing variance at the $\sqrt{n}$ scale consequently cannot justify an extreme-tail test at threshold zero.

Proposition 17 (Finite-sample Cantelli null test). *For a frozen positive contraction L and $n \geq 2$ iid directions from Q_0 , the complete quartic statistic satisfies*

$$\mathbb{E}_0 U = 0, \quad \text{Var}_0(U) = \beta_0 = \frac{72\kappa_p^2 \text{tr}(L^2)}{n(n-1)}. \quad (92)$$

For any $\alpha \in (0, 1)$ the strict rule

$$U > t_\alpha, \quad t_\alpha = \sqrt{\frac{1-\alpha}{\alpha}} \beta_0 \quad (93)$$

has false-positive probability at most α . A valid upper bound on β_0 can replace its exact value. If the null is instead an angular law satisfying $L\mathcal{T} = 0$, one may use $C^2\beta_0$ as a variance upper bound.

Proof. For Q_0 , $\mathcal{T} = 0$ and $\Sigma = \kappa_p I$. The exact variance formula of Proposition 6 gives (92). Cantelli's inequality gives $\Pr_0(U > t) \leq \beta_0/(\beta_0 + t^2)$, proving the claim. If $\beta_0 = 0$, then $U = 0$ almost surely and the strict rule does not reject. For the stated angular null, positivity of L implies $\theta = 0$, and its canonical variance is bounded using $\Sigma \preceq C\kappa_p I$.

Weak convergence in (90) does not by itself give a finite- n tail bound. Hanson–Wright inequalities require a specified vector with independent centered subgaussian coordinates, a fixed matrix, and the appropriate distribution and absolute constants [24, Theorem 1.1]. The coordinates of a finite-sample quartic spherical feature need not be independent. Concentration for canonical U-statistics instead controls additional kernel and conditional-moment quantities [25, Theorem 3.3]. Under Q_0 the kernel eigenvalues here are already known from Theorem 13; this spectral information alone does not verify the hypotheses of a finite-sample concentration theorem.

Proposition 18 (A finite-sample failure of a limit-based threshold). *In the exact angular model there is an overlap-zero candidate with $p = 5, m = 4, r =$*

$1, n = 2, a = 1/4$ for which the following substitution does not yield false-positive probability 10^{-6} : use the limiting Gaussian Orlicz constant $K_G^2 = 8/3$ and the kernel spectrum Λ in

$$t_\alpha^{\text{lim}} = \frac{K_G^2}{n} \max \left\{ \| \Lambda \|_F \sqrt{\log(1/\alpha)}, \| \Lambda \|_{\text{op}} \log(1/\alpha) \right\}. \quad (94)$$

The same example applies under the clean spherical null.

Proof. Let the carrier be $\text{span}(e_1, \dots, e_4)$ and the candidate be e_5 . The bulk coordinate $t = q_5$ has density $3(1 - t^2)/4$ on $[-1, 1]$, since the angular map preserves the bulk coordinate and every radius. The feature expansion (26) gives

$$H_{5555}(q) = Y(t) := 35t^4 - \frac{70}{3}t^2 + \frac{5}{3}.$$

For $k = 4$ the kernel is $6Y(t)Y(t')$. Integration against that density gives $\mathbb{E}Y = 0$ and $\mathbb{E}Y^2 = 320/99$; its only nonzero integral-operator eigenvalue is $\lambda = 6\mathbb{E}Y^2 = 640/33$. For a standard Gaussian G , the convention $\mathbb{E} \exp(G^2/K_G^2) \leq 2$ gives the smallest $K_G^2 = 8/3$, by evaluating $(1 - 2/K_G^2)^{-1/2}$. At $\alpha = 10^{-6}$, (94) therefore is

$$t_{10^{-6}}^{\text{lim}} = \frac{2560}{99} \log(10^6) < \frac{35840}{99}.$$

The polynomial Y is positive and strictly increasing on $[47/50, 1]$, and is even. On the event that both independent coordinates have $|t| > 47/50$, the statistic $U_4 = 6Y(t)Y(t')$ is greater than

$$6Y(47/50)^2 = \frac{986481371706601}{2343750000000} > 420 > \frac{35840}{99}.$$

The probability of this event is exactly

$$\left(\int_{|t| > 47/50} \frac{3}{4}(1 - t^2) dt \right)^2 = \left(\frac{1323}{250000} \right)^2 = \frac{1750329}{62500000000} > 10^{-6}.$$

The bulk-coordinate marginal is also the marginal under Q_0 , so the same proof covers that null.

This example refutes the stated transfer of the limiting Gaussian constant to finite observations. It does not rule out a separately proved inequality with constants for the actual finite-sample representation, and does not measure false-positive probability at a larger n .

Proposition 19 (Power cost of one conservative null threshold). *Fix $p = 240, m = r = 16, k = 4, n = 65536$ and $\alpha = 10^{-6}$. Use the test (93) with $\tau_{240,16}^{(4)} = 209990519484/54479161$. Under the exact angular alternative with $a = 1/4$, every independently frozen rank-sixteen candidate has true-positive probability strictly less than 0.06% for this test.*

Proof. Here $c = 2/\pi$, $C = (1 + c)^2$, and $\theta \leq \theta_{\max} = 18c^2m$ by Theorem 12. Let β_0 be (92), $t = t_\alpha$, and

$$V_{\max} = \frac{24C\kappa_p}{n}\theta_{\max} + C^2\beta_0.$$

Proposition 6 gives $\text{Var}_1(U_4) \leq V_{\max}$. Substitution yields

$$t \approx 186.67535, \quad \theta_{\max} \approx 116.72200, \quad V_{\max} \approx 2.90885681.$$

In particular $t > \theta_{\max}$. Cantelli's bound is increasing in the variance and decreasing in the positive threshold-to-mean gap, so

$$\Pr_1(U_4 > t) \leq \frac{V_{\max}}{V_{\max} + (t - \theta_{\max})^2} \approx 0.00059408324 < 0.0006. \quad (95)$$

The strict final comparison does not depend on the displayed rounding: the exact rational trace and $3.14159 < \pi < 3.14160$ give $t > 186.67$, $\theta_{\max} < 116.73$, and $V_{\max} < 2.91$. Consequently the right side is less than $2.91/[2.91 + 69.94^2] < 0.0006$.

This is a power upper bound for the specified conservative Cantelli test at the displayed parameters. It neither bounds the power of every valid test nor supplies an empirical ROC curve. If false-positive control is instead supported by independent binary decisions with zero observed errors, the exact 95% one-sided binomial upper bound from N decisions is $1 - 0.05^{1/N}$. Making this at most 10^{-6} requires $N \geq \lceil \log(0.05)/\log(1 - 10^{-6}) \rceil = 2,995,731$. Here N counts independent decisions, whereas n counts directions inside one U-statistic. This calculation concerns that particular empirical validation design, not all methods of proving a false-positive guarantee.

