Supplementary Information (Online Resource 1) for "Prospective Completion Prediction for Independent-Trial Cluster Search with Compositional State: Measurement Burden, a Pre-Outcome Separation Gate, and Structural Limits"

Min Su Yoo[1] and Hyunglip Bae[1*]

[1*]Department of Big Data Convergence, Chonnam National University, Gwangju, South Korea.

*Corresponding author(s). E-mail(s): hlbae@jnu.ac.kr;
Contributing authors: dbalstn1822@jnu.ac.kr;

Abstract

This author-supplied Online Resource contains the supplementary methods, complete diagnostic tables, execution-environment details, frozen-evidence chronology, and scoped post-hoc analyses supporting the associated manuscript submitted to *Cluster Computing*.

Keywords: Cluster computing, Performance prediction, Supplementary information, Reproducibility

Supplementary Information

Supplementary Figure S1: cyclic-placement heterogeneity

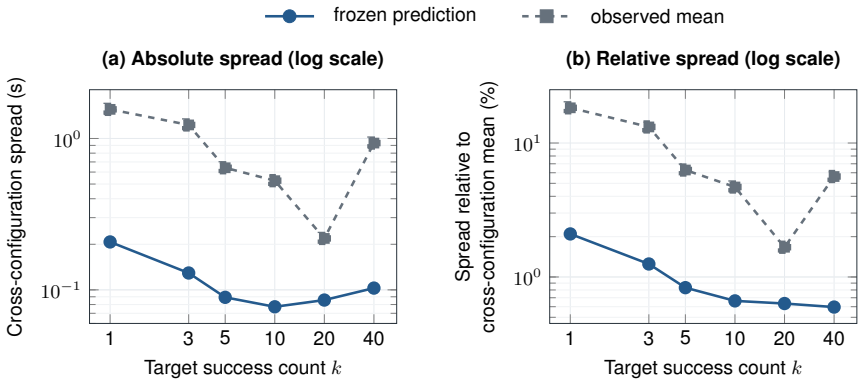
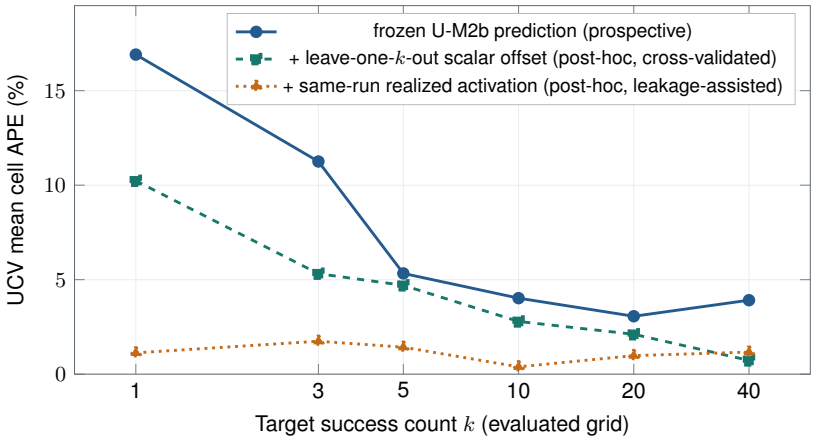


Fig. S1 Frozen-prediction and observed-mean cross-configuration spread across the 18 UCV cells. Absolute and relative spreads show that prediction-side configurations are closer than observed completion summaries at several target success counts. Cyclic placement identity is descriptive; the pattern does not identify a physical causal mechanism.

Supplementary Figure S2: UCV residual structure by k



Frozen prospective error is concentrated at the lowest k. Applying a held-out scalar offset per placement reduces error and decreases monotonically over the evaluated grid under the weighted-L1 estimator, but it consumes held-out cell outcome means and is a post-hoc cross-validated diagnostic rather than prospective validation. Substituting each evaluated run's own realized activation removes most of the remaining residual at every k; that series is leakage-assisted, non-deployable, and explanatory only. No series identifies a causal mechanism, and the comparison is not a causal decomposition of frozen error into activation and capacity parts.

Fig. S2 UCV mean cell APE by target success count k under three information conditions. The prospectively frozen U-M2b series is the only prospective curve. The leave-one-k-out scalar-offset series consumes held-out cell outcome means and is a post-hoc cross-validated diagnostic; under the weighted-L1 estimator it decreases monotonically over the evaluated grid. The same-run realized-activation series replaces the frozen activation input with the evaluated run's own activation and is post-hoc, leakage-assisted, non-deployable, and explanatory only. Frozen error is concentrated at the lowest k, and the scalar offset transfers least well there. No series identifies a causal mechanism and the comparison is not a causal decomposition of the frozen error.

Supplementary Table S1: complete UCV cells

Table S1: Complete UCV cell-level frozen prediction and outcome matrix.

Placement	k	Observed (s)	U-M1 (s)	U-M2b (s)	U-EQ (s)	U-M2b APE (%)
U6-A	1	7.680	9.777	9.763	9.763	27.12
U6-A	3	8.888	10.294	10.260	10.259	15.43
U6-A	5	9.828	10.726	10.673	10.673	8.60
U6-A	10	10.982	11.718	11.616	11.616	5.77
U6-A	20	13.076	13.673	13.469	13.469	3.01
U6-A	40	16.099	17.583	17.175	17.174	6.68
U6-B	1	8.588	9.946	9.934	9.934	15.68
U6-B	3	8.957	10.389	10.358	10.358	15.64
U6-B	5	10.266	10.786	10.736	10.736	4.58

Placement	k	Observed (s)	U-M1 (s)	U-M2b (s)	U-EQ (s)	U-M2b APE (%)
U6-B	10	11.142	11.764	11.664	11.664	4.69
U6-B	20	13.023	13.719	13.520	13.519	3.82
U6-B	40	16.629	17.629	17.230	17.230	3.61
U6-C	1	9.238	9.982	9.970	9.970	7.93
U6-C	3	10.117	10.417	10.389	10.389	2.68
U6-C	5	10.469	10.810	10.763	10.763	2.81
U6-C	10	11.509	11.787	11.693	11.694	1.60
U6-C	20	13.241	13.742	13.555	13.555	2.37
U6-C	40	17.032	17.653	17.278	17.278	1.45

Supplementary results: UCV stratified and model diagnostics

The frozen U-M2b pooled signed mean-ratio error by k was +16.32%, +10.89%, +5.27%, +3.99%, +3.06%, and +3.86% over the registered grid. This is a pooled signed mean-ratio series, not MPE or MAPE. The corresponding mean-cell APE values were 16.91%, 11.25%, 5.33%, 4.02%, 3.06%, and 3.91%. Across cyclic configurations, displayed prediction spread ranged from 0.60% to 2.10%, whereas observed spread ranged from 1.67% to 18.32%. These are descriptive patterns and do not identify a physical placement mechanism.

In a post-hoc descriptive paired comparison, M2b had lower APE than M1 in all 18 cells. The mean difference was 0.96 percentage points with interval 0.62–1.33 points. This does not establish prospective model superiority.

Supplementary Table S2: activation diagnostics

Table S2: Exported UCV activation diagnostics.

