Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Prospective Completion Prediction for Independent-Trial Cluster Search with Compositional State: Measurement Burden, a Pre-Outcome Separation Gate, and Structural Limits

Min Su Yoo
Chonnam National University
Hyunglip Bae
hlbae@jnu.ac.kr
Chonnam National University

Research Article

Keywords:

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10961374/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

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 telephone: +82-62-530-3445.

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

Abstract

Completion prediction for independent-trial cluster search requires productive capacity and worker activation, and the question is acquisition cost as well as accuracy. Under an ideal-hash model, we combine independently characterized state with a known per-trial success probability in a zero-fit analytic predictor. For the compositional representations in two measured configuration regimes, total-completion-time cell-mean absolute percentage errors were 1.56% and 2.54–2.67%, with no accuracy ordering established. For a nine-target acquisition geometry, node-total composition reduced total characterization burden by 35.52%. An unseen-family campaign remains descriptive after a post-outcome corrective. In a fresh characterization-only campaign, predicted occupancy-aware/equalized separation was 0.164%: it passed the prespecified characterization-uncertainty relative-precision heuristic but remained far below the 5% experimental-design continuation threshold, so the frozen protocol authorized no target collection. Structural analysis identifies rate–activation interaction and full-path invariance conditions; an outcome-free diagnostic additionally reveals material occupancy-four within-node worker-rate heterogeneity, limiting physical interpretation of equal-share worker rates.

Keywords: Cluster computing, Performance prediction, Kubernetes, Compositional characterization, Prospective evaluation, Structural invariance

1 Introduction

For independent-trial distributed search with a known per-trial success model, a job completes when a cluster accumulates its kth success. Nominal allocation alone does not determine that completion process. Workers can deliver different productive attempt rates under different local occupancies, and they become operational asynchronously. Two placements with the same nominal worker count can therefore expose different productive capacity over time.

This turns performance prediction into an information-selection problem. A predictor can

acquire highly target-specific state, but repeating aggregate-capacity and activation measurements for every candidate placement can erase the operational value of prediction. Coarser or more reusable state may reduce characterization burden depending on the required-record acquisition geometry, but may also omit placement structure. The systems question is therefore not only whether a predictor is accurate, but which characterization and composition rules can be reused, what that state costs to acquire, and which state must be refreshed.

Composition-rule reuse refers here to reapplying the characterization and composition framework, not to transferring measured state values or to transfer learning. Each experimental phase acquires the capacity and activation values required by its own frozen protocol.

We study this question with a zero-fit analytic predictor. Here, zero-fit means that target completion outcomes are not used to fit or calibrate the prediction; it does not mean measurement-free prediction. The model consumes a model-specified per-trial success probability together with independently characterized productive capacity and operational activation. The known-p identification advantage is central to the design: it isolates characterization and completion-model error without estimating a success process from target outcomes. This separation permits predictive fidelity to be evaluated alongside the information and measurement burden required to obtain the predictive state, but it does not extend the evidence to arbitrary workloads.

The evidence is deliberately layered, and the layering is an information boundary rather than a ranking (Fig. 1). P3 and corrected P4-A evaluate frozen analytic predictions in measured configuration regimes. P4 additionally quantifies node-local capacity composition and characterization burden. UCV froze predictions before observing outcomes for the previously unused $\{3, 2, 1\}$ family, but the two capacity representations were nearly indistinguishable ex ante; its subsequent diagnostics are historical post-outcome evidence. X411 therefore used a fresh $\{4, 1, 1\}$ family and a characterization-only gate computed after prediction freeze but before any target outcome. That gate, rather than an incomplete experiment, ended the target campaign according to the prospective protocol.

The structural novelty must be stated precisely. N=6 itself was not new: P3 included P3-E6=(2,2,2), P3-B6=(3,3,0), and P3-P6=(4,2,0). The UCV family was structurally unseen in the narrower sense that the $\{3, 2, 1\}$ occupancy composition and its cyclic assignments had not appeared as prior target configurations in the preserved study history. Claims about "unseen" placements in this paper are bounded to that single occupancy-composition family inside an otherwise homogeneous environment.

