A Technique Census of Phase 1 — Consolidated Update at Phase 1 Close

jamesrahenry (with LLM-assisted digestion, human-directed)

2026-08-18 · ARC White-Box Estimation Challenge 2026 · forum topic 18157

This document is the paste-stable PDF of the consolidated census update posted as replies on forum topic 18157 (the census of 2026-08-13). It covers the fourteen write-up topics published after the census's corpus cutoff, the final public rescore that closed Phase 1, and a restatement of the census at a glance and the open mathematics. Claims carry their source topics; corrections from the authors are welcome.

The consolidated update pass, as promised — with Phase 1 now closed. The challenge page marks Phase 1 concluded following a final public rescore, though the held-out results have not yet been announced. The rescore's visible effect on the public board was surgical: the two anomalous entries at the extreme front — the 9.5e-11 and 1.5e-9 adjusted scores whose raw MSEs sat below anything any published technique can explain — were **removed**, one author's fallback entry landing mid-board at an ordinary 1.34e-7; one other entry moved ~3%; and every remaining score is unchanged (147 of 147 overlapping rows within 2%, several spot-checked exactly). So the board this update reads is the final public state of Phase 1, and the census's biggest open question — what the silent front was — has an administrative answer rather than a technical one: *it was removed, not explained*. More on what that leaves standing at the front in the closing section of this document.

Fourteen write-up topics arrived after the corpus cutoff. The census (posted 2026-08-13) covered the 53 topics then in the category. Fourteen more landed 2026-08-14 → 18: mliston (18170, #327723), the baltsat visual notebook (18163) and bgrubbs1984's companion site (18169), then the wave of 2026-08-16/17 — pranay212 (18173, #327651), kaileh57 (18171, #327786), nitiz_khanal (18174, #326960), Team Puffi's two-paper drop (Cipo, 18175: #327950/#328046 plus the SSC estimator #327725), ely2sh (18176, #327749), keenanpepper (18177, #327838), qi_zhang5 (18179, #326954), violeta (18180), omer_kiraz (18181, #327792), trim_qewas (18182, #326725) — and finally andrei_bulzan's 82-page report (18183, #327801). They move specific census items. Deltas below, keyed to the original item labels.

New source: pranay212, topic 18173 (submission 327651; public adjusted 1.23e-7, final-layer MSE 1.55e-7). A hybrid of Techniques 3 and 4: a **BCH strength-4 orthogonal-array** angular design (coordinates indexed by GF(256), columns [1, a, a^3], antithetic quotient leaving 65,536 line representatives, of which a fixed target-independent prefix of **56,000 pairs** is used) mounted on the SOX dead/on/kink execution chassis (explicitly credited to 18106), with exact FWHT layer-0 evaluation, an exact antithetic layer-1 fold ($h_0(-u)W_1 = h_0(u)W_1 - u(W_0W_1)$), Winograd/Strassen, and layer-0 mean transport into a layer-1 bias (covariance transport helped an earlier sampling chassis but *hurt* the fixed-radius design). Item-level effects:

- **3a/4 (amended).** The best published **raw** MSE moves: 1.546e-7 (pranay212) vs the previous best 2.18e-7 (SOX, 18106). And **3c is now moot as stated**: the cubature family's edge over structure-aware sampling was "compute, not raw accuracy" — the

hybrid takes both, because it *is* both. This is also the first published estimator assembled substantially from other teams' published parts (SOX chassis + the census-documented cubature line): the Phase 1 literature is compounding, which is what a literature is for.

- **3b (amended).** "Under an adjusted score you never subsample the design" (nygaard's k^−1.2 ablation) gains a practical exception: pranay212 subsampled 65,536 → 56,000 pairs **for watchdog safety, not score** — the complete table was locally more accurate but left no budget margin on hard MLPs, and a 62k FWHT variant hit watchdog failures. His stated rule: "one failed MLP is much worse than a small average gain." So: never subsample for score; do subsample for the timeout tail.
- **6f (reinforced).** Another compiler-class result: a final exact-algebra pass (one-write assembly, Winograd formulas, sparse packing) cut billed FLOPs ~6.7% mean across four MLPs (192.566B → 178.923B on the representative one) with raw MSE unchanged to ~0.1%, worth **+7.5% adjusted** with no raw-MSE change across the checkpoint pair.
- **1e-adjacent.** His discovery path independently replays the closure story: started at cumulant propagation, found partial k=3 corrections can be worse than the Gaussian closure at depth — consistent with the census's C1 wall.

New source: kaileh57 (Kellen Heraty), topic 18171 (submission 327786; public adjusted 9.686e-8, raw 2.088e-7). A complete randomly-rotated 66,048-point Kerdock rule — the **third independent build of the C2 construction** — plus network-adaptive pruning (activation-column dropping at a 0.02 relative threshold; two-stage dead-output testing with 516 nomination + 516 disjoint confirmation rows and a 32-row sentinel), exact FWHT/Strassen/sparse-prefix algebra, and a final-layer-only output policy. The paper's center of mass is theory, and it upgrades two census verdicts from "measured" to "proved":

- **3 (verdict upgraded).** The census said the angular problem was "effectively solved"; there is now a **certificate**. Theorem: for the depth-32 NNGP (iterated arc-cosine) kernel, *every* fixed nonnegative spherical rule with ≤66,048 support points has normalized discrepancy within **0.0233188%** of the complete Kerdock rule's — i.e. fixed positive sampling at this support budget is nearly saturated. The engine is a Delsarte/Cohn–Kumar-style LP bound built on a genuinely new analytic fact (their Theorem 1): the **sixth derivative of every finite iterate of the dual-ReLU map is positive** on (−1,1), so a degree-5 Hermite minorant at three double nodes is global at every depth. The algebra is audited over the rationals and the depth-32 instance is certified with 256-bit interval arithmetic (Arb), with proof-check commands and SHA-256 manifests in a public release. This stands beside the closure floor (1a) as the field's second proved wall: **raw-risk reduction inside the fixed-positive class is closed** — what remains is cheaper evaluation, signed/adaptive/network-dependent rules, or hybrids. It also supersedes the quarantined near-optimality claim noted in 3b with a reviewable artifact. Two calibration numbers worth keeping: the Kerdock rule is **32.9% below** iid antipodal sampling's expected risk at equal support, and the public raw MSE (2.088e-7) runs ~14% *below* the idealized ensemble prediction (2.429e-7) — finite width, suite variation, numerical QR, and pruning all live in that gap.
- **5f (sharpened to a theorem).** The census's control-variate design rule was "under antithetic pairing, controls must be even." Their Theorem 3 (degree-five control inertness) is much stronger for the cubature family: on the complete Kerdock cloud, every fixed-radius, sphere-centered, input-polynomial control of total degree ≤5 is **pathwise identically zero** — not merely mean-zero — for every rotation and radius, and this survives their column pruning. Degree six is the first live harmonic band and carries only **13.93%** of the remaining discrepancy (consistent with amalgonim's 3d oracle argument); their 128-feature degree-six ridge dictionary measured a pooled-oracle R^2 of 0.2351% against a 4.05% cost surcharge — dead. The explicit carve-outs matter as much as the theorem: Gaussian-centered controls ($\|x\|^2$ has sampled value μ_{R^2} but Gaussian mean 256), row reweighting, and **controls of intermediate activations** are all outside it. Read together with 5d (kwon's 2.39× mid-network control), the control frontier for