Diagnostic	Stratum	Frozen MAPE (%)	Diagnostic MAPE (%)	Boundary
Same-run activation by k	1	16.91	1.13	Post-hoc, leakage-assisted, non-deployable
Same-run activation by k	3	11.25	1.74	Post-hoc, leakage-assisted, non-deployable
Same-run activation by k	5	5.33	1.43	Post-hoc, leakage-assisted, non-deployable
Same-run activation by k	10	4.02	0.39	Post-hoc, leakage-assisted, non-deployable
Same-run activation by k	20	3.06	0.98	Post-hoc, leakage-assisted, non-deployable

Diagnostic	Stratum	Frozen MAPE (%)	Diagnostic MAPE (%)	Boundary
Same-run activation by k	40	3.91	1.17	Post-hoc, leakage-assisted, non-deployable
Forward-10	Overall	9.82	5.05	Post-hoc temporal-transfer
Forward-10	U6-A	14.15	6.08	Post-hoc temporal-transfer
Forward-10	U6-B	11.13	6.05	Post-hoc temporal-transfer
Forward-10	U6-C	4.20	3.02	Post-hoc temporal-transfer
Same-run activation, pooled	All cells	7.42	1.14	Post-hoc, leakage-assisted, non-deployable; reduction 6.28 pp (84.63%)

Supplementary results: activation and environmental diagnostics

The same-run realized-activation values by k were 1.13%, 1.74%, 1.43%, 0.39%, 0.98%, and 1.17%. The last-activator distribution also shifted: slot 3 frequency changed from 98.33% in frozen characterization to 60.56% in realized completion, and worker 6 changed from 66.67% to 32.22%. These patterns are post-hoc and do not identify a scheduler, hardware, or transport mechanism.

The placement-pooled leave-one-run-out diagnostic was frozen as the sole cross-run rule before its performance was inspected. For each accepted run, it retains worker/slot identity and exact frozen U-M2b capacity while averaging the analytic expectation over all 119 other realized activation vectors from the same placement. The evaluated run is excluded, and completion-outcome magnitude does not enter construction of the mixture. Cell-mean MAPE was 1.46%; placement MAPEs for U6-A, U6-B, and U6-C were 1.78%, 1.43%, and 1.18%. Mean cell APE by $k = 1, 3, 5, 10, 20, 40$ was 2.06%, 2.43%, 1.85%, 0.48%, 0.58%, and 1.37%, respectively. All 18 cell predictions and APEs appear in Supplementary Table S12. The diagnostic remains post-hoc and non-deployable; it is not prospective evidence or a causal decomposition. No alternative cross-run pooling or single-vector rule was evaluated.

The two-by-two state diagnostic uses all 360 accepted completion runs. Each of their 2,160 worker records contains ten consecutive 0.5-second detector buckets, a fully populated trailing window, and a steady declaration before the target- completion stopping event. The capacity substitution is the preserved trailing median of those ten detector-bucket throughputs. It is a same-run steady- declaration-window input, not average throughput over the full productive search interval and not a quantity formed by dividing attempts by target stopping time. All four activation–capacity

combinations use the same analytic evaluator and cellwise aggregation. Their MAPEs are 7.42% (frozen activation, frozen capacity), 1.14% (same-run activation, frozen capacity), 7.35% (frozen activation, realized capacity), and 1.15% (same-run activation, realized capacity). The grid is post-hoc state sensitivity, not causal or percentage attribution and not a deployable estimator.

Productive-search counters and terminal-buffered trajectory buckets continue after steady declaration, but their availability depends on the stopping event. One accepted worker has no productive trajectory bucket, and 668 of 2,160 workers have no complete productive bucket. The campaign therefore does not preserve an independently predefined, outcome-independent fixed throughput window during productive search for every required worker and run. No new window or search-interval two-by-two diagnostic is defined post hoc, so the detector-window diagnostic cannot exclude later within-run throughput changes.

A distinct stopping-conditioned diagnostic remains computable from fields preserved for every worker. For worker w, its productive-attempt count is divided by the same-worker interval from steady declaration to receipt of the target-dependent stop. The six rates are summed for a run and divided by the exact frozen U-M2b capacity total for its placement. Supplementary Table S13 reports all 18 strata. Ratios are elevated and dispersed in several short low-k intervals, so they are not a fixed-window capacity estimate. In the post-hoc late-k subset, $k \in \{20, 40\}$, the mean/median ratios were 1.001/1.001 for U6-A, 1.001/1.003 for U6-B, and 1.005/1.008 for U6-C. This descriptive result does not support a large persistent U6-A average-rate excess of the magnitude suggested by the earlier two-endpoint slope point estimate. The registered endpoint-slope interval still includes zero; the two quantities are not mathematically equivalent, and neither supplies causal attribution or an independent capacity validation.

For descriptive context across separate phases, the P4-A same-run diagnostic changed M2b MAPE from 2.67% to 0.90%, while the UCV diagnostic changed it from 7.42% to 1.14%. The registered UCV/P4-A ratios were 2.78× under frozen activation and 1.26× under same-run activation. These outcome-conditioned ratios are not a controlled between-phase comparison, causal decomposition, corrected prospective result, or proof of capacity transfer.

The initial, preceding, and later non-target capacity checkpoints had means of 81,216,526.33, 81,784,301.80, and 81,446,372.11 attempts/s. The later checkpoint was −0.41% relative to the preceding checkpoint and +0.28% relative to the initial checkpoint. This is later environmental context over the raw observed interval; it does not establish absence of drift or explain UCV causally.

Post-hoc outcome-free transient-capacity sensitivity

The primary prediction assigns each modeled worker the equal share $C(\text{slot}, n_f)/n_f$ of independently measured node-total capacity at final local occupancy n_f, from activation onward. This is a representation by construction rather than a direct measurement of the physical per-worker rate vector. The reviewer-closure sensitivity instead updates the active workers on each node between activation events: at current active occupancy n, each is assigned the same equal-share construction $C(\text{slot}, n)/n$. The analysis used

12 measured slot–occupancy capacity states, 60 complete frozen startup vectors, the frozen target mapping, and the success probability 2^{-24}. It read zero target-outcome bytes and did not alter a frozen prediction or prospective result.

Across the 18 cells, the primary and alternative constructions produced 36 values. For the prespecified low-k set $\{1, 3, 5\}$, the largest change was at U6-B, $k = 1$: transient minus fixed was $-1.0221417205968919 \times 10^{-5}$ s, or $-0.00010288837017450792\%$. Thus the maximum magnitude was 10.221417205968919 microseconds. A separate adaptive-quadrature implementation checked all 36 values with zero failures; the maximum absolute disagreement was $6.02753402745293 \times 10^{-11}$ s. This is a *post-hoc outcome-free sensitivity*, not a prospective primary analysis, direct transient-throughput measurement, or universal transient-capacity bound. It neither excludes capacity-dynamic explanations nor identifies activation mismatch as the physical mechanism for the residual.

Supplementary Table S3: P4 secondary diagnostics

Table S3: P4 precision and A1-prime secondary diagnostics. A1-prime is the preserved post-hoc H1-minus-M1 contrast at the 20-repetition budget; its historical resolution threshold was not met.

Metric	Value	Unit	Evidence class
Relative SEM minimum	0.69	percent	Post-hoc outcome-independent
Relative SEM maximum	2.30	percent	Post-hoc outcome-independent
A1-prime effect	-2.62	percentage points	Post-hoc outcome-conditioned
A1-prime bootstrap SE	1.33	percentage points	Post-hoc outcome-conditioned
A1-prime interval lower	-5.30	percentage points	Post-hoc outcome-conditioned
A1-prime interval upper	-0.14	percentage points	Post-hoc outcome-conditioned
Absolute effect / SE	1.97	ratio	Post-hoc outcome-conditioned

Supplementary Table S4: capacity reuse and burden by scope

Table S4: P4 characterization reuse and burden detail, separated by burden **scope**. Worker-hours are scientific characterization measurement burden, not billed cloud cost. TOTAL REQUIRED CHARACTERIZATION is the union of unique acquisition records the representation requires over all nine P4-A target configurations; capacity-only is a strictly narrower scope. The total and capacity-only reductions are different quantities and must not be interchanged.

Metric	Value	Unit	Scope
M1 characterization burden	25.50	worker_hours	Total required
M2a characterization burden	25.50	worker_hours	Total required
M2b characterization burden	16.44	worker_hours	Total required
M1 capacity characterization	20.35	worker_hours	Capacity only
M2a capacity characterization	20.35	worker_hours	Capacity only
M2b capacity characterization	11.29	worker_hours	Capacity only
M1 activation characterization	5.15	worker_hours	Activation only
M2a activation characterization	5.15	worker_hours	Activation only
M2b activation characterization	5.15	worker_hours	Activation only
M2b reduction vs M2a	35.52	percent	Total required
M2b reduction vs M2a	44.51	percent	Capacity only
Maximum capacity-composition discrepancy	0.45	percent	Composition check
Relative SEM minimum	0.69	percent	Finite-repetition precision
Relative SEM maximum	2.30	percent	Finite-repetition precision

Supplementary Table S5: retrospective fitted grid

Table S5: Complete retrospective fixed-budget fitted-comparator grid. Budget is the number of frozen-order completion records used per training cell; burden is rounded to the nearest worker-second for display. Model and crossing definitions appear in Supplementary Methods.

