The missing basis was not the mechanism

ReLU-kernel-optimal angular quadrature and gate-support-preserving continuation for depth-32 means

ely2sh · ARC White-Box Estimation Challenge 2026 · Phase 1

GRADED ARTIFACT	ADJUSTED SCORE	RAW FINAL MSE	COMPLETION
#327749	1.08164×10^{-7}	2.30406×10^{-7}	**50/50 public rows**

Abstract

We estimate the final-layer mean of a width-256, depth-32, bias-free ReLU MLP under a 256-dimensional standard Gaussian input. The graded estimator combines exact radial conditioning, a 65,536-point antipodal rule formed from 128 flat mutually unbiased orthonormal bases, an exact first-layer ReLU-mean correction, a two-basis gate-support pilot, and fully metered compact continuation.

The carrier was initially justified incorrectly as a tight spherical 5-design: one coordinate basis is missing. Correcting that mistake led to a stronger, ReLU-specific result. For the normalized infinite-width He-ReLU kernel at every positive depth, we prove that pairwise mutually unbiased bases minimize expected antipodal quadrature MSE among all equal-weight unions of the same number of orthonormal bases. At depth 32 the theorem predicts a 1.00930× gain from adding the missing basis, versus 1.00907× on an untouched 200-network lockbox. It predicts the complete ordering MUB < Haar ≈ randomized-flat < Owen-Sobol; the finite-width study reproduces that ordering while amplifying the non-completion gaps.

The second contribution concerns computation after sampling. A 1,024-path pilot retained observed gate support through the suffix, removing about **8.3%** of effective compute with only **0.07%** raw-MSE change. More aggressive support removal exhibits a sharp error cliff. An exact Euler-Stein facet identity explains why path-moving moment compression is hazardous. The theorem is exact for plain antipodal NNGP quadrature; the finite-width, exact-H1-corrected experiment is a transfer test, not part of the theorem.

Central contribution. Mutual unbiasedness minimizes antipodal random-ReLU kernel redundancy at every depth; the resulting finite carrier becomes computationally useful when suffix work is removed only where observed gate support certifies it absent.

Claim boundary. We do not claim invention of Kerdock codes, MUBs, antithetic sampling, exact first-layer means, fast multiplication, or inactive-unit pruning. We prove one within-class kernel theorem, test its finite-width predictions, and connect it to one successfully graded artifact.

1. What needed explaining

For fixed weights $W_1, \dots, W_{32}$, let $f_W : \mathbb{R}^{256} \to \mathbb{R}^{256}$ be the bias-free ReLU MLP. The target is

$$\mu_W = \mathbb{E}_{X \sim \mathcal{N}(0, I_{256})}[f_W(X)]. \tag{1}$$

Three observations demanded one explanation:

1. In a matched full-width study, the shipped 128-basis carrier beat equal-path Haar-basis, Owen-scrambled Sobol, and randomized-flat controls by 1.286–1.391× corrected MSE.
2. Adding the one basis that makes the carrier an exact spherical 5-design improved lockbox MSE by only 1.0091×, nearly its 1.0078× path increase.
3. In the graded route, the statistically stronger carrier became useful only after a gate-support pilot removed suffix work without replacing nonlinear trajectories.

The first draft connected these facts through fourth frame potential and a plausible activation-facet story. That was incomplete. Quartic optimality is generic design theory, while the facet identity does not imply that MUBs minimize boundary discrepancy. The correct division is sharper: a *random-ReLU kernel theorem* explains the angular carrier; a *gate-support argument and measured cost/error ladder* explain downstream compression.

1.1 Score and evidence standard

For row i with final-row MSE M_i, analytical FLOPs F_i, and residual time R_i, the score is

$$S = \frac{1}{M} \sum_{i=1}^{M} M_i \max\left(0.1, \frac{F_i + 10^{11} R_i}{272 \times 10^9}\right). \tag{2}$$

Live results use this rowwise product. Post-phase experiments report paired network statistics and crossed uncertainty. Exact results, measurements, and interpretations are kept separate throughout.

2. ReLU-kernel mechanism

2.1 From random networks to kernel discrepancy

For normalized He-ReLU propagation on the unit sphere, the infinite-width correlation recursion is [2, 3, 4]

$$T(\rho) = \frac{\sqrt{1-\rho^2} + (\pi - \arccos \rho)\rho}{\pi}, \qquad K_L = T^{\circ L}. \tag{3}$$

Let $P = \{\pm u_1, \dots, \pm u_N\}$ be an equal-weight antipodal rule and U, V be independent uniform sphere points. For one scalar random-network output, the expected squared angular error is

$$\mathcal{D}_L(P) = \mathbb{E}_W\left[\left(\frac{1}{2N} \sum_{i=1}^{N} \{f_W(u_i) + f_W(-u_i)\} - \mathbb{E}_U f_W(U)\right)^2\right]. \tag{4}$$

Expanding the square and using rotational invariance gives

$$\mathcal{D}_L(P) = \frac{1}{N^2} \sum_{i,j=1}^{N} J_L(u_i^\top u_j) - q_L, \tag{5}$$

where

$$J_L(\rho) = \frac{K_L(\rho) + K_L(-\rho)}{2}, \qquad q_L = \mathbb{E}_{U,V} K_L(U^\top V). \tag{6}$$

Only the second-moment kernel is required; final ReLU outputs need not themselves form a Gaussian process. Exact radial conditioning multiplies every risk by $(\mathbb{E}R)^2$, preserving all ratios.