cubature carriers is now provably *not* at the input — the mid-network channel is the only live one.

- **5h (confirmed again).** Their pruning-threshold campaign is textbook transfer-trap discipline: α = 0.035 saved 2.3% arithmetic but cost 13.3% MSE (product 10.6% worse); a frozen screen on 24 fresh networks showed thresholds 0.022–0.025 produced development-only gains that reversed held-out; 0.020 retained. Their negative-results table (a second Kerdock rotation: <0.26% modeled headroom; two added optimized lines: ratio 1.0000203, with the optimizer returning **no negative weights**; shared-prefix multifidelity: correlation arrives more slowly than work is spent) all dies honestly.
- **6-adjacent.** One more compiler-class item: caching contiguous FWHT workspaces with ping-pong reuse removed 319,029,248 metered operations per test with bitwise-equal outputs. And a provenance observation the metering crew should see: **AIcrowd publishes no server-side artifact hash**, so a retained local upload cannot be cryptographically tied to the graded copy — their SHA-256 manifest is the best a participant can currently do.
- **Scoping honesty worth naming:** the certificate was derived after submission, covers only the unpruned core, and "does not explain the leaderboard score." The write-up also connects the challenge to the external literature (Wu, Lecomte, Winer, Robinson, Hilton & Christiano, arXiv:2605.05179) — the first in-forum write-up to do so.

New source: nitiz_khanal, topic 18174 (submission 326960; official adjusted 1.7275e-7, 50/50 graded, zero failures). A structure-aware Monte Carlo sampler — centered/whitened antithetic Gaussian cloud, the exact first-layer antithetic identity ReLU(−z) = ReLU(z) − z, a pilot-frozen dead-row mask (1,024-sample lead block; rows kept iff they fired at least once), SOX-style terminal stable-on/stable-off/kink classification at a 1.0-column-σ margin, and two-level Strassen/Winograd applied only above residual-aware size cutoffs. Individually these are census-known parts; the write-up's value is in three places:

- **A new design framing: "private-safe."** This is the first write-up organized around surviving the fresh-draw private re-evaluation rather than the public board: no public-instance identification or answer tables, the pruning mask frozen *before* production samples, sample count sized against a conservative **no-pruning** worst-case cost model, and a terminal classification that deliberately retains uncertain columns rather than forcing a call. It is 5h's transfer-trap discipline promoted from a validation rule to the architecture itself.
- **6a/6c (reinforced with a clean instance).** A level-three Strassen recursion reduced nominal FLOPs but lost in paired tests on residual runtime — and his graded checkpoint pair makes the same point at the score level: a more conservative recursion cutoff (#326963) improved the local FLOP proxy yet graded *worse* than the selected build (1.7292e-7 vs 1.7275e-7). "A lower local cost proxy did not transfer to a better official score."
- **4 (the chassis is now standard).** With explicit citation of "publicly documented structural pruning and terminal classification, independently integrated and validated," this is the second post-cutoff write-up to build on the SOX execution pattern — dead/on/kink routing has become the field's shared execution layer, not a team's technique.

His rejected-ablation table is the familiar honest ledger (bigger lead block: no gain; tighter margin: weaker conservatism for negligible benefit; earlier split: worse held-out; Cholesky whitening: cheaper and plausible but unshipped), and control variates, higher-order closures, and alternative sampling tables were all tried and not promoted — consistent with the census's Technique 1 and 5 verdicts.

New source: Team Puffi (Cipo), topic 18175 — two papers. The census cited Cipo's in-forum notes (18152); these are the formal write-ups, and they are the most substantial of

this batch.

Paper 1 — the score stack (submissions #327950, adj 9.0914e-8 / raw 1.8945e-7, and a lower-compute alternative #328046). The carrier is the complete MUB-129 bank (128 Kerdock bases + identity, 66,048 antipodal directions — the **fourth complete build of the C2 construction**), with a three-basis probe (1,536 rows) for pruning and omitted-mean fill, prefix-based sparse execution, and a positive-stable terminal bypass. Item-level effects:

- **3a (extended with the first published quadrature-rule ladder).** Measured final-layer MSE relative to Kerdock frames on a matched panel: i.i.d. Gaussian 2.36×, Owen–Sobol + antipodes 1.39×, rotated simplex 1.35×, signed-Hadamard frames 1.02×, Kerdock 1.00×, **complete real MUB-129 0.81×** — the identity basis is worth 19% and is nearly free at layer 0. No previous write-up ranked the alternatives head-to-head.
- **5c/5d (the mid-network control variate, now fully documented at the published frontier).** The census recorded Cipo's "~2×" inter-layer fluctuation-map control from forum notes; the paper gives the deployed construction: a shallow analytical reference ("A716" — a fitted finite closure through post-ReLU layer 9 carrying mean, full covariance, marginal third/fourth cumulants, a paired (2,2) fourth, and an eight-generation mixed (2,1) history, with 90 frozen fitted coefficients; layer-9 MSE ratio **0.511 vs Wu K=3**) supplies the layer-9 discrepancy, a ridge-fitted per-network response transports it to the final layer, and a 0.75-shrunk correction is subtracted from the final sampled mean. Measured: **17.6% final-MSE reduction** (paired-bootstrap 95% CI [7.8%, 26.3%]), winning on 34/50 networks, deployed at the best published-write-up-family rank. Their own framing is exactly the census's: "model-assisted rather than a textbook unbiased control variate," with the error identity $\|e\_CV\|^2 = \|e_{31}\|^2 - 2\rho\cdot(\text{alignment}) + \rho^2\cdot(\text{quadratic})$ stated rather than assumed. This is 7a (fit at scale, freeze) and 5d (mid-network control) composed and graded.
- **6 (a new score identity).** With per-network error e and multiplier q, the published score is $S = \bar{e}\cdot\bar{q} + \mathrm{Cov}(e,q)$ — the error–multiplier covariance is a real score channel (for #327950: $S/\bar{e} = 0.4799$ vs $\bar{q} = 0.4830$). Nobody had written that term down.
- **An instrument fact the metering crew should have:** the grading platform executed **100 MLPs; fifty formed the public scoring panel; the other fifty were gated evaluation cases not included in the public score.** That is a direct statement that a hidden 50-network panel is already being run per submission — relevant to 6h's leaderboard-population analysis and to what the private re-evaluation will draw from.
- **New open problem (frame selection).** Retrospective oracles show small Kerdock subsets can be excellent (a greedy target-using 38-frame oracle hits 5.47e-8), but every target-free selection rule tried — including herding on observable features — failed to transfer: herding reduced the discrepancy it optimized without consistently reducing final MSE, because "later nonlinear layers can rotate or regenerate error that was reduced at the selection layer." A mechanistic, fixed-before-evaluation frame-selection rule is named as their main open problem (forum note: they hope it reaches 0.1–0.2 multiplier).