Model	Budget label	Mean fold burden (worker-s)	LOCO MAPE (%)
C0	1	4,682	11.41
C0	2	9,386	11.44
C0	5	23,478	11.24
C0	10	46,745	11.16
C0	20	93,350	11.21
C1	1	4,682	7.67
C1	2	9,386	7.69
C1	5	23,478	7.58
C1	10	46,745	7.52
C1	20	93,350	7.57
C1-I	1	4,682	7.40
C1-I	2	9,386	8.01
C1-I	5	23,478	7.44
C1-I	10	46,745	7.59
C1-I	20	93,350	8.03
H1	1	4,682	3.21
H1	2	9,386	2.80
H1	5	23,478	2.75
H1	10	46,745	2.61
H1	20	93,350	2.63
Normative M2b	–	–	2.67

C1: no evaluated crossing; none of its five tested budget points matched the M2b reference.
C1-I: no evaluated crossing; none of its five tested budget points matched the M2b reference.

Supplementary Table S6: completion finite-repetition uncertainty

Table S6: Completion finite-repetition uncertainty by phase. All quantities are descriptive summaries of within-cell variation across the 20 accepted completion repetitions per cell; they are not confidence statements about prediction error and no significance test is attached. Cell means are therefore not uncertainty-free, and the interpretation classification for this evidence is QUALITATIVE WITH UNCERTAINTY OVERLAY.

Population	Quantity	Minimum	Maximum	Unit
P3 (36 cells)	Coefficient of variation	2.30	17.59	percent
P3 (36 cells)	Relative SEM of cell mean	0.51	3.93	percent
P4-A (36 cells)	Coefficient of variation	3.08	10.30	percent
P4-A (36 cells)	Relative SEM of cell mean	0.69	2.30	percent
UCV (18 cells)	Coefficient of variation	6.03	25.46	percent
UCV (18 cells)	Relative SEM of cell mean	1.35	5.69	percent
All-placement UCV endpoint-slope overlays, $k = 10 \rightarrow k = 40$ (block-aware paired resampling, 50,000 draws)				
U6-A observed slope	Point; descriptive 95% interval	0.1706	[0.1481, 0.1934]	s per unit k
U6-A frozen predicted slope	Point		0.1853	s per unit k
U6-A observed minus predicted	Point; descriptive 95% interval	−0.0147404	[−0.0372, 0.0081]	s per unit k
U6-B observed slope	Point; descriptive 95% interval	0.1829	[0.1549, 0.2136]	s per unit k
U6-B frozen predicted slope	Point		0.1855	s per unit k
U6-B observed minus predicted	Point; descriptive 95% interval	−0.00260289	[−0.0306, 0.0281]	s per unit k
U6-C observed slope	Point; descriptive 95% interval	0.1841	[0.1547, 0.2136]	s per unit k
U6-C frozen predicted slope	Point		0.1861	s per unit k
U6-C observed minus predicted	Point; descriptive 95% interval	−0.00205142	[−0.0315, 0.0275]	s per unit k
All three slope-mismatch intervals include zero. No statistically established capacity-side anomaly is claimed, and these overlays are not hypothesis tests.				

Across completion cells, the coefficient-of-variation ranges were 2.30–17.59% in P3, 3.08–10.30% in P4-A, and 6.03–25.46% in UCV. The corresponding relative-standard-error ranges for the cell means were 0.51–3.93%, 0.69–2.30%, and 1.35–5.69%. For the U6-A, U6-B, and

U6-C endpoints from $k = 10$ to $k = 40$, the observed-minus-predicted slope mismatches were -0.0147404, -0.00260289, and -0.00205142 s/success, respectively. All three descriptive mismatch intervals include zero, so none establishes a capacity-side anomaly or statistical significance.

Post-hoc completion-level interpretation diagnostic

Every result in this subsection is a post-hoc interpretation diagnostic. For each accepted completion run, the frozen model was evaluated at the observed activation vector and again after setting all activation times to zero. Their difference is the model-implied all-active activation effect. Across target levels, its median fraction of predicted completion time ranged from 86.0–87.8% across phases and placements; the minimum-k range was 98.0–99.1% and the maximum-k range was 55.6–62.0%. This is a counterfactual within the frozen model, not an additive decomposition of observed completion time. Prospective total-time MAPE is unchanged and remains the performance authority.

For each placement, increments were formed relative to its smallest target, $\Delta T_k = T_k - T_{k_0}$, for both observed cell means and frozen predictions. Median increment-error ranges were 8.07–8.13% in P3, 4.43–5.95% in P4-A, and 27.80–31.32% in UCV; all 216 observed increment denominators were nonzero. A common shift $a_i \mapsto a_i + c$ shifts the modeled mean and distribution by c while preserving relative staggering, a property verified numerically over the diagnostic rows with maximum residual $6.661338147750939 \times 10^{-16}$ s. Thus increments remove a common timing origin but retain differences in stagger. These outcome-informed diagnostics are interpretive only: they are not prospective validation, causal attribution, or replacements for any primary result.

Supplementary Table S7: post-hoc offset calibration and leave-one-k-out

Table S7: Post-hoc residual calibration and leave-one-k-out diagnostics on the 18 UCV cells. Every row consumes held-out cell outcome means and is therefore a post-hoc diagnostic, never prospective validation and never a corrected prospective performance figure. Offsets are additive scalars in seconds obtained as $1/T$-weighted medians of the residuals under the SMALLEST MINIMIZER tie rule; the in-sample rows fit and evaluate the offset on the same cells, whereas the leave-one-k-out rows evaluate a held-out k within each placement.

Diagnostic	Quantity	Value	Evidence class
Global scalar offset (single offset for all 18 cells)			
Global weighted-L1 offset	Offset (s)	0.6338	Post-hoc in-sample
Global-offset corrected error	MAPE (%)	3.95	Post-hoc in-sample
Global-offset corrected error	MPE (%)	1.61	Post-hoc in-sample
Placement-level offsets (one offset per cyclic placement, in-sample)			
U6-A offset	Offset (s)	1.0760	Post-hoc in-sample
U6-B offset	Offset (s)	0.6007	Post-hoc in-sample
U6-C offset	Offset (s)	0.2941	Post-hoc in-sample
Placement-corrected error, all cells	MAPE (%)	3.04	Post-hoc in-sample
Placement-corrected error, U6-A	MAPE (%)	4.67	Post-hoc in-sample