2.2 MUBs minimize risk at every depth

The normalized ReLU map has the convergent series

$$T(\rho) = \frac{1}{\pi} + \frac{1}{2}\rho + \frac{1}{\pi} \sum_{k\geq 1} \frac{\binom{2k-2}{k-1}}{4^{k-1}(2k)(2k-1)} \rho^{2k}. \tag{7}$$

Every coefficient is nonnegative. Composition preserves this property; antipodal evenization removes odd terms:

$$J_L(\sqrt{s}) = \sum_{m\geq 0} a_{2m} s^m, \qquad a_{2m} \geq 0. \tag{8}$$

At every positive depth, the coefficient of s^2 is positive, hence this function is strictly convex on $[0,1]$.

Theorem 1

MUB optimality for antipodal ReLU-kernel quadrature. Let P be the equal-weight antipodal union of $B \geq 2$ orthonormal bases in $\mathbb{R}^d$. For every positive depth L, its infinite-width He-ReLU quadrature risk $\mathcal{D}_L(P)$ is minimized exactly when every pair of bases is mutually unbiased, whenever such a family exists:

$$|q_{bi}^\top q_{cj}|^2 = \frac{1}{d} \quad \text{for all } b \neq c \text{ and all } i, j.$$

Proof. Within each basis, all kernel arguments are fixed. For two distinct bases Q, R, their cross-Gram matrix $G = Q^\top R$ is orthogonal, so

$$\frac{1}{d^2} \sum_{i,j} G_{ij}^2 = \frac{1}{d}. \tag{9}$$

Strict convexity and Jensen's inequality yield

$$\frac{1}{d^2} \sum_{i,j} J_L(G_{ij}) \geq J_L(1/\sqrt{d}). \tag{10}$$

Equality holds exactly when every $G_{ij}^2 = 1/d$. The full kernel energy is a fixed within-basis contribution plus the sum over cross-basis pairs; q_L is point-set independent. Thus pairwise MUBs minimize the complete risk. ∎

The fourth-frame-potential bound is now a corollary, not the paper's center: it isolates one positive even coefficient of the same ReLU kernel, whereas Theorem 1 controls the entire nonnegative mixture at every depth.

2.3 Quantitative predictions

We computed q_L by 768-node Gauss-Jacobi quadrature. MUB energies use the exact inner-product spectrum. Haar is an exact ensemble expectation. Randomized-flat energies are exact for all 16 fixed sign realizations. Owen-Sobol uses 1,048,576 fixed off-diagonal pairs per scramble with an exact second-moment control; its depth-32 gain has mean Monte Carlo standard error 0.0079.

At depth 32, the normalized angular discrepancy is 2.45629×10^{-7} for MUB-128 and 2.43366×10^{-7} for MUB-129. The predicted completion gain is 1.0092977×; the untouched lockbox gives 1.0090699×. No network weights or truth values enter this prediction.

The broader fit is directional, not exact. Predicted versus measured lockbox gains are completion 1.0093/1.0091, Haar 1.1869/1.2864, random-flat 1.1884/1.2896, and Sobol 1.2355/1.3582. The model reproduces the complete ordering but understates non-completion gaps by roughly 8–10% in gain. Finite width, exact-H1 correction, and higher joint structure may contribute to the residual.

3. Controlled finite-width study

3.1 Multi-realization design

A single global orientation was not trustworthy: one network's MSE varied by roughly 2.9–5.4× across ten rotations, while split-half rank correlation of rotation rankings was −0.358. Arms, seeds, estimands, bootstrap, and gates were fixed before opening protected truth.

- 200 validation networks (indices 600–799);
- an untouched 200-network lockbox (indices 800–999);
- 16 global realizations per arm, or 3,200 network-realization evaluations per partition;
- identical full-width continuation, exact radius, antipodes, and exact H1 fusion in every arm.

The same realization is shared across networks, so intervals use 100,000 crossed bootstrap resamples of networks and global draws. Draw 0 contains the submitted alignment; analyses are repeated without it.

3.2 Matched arms

The fixed arms were MUB-128; paired MUB-129 completion; 128 independent Haar bases; 32,768 Owen-scrambled Sobol sphere points; and 128 independently signed Walsh-Hadamard bases. MUB, Haar, and random-flat all have 32,768 rays, exact finite second moment $I/256$, and complete orthonormal blocks. Random-flat preserves coordinate flatness but disperses cross-basis inner products.