Conditional selection of an unknown carrier rank

Proposition 20 (Rank selection with explicit signal conditions). *Let $G \succeq 0$ have rank $m < p$ and let $\hat{G}$ be symmetric with $\|\hat{G} - G\|_{\text{op}} \leq \eta$. If $G = \lambda P_U$ and $\eta < \lambda/4$, the largest adjacent eigenvalue gap of $\hat{G}$, among indices $1, \dots, p-1$, occurs uniquely at m . For a general positive spectrum, if $\lambda_m(G) > 3\eta$, then the number of eigenvalues of $\hat{G}$ strictly greater than 2η equals m .*

Proof. Order all eigenvalues decreasingly. In the flat-spectrum case, Weyl's inequality gives

$$\hat{\lambda}_m - \hat{\lambda}_{m+1} \geq \lambda - 2\eta > 2\eta, \quad \hat{\lambda}_j - \hat{\lambda}_{j+1} \leq 2\eta \quad (j \neq m).$$

For a general spectrum with $\lambda_m(G) > 3\eta$, the first m estimated eigenvalues exceed $\lambda_m(G) - \eta > 2\eta$, and the remaining ones are at most η . This proves the threshold claim.

Flat spectrum is sufficient for the gap rule, not necessary. More generally it suffices that the population boundary gap exceed every internal gap by more than 4η . Without such a condition, a largest gap can precede the true rank even in this angular model. For $p = 5, m = 4$, take the single-block public probe $A = \text{diag}(1, 0, 10, 0, 0)$. The cumulants (14) give $C(A) = c \text{diag}(-5, 5, -1/2, 1/2, 0)$. Thus $C(A)^2$ has spectrum $c^2(25, 25, 1/4, 1/4, 0)$, whose largest gap is at two, although its rank is four.

A fixed GOE probe gives a separate consistent route. By Theorem 8, $\text{rank } C(A) = m$ almost surely. For any realized such probe and fixed parameters, suppose $t_n \rightarrow 0$ and $\|\widehat{C} - C(A)\|_{\text{op}} = o_{\text{Pr}}(t_n)$. With probability tending to one, the error is at most t_n . For all sufficiently large n , the smallest nonzero singular value of $C(A)$ is greater than $3t_n$; Weyl's singular-value inequality then shows that exactly m estimated singular values exceed $2t_n$. This requires the stated estimation-error control and a specified threshold. At $a = 1/2$, Theorem 9 gives identical clean observation laws for all carriers, so rank is not identifiable in that model.

The results above supply explicit conditional guarantees and counterexamples within the stated latent family. They do not depend on unreplicated ROC, BER, geometric-warp robustness, or runtime measurements, and imply no convergence or deployment guarantee for an adaptive neural procedure.

E Rank-dependent cubic noise and Gaussian-chaos analysis

E.1 Exact null moments and angular covariance comparison

Set $\alpha = p(p+2)$, $d = p(p+2)/(p+4)$, $e = 3p/(p+4)$, $s = q^T v$ and $t = \|v\|^2$. The score is

$$h_v(q) = (\alpha s^3 - 3dts)q + (-3ds^2 + et)v.$$

Under the uniform spherical law it has mean zero. For $v = e_1$ and $x = q_1$, its first component is $\alpha x^4 - 6dx^2 + e$, while an orthogonal component is $(\alpha x^3 - 3dx)q_j$. Use the exact moments

$$\mathbb{E}_0 x^{2k} = \frac{(2k-1)!!}{p(p+2) \cdots (p+2k-2)}, \quad \mathbb{E}_0(q_j^2 | x) = \frac{1-x^2}{p-1}.$$

Expanding the two squares gives

$$a_p = \frac{6p(p+1)(p+2)}{(p+4)^2(p+6)}, \quad b_p = \frac{18p(p+1)(p-2)}{(p+4)^2(p+6)}.$$

Orthogonal equivariance and degree-six homogeneity then give, for every fixed v ,

$$\mathbb{E}_0[h_v h_v^T] = a_p \|v\|^6 I + b_p \|v\|^4 v v^T. \quad (96)$$

For $v \sim N(0, P_W)$ independent of q , $\mathbb{E}\|v\|^6 = r(r+2)(r+4)$ and $\mathbb{E}\|v\|^4 v v^T = (r+2)(r+4)P_W$. These identities prove (11). They were checked with exact rational spherical moments, independently of the matrix accumulation code.

Let $b(v) = \mathbb{E}h_v = \mathcal{T}(\cdot, v, v, v)$ under the angular law. Theorem 4, applied to the fixed contraction map, proves

$$\mathbb{E}_v \text{Cov}_q(h_v) \preceq C_a L_{p,r} [(p+2)rI + 3(p-2)P_W].$$

The same statement holds conditional on a work frame chosen using independent training data. It does not hold by that argument conditional on a frame selected using the very observations being averaged.

For two fixed probes an additional exact public identity is

$$\mathbb{E}_0[h_v^T h_w] = \frac{6p(p+1)}{(p+4)^2} [(p+2)(v^T w)^3 - 3\|v\|^2\|w\|^2(v^T w)]. \quad (97)$$

Orthogonal invariance leaves these two bidegree- $(3,3)$ invariants. The score is harmonic as a polynomial of ambient v , so its Laplacian fixes the coefficient ratio to $-3/(p+2)$. The diagonal trace of (96) fixes the remaining coefficient. If J_{kl}^0 denotes (97) and $X(q) = K^{-1/2}[h_{v_1} - b(v_1), \dots, h_{v_K} - b(v_K)]$, then

$$\mathbb{E}_P[X^T X] \preceq (C_a/K)J^0.$$

To see this, apply the centered covariance comparison to every linear combination of the fixed contractions and sum its output-coordinate variances. This accounts for shared observations; it assumes no independence between different probes' observation noise.

The expectation of the uncorrected empirical Gram must also be centered correctly. For $G_K = K^{-1} \sum_k b(v_k)b(v_k)^T$ and fixed probes,

$$\mathbb{E} \frac{1}{K} \sum_k \bar{h}_{v_k} \bar{h}_{v_k}^T = \left(1 - \frac{1}{n}\right) G_K + \frac{1}{nK} \sum_k \mathbb{E}[h_{v_k} h_{v_k}^T] = G_K + \frac{1}{nK} \sum_k \text{Cov}(h_{v_k}).$$

The work-projector shape in (11) is exact under the null, whereas the angular increment is generally only bounded by its C_a multiple. This is why the original implementation removes the sample diagonal instead of treating null-law bias as an exact marked-law correction.