Our central research question is: *For independent-trial workloads with a known per-trial success model, which independently characterized state supports prospective completion prediction in measured configurations, when does occupancy-aware rate redistribution add design-relevant discrimination beyond equal aggregate capacity, and what structural conditions limit that discrimination?*

Our contributions are:

1. *A prospective predictor and the compositional state it runs on.* A zero-fit completion predictor for the studied independent-trial workload separates productive capacity from protocol-defined operational activation, with within-configuration accuracy evaluated after target outcomes were unsealed. What makes its compositional state usable is measured rather than assumed: occupancy-conditioned productive capacity is stable across the measured homogeneous nodes yet changes materially at occupancy four, while node-local capacity still composes closely to independently measured placement totals.
2. *What that state costs.* Characterization-burden accounting separated by scope quantifies the acquisition cost of each representation, including a corpus-specific capacity-acquisition amortization threshold and phase-specific required-record definitions.
3. *When finer state is not worth acquiring.* This limit is established prospectively and structurally rather than by accuracy comparison. An outcome-free continuation decision on a fresh $\{4, 1, 1\}$ occupancy family placed characterization and prediction freeze before target-outcome access; the small X-M2b–X-EQ separation passed the prespecified characterization-uncertainty relative-precision heuristic but

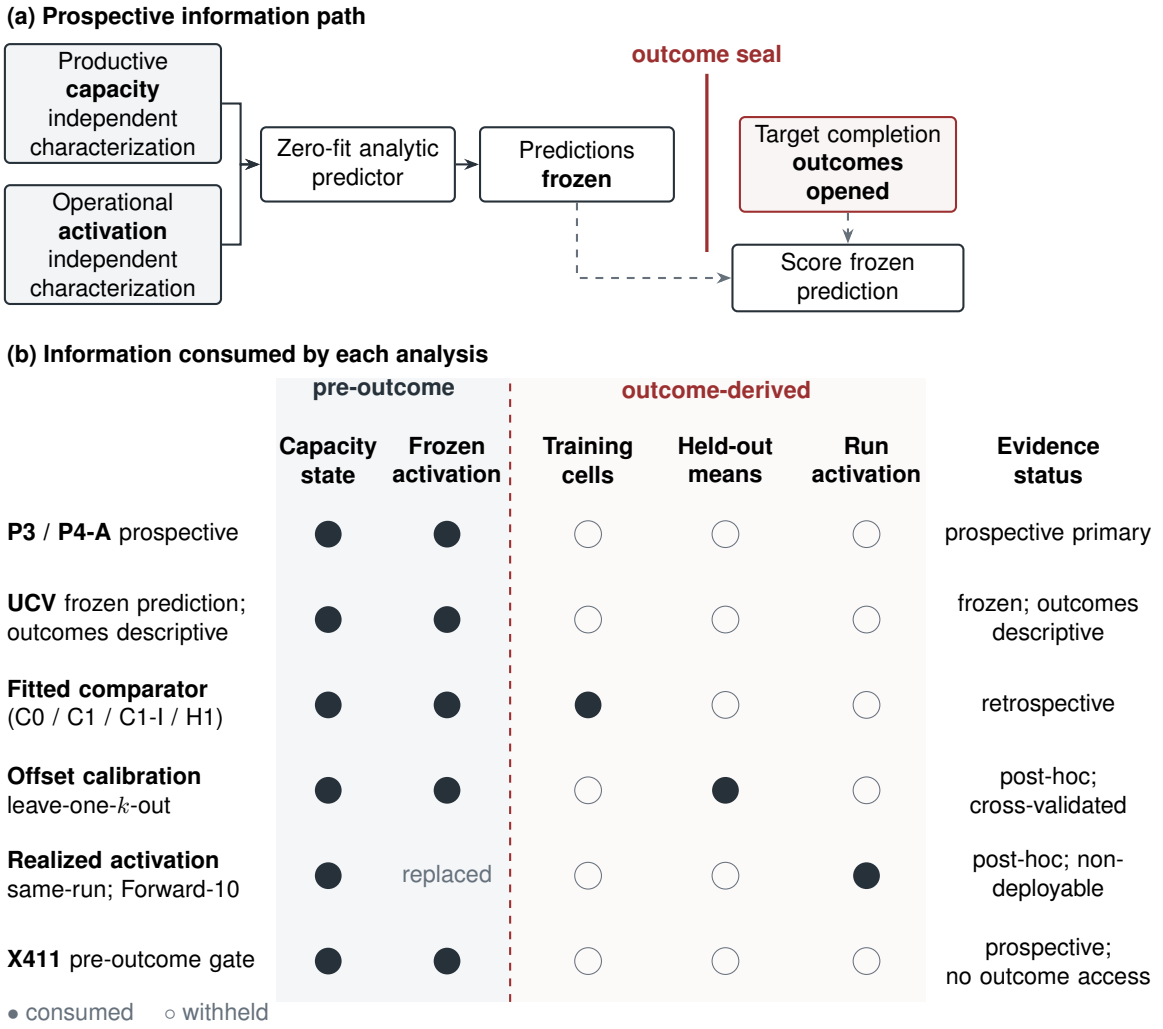


Fig. 1 Information-access and leakage boundaries for the completion-accuracy analyses and the X411 gate. (a) Productive capacity and operational activation are characterized independently of target outcomes; predictions are frozen before outcomes are opened. (b) Filled marks show consumed information and open marks show withheld information. The fitted comparator and UCV diagnostics cross the outcome boundary as described in the text. The X411 pre-outcome-gate row remains entirely on the no-outcome-access side: its gate stopped before target collection and is reported separately in Table 7.

failed the separate prespecified continuation criterion, and the target-completion outcome count remained zero. An exact structural characterization of the rate–activation interaction and full-path invariance boundary, kept distinct from the empirical X411 result, together with bounded seven-family screening, closes the frozen six-worker, three-node design program without elevating finite enumeration to a separate contribution.