Paper 2 — the mechanistic estimator (submission #327725, "SSC": separate-universe response closure; adj 1.6533e-6 / raw 2.6597e-6). The first structural advance in Technique 1 since the census closed the family:

- **1 (the projection step, not the order, is the live variable).** SSC keeps the Wu-et-al. K=3 cumulant backbone but replaces the fourth-cumulant projection: κ_4 is regenerated each layer from ≤4 live covariance-response modes coupled by a small fitted (indefinite) Λ — the super-sample-covariance algebra from cosmology (Takada–Hu), with the n^4 tensor never formed. On a matched 30-network panel it cuts layer-32 MSE **77.2% vs factorized K=3 and 48.2% vs factorized K=4**, at 0.501B ledger versus K=4's ~10^2B, and flattens the depth-error exponent to **2.21 (vs 2.69 for K=3, 4.13 for K=4)**. A matched ablation isolates the response branch: disabling it raises layer-32 MSE 2.728e-

6 → 7.698e-6.

- **The wall still stands — and they say so.** SSC's layer-32 MSE remains 7.94× worse than Kerdock QMC on the same panel, and the graded submission (1.65e-6 adjusted) does not approach the closure family's ~8.6–9.3e-7 readout floor (1a), which SSC's compressed fourth order does not carry enough joint state to break (compare O4's "carried joint object of dimension ≥64" — SSC carries four modes). But 1e said higher orders don't rescue the family because cost explodes and the series is asymptotic; SSC is the first measured answer of a different kind: **compress the order you keep instead of adding one**, beating full factorized K=4 at roughly 1/200th its cost. The census's "closed as a competitor" verdict survives; its "mechanism understood, nothing left to try" reading does not.
- **Useful taxonomy.** The paper names the whole Technique-1 family "deduction–projection estimators" (exact deduction advances the state, a projection compresses it) — a cleaner frame than "closures" for where the error actually enters, and it locates 1b–1e as facts about particular projections.
- **7a discipline, again:** coefficients fitted on 160 public networks, checked on 16 held-out, evaluated on 30 fresh independently generated networks; everything frozen before evaluation. A post-submission audit (refitting the response to target the consumed mean correction rather than κ_4 marginals: 13.6% further MSE reduction on the predecessor baseline) is reported as unintegrated future work rather than folded into claims.

New source: ely2sh, topic 18176 (submission 327749; adjusted 1.08164e-7, raw 2.30406e-7). The best-titled paper of the batch — "The missing basis was not the mechanism" — and the second theorem-bearing one. The author audited the standard 128-basis Kerdock carrier, found it is **not** a tight spherical 5-design (the complete construction in dimension 256 has 129 bases; the coordinate basis was missing from their asset), and then discovered the correction barely matters — which forced a better explanation of why the carrier works at all:

- **3 (a second, complementary theorem).** Theorem 1: for equal-weight antipodal unions of B ≥ 2 orthonormal bases in R^d, under the infinite-width He-ReLU (NNGP) kernel at **every positive depth**, quadrature risk is minimized exactly when every pair of bases is mutually unbiased (proof via the ReLU kernel's nonnegative power series, strict convexity, and Jensen — equality iff all cross-Gram entries square to 1/d). Read with kaileh57's result, Technique 3 now has both halves: **within the basis-union class, MUBs are exactly optimal at every depth (ely2sh); over all fixed nonnegative rules at this support, the complete construction is within 0.0233% of optimal (kaileh57)** — posted within a day of each other. The mechanism is *pairwise mutual unbiasedness*, not 5-design exactness: adding the missing basis (making the rule a true 5-design) buys only a predicted 1.00930× / lockbox-measured 1.00907×, while the MUB property over Haar / random-flat / Owen–Sobol controls is worth a measured 1.286–1.391× at matched paths.
- **A public discrepancy worth the authors' attention (the batch's version of 6g's label-floor disagreement):** Team Puffi's quadrature ladder (18175) prices completing MUB-129 at **~19%** over bare Kerdock frames; ely2sh's paired completion study prices the same missing basis at **~0.9%** (and only ~0.1–0.5% path-normalized). Different panels, corrections, and protocols — but the two numbers cannot both describe the same object, and both teams have the artifacts to resolve it.
- **2e (strengthened).** Removing antipodes was measured **119× worse** in a matched raw-cost product — the strongest published number for why antipodal pairing is non-negotiable on deterministic carriers.
- **4/6 (a causal cost ladder and a support cliff).** A 1,024-path gate-support pilot removes ~8.3% of effective compute at **0.07%** raw-MSE change (isolated as its own causal step in a live A→B→C ladder, +1.089× adjusted). But a threshold sweep shows a **sharp support-removal cliff**: pilot firing thresholds 1→8 leave raw MSE near 2.3e-7

while trimming width ~215→205; threshold 16 costs 2.08× the MSE for only 7.2% more narrowing, and 64 is catastrophic (1.3e-5). "Rarely firing units carry disproportionate downstream value" — the routing family's version of 4c's kink-mass concentration, and a measured warning for every pruning-style estimator in this census.
- **A new mechanism tool (path compatibility).** An Euler–Stein facet identity (E h(X) as a sum over activation-cone facets) explains *which* interventions are safe: exact identities and support-certified removal preserve trajectories; anything that **moves paths** — covariance whitening (2.65× worse), gate-level cubature replacement (~26× worse), barycentric fusion (catastrophic) — crosses facets and dies. This unifies several previously separate negatives, including pranay212's covariance-transport regression (18173) and the census's 1c/5-family chain-surgery failures.
- **A measurement-noise warning every small ablation should heed:** one network's MSE varied **2.9–5.4× across ten global orientations**, with split-half rank correlation of rotation rankings −0.358. Their answer — 16 fixed global realizations × 400 networks (200 validation + 200 untouched lockbox), gates predeclared — is the strictest experimental protocol in the published field, and it implies single-rotation ablation numbers elsewhere carry more noise than their error bars admit.