E.2 Third-chaos projection and retained signal

Write $v = Bg$ with $B^T B = I_r$, $g \sim N(0, I_r)$, and put $w = P_W q$, $u = \|w\|^2$. For fixed q , h_{Bq} is homogeneous cubic in g . Its first and third Hermite components are $h^{(1)} = \frac{1}{2} \Delta_g h$ and $h^{(3)} = h - h^{(1)}$. Termwise differentiation gives the explicit formula

$$\begin{aligned} h_v^{(3)}(q) &= \alpha q(s^3 - 3us) \\ &\quad - 3d[(s^2 - u)v - 2sw + (t - r - 2)sq] + e(t - r - 2)v. \end{aligned} \quad (98)$$

Here Δ_g is the Laplacian, not the full Ornstein–Uhlenbeck generator.

Gaussian first and third chaoses are orthogonal, both before and after subtracting the mean over q . Consequently

$$\mathbb{E}_v \text{Cov}_q(h_v) = \mathbb{E}_v \text{Cov}_q(h_v^{(3)}) + \mathbb{E}_v \text{Cov}_q(h_v^{(1)}).$$

This is a positive-semidefinite comparison after Gaussian probe averaging. It is not a pointwise comparison for every finite collection of probes. For the population signal, the Wick decomposition gives

$$Q^T S_W^{(3)} Q = 6FM^{\otimes 3}F^T, \quad Q^T S_W Q = 6FM^{\otimes 3}F^T + 9D_M M D_M^T.$$

If $M = LL^T$ and $A = FL^{\otimes 3}$, the second term is $9AJJ^T A^T$ with the contraction J of Appendix C.5. Both signals therefore have range $\text{range}(A)$, which is all of the carrier only when A has row rank m . Removing first chaos can remove useful signal as well as noise. At $r = p$, the ambient trace is zero and $h^{(3)} = h$.

For $r < p$ and any unit $x_0 \perp W$, the deleted fraction of null noise is

$$\frac{\mathbb{E}_{0,v}(x_0^T h_v^{(1)})^2}{\mathbb{E}_{0,v}(x_0^T h_v)^2} = \frac{3(p-r+2)}{(p+1)(r+4)}. \quad (99)$$

Indeed $u = \|P_W q\|^2$ has beta law with parameters $(r/2, (p-r)/2)$. Conditional first-chaos energy in $W^\perp$ is $9(1-u)u[\alpha u - d(r+2)]^2$. Its expectation divided by $p-r$ is

$$\frac{18p(p+2)r(r+2)(p-r+2)}{(p+4)^2(p+6)}.$$

Divide by the outside eigenvalue in (11). The inside value similarly follows from beta moments through order four. There is no outside direction at $r = p$; a formal

Table 19. Exact null noise at $p = 240$. Percentages concern Gaussian probe-averaged noise and do not measure recovery improvement.

r	Original $W^\perp$ variance	Deleted outside	Deleted inside
16	33029.55	14.0664%	13.9662%
32	224600.92	7.2614%	7.1991%
64	1647073.44	3.2585%	3.2235%

algebraic outside coefficient at that endpoint has no observable interpretation.

The score (98) has the form $aq + bv + c_0 P_W q$. Its mean and diagonal square sum can be accumulated through three square terms and three cross terms. The resulting U-estimator remains

$$\frac{1}{K} \sum_k \frac{n \bar{h}_k \bar{h}_k^T - \overline{h_k h_k^T}}{n-1}.$$

No positive-semidefinite truncation is applied before eigenspace selection. The optimized implementation accumulates terms involving $P_W q$ in $B^T q$ coordinates. Its arithmetic remains $O(np(K+p)+p^2K)$ and storage $O(pK+p^2+\text{batch}(p+K_b))$ for feature-block width K_b . Direct small-array expansion checks the mean, diagonal square sum and full U-matrix. The maximum optimized-versus-reference discrepancy is 2.85×10^{-13} ; three independent null engineering checks match the first implementation within 3×10^{-15} . These are ordinary float64 checks, not directed enclosures.

E.3 Finite-feature variability and nonmonotone signal

For one carrier block, write the two complex Gaussian coordinates as z_1, z_2 and let $X = |z_1|^2, Y = |z_2|^2$. Direct contraction of the block tensor gives

$$\|b_j\|^2 = \frac{9c^2}{4}XY(X+Y), \quad \|b(v)\| \leq \frac{3|c|}{4}\|P_U v\|^3.$$

When $W \supseteq U$, X, Y are independent χ_2^2 variables, with $\mathbb{E}X^k = 2^k k!$. Thus

$$\mathbb{E}\|b_j\|^2 = 72c^2, \quad \mathbb{E}\|b_j\|^4 = 54432c^4.$$

For four independent carrier blocks, $\mathbb{E}\|b\|^4 = 279936c^4$ and $\|S_W\|_F^2 = 5184c^4$. At $K = 1024$, $c = 2/\pi$, the Frobenius root mean-square error of the Gaussian-feature average is

$$\left(\frac{\mathbb{E}\|b\|^4 - \|S_W\|_F^2}{K} \right)^{1/2} = 6.63866416,$$

or 22.7503% of $\|S_W\|_F$. This is a feature-moment calculation, not an empirical subspace-error prediction. Under $M = \lambda I_m$, b scales by $\lambda^{3/2}$ and this relative error is unchanged; partial coverage alone need not cause ill-conditioning.

Nor is the complete signal Loewner-monotone in a nested work space. For one block at $c = 1$, take $M_1 = \text{diag}(1, 0, 1, 0)$ and $M_2 = \text{diag}(1, 1/3, 1, 0)$. They are realized by $W_1 = \text{span}(e_1, e_3)$ and $W_2 = W_1 + \text{span}((e_2 + \sqrt{2}e_5)/\sqrt{3})$ in $p \geq 5$. Direct contraction gives

$$S_1 = \text{diag}(27/4, 0, 27/4, 0), \quad S_2 = \text{diag}(27/4, 9/4, 6, 3).$$

$S_2 - S_1$ has eigenvalue $-3/4$, although the six-pairing contribution is monotone and the signal rank grows from two to four. The common e_3 Rayleigh quotient decreases; the full-carrier minimum eigenvalue instead increases from zero to $9/4$. These are different spectral claims.