P4's M2a–M2b ordering was not Holm-significant, UCV is descriptive-only, the fitted comparator is retrospective, and the activation and calibration diagnostics are post-hoc. X411 has no target outcomes and is not an accuracy evaluation. Thus the clean prospective unseen-family completion-accuracy evidence count is zero. No result establishes universal superiority, statistical equivalence, causal activation mechanisms, a 5% impossibility in homogeneous GKE, deployment benefit, or arbitrary-workload generalization.

Table 1 Experimental phases, target populations, and evidence boundaries. UCV predictions were frozen prospectively; after the post-outcome corrective, its preserved outcomes are descriptive historical evidence. X411 ended prospectively at its characterization-only separation gate; its zero target outcomes are a protocol result, not missing data.

Phase / analysis	Purpose	Target family	Cells	Reps./cell	Outcomes	Evidence status
P3	Measured-regime prospective evaluation	Measured grid, including P3-E6=(2,2,2) and P3-P6=(4,2,0)	36	20	720	Prospective primary
P4-A	Broader measured-regime evaluation	Measured homogeneous placements	36	20	720	Prospective primary
Fitted	Outcome-fitted alternative	Preserved P4-A outcomes	–	–	–	Retrospective
UCV	Structurally unseen-composition evaluation	Cyclic U6-A=(3,2,1), U6-B=(1,3,2), U6-C=(2,1,3)	18	20	360	Frozen; outcomes descriptive
X411	Pre-outcome representation-discrimination gate	Cyclic $\{4, 1, 1\}$ family; 18 rows per model	18	–	0	Prospective gate; no accuracy claim

2 Related work

Randomized-runtime research treats parallel completion as a distributional object rather than a deterministic speedup. Luby et al. analyse optimal restart strategies for Las Vegas algorithms [1], and Truchet and colleagues predict and estimate parallel runtimes for randomized constraint solving from the runtime distribution of a single process [2, 3]. Truchet et al. model parallel runtime through order statistics of independent Las Vegas processes obtained from a sequential runtime distribution. Our target instead is the kth accumulated success under staggered worker activation and worker-specific productive rates; their framework therefore does not directly encode the present state-and-timing target and does not provide a directly comparable implementation baseline. This line supplies the stochastic foundation the present predictor rests on; an earlier study of the same workload family empirically analysed container-based hash-search waiting time and overhead using Poisson and Gamma distributions [4]. We claim no novelty for the accumulated-success process or the Poisson/Gamma identity itself. Our question is the systems one underneath it: which measured cluster state is available before a target run, what that state costs to acquire, and how much of its characterization structure survives a change of placement.