The flat control changes both mutual unbiasedness and code structure, so by itself it identifies only a family effect. Theorem 1 supplies the isolation: mutual unbiasedness is sufficient to minimize prior-average ReLU-kernel risk among all basis unions in this class.

3.3 Protected results

The predeclared gate required at least 1.10× conservative gain, a bootstrap lower endpoint above one, majority network

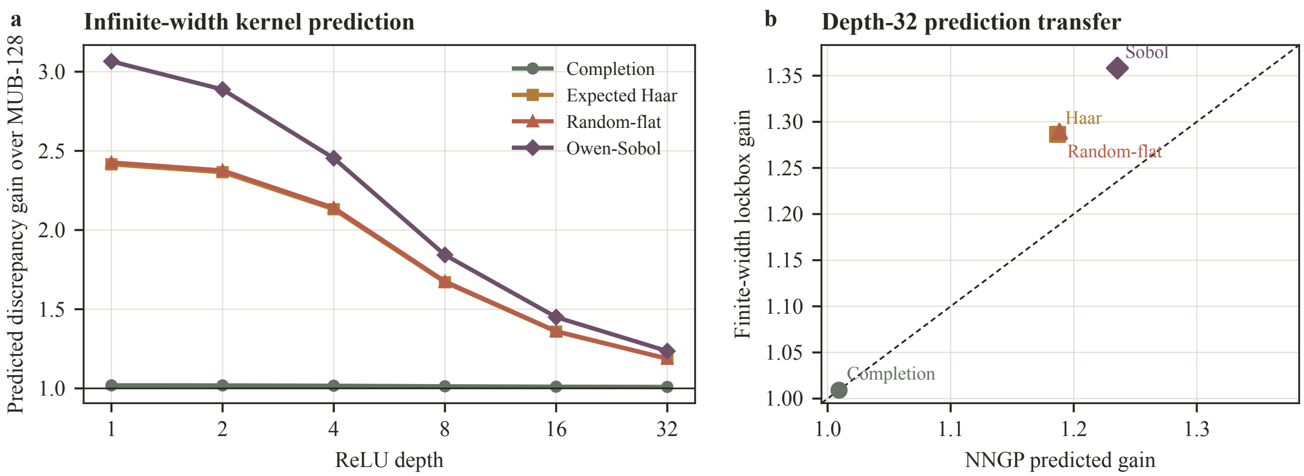

Figure 1. ReLU-kernel mechanism. (a) Exact or fixed-randomization discrepancy ratios across depth. (b) Depth-32 prediction versus the untouched finite-width lockbox. The theorem reproduces the ordering and nearly exactly predicts the missing-basis effect; width 256 and exact-H1 correction amplify the three non-completion gaps.

Table 1. ReLU-kernel discrepancy gains over MUB-128. Basis and MUB values are exact; Sobol reports the mean across 16 fixed scrambles.

ReLU depth	Completion/MUB	Expected Haar/MUB	Random-flat/MUB	Owen-Sobol/MUB
1	1.0192	2.4150	2.4266	3.0654
2	1.0188	2.3652	2.3764	2.8878
4	1.0169	2.1312	2.1405	2.4536
8	1.0132	1.6701	1.6756	1.8427
16	1.0107	1.3584	1.3614	1.4500
32	**1.00930**	**1.18688**	**1.18843**	**1.23553**

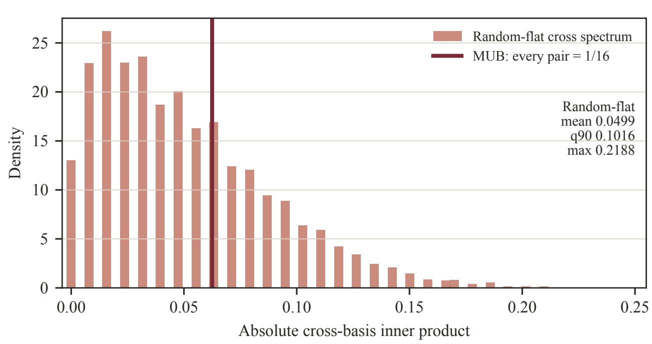

Figure 2. Cross-basis geometry. Both families contain 128 flat orthonormal blocks. MUB inner products are fixed at 1/16; the randomized-flat spectrum is dispersed. Convexity of $J_L(\sqrt{s})$ prices that dispersion.

wins, and non-regressing q90/q95. All equal-path comparisons passed on both partitions.

Validation gains were 1.0128, 1.3307, 1.3911, and 1.3488× in the same order. Excluding draw 0 leaves every comparator lower endpoint above 1.21 on lockbox. Deleting any one realization changes point gains only within 1.2774–1.2950 (Haar), 1.3498–1.3696 (Sobol), and 1.2731–1.2994 (flat).

3.4 The missing basis falsifies the 5-design explanation

The stored asset contains 128 flat pairwise MUBs. The complete real construction in dimension 256 contains 129 bases [5, 6]; adding its coordinate basis changes signed paths from 65,536 to 66,048, a factor 1.0078125, and makes the antipodal rule a tight spherical 5-design [7, 8].