E.4 Why an uncentered eighth-moment comparison is different

The dimension-independent bound concerns centered fourth-tensor covariance, not every nonnegative degree-eight observable. Let $\mathcal{T}_j^0$ be the one-block tensor

at $c = 1$, and define the actual harmonic quartic $f_b(Z) = \sum_{j=1}^{b_0} \langle \mathcal{T}_j^0, Z_j^{\otimes 4} \rangle$. Put $c_4 = \sin(4\pi a)/(4\pi a)$ with its continuous value at zero. Each summand equals $-3R_j^2 T_j^2 \cos(2\phi_j)$, so

$$\mathbb{E}_P f_j = 12c, \quad \mathbb{E}_P f_j^2 = 288(1 + c_4), \quad \mathbb{E}_0 f_j^2 = 288.$$

Independence and the common eighth radial moment give

$$\frac{\mathbb{E}_P f_b(q)^2}{\mathbb{E}_0 f_b(q)^2} = 1 + c_4 + \frac{(b_0 - 1)c^2}{2}.$$

This is a realizable nonnegative polynomial on the sphere, without unjustified divisions by individual pair radii. In contrast, the unnormalized centered variance ratio is $1 + c_4 - c^2/2$, with no block-count growth. Returning to $p(p+2)f_b(q)$ subtracts an additional squared-mean radial correction. Thus this example does not contradict Theorem 4.

Allowing a constant together with the quartic, the maximum eigenvalue of

$$\begin{pmatrix} 1 & c\sqrt{12b_0/\kappa_p} \\ c\sqrt{12b_0/\kappa_p} & 1 + c_4 + (b_0 - 1)c^2/2 \end{pmatrix}$$

is a valid lower bound for the unrestricted nonnegative degree-eight comparison constant. At $p = 240, b_0 = 4, a = 1/4$ it is 2.26837733; the pure-quartic-square ratio is 1.60792710. Neither number is asserted to be the optimum over all degree-eight polynomials. No such uncentered constant is needed by the public validation theorem.

E.5 What independent updates would certify

For a fixed population S_W of rank m , with carrier eigenvalues $s_1, \dots, s_m > \epsilon$, suppose a symmetric estimate satisfies $\|\widehat{G} - S_W\|_{\text{op}} \leq \epsilon$. For its top- m eigenspace $\widehat{W}$, a deterministic consequence is

$$\frac{\text{tr}(P_{\widehat{W}} P_U)}{m} \geq 1 - \frac{1}{m} \sum_{i=1}^m \left(\frac{\epsilon}{s_i - \epsilon} \right)^2.$$

To prove it, restrict $\widehat{G}$ to its invariant complementary space with projector R . Weyl gives $R\widehat{G}R \preceq \epsilon I$ there. For $S_W u_i = s_i u_i$, the equation $(s_i I - R\widehat{G}R)R u_i = -R(\widehat{G} - S_W)u_i$ gives the displayed bound by a resolvent norm at most $(s_i - \epsilon)^{-1}$. The condition requires an actual matrix error bound, including finite Gaussian features and both U-statistic terms. Using independent update data restores fixed-frame unbiasedness but does not supply those missing concentration steps. Splitting a fixed total sample budget also lowers the observations per stage, so a poorer split result alone cannot identify overfitting in the reused-data method.

E.6 Two independently frozen one-step pilots

Both panels use $p = 120, m = 16, a = 1/4$, 131,072 initial directions, a rank-32 initial frame and an independent 262,144-direction update. The data, key, initialization-probe and cubic-feature streams are new for each replicate. Code hashes, all eight stream groups, sample sizes and interpretation rules were fixed before each panel’s first new observation. The panels are local prospective pilots, not external registrations or the original $p = 240, n = 2^{20}$ same-data multipass experiment.

The first panel shares the candidate, update data and $K = 1024$ Gaussian features between the original and third-chaos scores. Timings below cover update accumulation and final eigenspace computation, excluding shared generation and initialization. Within a batch the original method runs before the third-chaos method; the order is not randomized. Mean original and third-chaos

Table 20. First pilot: every same-feature-count pair. Times are seconds.

Rep.	Original	Third chaos	Difference	Original time	Third time
1	0.512877	0.533746	+0.020869	1.3999	4.7072
2	0.496736	0.515098	+0.018362	1.4004	4.7114
3	0.454800	0.479096	+0.024295	1.4018	4.7097
4	0.524863	0.532780	+0.007917	1.3980	4.7055
5	0.494301	0.513758	+0.019457	1.4094	4.7109
6	0.485663	0.481615	-0.004048	1.4046	4.7101
7	0.464119	0.499990	+0.035871	1.3986	4.7065
8	0.445678	0.468178	+0.022499	1.4105	4.7158

overlap are 0.4848797 and 0.5030325. The paired mean is +0.0181528, sample SD 0.0118160, positive in seven of eight cases. The approximate Student 95% interval is [0.0082743, 0.0280312]. Mean update times are 1.4029 and 4.7096 seconds, so this panel cannot establish improved computation efficiency.

An algebraically equivalent implementation moves work-frame products to rank- r coordinates. A separate null-data timing calibration, performed before the second panel, measured approximately 1.44 times the original $K = 1024$ work. The second protocol therefore fixes original $K = 1536$ as the near-compute-budget comparator, while retaining original $K = 1024$ as a secondary feature-matched comparison. This is not a hard wall-clock stopping rule. No feature budget was retuned after viewing outcomes. The original $K = 1024$, original $K = 1536$ and third-chaos $K = 1024$ means are 0.4512942, 0.5058393 and 0.4712218, with mean times 1.4012, 2.0824 and 2.0396 seconds. All arms share the update data. Their accumulation order cycles across batches; the final fitting order is fixed. The first panel’s accumulation order was fixed instead. The primary paired difference is -0.0346175 , sample SD 0.0227265, negative in all eight cases, with approximate interval $[-0.0536173, -0.0156176]$. The secondary, feature-matched

Table 21. Second pilot: all eight pairs. The primary difference is third $K = 1024$ minus original $K = 1536$. Times concern these two primary arms.

Rep.	Orig. 1024	Orig. 1536	Third 1024	Primary diff.	Orig. time	Third time
1	0.511490	0.546803	0.536443	-0.010360	2.0827	2.0324
2	0.489602	0.570508	0.520254	-0.050254	2.0763	2.0282
3	0.378969	0.403708	0.396480	-0.007228	2.0837	2.0412
4	0.409752	0.444715	0.412174	-0.032541	2.0771	2.0403
5	0.459552	0.538591	0.472855	-0.065736	2.0860	2.0403
6	0.461820	0.542134	0.483068	-0.059066	2.0868	2.0537
7	0.450496	0.504625	0.466315	-0.038310	2.0792	2.0362
8	0.448671	0.495631	0.482186	-0.013445	2.0876	2.0443

difference is $+0.0199276$, positive in all eight cases, with approximate interval $[0.0115694, 0.0282858]$.