New source: qi_zhang5, topic 18179 (submission 326954; adjusted 1.46e-7 — forum text only, no PDF). A Sobol-family RQMC estimator whose write-up contributes calibration and two negative confirmations:

- **2/6 (the plateau, calibrated).** Working identity: raw × F = V_iid · c / g (per-sample IID variance scale × billed cost per effective sample ÷ variance gain vs same-N IID). Calibrating V_iid ≈ 0.0405 on public MLPs predicts the pure-sampling plateau at 6.24e-7 vs 6.47e-7 observed (3.7% off), with their production point consistent. This is the floor-bet thread's arithmetic (organizer-confirmed) now carrying a measured calibration.
- **5b/5e (independently confirmed).** A cross-fitted linear control variate worth 1.42× under IID retains only **1.04× under Sobol** — the RQMC carrier already absorbs the low-order variance the control would remove. An independent replication of kwon's fusion observation (18154), and the sampling-side twin of kaileh57's degree-≤5 inertness theorem: on a good carrier, the cheap channels are already spent.
- **6 (an arbitrage audit, negative).** A direct audit of the v0.10 cost model found **no billing arbitrage**: matmul/einsum/tensordot routes all bill at 1.000, fp16 = fp32, fp64 = 2×, and comparisons/where/take/concatenate are all priced. "Whatever the frontier is doing, it isn't metering tricks" — which further narrows the hypothesis space for the silent front (see below).

New source: trim_qewas, topic 18182 (submission 326725; adjusted 1.855e-7) — the heavyweight of the batch. Their score is mid-pack and they say so; the contribution is a **protocol**: replace one internal quantity of a running estimator with its ground-truth value and measure — the result upper-bounds every idea whose effect passes through that quantity, including ideas nobody has written yet. Fifty-four method families closed this way (each ~15 GPU-minutes), with a 27-page companion registry ("The Compendium") holding every stand, seal, and control; 396 recorded findings of which 113 correct earlier conclusions and 51 are outright self-refutations. The census-moving results:

- **1 (the closure story is now complete, and quantified at the top).** A *perfect* two-moment closure — true m and true C at every layer — scores **5.5e-8 adjusted, about 5th place**: the entire two-moment class is exhausted before the podium (×7.7 short of #2). A **full moment oracle** (exact m, C, per-neuron γ_3, γ_4) lands **8.16e-9 — level with #2 at best, ×5.4 short of #1**. So no moment-based state reaches the front, measured rather than argued. Their §4.1 also gives 1c its sharpest form yet: a model predicts the closure's per-step σ error at $R^2 = 0.836$, and correcting it buys *exactly nothing* (optimal coefficient 0.00); substituting the true covariance diagonal makes the answer **8.5× worse**. "A tuned closure is a balanced system of cancelling errors" — substituting

truth into one channel measures the other constants' staleness, not the substitution.

- **A smoothness ceiling that closes six approximation theories at once.** 10–16% of the variance sits at Hermite degrees ≥ 7 with a power-law tail (a kinked function's Gibbs phenomenon), so a method exact through *arbitrarily* high degree still saturates at **×7–9** — a ceiling on smoothness itself. And the natural counter-move (work *with* the singularities) is closed from the opposite side: the median distance from a sample to the nearest switching surface is **1.1e-4** and 99.1% of probability mass lies within 1e-3 of one — **the kinks are a dense medium, not localizable features**. Softplus-β control: at β = 4 the function is still 55–66% nonlinear but the high-degree tail has already collapsed 3–7×.
- **4/7-adjacent mechanism results:** the answer is **born early** (layers 0–7 carry 50.1% of the final mean via an exact per-layer covariance identity) while the **error is made late** (84% is accumulation; the backward operator has effective rank 2.74, condition number ~1e31 — state reconstruction from the output is numerically impossible); 95% of a step's error is the previous step's error transported (one object on one shaft — which kills per-layer independent corrections structurally); 68% of the final error lives in a **6-dimensional Lyapunov subspace computable from the weights alone**; and the closure error is a **finite-width 1/n fluctuation of the particular W** (exponent ≈ −1, three independent constructions), not a missing term in any moment expansion.
- **A flat-budget theorem (6-family).** Above the floor, a pure Monte-Carlo estimator's score is *budget-independent* (their seal: N = 65,000 at u = 1 and N = 6,500 at u = 0.1 grade identically), the better your deterministic branch the stricter the optimal budget floor, and the analytic advantage is a **small-sample phenomenon**: honestly re-fitted at ×64 budget the optimal analytic weight is exactly zero and the method degenerates to Monte Carlo. The competition sits exactly in the regime where analytics pay — a design property of the benchmark, worth reading against 6h.
- **§7: what the numbers exclude about the top of the board.** From oracle measurements plus a leading entry's public panel: the front is **not moment-based** (the full oracle is ×5.4 above #1), **not layer-wise** (its layer-0 error equals the ground truth's own layer-0 noise floor, 6.77e-10 — layers 1–30 are not computed at all), **not pair-coordinate**, is **deterministic** (vs-sampling ratio 1782×, per-network spread 9.6), and **can afford algebra** (2.61e10 effective FLOPs — 15.7× their closure). This is the strongest public forensics on the silent front since amalgonim's appendix, and it points the same direction: a deterministic, non-moment, angular/algebraic object. Their honest hint: an oracle on the *angular* factor alone reaches 0.2778% closure error — 1.83× better than their entire closure — "if a better class exists, its state is probably angular rather than moment-based."
- Plus a per-instance impossibility (§4.7: the optimal per-network constant depends linearly on the realised noise; every statistic computable inside the estimator is quadratic in it — the signal is **unobservable from inside**, which upgrades 5h's transfer-trap from "measured" to "mechanistic"), GOE level statistics ruling out hidden conserved structure, a weight-quantization rate-distortion curve (b ≥ 5 bits suffices; transfer constant D/Δw = 3.05), and the field's most complete instrument-failure taxonomy (§6.1: six ways a measurement lies, with counts — a companion piece to the census's production-modes aside).