Path-normalized completion gains are only 1.0050× on validation and 1.0012× on lockbox. Degree-five exactness supplies almost no efficiency. The kernel theorem explains why: MUB-128 already minimizes the relevant pair energy for 128 bases; completion mostly adds 0.78% more optimal paths.

4. The graded estimator

The theorem explains one ingredient of #327749; it does not claim that the artifact executes an NNGP calculation. The submitted route evaluates the actual finite network.

4.1 Exact radial Rao-Blackwellization

Write $X = RU$, where U is uniform on the sphere and $R \sim \chi_{256}$ is independent. Positive homogeneity gives

$$\mathbb{E} f_W(X) = \mathbb{E}R\, \mathbb{E} f_W(U), \qquad \mathbb{E}R = \sqrt{2}\,\frac{\Gamma(257/2)}{\Gamma(256/2)}. \tag{11}$$

The estimator samples no radii; every direction is scaled by the exact mean chi radius.

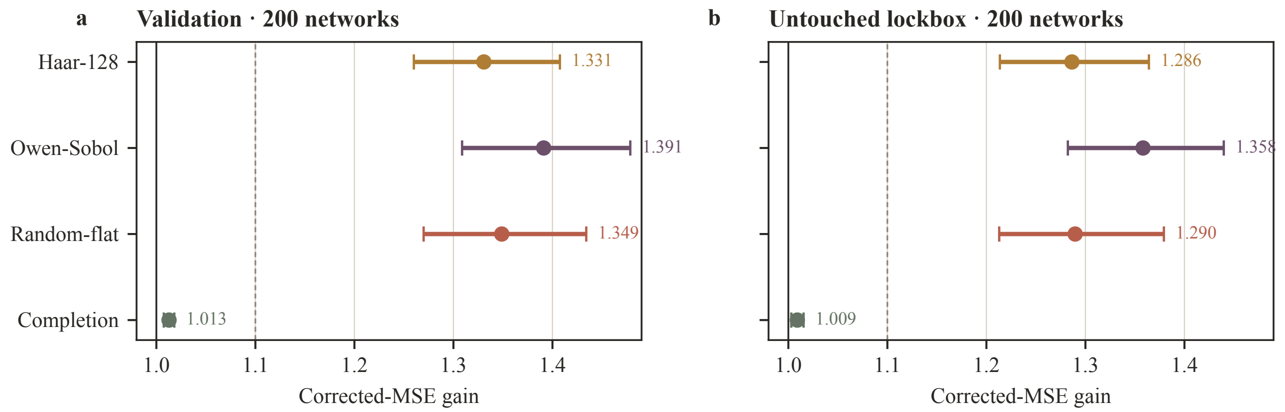

Figure 3. Protected finite-width results. Grand-mean corrected-MSE gains with 95% crossed network-by-realization intervals. The three equal-path controls clear the 1.10× materiality gate on validation and lockbox; exact degree-five completion remains near path-count parity.

Table 2. Untouched lockbox results after averaging 16 global realizations. MSE is corrected for reference noise.

Arm	Mean MSE	Median	q90	q95	Gain over MUB (95% CI)	Lower-MSE wins
MUB-128	2.6429e−7	2.2203e−7	4.4739e−7	5.0937e−7	reference	–
Completed-129	2.6192e−7	2.1924e−7	4.3700e−7	5.0843e−7	1.0091 [1.0030, 1.0154]	completion 135/200
Haar-128	3.3998e−7	2.9153e−7	5.7845e−7	6.7640e−7	1.2864 [1.2136, 1.3641]	MUB 179/200
Owen-Sobol	3.5896e−7	3.0172e−7	6.3021e−7	6.9784e−7	1.3582 [1.2823, 1.4397]	MUB 190/200
Random-flat-128	3.4084e−7	2.9179e−7	5.9154e−7	6.4677e−7	1.2896 [1.2129, 1.3792]	MUB 177/200

4.2 Carrier, antipodes, and exact H1

The float32 asset contains 32,768 unit rays in 128 consecutive orthonormal bases. Stored-array certificates give zero within-block Gram error to displayed precision, cross-block absolute inner product 1/16, and coordinate magnitude 1/16. Runtime propagates both signs. At H1,

$$\mathrm{ReLU}(z) + \mathrm{ReLU}(-z) = |z|. \tag{12}$$

Removing antipodes changes the target and was 119× worse in matched raw-cost product.

For column w_j of W_1,

$$m_j = \mathbb{E}[\mathrm{ReLU}(w_j^\top X)] = \frac{\|w_j\|_2}{\sqrt{2\pi}}. \tag{13}$$

If $\hat{m}$ is the sampled H1 mean, the correction is fused into H2:

$$(H_1 + \mathbf{1}(m - \hat{m})^\top)W_2 = H_1 W_2 + \mathbf{1}((m - \hat{m})^\top W_2). \tag{14}$$

This retains every path. A powered retrospective study measured roughly 3% local raw benefit; live isolation measured about 0.8%. It is a correct, cheap refinement, not the main gain. Full covariance whitening moved the cloud and worsened raw MSE by 2.65×.