All intervals here use independent replicate groups and a Student approximation for paired differences; they are not the nonasymptotic confidence guarantee of Theorem 5. Execution-order effects and single-machine timings also limit fine runtime comparisons. The observed negative primary comparison is retained and blocks a near-compute-budget superiority claim for this variant in this configuration. It does not negate the proved noise covariance reduction. The original cubic estimator remains the main algorithm. Complete seeds, per-arm times, code hashes and protocols are in chaos-feature-control and chaos-compute-control.

F Public radial lifting and an executed NGCA comparison

The unit vector q is not an isotropic Euclidean Gaussian observation, so it cannot simply be inserted into a Gaussian Stein operator. In this angular model, however, the positive homogeneous endpoint gives $Z = Lq$ with $L \sim \chi_p$ independent of q . Publicly drawing an independent $L' \sim \chi_p$ therefore gives $L'q \stackrel{d}{=} Z$. No carrier basis or bit enters this distributional identity. It provides valid inputs for NGCA methods such as [13,17,18], subject to each method’s own assumptions.

For a symmetric trace-zero matrix A , the appropriate fourth-Hermite contraction is

$$\Psi_A(z) = (z^T A z) z z^T - (z^T A z) I - 2(A z z^T + z z^T A) + 2A.$$

Since $\mathbb{E} Z Z^T = I$, its expectation is $C(A)$. Conditional averaging over the artificial radius gives

$$\begin{aligned} \mathbb{E}_{L'}[\Psi_A(L'q) \mid q] &= p(p+2)(q^T A q) q q^T - p(q^T A q) I \\ &\quad - 2p(A q q^T + q q^T A) + 2A. \end{aligned}$$

The averages of L'^2 and L'^4 are p and $p(p+2)$, respectively. This is exactly the trace-zero control-variate summand in (2). For its vectorization, the conditional covariance decomposition proves that this analytic average has no larger covariance than the particular estimator using the sampled artificial radius. It does not show global minimum variance among NGCA methods or establish a runtime or recovery-factor advantage.

The comparison motivates the executed MIPP baseline in Appendix F.1: it uses the same public observations and charges radial lifting and repeated passes explicitly. For a cubic scalar test $h(z) = (v^T z)^3$, the Stein vector is $\mathbb{E}[Zh(Z) - \nabla h(Z)] = \mathcal{T}(\cdot, v, v, v)$; thus the adaptive cubic learner is also an NGCA-type construction. The conditional variance ordering above concerns one specified polynomial estimator, not the non-polynomial MIPP comparator. ICA methods assuming independent scalar coordinates have stronger structural assumptions than our independent four-dimensional blocks and are secondary empirical controls. The full-carrier result of [16, Theorem 9] assumes a non-Gaussian moment in every carrier direction, whereas our individual Gaussian pairs violate that condition. This prevents a direct transfer of its full-recovery guarantee; it does not prove that its algorithms fail in every concrete instance. The executed comparison does not include IMAK or a log-density-gradient estimator, and does not order the performance of these method families.

Stein moments and hard support. The angular law has independent standard Gaussian coordinates on $U^\perp$. For every $v \in U^\perp$, Gaussian integration by parts along that coordinate gives $\mathbb{E}[(v^T Z)h(Z)] = \mathbb{E}[\partial_v h(Z)]$ for smooth tests with integrable terms. Consequently $\mathbb{E}[Zh(Z) - \nabla h(Z)] \in U$. This argument integrates only across the Gaussian complement and does not require differentiating the carrier density across its boundary. It proves the population subspace inclusion used by MIPP, not a finite-dictionary span guarantee or convergence of empirical adaptation.

For $0 < a < 1/2$, the support identity (77) and the independent radius imply the lifted density

$$p_F(z) = r_a^{-1} \phi_p(z) \mathbf{1}_{C_F}(z), \quad r_a = (2a)^{m/4}, \quad C_F = \{z \neq 0 : z/\|z\| \in E_F\}.$$

Its classical interior score equals $-z$ almost surely; hence the classical vector $\nabla \log p_F(z) + z$ is zero almost surely. The distributional derivative of the density still has a boundary term. Smooth-score NGCA guarantees such as those motivating [18] therefore require a separate regularity check here. This is not an empirical failure result for finite-basis LSNCA: a regularized finite-basis objective can respond to boundary structure. It also does not invalidate the complement-wise Stein identity above.

F.1 The fixed-threshold non-polynomial NGCA comparison

This subsection reports the fixed-threshold comparison and its implementation controls. It is separate from the eight-candidate validation family and does not

change its coverage allocation. The reference is the MIPP/NGCA procedure of [13], especially its Stein vectors, variance normalization and fixed-number direction adaptation. Our streaming code is a budgeted reimplementaion, not author-supplied code. The original 4,000-function budget, $\varepsilon = 1.5$ and $T = 10$ are specified in its Section 4, page 262. Section 3.2, page 255, recommends calibrating retention on independent Gaussian data, for example by a 95th percentile; Section 4 explains that 1.5 came from preliminary calibration and that function-specific calibration would be preferable. We borrow that numerical value without recalibrating for our dimension, sample size, dictionary, empirical whitening or same-data iterations. Its null quantile in this pipeline is unmeasured. Trace normalization does not make the Stein-vector covariance spherical, and adaptive directions are not independent of their fitting sample. Neither a χ_p^2/p approximation nor a causal attribution of the rank failures to calibration is established here. The zero-threshold arm is a prespecified sensitivity, not an estimated Gaussian cutoff.

Explicit finite-sample procedure. Let $Z_i = L_i q_i$, where $L_i \sim \chi_p$ are independent public draws replayed identically on every pass. Let $\hat{\mu}$ and $\hat{\Sigma} = n^{-1} \sum_i (Z_i - \hat{\mu})(Z_i - \hat{\mu})^T$ be the empirical mean and covariance, and $Y_i = \hat{\Sigma}^{-1/2}(Z_i - \hat{\mu})$. For each unit direction w and scalar function f , compute

$$b_i(w, f) = Y_i f(w^T Y_i) - w f'(w^T Y_i), \quad \hat{\beta} = \frac{1}{n} \sum_i b_i, \quad N = \frac{1}{n} \sum_i \|b_i\|^2 - \|\hat{\beta}\|^2.$$

Starting from independent normalized Gaussian directions in whitened coordinates, apply $w \leftarrow \hat{\beta}/\|\hat{\beta}\|$ ten times and then recompute the final $\hat{\beta}, N$. This follows the extra final application of (8) described in the reference’s Section 3.3, page 257; Figure 4 on page 259 instead stores the final moment at w_{T-1} . There is no convergence-based stopping or deflation. Following the variance normalization in that Figure 4, use $v = \sqrt{n/N} \hat{\beta}$ when $N > 0$, retain $\|v\| \geq \varepsilon$, and take the top m eigenvectors J of $\sum v v^T$. The reported Euclidean frame is $\text{qr}(\hat{\Sigma}^{-1/2} J)$. Using the positive square root here would map the wrong index space. Neither empirical N nor the retention threshold is a confidence certificate for directions adapted on these same data.