Diagnostic	Quantity	Value	Evidence class
Placement-corrected error, U6-B	MAPE (%)	3.40	Post-hoc in-sample
Placement-corrected error, U6-C	MAPE (%)	1.06	Post-hoc in-sample
Leave-one-k-out weighted-L1 offset (held-out k within placement)			
Held-out error, all cells	MAPE (%)	4.31	Post-hoc cross-validated
Held-out error, all cells	MPE (%)	1.12	Post-hoc cross-validated
Held-out error, U6-A	MAPE (%)	6.79	Post-hoc cross-validated
Held-out error, U6-B	MAPE (%)	4.99	Post-hoc cross-validated
Held-out error, U6-C	MAPE (%)	1.16	Post-hoc cross-validated
Held-out error, $k = 1$	Mean APE (%)	10.23	Post-hoc cross-validated
Held-out error, $k = 3$	Mean APE (%)	5.32	Post-hoc cross-validated
Held-out error, $k = 5$	Mean APE (%)	4.70	Post-hoc cross-validated
Held-out error, $k = 10$	Mean APE (%)	2.79	Post-hoc cross-validated
Held-out error, $k = 20$	Mean APE (%)	2.11	Post-hoc cross-validated
Held-out error, $k = 40$	Mean APE (%)	0.73	Post-hoc cross-validated
Residual-mean estimator sensitivity (same procedure, arithmetic-mean residual instead of weighted median)			
Held-out error, all cells	MAPE (%)	3.86	Post-hoc cross-validated sensitivity
Held-out error, all cells	MPE (%)	0.66	Post-hoc cross-validated sensitivity
Held-out error, $k = 1$	Mean APE (%)	9.50	Post-hoc cross-validated sensitivity
Held-out error, $k = 3$	Mean APE (%)	4.30	Post-hoc cross-validated sensitivity
Held-out error, $k = 5$	Mean APE (%)	2.39	Post-hoc cross-validated sensitivity
Held-out error, $k = 10$	Mean APE (%)	3.14	Post-hoc cross-validated sensitivity
Held-out error, $k = 20$	Mean APE (%)	3.09	Post-hoc cross-validated sensitivity
Held-out error, $k = 40$	Mean APE (%)	0.74	Post-hoc cross-validated sensitivity

Under the weighted-L1 estimator the held-out by-k mean APE decreases monotonically over the evaluated grid. The residual-mean estimator reproduces the same concentration of error at the lowest k without a monotone ordering, so the low-k concentration is the estimator-robust feature and the monotone ordering is specific to the weighted-L1 estimator. Neither row set identifies a physical mechanism.

The global in-sample weighted-median offset was 0.6338 s, reducing MAPE from 7.42% to 3.95% and leaving MPE at +1.61%. Placement-specific offsets for U6-A, U6-B, and U6-C

were 1.0760, 0.6007, and 0.2941 s, respectively, and their pooled corrected MAPE was 3.04%. Leave-one-k-out evaluation produced 4.31% MAPE and +1.12% MPE overall; configuration MAPEs were 6.79%, 4.99%, and 1.16%, and the by-k MAPEs were 10.23%, 5.32%, 4.70%, 2.79%, 2.11%, and 0.73%. The arithmetic-mean-offset sensitivity result was 3.86% MAPE. All of these results consume completion outcomes and remain post-hoc; the in-sample variants do not estimate transfer, and the held-out variant is not prospective validation.

Supplementary methods: finite-repetition uncertainty

Each completion cell contains 20 accepted repetitions. Cell means are formed by summing authoritative integer completion nanoseconds and dividing once by $n \times 10^9$, then converting to binary64; consequently the values reported here differ from earlier serializations of the same cell means only at floating-point precision, and the historical exact values remain the registered authority for those cells. The sample standard deviation uses denominator $n - 1$. The coefficient of variation is the sample standard deviation divided by the cell mean, the standard error of the mean is the sample standard deviation divided by $\sqrt{n}$, and the relative standard error is that quantity divided by the cell mean.

These completion-only finite-repetition overlays are distinct from the primary UCV joint bootstrap described in the main Methods. They use 50,000 nonparametric resampling draws with the 2.5th and 97.5th percentiles taken by linear quantile interpolation, under a resampling contract frozen before any interval was inspected. The resampling unit is block aware and differs by phase. In P3 and P4-A each draw independently resamples the 20 accepted repetitions with replacement within each cell. In UCV each draw samples the ordered set of 20 complete protocol blocks once with replacement and applies that identical sampled block-index vector to every UCV cell, so completion runs are never treated as independent across cells and the paired block structure of the UCV campaign is preserved. The U6-A, U6-B, and U6-C endpoint-slope and slope-mismatch overlays reuse the same paired UCV block draws that produced each placement's $k = 10$ and $k = 40$ cell-mean overlays, so every endpoint pair is resampled jointly rather than independently. The overlay uses NumPy `Generator(PCG64)` with fixed seed `2026081401` and a fixed phase and cell consumption order, so it is reproducible from the archived contract.

These overlays are descriptive. They quantify how much a cell mean or an endpoint slope moves under resampling of the observed completion structure. They are not confidence intervals for prediction error, they are not hypothesis tests, no formal significance is computed from them, and they are not used for model comparison. Every endpoint-slope mismatch interval contains zero, so the corresponding statements in the main text remain qualitative.

Supplementary methods: post-hoc offset calibration

For cell residuals $e_i = \widehat{T}_i - T_i$ with observed cell means T_i, an additive offset δ minimizes the $1/T$-weighted absolute deviation

$$\sum_i |e_i - \delta| / T_i.$$

The minimizer is the $1/T$-weighted median of the residuals. When the weighted median is attained on an interval rather than at a point, the frozen tie rule selects the smallest minimizer, so the reported offset is uniquely determined. The weighted median is used rather than the mean because it is the natural minimizer of the weighted absolute-error objective and is insensitive to a small number of extreme cells; an arithmetic-mean residual variant is reported as a sensitivity check, and it minimizes a weighted squared-error objective instead.

Three scopes are reported. The global variant estimates one offset from all 18 cells and evaluates on the same cells. The placement variant estimates one offset per cyclic placement

from that placement's six cells and evaluates on the same six. Both are in-sample and therefore measure how much structure an additive correction could absorb, not how well it would transfer. The leave-one-k-out variant estimates the offset within a placement from the five retained k cells and scores the single held-out k cell; each of the 18 cells is held out exactly once, and the reported held-out MAPE and MPE aggregate those 18 held-out scores. No corrected value is substituted for any prospective result in the main text, and none of these diagnostics is prospective validation.

Supplementary methods: post-hoc UCV M1–M2b paired comparison

The paired unit is one of the 18 UCV configuration–k cells. The statistic is the mean across cells of M1 APE minus M2b APE. Its descriptive interval uses 50,000 nonparametric draws that resample the 18 paired cells with replacement, preserving the two model errors within each sampled cell, with the 2.5th and 97.5th percentiles obtained by linear quantile interpolation. This analysis was specified after the prospective UCV result and supports neither a prospective superiority claim nor a population-level inferential claim.

Supplementary methods: characterization burden accounting

Burden is computed as a union over unique acquisition records. For each representation and phase we enumerate the acquisition records that representation requires, take the set union so that a record shared by several requirements is counted exactly once, and sum the measured worker-time of that union. Burden is therefore not additive across representations, and no representation's burden may be obtained by summing component burdens of another.

Every burden figure carries a scope label. Total required characterization is the union over all records the representation requires. Capacity-only characterization is the union restricted to capacity acquisition records, and activation characterization is the union restricted to activation acquisition records. Reductions are computed within a single scope and are not comparable across scopes: in P4-A the total-scope reduction and the capacity-only reduction differ precisely because the activation requirement is common to all three representations and therefore does not shrink. Burden is phase specific; UCV burden is computed from UCV acquisition records and P4 figures are never substituted for it.

For the retrospective hybrid H1 two scopes are reported because H1 consumes the frozen M2b prediction. The incremental scope counts only the completion-outcome acquisition H1 adds, conditional on M2b state already being available. The standalone scope counts the total acquisition required to reach the same configuration from scratch, and therefore includes the prerequisite M2b characterization. Quoting the incremental figure as if it were a standalone acquisition cost would understate the information H1 requires.

Supplementary methods: retrospective fitted comparator

The retrospective analysis uses leave-one-configuration-out folds: all k cells of one configuration form the test fold, and all cells of the remaining configurations form the training set. Within each training cell, the fixed budgets $r \in \{1, 2, 5, 10, 20\}$ use the first r completion records in the frozen run order. Fits use original-scale completion time and equal total weight per represented training cell. The burden reported for a budget is the mean across folds of the worker-seconds consumed by those selected training records.