4.3 Gate-support continuation

The dominant cost is propagating 65,536 states through 31 nonlinear dense layers. The route reuses two complete carrier bases and antipodes as a 1,024-path full-width pilot. At each hidden layer it retains every unit firing more than once, pads width for the selected multiplication plan, and compacts the corresponding suffix matrices.

We do *not* claim that complete bases are a uniquely optimal 1,024-point pilot; that comparison was not run. The production fact is narrower: these fixed pilot paths preserved full-carrier statistics while removing material suffix work. Recursively compacting the pilot lost more accuracy than its 0.19% compute saving justified.

The exact-H1 bias is added after pilot orders are formed. A protected diagnostic found the mismatch negligible: 11/20 networks had identical orders at all 30 layers; mean selected-set Jaccard was 0.999675; only three networks changed one padded width; and the live-anchored cost ratio was 0.999895.

4.4 Metered continuation

The compact suffix uses guarded recursive Winograd/Strassen multiplication only when the analytical model wins for the actual rectangular shape. Odd dimensions are peeled exactly; unfavorable shapes use ordinary metered products. Removing grouped routing from the small pilot raised analytical work slightly but reduced residual dispatch enough to improve effective compute while preserving the MSE fingerprint.

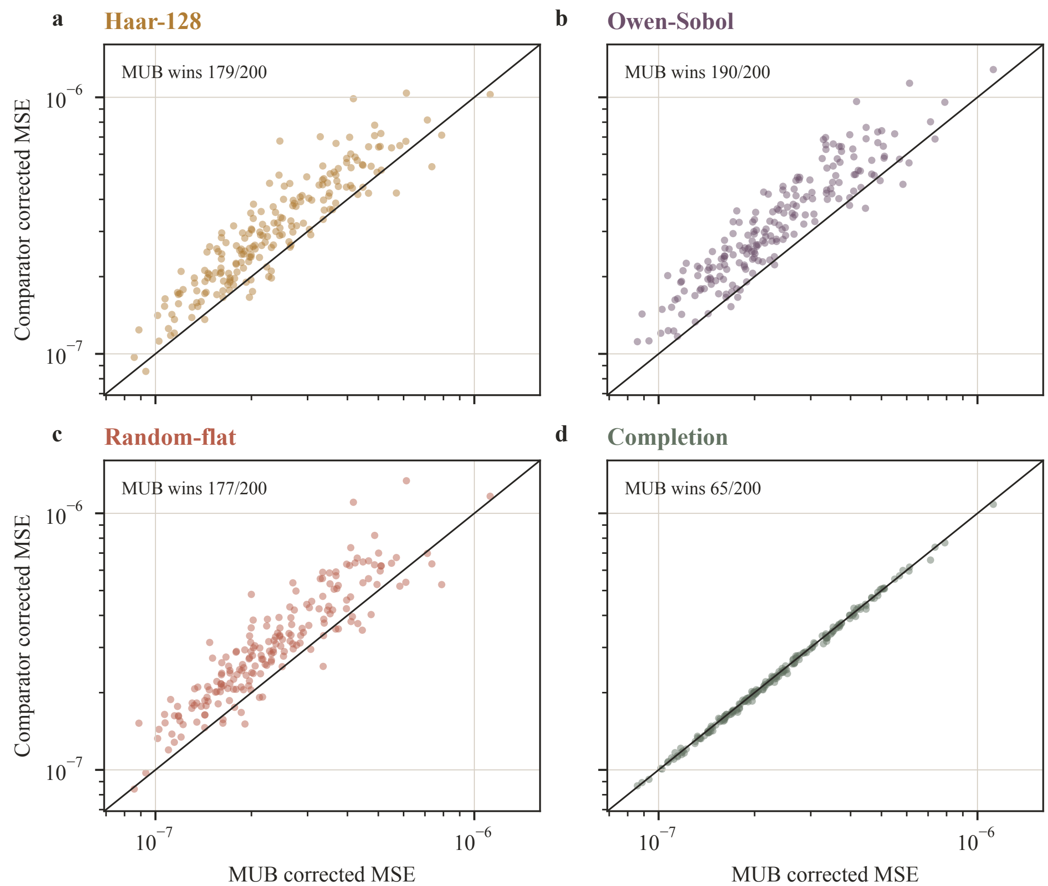

Figure 4. Network-level lockbox transfer. Each point is one network after averaging 16 draws. Points above the identity line favor MUB. Completion remains close to parity; the three equal-path controls separate consistently.

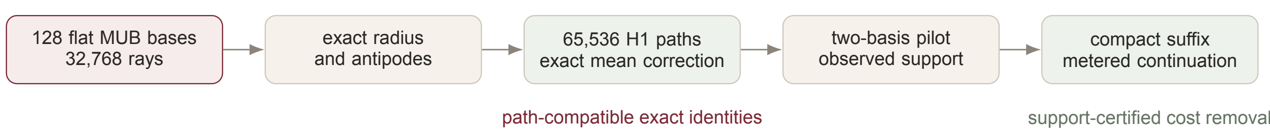

Figure 5. Submitted route. The finite estimator evaluates the actual network. The post-phase theorem explains the carrier; the pilot and metered algebra reduce the price of propagating it.