The dictionary has four families with 16 deterministic parameters and four random initializations each:

$$s^3 e^{-s^2/(2\sigma^2)}, \quad \tanh(bs), \quad \sin(as), \quad \cos(as),$$

where $\sigma^2 = 0.5 + 4.5(j - 1)/15$, $b = 5j/16$ and $a = j/4$, $j = 1, \dots, 16$. Zero constant/linear limits are omitted. The main threshold is $\varepsilon = 1.5$; the prespecified sensitivity uses $\varepsilon = 0$ on the same final moments. The $T = 0$ ablation uses identical initial directions, parameters, empirical whitening and radii, with no updates. Cosine’s population Stein moment is zero under exact central symmetry; empirical adaptation can still generate a nonzero sample moment. All four families are retained in the experiment.

A covariance eigenvalue below 10^{-12} times the largest causes explicit failure; no ridge is introduced. Update norms at most 10^{-14} leave the previous direction unchanged and are counted; nonfinite updates fail. Invalid variance estimates are discarded. Fewer than m retained independent directions gives *insufficient retained rank*, with no arbitrary tail completion. Every retained count, effective rank, spectrum and threshold status is saved. Failed rank- m fits have no rank- m overlap; they are not assigned zero or omitted from the denominator when reporting success.

Observations, costs and prospective binding. The eight data seeds are 32101–32108. The complete generation, learner and lift seeds, all parameter values, method order and analysis rules are in ngca-fixed-protocol. Its transitive code/input hashes were recorded in ngca-fixed-binding before generating the first full-size array. This is a local prospective freeze, not externally timestamped preregistration. Public configurations supplied to the isolated learner processes omit the generation and key seeds. Hidden-carrier overlap is computed by the controller only after each worker has saved its public-data fit. No full-panel outcome changes the parameters, seeds or stopping rule.

Each group generates one float64 read-only memory-mapped array of $2^{20} \times 240$ public unit directions. All three methods read that identical array; MIPP verifies its SHA256 on every pass and uses the same per-row radial draw on every replay. Array hashes and generator seeds are retained; the large temporary arrays are removed after that group’s three fits. Each method runs sequentially in a separate process with numerical thread limits of two. Method order cycles across groups; eight groups do not give perfectly balanced orders. Times include initialization/whitening, public lifting, repeated reads, checks, nonlinear evaluation and eigensolves, and exclude shared data generation, oracle scoring and saving. Peak process RSS includes resident pages of the memory-mapped observations. These are fresh timings under this two-thread worker protocol; the equal-feature-budget times in Table 4 are separate measurements and should not be used to infer a cross-panel speedup.

Adaptive cubic uses four passes, 24 matrix probes and three blocks of 1024 cubic features. MIPP uses twelve passes and $2816n$ scalar-function evaluations, plus the same number of derivative evaluations. The $T = 0$ ablation uses two passes and $256n$ of each. Threshold variants reuse the same fit and share its recorded cost. This is a comparison at equal unique observations; neither feature count nor measured time is constrained to match. The adaptive schedule was selected using oracle-scored development seeds, while MIPP’s budgeted parameter choices have not received an equivalent tuning history.

All primary and sensitivity outcomes. The original threshold returns 0/8 rank-16 MIPP fits and 0/8 fixed-direction fits; insufficient retained rank is a reported outcome, not replaced by arbitrary directions. At the prespecified zero threshold, all eight fits exist: MIPP’s mean overlap is 0.1825, versus 0.6519 for adaptive cubic and 0.0649 for fixed directions. The adaptive-minus-MIPP paired

Table 22. Fixed-threshold paired NGCA panel at $p = 240$, $n = 2^{20}$. Counts refer to returned rank-16 frames; 0/8 means insufficient retained rank in every attempt. Overlap summaries use returned frames only. The borrowed threshold 1.5 was not recalibrated; 0 is a prespecified sensitivity. Times average eight attempts under a two-thread worker protocol, include preprocessing and are shared by threshold variants.

Method	ε	Rank-16 frames		Overlap mean (SD)	Seconds
		returned			
Adaptive cubic	—	8/8	0.6519 (0.0913)	42.8	
MIPP, $T = 10$	1.5	0/8	—	56.2	
MIPP, $T = 10$	0	8/8	0.1825 (0.0843)	56.2	
MIPP, $T = 0$	1.5	0/8	—	7.5	
MIPP, $T = 0$	0	8/8	0.0649 (0.0059)	7.5	

difference averages 0.4694 (sample SD 0.0976), positive in 8/8 groups. This comparison concerns the declared zero-threshold sensitivity, not successful recovery at the original threshold. The primary-threshold retained counts range from 0

Table 23. Every fixed-threshold-panel seed. MIPP columns show overlap at $\varepsilon = 0$ and the retained count R at the primary threshold 1.5. Primary success status and every threshold output are separately retained in the raw records.

Seed	Adaptive	MIPP $T = 10$	R_{10}	MIPP $T = 0$	R_0
32101	0.518204	0.062956	0	0.072882	0
32102	0.683739	0.348744	5	0.058351	0
32103	0.602707	0.112847	1	0.067041	0
32104	0.653351	0.158351	1	0.072808	0
32105	0.583130	0.227782	3	0.063647	0
32106	0.743054	0.194390	3	0.062748	0
32107	0.626492	0.185074	1	0.056772	0
32108	0.804525	0.170202	2	0.064776	0

to 5 for $T = 10$, and 0 to 0 for $T = 0$. The unthresholded MIPP overlap range is $[0.062956, 0.348744]$, compared with $[0.518204, 0.804525]$ for adaptive cubic. The eight zero-threshold paired differences range from 0.334995 to 0.634324. The public covariance deviation $\|\hat{\Sigma} - I\|_{\text{op}}$ ranges from 0.029799 to 0.031261. All saved frames have been independently rescored from the frozen generator keys; raw and regenerated overlap values agree within 10^{-12} . All successful rank- m frames are scored. Descriptive means and sample standard deviations have their success denominators stated; an eight-seed paired comparison is reported only when all eight paired frames exist. There is no post-hoc best-threshold selection, significance test or general performance ordering of NGCA families.