The frozen model family is:

$$\text{C0}: \quad T = \beta_0 + \beta_1(k/N),$$

$$
\begin{aligned}
\text{C1}: &\quad T = \beta_0 + \beta_k k + \beta_N \log N + \text{placement indicators},\\
\text{C1-I}: &\quad T = \text{C1} + k \log N + k \times \text{placement} + \log N \times \text{placement},\\
\text{H1}: &\quad T = \alpha + \beta \widehat{T}_{\text{M2b}}.
\end{aligned}
$$

C0 is a sanity model, C1 is an additive completion-only model, C1-I is the prespecified limited full two-way completion-only model, and H1 is a hybrid calibration rather than a standalone completion-only substitute. Continuous features are standardized using only the represented training cells in each fold; held-out outcomes do not enter fitting, budget selection, or scaling. The comparison is deliberately information-asymmetric: C1 and C1-I extrapolate to the held-out configuration from completion outcomes in the other represented configurations, whereas M2b receives independently characterized state required for the target configuration but no target completion outcome. It is an information-source substitution comparison, not a same-information model contest.

The analysis unit is the held-out-configuration fold, and there are nine folds per model and budget. Fold-level dispersion is reported as a descriptive distribution over those nine folds using the median, interquartile range, and range. Folds are not independent repetitions of a single cell and share training data, so fold spread is not converted into a repetition-level confidence interval and no significance is attached to it.

For C1 and C1-I, an observed evaluated crossing is the first tested budget whose LOCO MAPE is no greater than the frozen M2b reference. When no tested point satisfies that condition, the status is "no evaluated crossing": the crossing value and any downstream amortization crossing are undefined. No interpolation, extrapolation, or monotonic smoothing is used. Table S5 reports every evaluated point.

Supplementary Table S8: execution environment and acquisition design

Table S8: Execution environment and acquisition design of the three outcome-bearing phases. Every entry is taken from that phase's own frozen authority; no value is carried across phases, and the three phases ran on separate clusters with distinct control-plane versions. Entries marked *not recorded* are absent from the bound authority for that phase and are deliberately left blank rather than inferred from another phase. These are frozen method and environment parameters, not experimental results.

Item	P3	Corrected P4-A	UCV
Platform			
Managed platform	GKE, zonal	GKE, zonal	GKE, zonal
Zone	asia-northeast3-a	asia-northeast3-a	asia-northeast3-a
Control-plane / node version	v1.35.5-gke.1241004	v1.35.6-gke.1127000	1.35.6-gke.1250000
Node OS / container runtime	COS, containerd 2.1.7	COS, containerd 2.1.7	COS/containerd, regular channel
Hardware			
Worker machine type	n2-custom-6-24576	n2-custom-6-24576	n2-custom-6-24576
Worker nodes	3	9	3
Worker CPU platform	Intel Ice Lake	Intel Ice Lake	Intel Ice Lake (minimum)
Coordinator machine type	not recorded	n2-standard-2	n2-standard-2
Worker boot disk	not recorded	40 GB pd-balanced	40 GB pd-balanced
Max workers per node	4	4	4
Worker pod resources	one pinned CPU per pod (cpuset)	1 CPU, 256 MiB	not recorded
Workload			
Trial	SHA-256 leading zero nibbles	SHA-256 leading zero nibbles	SHA-256 leading zero nibbles
Difficulty d	6	6	6
Success probability $p = 16^{-d}$	2^{-24}	2^{-24}	2^{-24}
Completion design			
Worker counts N	4, 6, 8	12, 18, 24	6
Placements	6 registered vectors	3 classes per N	3 cyclic, $\{3, 2, 1\}$
Target success counts k	1, 3, 5, 10, 20, 40	1, 10, 40, 100	1, 3, 5, 10, 20, 40
Completion cells	36	36	18
Repetitions per cell	20	20	20
Accepted completion runs	720	720	360

continued on next page

Item	P3	Corrected P4-A	UCV
Characterization acquisition			
Capacity repetitions	not recorded	360 node-local, 90 placement-total runs	10 input rounds
Reference placements	not recorded	not applicable	$[4, 2, 0]$ and $[2, 2, 2]$
Activation repetitions	not recorded	180 startup-only runs	20 startup vectors
Steady measurement window	not recorded	40 s	40 s
Warm-up bucket / plateau window	not recorded	0.5 s / 5 s	0.5 s / 5 s
Steady gate	not recorded	dispersion ≤ 0.08	std/median ≤ 0.08
Maximum warm-up	not recorded	60 s	60 s

Supplementary Table S9: prospective-freeze provenance

Table S9: Prospective-freeze chronology for UCV, so that the information boundary can be inspected independently rather than taken on assertion. Each row names a preserved artifact and its SHA-256 identity, abbreviated to its first eight and last six hex digits; full digests are listed in the archived manifest. The ordering is established by the artifacts' own recorded information-boundary counters and embedded creation times, never by file naming. The decisive evidence is in the counters: at the moment the predictions were frozen, the number of target completion runs executed and the number of target completion outcomes accessed were both zero.

Artifact	Role	Identity (SHA-256)	Position in the chronology
Pre-outcome numerical-feasibility amendment	Uniform tightening of integration tolerances for all 54 prediction rows	`0ec13795...bb5f94`	Before the freeze. Recorded status IMMUTABLY FROZEN BEFORE U10 PREDICTIONS; target capacity, completion and success-trace access counters all zero
Prediction freeze authorization	Seals the frozen prediction set	`8297c298...85ecb7`	The freeze itself. Recorded status FROZEN BEFORE U11; 54 prediction rows emitted, target completion access count zero, completion runs executed zero
Frozen prediction table	The 54 model–cell predictions later scored	`97358207...8751aa`	Bound by the freeze authorization above; unchanged thereafter under semantic freeze-hash equality
Authoritative completion record set	The 360 accepted target completion outcomes	`e70a995a...bef303`	After the freeze. The first artifact in the chain that contains target completion outcomes
Post-outcome transport-parser corrective	Applies the predefined decoder and the earliest-valid-record rule	`30f5002f...fdad85`	After outcome access had begun. Executed no additional target completion run; no attempt had been accepted before it; the frozen predictions were verified unchanged

continued on next page

Artifact	Role	Identity (SHA-256)	Position in the chronology
Primary audit lock	Seals the primary UCV result and accepted-record membership	`0d9d2470...32fb8`	Last. Minted after acquisition and before the fitted-comparator workstream

Table S10 Post-hoc descriptive P3 M2b error for the matched-structure $N = 6$ subset (P3-E6 and P3-P6). Each row averages only two cells, and at several k values the error is comparable with the finite-repetition precision of the observed cell means; these values describe context rather than a precise cross-k shape or a controlled P3–UCV effect. The phases are separate and historical P3 startup vectors are not preserved.

k	MAPE (%)	Signed MPE (%)
1	1.28	-0.47
3	1.61	-1.61
5	0.39	+0.39
10	1.53	-1.09
20	0.96	+0.46
40	0.51	-0.17

Table S11 Retrospective fitted comparison at each family's best evaluated budget. All error columns are leave-one-configuration-out fold MAPE in percent. The analysis unit is the held-out-configuration fold, so the spread columns are a descriptive distribution over nine folds and not a repetition-level confidence statement. A crossing would be the first evaluated budget at which a standalone fitted family matched or improved on the prospectively frozen 2.67% M2b reference; NO EVALUATED CROSSING is bounded to the grid $r \in \{1, 2, 5, 10, 20\}$ and locates no crossing outside it. H1 is a hybrid that consumes the frozen M2b prediction, so its burden must be read with the information boundary in Table 5.

Family	Information used	Best r	Mean	Median	IQR / range	Grid outcome and status
C0	Training-cell outcomes and k/N	10	11.16	11.07	6.47 / 4.82–19.18	Context sanity model; retrospective
C1	Training-cell outcomes, k, cluster size, placement	10	7.52	7.63	0.62 / 5.78–8.58	No evaluated crossing; retrospective
C1-I	C1 inputs and prespecified two-way interactions	1	7.40	7.29	1.95 / 5.52–9.60	No evaluated crossing; retrospective
H1	Training-cell outcomes **and** the frozen M2b prediction	10	2.61	2.62	1.42 / 0.65–4.29	Not standalone; hybrid, retrospective
M2b	Independently characterized capacity and activation	–	2.67	–	–	Prospectively frozen analytic reference

Supplementary Table S12: placement-pooled leave-one-run-out activation cells

Table S12 Placement-pooled leave-one-run-out realized-activation sensitivity by UCV cell. Each repetition excludes its own activation vector and uses the 119 other accepted vectors from the same placement with exact frozen U-M2b capacity. Predictions are averaged over the 20 repetitions before cell APE is computed. The diagnostic is post-hoc and non-deployable, not corrected prospective performance or causal attribution.