New source: keenanpepper, topic 18177 (submission 327838; raw 1.50e-5 / adjusted 1.95e-6) — the bitter-lesson arm, measured to its ceiling. Co-credited to Claude Opus 5 as a listed author. A fully sampling-free learned propagator (mean, second moment, an $O(N^2)$ third-cumulant *plane* standing in for the $O(N^3)$ tensor whose exact transport bills 1.4–3.4× the entire budget, plus a recurrent latent), trained end-to-end by backprop through the 32-layer recursion on a 499k-MLP corpus:

- **The headline negative is self-ablation:** the learned attention closure the architecture was designed around is **numerically inert** — deleting it changes predictions in

the fifth significant figure while consuming **61% of billed cost**; the bit-identical lean rebuild lands under the multiplier floor and would have scored 1.50e-6 adjusted, a 1.30× improvement by deleting code. Roughly ninety numbered experiments went into a mechanism that trained itself to zero. ("Ablate the shipped artifact, not the design.")

- **The ceiling result the census should carry (O4-adjacent, agrees with 18182):** inject *true* means and covariance at every layer and the recursion bottoms out at **8.1e-7 raw** — an oracle state genuinely beats sampling at equal arithmetic (4.7×), but both methods sit near the multiplier floor, so it converts to only **1.4–2.0× on the scored metric**. "The entire remaining research program is worth about a factor of two." Note the convergence: keenanpepper's (μ, M_2) oracle at 8.1e-7 raw and trim_qewas's allm at 5.5e-8 adjusted are the same wall measured through two different instruments.
- **Wu et al.'s crossover, located:** Wu et al. beat sampling at sufficient width to depth 12; this measures the program losing at width 256 / depth 32. Whether that's a width, depth, or width/depth effect is named as the measurement they most want — **and it is now DONE** (jamesrahenry, in-thread reply 2026-08-17, analytic-baseline-vs-plain-MC version): a 7-depth × 5-width median-based grid (48 nets/cell deep rows, FLOP-matched at n = 2w) shows **both effects are real and width dominates monotonically**, with crossover depth d*(w) rising from <24 at w=64 to ≥48 at w=1024 — and the Phase 1 operating point (256, 32) sitting at **median parity** (ratio 1.04, bootstrap 95% CI [0.75, 1.38]). Consistency anchors hold: at d=12 the analytic floor wins at every width (matching Wu), and fitted corrections shift ratios 5–10× down, so the corrected-analytic crossover sits deeper than this floor contour. Code and per-cell data public (`arc-whitebox-replication/analytic_vs_sampling`).
- **Infrastructure (C4-class):** two public HF datasets — `whestbench-relu-mlp- means-500k` (500k MLPs, per-layer mean targets, seed-regenerable, validation noise ~7.5e-10 matching official eval targets) and `whestbench-relu-mlp- moments-10k` (weights + pairwise moments to 4th order at N=1e8). Cipo and bgrubbs1984 already credit them in replies. Also the second independent statement that the suite is **50 public + 50 sealed with per-MLP seeds identical across submissions**, and prizes decided on a fresh re-evaluation with unseen seeds.
- Five measured "walls" for learned correctors (teacher-forced R^2 is not a gate — fidelity and value dissociate in both directions; corrections convert only where the residual is low-rank; the one converting write is a co-adapted *compensator*, not a corrector — injecting truth into the trained stack makes it *worse*), which sharpen the census's 7b/7c with mechanism.

New source: mliston, topic 18170 (submission 327723; adjusted 1.334e-7, raw 1.218e-6 at 10.95% utilization). The census's 5d line (kwon's bias–variance control-depth tradeoff) built into a full submission, plus the batch's most important integrity disclosure:

- **5d, engineered:** a cumulant-aware closure through layer 18 (rank-capped third-order carriers + per-neuron κ_4) anchors a control at **layer 22 with strength 0.6**; the residual is transported through the actual weights and observed gate fractions. The oracle diagnostic that motivated it reproduces kwon's law exactly: with an *effectively exact* center, a layer-24 control removes 45.2% of raw MSE and deeper is monotonically better; with the plain closure center, deeper is monotonically *worse*. Centering accuracy, not correlation with the output, is the bottleneck (5e verbatim).
- **7a, upgraded to a sequence model:** a frozen ~52k-parameter GRU, trained offline on thousands of MLPs to predict the closure's centering error from the moment state the anchor already tracks (deliberately excluding the sampled-mean gap so it can't erase the control's signal), cuts raw MSE 17–19% held-out. A later diagnostic measures the residual contamination mechanism the census's 5g warned about: same-batch features let the learned correction align with layer-22 sample noise at ρ = 0.885, and an (unavailable) across-batch average would improve MSE a further 22.6% — the cleanest measurement yet of *why* controls must be deterministic.

- **6 (a real billing bug, disclosed):** `fnp.tensordot` with contracted rank > 52 bills multiplies only — the adds in each FMA are free, ≈ 2× effective sample budget. mliston's public-board best (8.62e-8, our census rank 11, sub 327734) **uses this bug, is disclosed as invalid, and was not among their selections**; AIcrowd/@mohanty are aware. Read with qi_zhang5's no-arbitrage audit (which checked standard routes at rank ≤ 52 and found ratio 1.000): the audit and the bug are compatible — the arbitrage lives past the audit's rank horizon. **Census consequence: the frozen public board's rank 11 is admitted-invalid**, and any "documented field reaches rank N" arithmetic should skip it.