4.5 Algorithm and live progression

Algorithm 1 · Structured angular mean

1. Load the fixed 128-basis MUB asset and deterministic alignment.
2. Propagate two bases and antipodes as a full-width support pilot.
3. Compact suffix matrices by retained identities, padding only where the metered plan benefits.
4. Scale rays by $\mathbb{E}\chi_{256}$, form antipodes, and propagate H1 in chunks.
5. Fuse the exact-minus-sampled H1 mean through W_2.
6. Propagate through the compact suffix with guarded metered Winograd levels and float64 accumulation.

Only our eligible, fully metered artifacts appear in Table 3. Labels describe causal state, not submission history.

A→B is confounded: it changes path count and early machinery as well as geometry. Its roughly twofold raw improvement is not evidence that MUB structure alone halved error. The matched full-width study supplies the causal geometry estimate. B→C is the clean production result: support compaction changes raw MSE by only 0.07% while improving score by 1.089×.

5. Why support-preserving compression differs

5.1 Euler-Stein facet identity

For one scalar output h, bias-free ReLU composition makes h continuous, degree-one homogeneous, and piecewise linear. Within activation cone C_r, $h(x) = g_r^\top x$ and Euler's identity gives $h(x) = x^\top \nabla h(x)$ almost everywhere. Gaussian integra-

Table 3. Live causal ladder. The A→B transition bundles path count and early machinery; B→C isolates support compaction.

Route state	Raw final MSE	Effective compute	Adjusted score	Causal interpretation
A · metered systems control	$\approx 4.82\text{e}{-7}$	$\approx$ 78.0B	1.38243e−7	frozen pre-carrier route
B · full B128 + fixed alignment	2.3554e−7	$\approx$ 150.4B	1.30270e−7	lower raw error, nearly twice the work
C · B plus support compaction	2.3571e−7	$\approx$ 138.3B	1.19608e−7	0.07% raw change; about 8.3% effective-cost removal
Final · exact H1 + metered refinements	2.3041e−7	127.987B	1.08164e−7	small exact correction plus systems savings

tion by parts, interpreted distributionally, yields

$$\mathbb{E}h(X) = \langle \Delta h, \varphi \rangle = \sum_F ((g_+ - g_-)^\top n_F) \int_F \varphi(x)\, d\mathcal{H}^{d-1}(x). \tag{15}$$

The classical Laplacian vanishes inside every cone; its distributional part lies on shared facets. For $h(x) = \mathrm{ReLU}(a^\top x)$ the formula reduces to $\|a\|/\sqrt{2\pi}$.

This is not the explanation for MUB optimality; Theorem 1 is. The facet identity instead distinguishes exact additive transfer from path surgery. Whitening, cubature replacement, or barycentric fusion can cross later facets. Support pruning is safe only while it retains rare activations defining relevant cones.

5.2 Measured support cliff

A ten-network sweep changed only the pilot firing threshold while keeping all 65,536 carrier paths. Thresholds 1, 2, 4, 8, 16, 32, and 64 retained mean widths 214.7, 212.4, 209.0, 204.6, 199.3, 192.4, and 184.3, respectively. Raw MSE stayed near 2.24–2.48×10^{-7} through threshold 8, then rose to 4.67×10^{-7}, 1.89×10^{-6}, and 1.30×10^{-5}.

This establishes a support-removal cliff, not superiority to every matched-width magnitude criterion. A support-versus-magnitude/RMS comparison at identical widths remains the clearest missing causal ablation.

6. Falsifications

The strongest failures located the missing object because each preserved an attractive summary while damaging the final expectation.

The common lesson is path compatibility. Exact radius and H1 identities remove variance without changing angular trajectories. MUB structure lowers prior-average pair redundancy. The pilot removes work while retaining surviving trajectory identities. Low-component moment replacement and averaging across different gate cones do not.

A separate diagnostic propagated a Gaussian closure's response to the exact-versus-sampled H1 perturbation. Across six networks its correlation with true final error averaged +0.057, ranging from −0.508 to +0.753. A same-data oracle coefficient appeared to gain 1.859×, but unstable sign identifies target fitting rather than a transferable control. We label this diagnostic, not population evidence.

7. Accounting and eligibility

The final implementation obeys the Phase 1 v0.10 accounting model [10]:

- all numerical and data-dependent Boolean operations influencing output use `flopscope.numpy`;
- the direction table and alignment are fixed, model-independent assets;
- every metered operation lies on returned-output dataflow;
- no native arithmetic lane, raw-array bridge, subprocess computation, dummy cover, truth lookup, or residual-time compute lane contributes;
- float64 is used only where accumulation stability warrants its 2× charge;
- every public and all-100 diagnostic row completed with finite output.