Placement	k	Diagnostic prediction (s)	Observed mean (s)	APE (%)
U6-A	1	7.984472	7.680346	3.96
U6-A	3	9.031946	8.888206	1.62
U6-A	5	9.780150	9.827962	0.49
U6-A	10	10.995286	10.982323	0.12
U6-A	20	12.910713	13.075899	1.26
U6-A	40	16.618266	16.098951	3.23
U6-B	1	8.526868	8.587659	0.71
U6-B	3	9.259373	8.957249	3.37
U6-B	5	9.896359	10.265790	3.60
U6-B	10	11.143573	11.141501	0.02
U6-B	20	13.042468	13.022578	0.15
U6-B	40	16.752531	16.629476	0.74
U6-C	1	9.378731	9.237895	1.52
U6-C	3	9.883184	10.117192	2.31
U6-C	5	10.316116	10.468589	1.46
U6-C	10	11.358979	11.508789	1.30
U6-C	20	13.284908	13.241234	0.33
U6-C	40	17.006525	17.031511	0.15

The largest U6-A late-k cell residual in this diagnostic is 3.23% at $k = 40$. Finite-repetition resolution remains limited, and the registered U6-A endpoint-slope mismatch interval includes zero; this cell does not identify a capacity anomaly or a productive-search-rate mechanism.

Supplementary Table S13: stopping-conditioned productive-interval rates

Table S13 Post-hoc stopping-conditioned productive-interval mean-rate sensitivity. Each run sums six worker rates computed from productive attempts divided by the same-worker steady-to-target-stop interval, then normalizes by exact frozen U-M2b placement capacity. This is not a fixed-window capacity measurement and is not causal. Duration is the across-run mean/median of each run's mean worker interval. The final block is the post-hoc late-k subset defined after the primary diagnostic.

Placement	k	n	Ratio mean/median	Duration mean/median (s)
U6-A	1	20	1.183/1.174	0.239/0.152
U6-A	3	20	1.172/1.140	0.515/0.438
U6-A	5	20	1.102/1.022	0.817/0.718
U6-A	10	20	1.018/1.004	1.899/1.876
U6-A	20	20	1.002/1.001	3.782/3.835
U6-A	40	20	1.000/1.000	6.979/7.034
U6-B	1	20	1.132/1.122	0.253/0.104
U6-B	3	20	1.112/1.075	0.490/0.501
U6-B	5	20	1.083/1.011	0.928/0.717
U6-B	10	20	1.005/1.006	1.783/1.817
U6-B	20	20	1.001/1.002	3.847/3.933
U6-B	40	20	1.001/1.004	7.301/7.168
U6-C	1	20	1.145/1.168	0.281/0.177
U6-C	3	20	1.014/1.009	0.609/0.530
U6-C	5	20	1.017/1.010	0.965/0.686
U6-C	10	20	1.012/1.005	1.836/1.797
U6-C	20	20	1.003/1.006	3.908/3.803
U6-C	40	20	1.008/1.008	7.443/7.735
Post-hoc late-k subset defined after the primary diagnostic, $k \in \{20, 40\}$				
U6-A	20,40	40	1.001/1.001	5.380/5.029
U6-B	20,40	40	1.001/1.003	5.574/5.521
U6-C	20,40	40	1.005/1.008	5.676/5.227

Supplementary Table S14: two-by-two state-input sensitivity

Table S14 Post-hoc UCV state-input sensitivity under one common analytic evaluator (MAPE, %). Realized activation is taken from the evaluated run and is leakage-assisted. The capacity substitution is the trailing-median attempt rate from the ten fixed detector buckets ending at steady declaration before the target-completion stopping event; it is not average throughput over the full productive search interval. The grid is explanatory only: it is not a causal decomposition, corrected prospective performance, or a deployable estimator.

Capacity input	Frozen activation	Realized activation
Frozen U-M2b capacity	7.42	1.14
Detector-window capacity	7.35	1.15

Supplementary methods: execution environment and acquisition protocol

Table S8 records the execution environment and acquisition design of the three outcome-bearing phases. Each column is populated only from that phase's own frozen authority. The three phases ran on separate managed clusters whose control-plane versions differ, so no environment value is shared between columns by assumption; where a phase's authority does not record an item, the cell is left explicitly unrecorded rather than filled from a neighbouring phase.

Workload and probabilistic assumption.

Each worker repeatedly hashes distinct counter-derived candidates. A success is *defined* as a SHA-256 digest beginning with six zero hexadecimal digits. Under uniform digest behavior, the success probability is *calculated* as $p = 16^{-6} = 2^{-24}$; it is not a measured cluster quantity. The distinct deterministic inputs and recorded seeds make execution reproducible, but determinism does not prove probabilistic independence. The completion model therefore makes the standard ideal-hash/random-output *modeling assumption* that distinct counter-derived SHA-256 digests behave as independent uniform outputs for this predicate. Conditional on that assumption, attempts are modeled as independent Bernoulli trials with fixed rare p, and accumulated successes are modeled using the non-homogeneous Poisson rare-event approximation of Section 3. Conditional on exact integer cumulative attempts $N(t)$, the exact count is Binomial$(N(t), p)$ and the model count is Poisson$(N(t)p)$. The Barbour–Hall bound is $d_{\mathrm{TV}} \leq (1 - e^{-N(t)p})p < p$, so at $p = 2^{-24}$ every fixed-exposure count event, including a fixed-time kth-completion CDF event, has absolute probability error strictly below $5.9604644775390625 \times 10^{-8}$. This is a mathematical count-distribution bound, not an empirical runtime, MAPE, or expected piecewise-clock-time bound. The discrete negative-binomial and continuous Gamma mean attempt exposures are both exactly k/p; converting attempt exposure into actual clock time additionally requires the realized integer attempt chronology.

Capacity characterization.

Capacity acquisition measures productive attempt throughput without any target completion outcome. A run discards a warm-up transient and then counts attempts over a fixed steady window; the per-instance rate is the attempt count divided by the actual window duration, and a placement or cell value is the sum over participating instances, averaged arithmetically over replicates. The window does not extend adaptively and is never shortened or lengthened in response to observed variance, so the estimator cannot be tuned by what it measures. Warm-up is judged by bucketing the attempt curve and requiring the dispersion over the trailing plateau to fall below the steady gate; a run that never reaches a plateau is marked as a structural infrastructure failure rather than being excluded on scientific grounds. In UCV the node-local family instantiates n co-located instances pinned to a single canonical slot for $n \in \{1, 2, 3, 4\}$, while reference and drift-anchor measurements distribute six instances over the three slots according to a placement vector; node-local, reference, and anchor measurements use the same scalar estimator and differ only in deployment geometry.

Activation characterization.

Activation is measured on the coordinator alone. The clock origin is taken immediately after the coordinator binds its listening socket, and a worker's activation time is the arrival of its steady message at the coordinator, relative to that origin. A single coordinator-side monotonic counter is used and no cross-node wall clock enters any scientific value, so the measurement

carries the coordinator's queueing and transport delay identically for every worker. This is why activation is described throughout as protocol-defined rather than as physical startup latency. A startup vector is emitted only when every expected worker has reported steady; UCV averages 20 complete startup vectors, and vectors are retained whole rather than being recombined across runs.

Completion runs.

A completion run executes one configuration at one target success count and yields one elapsed time from the listen-start origin to the kth accepted success. Completion runs are executed strictly one at a time, so no two scientific runs contend for the cluster. Each phase records 20 accepted repetitions per cell.