New source: violeta, topic 18180 (no PDF — post + GitHub repo `01-1/arc-wbe`). Sixty-nine closed research directions with the measurements that closed them, and three census-grade items:

- **amalgonim's score identity (18151 item 5), finally tested rather than assumed:** four paired 100-MLP ladders at n = 4k…32k fit MSE = F + c·n^(−α) with **α = 1.039** (90% CI [0.840, 1.240]) and floor F bounded below 1.5e-7, consistent with zero — *for a sampler whose only closure sits at layer 1*. Read against amalgonim's 8.8e-7 closure-family floor, the synthesis is sharp: **the floor tracks where the analytic approximation sits — closure at the readout floors; closure at the first layer inside a sampler doesn't.** Predictions were computed from the three-point fit before the fourth point ran; three points admitted three different verdicts ("fitting three parameters to three points identifies nothing").
- **A screening rule cheaper than any gate:** convert a candidate's variance saving into the bias budget it can afford (3.2% variance saving tolerates $b^2 < 8.9\text{e-}9$ at full budget). Their odd-state Rao–Blackwell had genuinely 0.564× the variance and still died at 2,000× its bias allowance.
- **A bug warning for everyone running replicate gates:** three of their gate aggregators mislabeled M_3 (MSE of the replicate mean) as b^2; the correct three-replicate decomposition is $\sigma^2 = (3/2)(M_1 - M_3)$, $b^2 = (3/2)M_3 - (1/2)M_1$. At least one gate verdict inverted. (Their repo ships raw rows beside every verdict for exactly this reason.)
- Also: tools (a GPU-accelerated flopscope fork; a fly.io 100-machine parallel evaluator), and a documented reward-hacking incident log — GPT 5.5 fitting to the public test cases, caught by a second LLM audit; local test-case generation banned thereafter. Production-modes material.

New source: omer_kiraz, topic 18181 (submission 327792; adjusted 3.327e-7, raw 1.062e-6). A deliberately simple 20,000-point sqrt-prime rank-1 RQMC lattice, characterized across a 28-submission campaign. Three useful instrument points:

- **The compute–accuracy frontier has an interior optimum:** a 12k→40k sample sweep shows raw MSE falling monotonically while adjusted score is minimized near 20k (40k: raw −42%, adjusted +15%) — the cleanest published demonstration that the official metric is an objective in its own right.
- **2e gets a genuine counterexample on lattices:** replacing half the lattice points with exact antithetic mirrors *worsened* raw MSE 8.49e-7 → 1.09e-6 at equal compute — on a rank-1 lattice, distinct low-discrepancy points beat mirror pairs (the odd-order cancellation is already largely provided by the lattice). Consistent with qi_zhang5's CV collapse and the census's "the cheap channels are already spent on good carriers," but new as a *negative* for antithetics specifically — the census's 2e range (1.007–1.05×, on plain MC) does not transfer to lattice carriers.
- **Generator sensitivity is first-order:** a Roberts generalized-golden-ratio vector at identical N and compute scored 2× worse raw than sqrt-prime — rank-1 lattice quality is strongly generator-dependent at finite N.

Two companions, for the record. konstantin_baltsat turned the 18149/18159 atlas into an

executable Kaggle visual notebook (18163) — the C2/6f author's full idea map, runnable without challenge data. bgrubbs1984 posted a public companion site for submission #326948 (18169; "a 332-submission campaign case study") with figures, notebooks, and history, offered free for reuse. And in the skibidi thread (18166), Cipo reports independently reproducing the hot/cold approach for a **6% multiplier reduction at similar MSE on 100 fresh held-out nets** — the batch's one new cross-team replication of a published technique.

New source: andrei_bulzan, topic 18183 (submission 327801; adjusted 1.14330e-7, raw 2.20682e-7, C/B 0.51729) — an 82-page report, the corpus's second-largest. Built as a from-first-principles exposition (readable by a non-specialist through page 25) that lands on a genuinely distinctive construction: a **particle-repair hybrid** — the complete Kerdock/MUB-129 cloud propagated through the realized network, with the analytic moment machinery demoted to making *small, trusted, in-flight repairs* to the cloud rather than predicting anything itself. The repair schedule is graded by evidence quality: layer 1 exact (affine-standardize each neuron's particle distribution to its closed-form Gaussian moments), layer 2 cautious (8.75% of the mean gap, 11.25% of the spread gap toward moment-matched targets), layer 4 an Edgeworth-style shape correction, then a width schedule (256 → 232 → 216 → 200 coordinates, energy-screened per network with the omitted coordinates' means folded back in to preserve gate operating points), and at the endpoint a **1.9% pull** of the sampled signed preactivation mean toward an analytic proxy, applied *before* the final ReLU so particles can cross the gate. Item-level effects:

- **C3/5 (a new job for the analytic chain: repair, not predict).** The census recorded the field reassigning closures to router/control/null roles; this adds a fourth — *in-flow repair target* — with the trust dial made explicit and measured (exact where exact, 1.9–11% where approximate). The endpoint pull + one global output scale bought a cross-half-validated **1.367%** pooled raw reduction at zero added counted work. And the proxy's calibration is an independent refit of the census's O1 constant: **0.9927832064** (pscamillo's 0.992 scale bias, third appearance).
- **3a (the field's third full variance ladder).** IID 7.4115e-7 → antipodal 6.4919e-7 (1.14×) → matched orthonormal 5.9635e-7 (1.09×) → complete production system 2.0905e-7 (2.85×, jointly: Kerdock arrangement + repairs + readout).
- **6 (three instrument facts).** (i) The grader's remote report exposes a `sampling_mse` field — the grader **runs its own reference Monte-Carlo comparator per submission** (mean 6.46947e-7 on the public 50, reproduced within 0.3% by ~74,752 antipodal rows) — which independently confirms qi_zhang5's observed pure-sampling plateau (6.47e-7) from the instrument side. (ii) The per-net timing split: 40.8s trusted backend + 5.09s FlopScope overhead + **0.171s residual — 0.371% of wall time yet 12.1% of effective compute** at the 1e11 conversion; "large enough to conceal a sub-one-percent source improvement in a noisy grading window." (iii) A **rank-40 rectangular fast- multiplication identity (Tichavský, arXiv:2104.05323)** enters the field's toolbox beyond everyone else's 2×2 Strassen/Winograd: 54 → 40 block products, nested twice over rank-7 leaves = 48% of ordinary leaf products (82.08B → 49.45B including transforms), with the whole execution layer (cells, waves, views, destination-backed writes) delivering bit-identical predictions at 0.65–0.79% lower effective compute.
- **The basis-reweighting oracle (O6/O8, now measured twice).** Their page-70 experiment tests per-network *unequal weights* over the 129 completed basis endpoints: a target-aware oracle removes **95.64%** of pooled error, held-output cross-fitting retains **81.82%** of that (≈ 5.5× equivalent) — but the lawful analytic-proxy direction is **cosine 0.018** to the oracle direction and removes only **1.115%**. Their stated open problem: "completed basis endpoints contain contrast directions with substantial target-aware capacity; our lawful analytic observables did not identify the target-side contraction… solving that contraction, or replacing it with an unbiased low-variance audit, remains the most direct route to a larger accuracy gain."
- **Retired branches worth the census's memory:** a complete complementary dual

frame was the strongest raw-accuracy branch observed anywhere in the published field (2.565e-7 → **1.395e-7**, 1.84×) and died purely on cost (246.8B ≈ 2× budget; a compressed variant at 212.4B still didn't pay); a weight-adapted Schur-pair frame gained 8.24% in development and **failed transfer** (54/100 rows, remote regression) — another O8 casualty; a 64-regime late-state atlas missed its error budget by **7,799×** ("common late states were easy to describe; the tiny signed differences between them still controlled the final mean" — trim_qewas's cell-conditioning negative, independently rediscovered); and an oracle-first research rule ("try oracles before dedicating compute") matching 18182's protocol.