A second line predicts performance in order to choose a configuration, and the systems in it differ in the evidence they consume. Performance skeletons reproduce an application's execution signature synthetically, whereas Ernest fits a parametric computation/communication model from small-input job runs [5, 6]. For configuration choice, CherryPick adaptively samples candidates, PARIS combines offline benchmark profiles with online application fingerprinting, and Selecta applies collaborative filtering to sparse training-workload measurements [7–9]. Lotaru is especially close in its characterization-first motivation: it combines microbenchmarks with reduced-input local workflow profiling and Bayesian regression to predict task runtimes before target workflow execution on heterogeneous infrastructure without historical executions [10]. Its workflow and hardware validation is broader than ours; our narrower target is an analytically composed independent-trial k-success endpoint with an explicit prospective target-outcome boundary and state-component burden/reuse analysis. More recently, Yildirim et al. use monitored job and system information with neural networks to predict runtime and resource utilization across integrated cloud and HPC systems [11]; this further confirms that learned runtime prediction is not the novelty claimed here, while their prediction target and information inputs differ from our known-p independent-trial target and prospective outcome boundary. CrystalPerf instead learns resource-time behavior from source code and past execution traces, then predicts dataflow runtime under resource-configuration changes [12].

These methods establish that pre-execution and characterization-based runtime prediction are not themselves novel. They remain relevant because each selects a compact measured representation from which to extrapolate performance, but differences in prediction targets, estimators, and information inputs preclude a direct numerical comparison with our predictor.

Configuration-selection systems have also treated measurement cost and changing operating conditions explicitly: Micky shares exploration across workloads to reduce cloud-instance selection cost [13], while Tuneful studies whether tuning should be amortized or repeated as workloads or environments change [14]. Historical-runtime reuse operates at distinct data and model levels: C3O shares records across users and contexts for cluster-configuration prediction, whereas Bellamy reuses an algorithm-specific runtime model across encoded contexts and can adapt it with new-context runtime samples [15, 16]. Garralda-Barrio et al. likewise use adaptive incremental transfer learning to carry historical Spark-workload information into execution-time models for new workloads and configurations [17]; that learned transfer object differs from reusing a composition rule while reacquiring fresh numeric capacity and activation values here. Together, these works establish measurement cost, reuse, and adaptation under changing conditions as prior systems concerns; our narrower question is which prospectively characterized predictor-state components can be reused while target completion outcomes remain withheld. In the evaluated regime only, capacity representation was comparatively reusable whereas activation information was more freshness-sensitive; this is not a general cluster-state or refresh-policy claim.

A third line manages resources using measured or learned runtime state. Bubble-Up characterizes an application's sensitivity to memory-subsystem pressure with a tunable stress probe, and its contentiousness towards a co-runner, in order to predict the degradation caused by co-location [18]. Quasar classifies an unknown incoming workload with respect to heterogeneity and interference by collaborative filtering against previously profiled applications, and allocates resources from that classification [19]. Sinan learns the performance impact of dependencies between microservice tiers and allocates per-tier resources online so that an end-to-end tail-latency target is preserved [20]. For GPU ML jobs, Bin Faisal et al.'s PCS couples completion-time estimates to scheduling and explicitly balances predictability, performance, and fairness [21]; our fixed-configuration predictor neither controls scheduling nor solves that optimization problem. In storage, Li and Wang monitor network and load state in real time and combine an object-storage-device performance prediction model with multi-attribute decision making to rebalance placement in Ceph [22], while Zanussi et al. forecast incoming workload in order to scale resources ahead of demand, evaluated in a CloudSim simulation environment rather than in production [23]. Rossi et al. instead forecast future cloud workload demand with uncertainty-aware Bayesian deep models and evaluate transfer learning on Google and Alibaba cluster traces [24]; this demand-forecasting task is distinct from fixed-target stochastic completion prediction. Each of these characterizes state in order to reach a resource-management decision: which machine, which tier, which replica, how many instances. The present study holds the target configuration fixed and asks a different question – whether independently characterized pre-outcome state predicts the stochastic completion time of that fixed configuration. Accordingly, we do not treat these systems as implementation baselines for the predictor, nor do we evaluate a placement or scaling policy.