Table 24. Per-seed total fit seconds and peak worker RSS in GiB. RSS includes the memory-mapped observation pages; it is not an estimate of feature-only workspace.

Seed	Fit seconds			Peak RSS (GiB)		
	Adaptive	MIPP 10	MIPP 0	Adaptive	MIPP 10	MIPP 0
32101	41.76	56.25	7.48	2.003	1.934	1.930
32102	42.00	56.28	7.49	2.001	1.933	1.932
32103	42.75	56.23	7.45	2.001	1.933	1.930
32104	42.80	56.24	7.45	2.001	1.933	1.930
32105	42.78	56.30	7.44	2.001	1.933	1.930
32106	43.33	56.31	7.43	2.003	1.937	1.930
32107	43.44	56.09	7.47	2.001	1.933	1.930
32108	43.67	56.16	7.43	2.001	1.933	1.930

Independent implementation controls. Six groups of independent checks cover functions and derivatives; Stein moments and variances; non-diagonal covariance pullback; rescaling and rank failure; batch replay; and four-dimensional angular quadrature. The quadrature verifies a nonzero sine moment on a mixed carrier direction and zero Gaussian/cosine controls. Maximum derivative error was $1.10 \cdot 10^{-10}$, covariance-pullback projector error $1.05 \cdot 10^{-15}$, and the two angular quadratures differed by $9.77 \cdot 10^{-16}$. These are float64 checks, not directed interval proofs.

Before freezing the main panel, a same-array three-worker integration check used $p = 80, n = 512$. A further fixed positive control used $p = 60, m = 16, n = 32768, a = 1/4$ and the same MIPP settings: its prespecified zero-threshold overlap was 0.734739 against random expectation 0.266667. The original threshold retained 23 vectors; post-hoc scoring of its saved rank-16 frame gives overlap 0.732895. This additional score uses the existing frame and is not new training. The frozen summary’s `oracle_overlap` belongs to `sensitivity_0_frame`; `primary_retained=23` belongs to `primary_frame`. Their mapping and source hashes are in `ngca-positive-control`; the original summary and saved frames remain unchanged. These checks did not select or modify the main protocol and are not pooled with its performance results. Code, raw diagnostics and formulas are in `ngca-fixed-estimator`, `ngca-fixed-controller` and `ngca-fixed-checks`.

Reproduction and replay controls. The fixed-threshold tables use the frozen implementations `ngca-fixed-estimator` and `ngca-fixed-controller`. Their NumPy version and observation batches are recorded with every fit; reproduction of this panel uses those implementations. The replay-controlled implementation `ngca-replay-estimator` uses explicit PCG64 generation in fixed blocks of 4096 radii, independently of observation-batch cuts. Each pass checks both public-direction and radius digests. This fixes random-call shapes within a specified software environment; bitwise agreement across software environments is not implied. Its controller `ngca-replay-controller` checks an explicit public-input schema before reading observations, including under Python’s optimization

flag. Each execution must bind the implementation files actually used. Optional MIPP counts must be integers (iterations may be zero), tolerances finite and strictly positive, and thresholds finite and nonnegative; booleans are not numeric configuration values. Invalid optional fields are rejected before observations are opened in both execution modes. Small-sample controls cover radius-block boundaries, changed-stream rejection, fitted-subspace agreement up to rounding with the direct replay comparator, and rejection of undeclared fields in normal and optimized execution. The records in ngca-replay-checks are implementation checks, not main-scale performance measurements. The fixed-threshold configurations omit undeclared secret fields, but their Python optimization flag was not recorded; normal-mode execution cannot be established from those records. The Gaussian-calibrated panel explicitly records this flag and rejects extra fields in either mode. The fixed-threshold feature seeds are 332101, 332201, ..., 332801: each three-stage refinement uses its base plus 0, 1, 2, so the stride 100 leaves disjoint stage ranges. These are the actual executed seeds.

F.2 Independent calibration and a shared initialization control

The Gaussian-calibrated panel measures a null cutoff and controls initial public information. It uses independent seeds and a separately frozen implementation; its results are not pooled with the fixed-threshold training panel or the validation panel. The carrier dimension, angular parameter, observation count, dictionary and ten updates remain $p = 240, m = 16, a = 1/4, n = 2^{20}$, 256 function/start pairs and $T = 10$.

Prospective calibration and test protocol. The local pre-execution record ngca-calibrated-binding binds the transitive code, full protocol and all seeds before full-size data generation. This is local prospective binding, not independently timestamped preregistration. Data seeds are 62501–62504 for Gaussian calibration, 72501–72502 for Gaussian holdout diagnostics, and 32501–32508 for angular tests. They are disjoint from the fixed-threshold 32101–32108 NGCA panel and the eight-candidate validation family. The full production, key, probe, feature, direction and radial-lift seeds are in ngca-calibrated-protocol; only the declared public learner inputs enter worker configurations.

Each null sample consists of normalized independent standard Gaussian vectors. An independent χ_p lift therefore has the standard Gaussian law. Both null arms execute empirical centering/whitening, ten updates and the final moment calculation on that same lifted sample. The spectral arm also estimates its public initializer on that null sample. For each arm, pool the $4 \cdot 256$ final variance-normalized Stein-vector norms and take the empirical 95th percentile with linear interpolation. The file ngca-thresholds records both resulting thresholds and the calibration-record hashes before generating either holdout or angular test data. No oracle score or test outcome selects a threshold. Two further independent null groups measure retained-vector fractions and rank-return statuses descriptively. Heterogeneous functions and dependent same-data updates preclude interpreting these fractions as independent Bernoulli trials or as a proved 5% error rate.

Matched public spectral initialization. Let B_0 be the orthonormal rank-64 frame from the same 24 GOE probes and same directions used by the adaptive cubic initializer. In raw coordinates, B_0 spans covectors. If $W = \widehat{\Sigma}^{-1/2}$ is empirical whitening, form $J_0 = \text{qr}(W^{-1}B_0)$, project the shared standard Gaussian initial-direction matrix G onto J_0 , and normalize its columns:

$$w_{0,j} = \frac{J_0 J_0^T G_j}{\|J_0 J_0^T G_j\|}.$$

Then the raw covectors $Ww_{0,j}$ lie in $\text{range}(B_0)$. Using WB_0 instead would generally select the wrong covector span. The ambient arm normalizes those same G_j without projection. Subsequent updates use the full whitened coordinates in both arms; this is an initialization control, not permanently restricted optimization. The spectral arm recomputes B_0 in its own worker, with its own fitting cost included. Saved frames verify agreement with the adaptive initializer. No hidden carrier enters initialization or fitting.