- **Production-modes material:** the LLM section is the field's most candid — "billions of tokens" across model families, "I have not touched one line in Visual Studio Code throughout this competition," the orchestrator-vs-sidekick meditation, a warning that LLM loops stall in local minima, and the observation that the field's convergence on similar constructions may itself be a model-driven effect. Proposes the challenge as a standing LLM eval ("lowest adjusted score attained in 4 hours").

Convergence and open-problem deltas.

- **C2 (amended, with a wrinkle).** The 66,048-point construction now has **four** independent complete builds (nygaard 18145, baltsat 18149, kaileh57 18171, Team Puffi 18175 — upgrading what 18152's forum notes had only sketched) plus a near relative (pranay212's BCH strength-4 OA, 18173) — the field's standard carrier, and per kaileh57's bound, nearly the *optimal* one in its class. The wrinkle: ely2sh (18176) shipped the **128-basis** variant, audited it, and proved the convergence was never about 5-design exactness at all — the 129th basis is worth ≤1% (their measurement) or ~19% (Puffi's ladder; see the discrepancy above), while the MUB property itself is the theorem-certified mechanism. "Two teams built the identical construction" survives; *why* it wins is now a theorem instead of a heuristic.
- **New convergences.** The oracle ceiling was measured twice through different instruments and agrees: keenanpepper's true-(μ, M_2)-at-every-layer recursion bottoms at 8.1e-7 raw / worth ~2× adjusted (18177), and trim_qewas's allm substitution prices the perfect two-moment closure at ~5th place, ×7.7 short of #2 (18182) — the two-moment class is exhausted below the podium, by agreement. Second: the balanced-cancellation account of tuned closures now has three independent statements (jamesrahenry's 16:1 in 18097 → census 1c; trim_qewas's $R^2 = 0.836$-predictor-worth-nothing and 8.5×-worse true diagonal, 18182; keenanpepper's Wall 5 compensator result, 18177). Third: violeta's no-floor measurement ($\alpha = 1.039$) combined with amalgonim's 8.8e-7 closure floor yields a synthesis neither team stated alone: the floor tracks where the analytic approximation sits.
- **O4 (partially answered).** The census asked what could satisfy the oracle- derived conditions for a deterministic method at the front. 18182's §7 now *measures* the front: not moment-based (a full moment oracle is ×5.4 above #1), not layer-wise, not pair-coordinate, deterministic, and able to afford ~2.6e10 FLOPs of algebra. Combined with kaileh57's saturation bound (fixed positive angular rules are exhausted) and 18182's angular-oracle hint (an oracle on the angular factor alone beats their entire closure 1.83×), the surviving hypothesis space for the front is narrow: a deterministic, algebra-heavy, *adaptive or signed* angular object — exactly the O6/O8 direction.
- **New open problems. O6:** the LP bound covers fixed *nonnegative* rules — what do signed-weight or network-adaptive rules buy? (kaileh57's small search found no negative weights wanted; that is one parameterization, not a bound.) **O7:** how does the 0.0233% gap scale with support count — where is the support-optimal frontier for this kernel family? **O8:** a transferable, fixed-before-evaluation frame-selection rule — Puffi's oracles show small network-specific subsets with ~4× headroom exist, and their herding results show why observable-feature matching at the selection layer doesn't survive the later nonlinear layers (18175). Note O6 and O8 are the two faces of the same escape

route from kaileh57's saturation bound. **O9 (from 18177) — ANSWERED 2026-08-17** (jamesrahenry, in-thread on 18177): both effects real, width dominates monotonically; d*(w) ≈ <24/24/32/32–48/≥48 across w = 64…1024; Phase 1's (256, 32) sits at median FLOP-parity (1.04, CI [0.75, 1.38]). **O10 (from 18182):** does any deployable statistic reach the 6-dimensional Lyapunov error subspace that is computable from the weights alone and carries 68% of the final error?

- **What the census cannot see — RESOLVED BY THE RECOMPUTE (2026-08-18).** The public board was recomputed shortly before this update posted, and the change is surgical: the two anomalous front entries — the 9.5e-11 and 1.46e-9 adjusted scores whose raw MSEs sat in the low 1e-9s — are **gone from the board**, one of their authors reappearing at rank 32 with an ordinary 1.34e-7 entry; one other entry moved ~3% and every remaining score is unchanged. So the census's closing question ("is the silent front a missing technique or something the re-evaluation will reframe?") has an administrative answer: **the extreme front was removed, not explained.** The forensic literature reads correctly in hindsight — amalgonim's appendix ruled out every legitimate construction at that accuracy, qi_zhang5's audit cleared the standard metering routes, and mliston's disclosure proved at least one >2× arbitrage existed beyond the audit's horizon.
- **Which makes the new #1 the field's most interesting object.** ednacob's #327208 (adjusted 1.845e-8, **raw 3.63e-8**, ~0.51 multiplier, 100/100 clean, final-layer-only, cubature-class cost signature) was rank 3 all along and survived the recompute. Read it against this batch's theory: raw 3.63e-8 is **5.2× below the Kerdock-class raw (~1.9e-7)** — below what kaileh57's certificate permits any fixed *nonnegative* rule at that support, at a multiplier that rules out brute support scaling. And 5.2× is almost exactly the **signed per-network reweighting oracle ceiling measured twice in this batch** (bulzan's 81.8% cross-fitted retention, 18183 p70 ≈ 5.5×). If #327208 is what it appears to be, someone has found a lawful observable that solves the target-side contraction problem bulzan names as open — the single most valuable unpublished fact in the field. O6/O8 are no longer hypothetical escape routes; the board now contains an existence proof.