Freeze and prediction flow.

Capacity and activation are acquired first and sealed; predictions are then computed from that sealed state and frozen; only afterwards are target completion outcomes opened, and they are used solely to score the frozen predictions. Table S9 gives the artifact-level chronology for UCV.

Retry, exclusion, and accepted records.

The acceptance rules were fixed before outcomes and are outcome-independent. There is no automatic same-run retry of a scientific attempt, and a timeout on the target success cannot be silently replaced. Replacement of a failed attempt is capped per cell and per block, and an accepted replacement remains attached to its original scientific position. No attempt is ever excluded because of the value it produced: a valid but slow or extreme run is retained, invalid attempts remain visible in the record rather than being deleted, and outcome-derived replacement is prohibited. Exactly one accepted observation satisfies each planned entry.

Supplementary definitions for secondary diagnostics

The last activator is the worker with the latest protocol-defined activation time in a run; a slot frequency records the placement slot occupied by that worker. A1-prime is a preserved post-hoc, cardinality-matched P4 contrast between the hybrid H1 model and M1 at the 20-repetition budget. Its historical resolution rule compared the absolute effect with its bootstrap standard error and required a ratio of at least 3; the observed ratio was 1.97, so the threshold was not met. It is retained only as a secondary diagnostic and is not part of the primary comparison family.

Protocol and provenance disclosures

The earlier invalid P4 campaign is excluded because its receive-order semantics violated the intended protocol; it is named only as excluded provenance and supplies no scientific input. A uniform numerical-feasibility amendment was completed before UCV predictions were frozen and was verified as pre-outcome. After outcomes were opened, a transport/parser corrective applied the predefined decoder and a deterministic earliest-valid chronology. The archive preserves 362 physical attempts: the earliest valid attempt at the affected planned position was admitted, 360 attempts form the accepted analytical corpus, and two later attempts at that position are permanently excluded as redundant. Admission used neither outcome magnitude nor prediction-residual desirability, and the frozen predictions were verified unchanged under semantic freeze-hash equality. Table S9 lists the preserved artifacts and their identities in

chronological order, so that a reader can check the information boundary against the record rather than accepting it as an assertion.

The classification of the corrective's scientific effect is a bounded effect with an explicit evaluability boundary. Whether the pre-corrective accepted record set, the observed cell means, the primary UCV metric, the locked bootstrap interval, or the post-closure same-run and Forward-10 diagnostics would have differed is not evaluable from the frozen record, and this supplement does not assert that any of them was unchanged. The evidence record establishes that the frozen predictions did not change. Although all preserved raw attempts are decodable, immediate admission of the first affected attempt would have prevented two later retries and changed the subsequent execution timing and cluster chronology; re-decoding alone therefore cannot reconstruct that counterfactual campaign. The current manuscript reports the locked post-corrective analytical corpus; the bound provenance identifies no further numerical correction to those locked values. Full retry, exclusion, hash-chain, and byte-level records remain in the archived reproducibility materials.

Post-X411 authoritative package and checker audit

The post-X411 integration used sealed evidence only and did not rerun any cloud or scientific experiment. The live-characterization package contains 704,479 bytes and has SHA-256 `ebdc2b239e2588c58fc2228dc4057e10141011b90068cdf50e18f344cd80fc53`. The prediction/gate package contains 133,220 bytes and has SHA-256 `0f0e32b83f10d9c21b3d57dfd2185cec112ef8c467b7f099fe9924c9202e6b6c`. The corrected structural package contains 145,632 bytes and has SHA-256 `63e61ef150f46dcaf008e3b85c49faa94c50a8fe29d83097758c42c68df1165f`; it supersedes the original package with SHA-256 `ae10444c558b2ec27937225ae6e7ad6c23c74ac100efcc0f2ece1766d72da0c7`. The corrected successor package contains 170,170 bytes and has SHA-256 `ad870cec7619b12307bc9f9ad4c89d5dd2e914a10dfbfaa8ce32f0a7d74100f6`; it supersedes the original package with SHA-256 `b42a06ed57e9614034843e3221b058a41dd39fc67da352954f34914ca8ed334b`. The full runtime archive contains 26,718,278 bytes and has SHA-256 `e5d18f79ba086c9ef5f0f8526bc2c1204cea5fc875316adc8e1734fd3cb2b9ce`.

All ZIP CRC checks and all internal SHA-256 manifests passed. The packages bind production commit `9fd8e52eaf470e77843d1abfb62a37e1abe380e0` and tree `32f04c5099f8ce3e8882da9810e46e8f32765e5b`; the packaged predictor, discriminability, and contract source hashes match that checkout. Independent offline recomputation matched all 36 frozen model-cell predictions. The corrected structural checker passed 61 of 61 checks and the corrected successor checker passed 90 of 90 checks. The runtime archive contains no successor target directory, and every attempt-ledger record declares scientific completion outcome unavailable.

X411 execution scope and characterization

X411 used three n2-custom-6 worker nodes with Intel Ice Lake CPUs and one n2-highmem-2 coordinator in zone `asia-northeast3-b`, under GKE 1.35.6-gke.1250000 and COS/containerd 2.1.7. The sealed characterization counts are 12 warmups, 240 node-capacity records, 120 direct-capacity records, and 120 activation records. Warmups are outside the 480-record characterization total. The direct-capacity medians for X411-A, X411-B, and X411-C were 82.186, 82.181, and 82.077 Mhash/s. Their absolute discrepancies from composed node-local totals were 0.0470%, 0.1031%, and 0.0667%, respectively. These direct measurements support close node-total composition agreement and had zero information flow into primary prediction. No equivalence margin or procedure was prespecified.

Occupancy-four within-node diagnostic.

A post-hoc outcome-free diagnostic reconstructed physical worker rates only for records byte-equal to the accepted 240-record capacity subset. The occupancy-four portion contains 120 accepted node-rounds (40 on each physical node) and 480 simultaneous worker observations. The four physical rates sum exactly to the accepted aggregate node rate. Population within-node CV had median 0.1556898538, IQR 0.1247777045–0.1687068030, and maximum 0.2154730412; the max/min rate ratio had median 1.4092691871 and maximum 1.6112507974. This material heterogeneity means that $C(\text{node}, 4)/4$ is not empirically established as the physical rate vector. Worker ordinal does not provide a stable fast/slow identity, and no CPU affinity, sibling map, or pod-to-core topology was preserved. The diagnostic is post-hoc and outcome-free; it changes neither node-total composition nor a frozen prediction, gate statistic, or decision.

The prediction freeze contains 18 placement–k rows and 18 predictions for each of X-M2b and X-EQ. The gate uses 10,000 characterization-only bootstrap replicates. The complete executed resampling contract is as follows. Its input contains 240 accepted aggregate node-capacity records and 120 accepted activation records; the 120 direct placement-capacity records are excluded. Capacity is blocked into six fixed node-by-occupancy strata $(A, 1), (A, 4), (B, 1), (B, 4), (C, 1), (C, 4)$ with 40 observations each. Within each stratum, 40 aggregate records are sampled with replacement and reduced by the ordinary sample median. Activation is blocked into X411-A, X411-B, and X411-C, with 40 ordered six-worker vectors per placement. Forty complete vectors are sampled with replacement inside each placement; vectors are never split or recombined, and worker positions remain in serialized order.

Capacity and activation are resampled separately rather than paired or jointly, using successive draws from one deterministic stream. Node, occupancy, and placement identities remain fixed. One replicate's six capacity medians and three resampled 40-vector activation mixtures feed all 18 cells and both models. For each placement, X-M2b selects the three relevant node medians, assigns each worker $C(\text{node}, n)/n$, evaluates exact piecewise non-homogeneous Poisson-process expected time for every resampled activation vector, and averages the 40 expectations. X-EQ uses the same three node totals and activation mixture, but assigns one-sixth of their sum to every worker. Cells are evaluated in placement order A, B, C and, within placement, at $k = 1, 3, 5, 10, 20, 40$. The global statistic is the relative RMS over those 18 paired predictions, as defined in the main text.