Container and Kubernetes studies show that delivered performance depends on resource layout and runtime conditions [25–27]. Closest to the timing component of our state, Beni et al. address the cold-start delay incurred when Kubernetes bootstraps additional containers during elastic scaling, combining pre-creation of network containers, images that share linked libraries in memory, and parallel imperative creation of application containers [28]. That literature is adjacent to, but not identical with, the quantity measured here: our operational activation is the protocol-defined time at which the coordinator receives the worker's steady-state indication, and it may combine provisioning already in progress, container startup, warm-up, steady detection, and transport. Cold-start work seeks to reduce container bootstrap latency, whereas we treat the realized activation profile as a characterized state input to a completion-time predictor; the two are not the

same measurement and no equivalence is claimed. Tail-latency work likewise shows how response-time variability in service components shapes realized end-to-end latency at scale [29]. Against this background, the contribution is not new Poisson/Gamma mathematics, pre-execution runtime prediction itself, or a universal alternative to learned performance models. We study the combination of a prospective target-outcome information boundary, a productive-capacity/activation decomposition, explicit measurement-burden and reuse accounting by state component, a pre-outcome discriminability gate, and structural conditions governing when finer rate–activation state can change predicted completion time. The retrospective fitted comparator is a bounded parametric robustness check on our own corpus and budget grid; it does not implement a prior system or serve as a general learned-model baseline.

3 Methods

3.1 Problem formulation and metrics

Each active worker repeatedly evaluates distinct counter-derived SHA-256 candidates. A success is a digest with six leading zero hexadecimal digits. Under the standard ideal-hash/random-output modeling assumption that the resulting digests behave as independent uniform outputs, each trial has $p = 16^{-6} = 2^{-24}$. This is a probabilistic modeling assumption, not a consequence proved by deterministic SHA-256 execution. Let T_k be elapsed time from the coordinator's listen-start origin to the kth accepted success. Worker j becomes operationally active at time a_j and supplies steady attempt rate c_j. Productive success exposure by time t is

$$\Lambda(t) = p \sum_j c_j (t - a_j)_+, \tag{1}$$

where $(x)_+ = \max(x, 0)$. Under the non-homogeneous Poisson approximation,

$$\Pr(T_k > t) = e^{-\Lambda(t)} \sum_{r=0}^{k-1} \frac{\Lambda(t)^r}{r!}, \tag{2}$$

and

$$\widehat{\mathbb{E}[T_k]} = \int_0^\infty \Pr(T_k > t)\, dt. \tag{3}$$

Conditional on the ideal-hash assumption, the approximation replaces the exact conditional count Binomial$(N(t), p)$ by Poisson$(N(t)p)$; under constant attempt cadence this is equivalently the negative-binomial-to-Gamma waiting-exposure approximation. For the modeled independent trials, the Barbour–Hall Chen–Stein bound gives

$$\begin{aligned} d_{\mathrm{TV}}\{\mathrm{Bin}(N, p), \mathrm{Pois}(Np)\} &\le \left(1 - e^{-Np}\right) p \\ &< p. \end{aligned} \tag{4}$$

With the frozen $p = 2^{-24}$, the absolute error of any fixed-exposure success-count event, and hence of each fixed-time kth-completion CDF conditional on exact integer $N(t)$, is uniformly less than $5.9604644775390625 \times 10^{-8}$ [30]. This mathematical probability bound is not an empirical runtime, MAPE, or clock-time expected-value bound. The exact negative-binomial and Gamma mean attempt exposures are both k/p, but mapping attempt exposure to actual piecewise clock time requires the realized integer attempt chronology; the continuous measured-rate model does not supply a rigorous bound for that separate approximation.

A cell is a fixed configuration–k combination. One completion repetition is one accepted execution of that configuration and target k, producing one elapsed time from listen start to the kth accepted success. P3, corrected P4-A, and UCV each use 20 accepted completion repetitions per cell. The observed value T_i is the arithmetic mean of those 20 repetitions, and model MAPE is the unweighted mean across cell-level errors (36 P3 cells, 36 P4-A cells, or 18 UCV cells). Repetitions quantify within-cell outcome variation; they are not treated as independent model-comparison cells, because all repetitions in a cell share the same configuration, k, and frozen prediction. For observed cell mean T_i and frozen prediction $\widehat{T}_i$, we retain the predefined definitions

$$\mathrm{APE}_i = \left| \frac{\widehat{T}_i - T_i}{T_i} \right| \times 100, \tag{5}$$

$$\mathrm{MAPE} = \frac{1}{m}\sum_{i=1}^{m} \mathrm{APE}_i, \tag{6}$$

$$\mathrm{PE}_i = \frac{\widehat{T}_i - T_i}{T_i} \times 100, \tag{7}$$

$$\mathrm{MPE} = \frac{1}{m}\sum_{i=1}^{m} \mathrm{PE}_i, \tag{8}$$

$$\mathrm{WAPE} = \frac{\sum_{i=1}^{m} |\widehat{T}_i - T_i|}{\sum_{i=1}^{m} T_i} \times 100. \tag{9}$$

Positive MPE denotes overprediction. The primary phase summary is cell-mean MAPE; PE/MPE and WAPE are companion metrics where predefined outputs exist. Pooled signed mean-ratio errors are distinct from MPE, which averages per-cell signed percentage errors, and MAPE applies the absolute value at the cell level. The quantities therefore differ by aggregation order even though they derive from the same frozen predictions and observations.