Complete records and cost accounting. A single public float64 memory-mapped array is generated per group and removed after that group’s sequential fits. Workers use two numerical threads; order rotates across groups. Every record includes software versions, actual Python optimization mode, public configuration hash, observation and radius digests, sample passes, full spectrum, retained count, rank status, elapsed fitting time and process peak RSS. Configuration validation precedes observations and uses explicit exceptions in both normal and optimized Python. The controller scores hidden overlap only after each public-data fit has completed. Failed rank-16 fits have no rank-16 overlap and remain in every success denominator.

The fixed-threshold archives contain frames without the full final moments. The Gaussian-calibrated MIPP archives retain $\widehat{\beta}$, variance estimates, empirical mean and whitening, final and initial directions, dictionary parameters and family identifiers, sample size and target rank. They therefore support independently checked threshold reaggregation. The primary Gaussian-calibrated result and the prespecified zero-threshold sensitivity reuse these identical moments; no extra fit is charged for the sensitivity. Calibration and holdout costs are reported separately from the eight training-group times. This protocol does not impose equal time, equal feature counts or symmetric hyperparameter search. Four null calibration samples are limited empirical calibration, not a theorem or a substitute for function-specific tuning.

Calibration rows report retained counts at the derived cutoff, while their archived fit used zero threshold to preserve every moment; no rank decision at that derived cutoff is claimed for those rows. Each arm pools 1024 norms and retains 52 above its own empirical cutoff. The two holdout datasets are not a precise assessment of extreme-tail or familywise error. The ambient holdouts retain 12 and 11 vectors, a pooled fraction 0.044922, and return 0/2 rank-16 frames. The spectral64 holdouts retain 9 and 15 vectors, a pooled fraction 0.046875, and return 0/2 rank-16 frames. Calibration, holdout and angular-test fitting totals

Table 25. All Gaussian calibration and independent holdout diagnostics. Calibration norms use the full same-data fitting pipeline; thresholds are fixed before holdout generation. Retained fractions are descriptive.

Phase/seed	Initialization	Cutoff	Retained/256	Rank-16	Seconds
Cal. 62501	ambient	1.045613	15	n/a	57.6
Cal. 62501	spectral64	1.040444	16	n/a	76.4
Cal. 62502	ambient	1.045613	11	n/a	57.6
Cal. 62502	spectral64	1.040444	15	n/a	76.6
Cal. 62503	ambient	1.045613	18	n/a	57.6
Cal. 62503	spectral64	1.040444	15	n/a	77.4
Cal. 62504	ambient	1.045613	8	n/a	57.6
Cal. 62504	spectral64	1.040444	6	n/a	77.1
Hold. 72501	ambient	1.045613	12	no	57.6
Hold. 72501	spectral64	1.040444	9	no	77.3
Hold. 72502	ambient	1.045613	11	no	57.6
Hold. 72502	spectral64	1.040444	15	no	77.6

Table 26. All eight angular outcomes: Adaptive cubic primary. Failed rank-16 fits have no overlap.

Seed	Retained	Effective rank	Overlap	Seconds
32501	–	–	0.555079	46.16
32502	–	–	0.742941	47.69
32503	–	–	0.702796	48.08
32504	–	–	0.800231	47.35
32505	–	–	0.633411	46.97
32506	–	–	0.690481	47.44
32507	–	–	0.693555	47.91
32508	–	–	0.755661	47.73

Table 27. All eight angular outcomes: MIPP ambient primary. Failed rank-16 fits have no overlap.

Seed	Retained	Effective rank	Overlap	Seconds
32501	18	18	0.125414	57.54
32502	23	23	0.359319	57.62
32503	19	19	0.268309	57.68
32504	10	10	–	57.62
32505	13	13	–	57.57
32506	19	19	0.289962	57.51
32507	12	12	–	57.54
32508	10	10	–	57.53

Table 28. All eight angular outcomes: MIPP spectral64 primary. Failed rank-16 fits have no overlap.

Seed	Retained	Effective rank	Overlap	Seconds
32501	20	20	0.286593	77.91
32502	17	17	0.341039	77.73
32503	17	17	0.329481	78.10
32504	21	21	0.408633	78.36
32505	19	19	0.398057	78.06
32506	27	27	0.691104	78.17
32507	17	17	0.341359	78.07
32508	23	23	0.530851	78.05

Table 29. All eight angular outcomes: MIPP ambient zero-threshold sensitivity. Failed rank-16 fits have no overlap.

Seed	Retained	Effective rank	Overlap	Seconds
32501	256	240	0.109656	57.54
32502	256	240	0.361949	57.62
32503	256	240	0.254830	57.68
32504	256	240	0.122539	57.62
32505	256	240	0.177828	57.57
32506	256	240	0.276185	57.51
32507	256	240	0.120764	57.54
32508	256	240	0.218869	57.53

Table 30. All eight angular outcomes: MIPP spectral64 zero-threshold sensitivity. Failed rank-16 fits have no overlap.

Seed	Retained	Effective rank	Overlap	Seconds
32501	256	240	0.295305	77.91
32502	256	240	0.326721	77.73
32503	256	240	0.332592	78.10
32504	256	240	0.400376	78.36
32505	256	240	0.371488	78.06
32506	256	240	0.693799	78.17
32507	256	240	0.335158	78.07
32508	256	240	0.503956	78.05

are respectively 538.0 s (8 fits), 269.9 s (4 fits), 1464.4 s (24 fits). These exclude common data generation, saving and oracle scoring. Shared initialization changes only the starting public span; dictionary size, same-data adaptation and unequal development effort remain. The complete archive includes every weak result and does not convert a high overlap into a convergence theorem.

Implementation and analysis checks. The Gaussian-calibrated implementation `ngca-calibrated-estimator` uses the fixed-block replay controls above and adds the shared spectral initializer. The fixed-threshold implementation remains separately archived. Small-sample controls compare unchanged ambient fits, verify the initial raw covector span, independently reconstruct aggregation with an SVD, check insufficient rank and invalid public inputs, and exercise all three workers in normal and optimized Python. These checks are in `ngca-calibrated-checks` and its optimized counterpart. The post-execution analysis reaggregates every saved MIPP archive, checks all eight shared spectral initializers and binds each table to the complete results. These are implementation checks and descriptive experiments, not directed-rounding enclosures or proofs of adaptive convergence.