The interval is the nonparametric 0.025/0.975 percentile interval. Values are sorted and linearly interpolated at index $(n - 1)p$ between its floor and ceiling (NumPy type-7 convention); the reported center is Python `statistics.median`. The custom version-independent `HashPRNG` hashes `seed_bytes || counter`, where the counter is 16-byte big-endian, and uses 256-bit rejection sampling to avoid modulo bias. The exact seed is `f46fe545a9ed454eac71848f27759c6decf81e42a913a466c15aa1889d504e40`. Capacity keys are lexicographically sorted; placements follow frozen A/B/C insertion order; k follows the frozen grid; activation serialization order is preserved. Exactly 10,000 production replicates are required. This specification documents the already sealed computation; the bootstrap was not rerun.

Its median, interval, half-width, and resolvability ratio are 0.001637256655444657, [0.001040761885476258, 0.002375509589593516], 0.000667373852058629, and 2.4532825947469448. The resolvability threshold is 2.0 and the design separation threshold is 0.05. Resolvability passed, design separation failed, and the overall gate failed. The sealed adequate-completion-power claim is `NO`. The gate is not a hypothesis, significance, equivalence, confidence, or calibrated power analysis.

The threshold language is design-based. The 2.0 ratio is a prespecified characterization-uncertainty resolvability heuristic, and 0.05 is a separate prespecified experimental-design continuation threshold. Neither has a frozen external, application-level, decision-theoretic, or inferential calibration. Earlier archived closure records used a resolvability value of 1.0; the executed 2.0 rule was fixed before characterization, but no frozen rationale for the 1.0-to-2.0 change was found. This provenance inconsistency is disclosed rather than retrospectively justified. Neither ratio is a hypothesis, significance, equivalence, confidence, or power test. The predicted separation passed the prespecified characterization-uncertainty relative-precision heuristic but remained far below the prespecified experimental-design continuation threshold; the frozen protocol therefore did not authorize target-completion collection.

Post-hoc threshold transparency.

Table S15 applies alternative display cutoffs to the already frozen statistics. It is a post-hoc transparency analysis, not threshold reselection or confirmatory inference. No threshold, statistic, or historical decision is recomputed or changed.

Table S15 Outcome-free threshold transparency for the frozen X411 statistics.

Axis	Tested cutoffs	Frozen decisions
Resolvability ratio	1.0, 1.5, 2.0, 2.5, 3.0	pass, pass, pass, fail, fail
Design separation	0.1%, 0.25%, 0.5%, 1%, 2%, 5%	pass, fail, fail, fail, fail, fail

The frozen ratio is 2.4532825947 and the frozen separation is 0.1637256655%. Thus the resolvability subdecision depends on a cutoff near 2.5, whereas the magnitude shortfall from the historical 5% criterion is much larger. The historical executed cutoffs remain 2.0 and 5%, and the gate decision is unchanged.

Corrected structural and successor scope

The corrected structural package proves the cumulative-intensity identities reported in Section 3.5. Its complete invariance criterion is groupwise over identical activation values; zero covariance alone is only a post-full condition. A full expected-time shift additionally requires negligible pre-full kth-arrival mass for both models. The corrected historical observed-region maximum is 0.003766238072740347. The value 0.049994522077259754 belongs to a factor-200 activation extrapolation and is not a historical result. For completeness, consider the normalized two-level six-worker rate vector $r = (\rho, \rho, \rho, \rho, 1, 1)$ with $0 < \rho < 1$, positive rates, nonnegative activations, frozen $\{4, 1, 1\}$ construction, frozen k grid and p, and the stated nonhomogeneous Poisson-process semantics. Its mean is $(2\rho+1)/3$. Four deviations are $(\rho-1)/3$ and two are $2(1-\rho)/3$, hence

$$\mathrm{Var}(r) = \frac{2}{9}(1-\rho)^2, \qquad \mathrm{CV}(r) = \frac{\sqrt{2}(1-\rho)}{2\rho+1}. \tag{1}$$

For the post-full surrogate means $\widetilde{T}_k^M = \mu_{ar} + q_k$ and $\widetilde{T}_k^E = \bar{a} + q_k$, with $q_k \geq 0$ and both denominators at least $a_{\max}$, their exact difference is $\delta = \mu_{ar} - \bar{a} = \mathrm{Cov}(a, r)/\bar{r}$. Cauchy–Schwarz and Popoviciu give $|\delta| \leq \mathrm{CV}(r)\sigma_a \leq \mathrm{CV}(r)(a_{\max} - a_{\min})/2 \leq \mathrm{CV}(r)a_{\max}/2$, so the symmetric surrogate relative shift is at most $\mathrm{CV}(r)/2$. This statement is exact only for the stated post-full surrogate. For actual NHPP means, $\mathbb{E}[T_k^M] - \mathbb{E}[T_k^E] = \delta + \epsilon_k$; controlling

ϵ_k requires negligible pre-$a_{\max}$ kth-arrival mass under both models. That mass was not quantitatively bounded, and $\mathbb{E}[T_k] > a_{\max}$ is insufficient. Therefore no exact $\mathrm{CV}(r)/2$ bound on the actual means is claimed. The $\mathrm{CV}(r)/2$ envelope therefore does not bound the executed X411 gate statistic G_b: G_b is computed from the full NHPP frozen mean predictions, whereas the envelope applies only to the stated post-full surrogate.

The certificate uses the X411-A-derived literal normalized input $\rho = 0.853516$, for which $\mathrm{CV}(r)/2 = 0.0382632454$. The underlying placement-specific capacity ratios are 0.8535155549558981 (X411-A), 0.8531580782754280 (X411-B), and 0.8506855023588270 (X411-C), with pooled ratio 0.8524521016169517; only X411-A rounds to the certificate literal at six decimals. The envelope is not a tight prediction: it is 23.3703 times the frozen global separation median 0.0016372567 and 11.7854 times the maximum frozen cell separation 0.0032466604. Using the non-rounded point global value 0.0015970698265208806 would give 23.9584 instead; the reported 23.3703 comparison intentionally uses the frozen median authority.

The boundary at which this surrogate envelope ceases to exclude 5% solves $\sqrt{2}(1-\rho) = (2\rho+1)/10$, giving

$$\rho_* = \frac{10\sqrt{2}-1}{10\sqrt{2}+2} = \frac{101-15\sqrt{2}}{98} = 0.8141509853510569. \tag{2}$$

For $\rho \le \rho_*$ the envelope is at least 5%; this is only the exact algebraic boundary where that specified surrogate envelope stops excluding 5%. It is not a physical threshold and does not prove that the actual system or actual X411 prediction difference reaches 5%. At the X411-A-derived certificate input $\rho = 0.853516$, the envelope remains below 5%. A separate extreme witness shows attainability only in the minimally constrained formal model class, not in the studied homogeneous GKE environment.

The corrected successor package enumerates seven canonical families and 28 physical permutations. Five families were already scientifically used, but historical use is distinct from present numerical evaluability and rankability: $\{4,1,1\}$, $\{3,2,1\}$, $\{4,2,0\}$, $\{3,3,0\}$, and $\{2,2,2\}$. The active predictor requests $C(\mathrm{node}, n)$ for all three nodes without a zero-occupancy guard, so every zero-containing family needs the undefined $C(\mathrm{node}, 0)$. The only unused families, $\{6,0,0\}$ and $\{5,1,0\}$, additionally need unmeasured $C(\mathrm{node}, 6)$ or $C(\mathrm{node}, 5)$, exceed the measured design limit of four workers per node, and violate the common-total-rate premise used by the post-full comparison. These are independent blockers. Measured node capacities are not exactly equal merely because machine types match. The resulting eligible rankable count is zero, so no unused family was selected, ranked, or authorized for execution and no successor target campaign or target run was created under the stated model and measurement constraints.