Percentage reductions, ratios, and other derived quantities are computed from the registered full-precision values, while the component values shown in the text, tables, and figures are rounded independently for presentation. A reader who recomputes a derived quantity from the rounded components may therefore obtain a slightly different final decimal; the registered full-precision computation is authoritative in such cases.

3.2 Scope of claims and evidence classes

The phases are distinct populations and are never pooled (Table 1). P3 and corrected P4-A provide prospective completion-accuracy evidence within measured configuration regimes. UCV froze predictions prospectively but underwent a post-outcome execution/transport corrective; its unseen-family outcomes are descriptive historical evidence, and the correct-decoder-from-start chronology cannot be reconstructed. X411 provides clean prospective characterization and a pre-outcome decision, but zero target outcomes and therefore no completion-accuracy evidence. Consequently, no phase supplies clean prospective unseen-family completion-accuracy evidence. Capacity-composition results are a separate evidence class at the node-total characterization level. The fitted comparator is retrospective, and post-hoc diagnostics cannot certify prospective performance. We use *prospective* when all information entering the stated decision was frozen before target outcomes; *retrospective* when target-adjacent outcomes entered fitting; and *post-hoc* when an analysis was defined or executed after the prospective result.

3.3 Productive capacity, operational activation, and state representations

Capacity is productive attempt throughput characterized independently of target completion outcomes. Activation is the protocol-observed time at which a worker is received as steady relative to the coordinator's origin. It may combine provisioning already in progress, startup, warm-up, steady detection, and transport, and must not be interpreted as pure physical startup latency.

The estimators differ only in the state supplied to the same analytic completion calculation (Table 2). M1 is coarse and pooled. M2a uses target-placement aggregate capacity and target-specific activation. M2b composes node-local capacities while retaining target-specific activation. U-EQ equalizes capacity under the frozen UCV activation input. M2b therefore evaluates capacity-side reuse, not complete-state reuse; U-EQ is a discriminability comparator, not an equivalence test.

Two senses of "reuse" must be kept apart, because only one of them is evidenced here. *Composition-rule reuse* means reapplying a characterization scheme – node-local capacity measurement together with the rule that composes those measurements into a placement estimate – to a new target. *Value reuse* would mean carrying previously measured capacity coefficients forward as prospective inputs for a new target. This paper evidences representation and composition-rule reuse only; it does not study transfer learning. Each phase characterizes its own capacity and activation state afresh, and no capacity coefficient measured in one phase enters another phase's prospective prediction.

The primary M2b predictor measures node-total capacity $C(\mathrm{node}, n)$ at final local occupancy n and assigns each modeled worker $r_i =$

$C(\text{node}, n)/n$ from activation onward. This equal-share vector is defined by construction; it is not a directly measured physical per-worker rate vector. A post-hoc, outcome-free sensitivity replaces that fixed representation between activation events with $C(\text{node}, n_a)/n_a$ at the currently active count n_a. It therefore shares the equal-share limitation and does not test arbitrary within-node rate heterogeneity. The construction uses frozen capacity estimates, activation vectors, and target mapping only; it reads zero target-outcome bytes.

3.4 X411 prospective separation gate

X411 targeted the cyclic six-worker $\{4, 1, 1\}$ family. Its frozen campaign accepted 12 warmups, 240 node-capacity records, 120 direct-capacity composition-agreement records, and 120 activation records, for 480 characterization records. Direct placement-capacity measurements had zero information flow into primary prediction. From the sealed characterization, the protocol froze 18 placement–k rows, each with an occupancy-aware X-M2b prediction and an equalized X-EQ prediction. Neither representation consumed a target completion outcome.