Public means were 117.583B analytical FLOPs, 0.1040s residual time, and 127.987B effective compute; 91.9% of the bill was directly instrumented analytical work. The package used Python 3.10.20, NumPy 2.2.6, WhestBench 0.14.0, and FlopScope 0.10.0. The complete artifact digest appears in Section 9.

8. Limitations and decisive next tests

1. **Theorem scope.** Theorem 1 is exact for plain equal-weight antipodal infinite-width ReLU-kernel quadrature. The controlled route also applies exact H1 correction, and #327749 is finite width.
2. **Magnitude mismatch.** The model predicts the correct depth-32 ordering but understates Haar, flat, and Sobol gaps.
3. **Reference-route transfer.** The 16-realization geometry study uses a full-width sibling route. Live B→C establishes support compaction's raw parity, but a carrier × continuation factorial was not run.
4. **Pilot identification.** We did not compare the two complete MUB bases against random carrier, Haar, or Sobol pilots at the same 1,024 paths.
5. **Pruning identification.** The threshold cliff does not compare support with magnitude, RMS activation, sensitivity, or random selection at matched width.
6. **Non-exhaustive controls.** Haar, Owen-Sobol, and random-flat are strong matched controls, not every optimized lattice or digital rule.
7. **Finite deterministic bias.** The estimator remains a finite angular quadrature for a frozen depth-32 network.
8. **Challenge-specific economics.** Width 256, depth 32, homogeneity, and v0.10 accounting all matter.

The most decisive next experiment is a two-by-two production-route factorial: MUB versus randomized-flat carrier, each with full-width versus frozen support compaction.

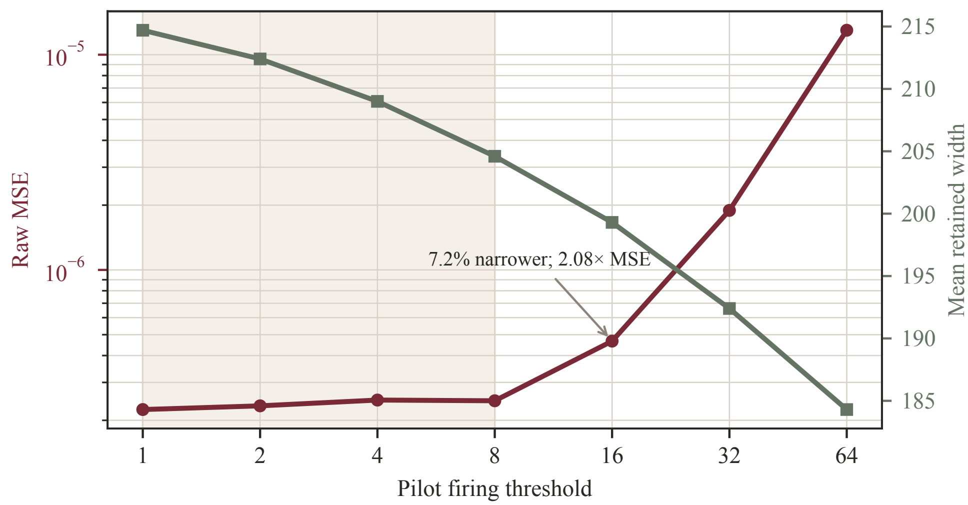

Figure 6. Support cliff. Threshold 16 is only 7.2% narrower than threshold 1 yet has 2.08× its MSE. Capacity falls smoothly; error does not. Rare gate participation carries disproportionate downstream value.

Table 4. Scoped falsifications. They reject the tested summaries, not every possible distributional approximation.

Intervention	Preserved or targeted object	Measured result
H1 covariance whitening/recoloring	exact H1 mean/covariance geometry	2.65× worse raw MSE
Four/eight-component gate cubature	conditional means and covariances	about 26× worse raw MSE
H2 dual-pool balanced thinning	H2 mean and coordinate/projected second moments to about 10^{-4}	1.53× worse raw MSE
H8 dual-pool balanced thinning	actual states plus matched moments	1.273× raw gain, but 1.392× cost; worse score
H8 barycentric fusion	exact paired mean	MSE near 1.6×10^{-3}; catastrophic
Exact degree-five completion	missing fourth-moment component	only 1.0091× lockbox MSE gain

The second is matched-width support versus magnitude/RMS selection. Neither unrun result is implied here.

9. Reproducibility

The theorem audit accesses no network weights or truth to form predictions. MUB, completion, and random-flat energies are exact; Haar is an exact ensemble expectation; Sobol reports fixed-pair uncertainty. Point-set generation is deterministic from recorded 64-bit seeds. Finite-width analysis rejects duplicate cells, changed seeds, missing identities, and nonfinite values.