The census at a glance — restated at Phase 1 close. The original seven verdicts, updated by fourteen write-ups and the final rescore:

1. **Parametric distribution propagation** — still closed as a competitor, now with the ceiling measured from two directions: a *perfect* two-moment closure prices at ~5th place and a full moment oracle lands level with #2 (18182); the whole moment-state program converts to ~2× on the scored metric (18177). But "nothing left to try" is dead: the projection step, not the order, is the live variable (SSC's covariance-response closure, 18175), the closure floor tracks *where* the analytic sits — readout floors, first-layer-inside-a-sampler doesn't (18180 + 18151) — and the family found a fourth working job: in-flow repair target at explicitly graded trust (18183).

2. **Monte Carlo + variance reduction** — the plateau is now calibrated (V·c/B with V_iid ≈ 0.0405 predicts 6.24e-7 vs 6.47e-7 observed, 18179) and instrument-confirmed (the grader's own `sampling_mse` comparator reads 6.47e-7, 18183). Two corrections to the census's 2-family: antithetic pairing does NOT transfer to rank-1 lattices (measured negative, 18181), and the metering-arbitrage question is bracketed — standard routes audit clean (18179) while one real >2× bug existed past the audit's horizon, disclosed by its finder (18170).

3. **Deterministic spherical cubature** — "effectively solved" upgraded to solved-with-theorems: pairwise mutual unbiasedness is exactly optimal in-class at every depth (18176), the complete construction is within 0.0233% of *every* fixed nonnegative rule at its support (18171), and the mechanism is MUB-ness, not 5-design exactness — the missing 129th basis is worth ~1% (18176; vs ~19% in 18175's ladder — an open

discrepancy between those authors). Four complete builds plus a BCH cousin (18173). The class is exhausted; the escape routes are signed/adaptive designs (O6/O8) — for which the post-rescore board now holds an apparent existence proof at #1.

4. **Structure exploitation** — the dead/on/kink chassis became the field's shared execution layer (explicitly reused by 18173 and 18174); best published raw moved to 1.55e-7 via the OA-cubature hybrid (18173), mooting the old "cubature wins on compute, not accuracy" reading; the support-removal cliff is measured — rarely-firing units carry disproportionate downstream value (18176); and coordinate-level width schedules with mean-folding protect gates while narrowing (18183).

5. **Control variates & blending** — the mid-network channel is now the *only* live one, by theorem (input-polynomial controls of degree ≤5 are identically zero on the standard carrier, 18171) and by three independent collapses of cheap controls on good carriers (18154, 18179, 18181). Deployed at the frontier three ways: fitted-closure transport at 17.6% (18175), a GRU-centered layer-22 control (18170), a 1.9% pre-ReLU endpoint repair (18183). The binding constraint everywhere is 5e, centering fidelity. The oracle headroom of per-network *reweighting* was measured publicly: a target-aware combination of the 129 basis endpoints removes ~96% of pooled error and holds ~82% under cross-fitting, while every lawful observable tried points almost orthogonally to it (cosine 0.018 — 18183 p70; same wall as 18175's herding failure).

6. **Metering, score, and benchmark mathematics** — the score decomposes as $S = \bar{e}\cdot\bar{q} + \mathrm{Cov}(e,q)$ (18175); above the floor a pure sampler's score is budget-independent and the analytic advantage is a small-sample phenomenon (18182; independently re-derived in the 18177 thread's crossover measurement, where Phase 1's (256, 32) sits at median FLOP-parity); the suite runs 100 nets per submission — 50 public, 50 gated, identical seeds across submissions (18175, 18177); residual wall time is 0.371% of the clock but 12.1% of effective compute (18183); and the final rescore removed the extreme front outright.

7. **Learned corrections** — the trajectory-fitted-works / local-patches-die split now has mechanism: fidelity and value dissociate in both directions, corrections convert only where the residual is low-rank, the one converting write is a co-adapted compensator that truth-injection makes *worse*, and teacher-forced R^2 is not a gate (18177 Walls 2–5); per-instance adaptivity is not just a transfer trap but *unobservable from inside* — the optimal constant is linear in the realised noise, every internal statistic quadratic (18182).

The open mathematics, restated.

- **O1 — the 0.992 mean-drift saturation.** Still open, now with its *third* independent refit (0.9921 in 18063, the census's original; 0.99278 as a global affine calibration in 18183). Three teams converge on the constant; nobody has derived it from the propagation recurrence.
- **O2 — the anticorrelated-error mechanism theorem.** Still open on the covariance channel; the mean channel's unit-gain result (18154) stands, and 18182's balanced-cancellation measurements (a per-step error predictor at R^2 0.836 whose correction is worth exactly zero) sharpen what the theorem must explain.
- **O3 — the joint law of near-threshold neurons.** Still open; 18176's support-removal cliff and 18183's mean-fold gate-preservation add evidence that the kink-adjacent joint structure is where compression dies.
- **O4 — conditions for a deterministic method at the front.** *Partially answered:* the front is deterministic, non-moment, non-layer-wise, and algebra-affording (18182 §7) — and post-rescore, the surviving #1 sits exactly in the gap those necessary conditions leave open.

- **O5 — the label-floor discrepancy.** Still open, and sharpened: the grader's own comparator field (18183) now gives a third instrument reading to reconcile.
- **O6 — signed-weight rules** (new, from 18171's bound covering only nonnegative rules): what do signed designs buy? The oracle answer is now public — large (18183 p70) — but no average-case construction exists.
- **O7 — support scaling** (new): how does the 0.0233% in-class gap move with support count; where is the support-optimal frontier for this kernel family?
- **O8 — transferable frame selection** (new, from 18175's herding failure and 18183's Schur-pair transfer failure): a fixed-before-evaluation rule that captures any of the network-specific subset/reweighting oracle. Every observable tried so far points nearly orthogonally to the needed correction.
- **O9 — the Wu crossover.** *Answered* in the 18177 thread: both effects real, width dominates monotonically; d*(w) rises from <24 (w=64) to ≥48 (w=1024); Phase 1's (256, 32) sits at median FLOP-parity (1.04, CI [0.75, 1.38]); fitted corrections shift the corrected-analytic contour systematically deeper.
- **O10 — the low-dimensional error subspace** (new, from 18182): most of a deterministic method's final error concentrates in a subspace of dimension ~6 computable from the weights alone. Does any deployable statistic reach it — or is the target-side contraction of O8 the same obstruction wearing different clothes?

What the census could not see is no longer a mystery — it is a removed artifact, an unexplained #1, and a measured gap between oracle capacity and lawful observables. That gap is Phase 2's inheritance.

As before: claims carry their source topics, corrections from the authors are welcome, and all fourteen topics' disclosed LLM-assisted/human-directed production mode matches the field's now-standard practice.