Key sources are `analyze_relu_nngp_mub_mechanism.py`, `paper_multirealization_geometry_study.py`, and `paper_flat_hadamard_control.py`. Frozen analyses are `relu_nngp_mub_mechanism.json`, the validation/lockbox geometry summaries, and the exact-H1 pilot diagnostic. Complete SHA-256 digests are:

Artifact #327749

`4349c2594d4b51d39016a347982721b7773cb0111ace1f6c95542a3ae04e696e`

Real B128 rule

`18f958a1f627b797a48e4b2f1b8579d0287f97435fb8e653b5562ab11b3677d1`

Sobol integer table

`fa029f6b387faa0e3a0dc16eeafa465549e6096d643e266c37884bb62c0cefca`

10. Conclusion

The original explanation was backwards. The carrier was not useful because it almost completed a spherical 5-design. It was useful because 128 bases already solve a ReLU-specific optimization problem: at every positive depth, mutual unbiasedness minimizes antipodal infinite-width random-ReLU quadrature risk among equal-sized unions of orthonormal bases.

At depth 32, the theorem forecasts 1.00930× from the missing basis; the untouched lockbox measures 1.00907×. It also forecasts MUB < Haar ≈ randomized-flat < Owen-Sobol. Finite width amplifies comparator gaps but does not reverse them.

The graded artifact supplies the computational half. The operational B128 transition lowered raw error but raised propagation cost and cannot be assigned causally to MUB structure because path count and machinery changed together. The matched study isolates geometry. In production, a conservative support pilot removed about 8.3% of effective work with 0.07% raw change; more aggressive removal crosses a sharp support cliff.

> **Design rule.** Minimize redundant sampled trajectories with structure matched to the random computation; then compress the finite network only where realized gate support certifies work absent.

Exact identities, kernel-optimal angular organization, and support-preserving continuation solve different parts of the

Table 5. Compact evidence manifest.

Claim	Evidence unit	Independence protection
MUB minimizes basis-union NNGP risk	exact kernel-risk proof	no network data
Predicted point-set ordering	exact spectra + 16 fixed flat/Sobol draws	no network weights/truth; Sobol SE reported
Finite-width ordering transfers	200 validation + 200 lockbox × 16 draws	crossed bootstrap; draw-0 exclusion; leave-one-out
Missing-basis effect is small	paired MUB/completion draws	same rotations; exact path ratio
Pilot compaction preserves statistics	adjacent live B/C routes	raw parity; rowwise scorer
Aggressive support removal fails	fixed threshold sweep	same carrier and continuation
Accounting route is complete	grader telemetry and package	exact runtime tuple; all rows finite

same expectation problem. Their mathematical prediction, finite-width test, and successfully graded implementation agree within explicit boundaries.

Human direction and LLM disclosure

The participant directed the campaign: defining objectives, selecting among human- and AI-generated hypotheses, choosing experimental pivots, stopping unproductive open loops, overseeing submissions and revisions, and making all final decisions. The participant supplied the critique that triggered the ReLU-kernel theorem audit and directed this revision.

ChatGPT/Codex and Claude were substantive research collaborators. They generated and screened hypotheses, searched the literature, co-developed the ReLU-kernel theorem and proof, designed and implemented estimators and falsification studies, built analyses, figures, and submission artifacts, interpreted results, and drafted the paper. Their work was iterative and supervised. Both systems had repeated the incorrect spherical-5 label before a direct audit falsified it; the participant directed the correction and adjudicated the final claims. The participant owns the submission and is responsible for every claim and limitation.

References

[1] W. Wu, V. Lecomte, M. Winer, G. Robinson, J. Hilton, and P. Christiano. Estimating the expected output of wide random MLPs more efficiently than sampling. *arXiv:2605.05179v2*, 2026.

[2] Y. Cho and L. Saul. Kernel methods for deep learning. *Advances in Neural Information Processing Systems*, 2009.

[3] J. Lee, Y. Bahri, R. Novak, S. Schoenholz, J. Pennington, and J. Sohl-Dickstein. Deep neural networks as Gaussian processes. *ICLR*, 2018.

[4] A. Matthews, M. Rowland, J. Hron, R. Turner, and Z. Ghahramani. Gaussian process behaviour in wide deep neural networks. *ICLR*, 2018.

[5] A. R. Calderbank, P. J. Cameron, W. M. Kantor, and J. J. Seidel. $\mathbb{Z}_4$-Kerdock codes, orthogonal spreads, and extremal Euclidean line-sets. *Proceedings of the London Mathematical Society*, 75(2), 1997.

[6] P. O. Boykin, M. Sitharam, M. Tarifi, and P. Wocjan. Real mutually unbiased bases. *arXiv:quant-ph/0502024*, 2005.

[7] P. Delsarte, J. M. Goethals, and J. J. Seidel. Spherical codes and designs. *Geometriae Dedicata*, 6, 1977.

[8] B. Reznick. Some constructions of spherical 5-designs. *Linear Algebra and its Applications*, 226–228, 1995.

[9] A. B. Owen. Monte Carlo variance of scrambled net quadrature. *SIAM Journal on Numerical Analysis*, 34(5), 1997.

[10] AIcrowd. ARC White-Box Estimation Challenge 2026: Phase 1 accounting update. 2026.