For bootstrap replicate b, global separation was the relative root-mean-square displacement

$$G_b = \left[\frac{1}{18} \sum_{i=1}^{18} \left(\frac{\widehat{T}_{i,\mathrm{M2b}}^{(b)} - \widehat{T}_{i,\mathrm{EQ}}^{(b)}}{[\widehat{T}_{i,\mathrm{M2b}}^{(b)} + \widehat{T}_{i,\mathrm{EQ}}^{(b)}]/2} \right)^2 \right]^{1/2}. \quad (10)$$

The frozen executed gate required both a median separation of at least 0.05 and a resolvability ratio of at least 2.0. The latter is the bootstrap median divided by the 95% interval half-width. The 2.0 value is a prespecified characterization-uncertainty resolvability heuristic; the 0.05 value is a separate prespecified experimental-design continuation threshold. No external, application-level, decision-theoretic, or inferential calibration was found for either value. The heuristic is not a hypothesis, significance, equivalence, confidence, or power test. Earlier archived closure records used 1.0 before the executed 2.0 rule was fixed pre-characterization; the record contains no frozen rationale for that change, which is disclosed rather than retrospectively justified. Passing the former and failing the latter means that the estimated difference passed the stated characterization-uncertainty relative-precision heuristic while its magnitude remained below the design threshold.

3.5 Structural discrimination boundary

Let r_i be the occupancy-aware worker rate, $\bar{r} = R/n$ its equalized counterpart, $R = \sum_i r_i$, and a_i the frozen activation. Exact subtraction of the two cumulative intensities gives the proven mathematical identity

$$\Delta\Lambda(t) = p \sum_i (r_i - \bar{r})(t - a_i)_+. \quad (11)$$

For $t \geq a_{\max}$, define $\mu_{ar} = \sum_i r_i a_i / R$ and $\bar{a} = n^{-1} \sum_i a_i$. Then

$$\Lambda_{\mathrm{M2b}}(t) = pR(t - \mu_{ar}), \quad (12)$$

$$\Lambda_{\mathrm{EQ}}(t) = pR(t - \bar{a}), \quad (13)$$

so both have slope pR and displacement $\delta = \mu_{ar} - \bar{a} = \mathrm{Cov}(r, a)/\bar{r}$, where covariance and standard deviations below are population quantities over workers. Equality after full activation follows from zero rate–activation covariance, but equality for the complete path is stronger: for every distinct activation value a, $\sum_{i:a_i=a}(r_i - \bar{r}) = 0$. Equal rates and equal activation are special cases.

The relative-shift envelope requires a separate post-full surrogate. Let $\widetilde{T}_k^M = \mu_{ar} + q_k$ and $\widetilde{T}_k^E = \bar{a} + q_k$ share the same post-full waiting term, and assume both surrogate means are at least $a_{\max}$. Then

$$|\delta| = \frac{|\mathrm{Cov}(r, a)|}{\bar{r}} \leq \mathrm{CV}(r)\sigma_a, \quad (14)$$

$$\sigma_a \leq \frac{a_{\max} - a_{\min}}{2} \leq \frac{a_{\max}}{2}, \quad (15)$$

where the first inequality is Cauchy–Schwarz and the second is the population range bound. Because the symmetric denominator $D_k = (\widetilde{T}_k^M + \widetilde{T}_k^E)/2$ is at least $a_{\max}$,

$$\frac{|\widetilde{T}_k^M - \widetilde{T}_k^E|}{D_k} \leq \frac{|\delta|}{a_{\max}} \leq \frac{\mathrm{CV}(r)}{2}. \quad (16)$$

This is an exact bound for the stated post-full surrogate. For the actual NHPP means, write

Table 2 Information-state and estimator definitions. Capacity-side reuse does not imply reuse of target-specific activation.

Estimator	Capacity representation	Activation representation	Intended interpretation
M1 / U-M1	Nominal-size pooled state	Pooled operational activation	Coarse analytic baseline
M2a	Target-placement aggregate	Target-specific operational activation	Full target-state comparator
M2b / U-M2b / X-M2b	Composed node-local capacity	Target-specific or frozen campaign activation	Capacity-side characterization reuse
U-EQ / X-EQ	Equalized capacity at the same aggregate rate	Same frozen activation as paired model	Discriminability comparator, not equivalence

$\mathbb{E}[T_k^M] - \mathbb{E}[T_k^E] = \delta + \varepsilon_k$, where ε_k is the full-path correction. Interpreting the surrogate bound as an approximation to actual mean completion therefore requires negligible pre-$a_{\max}$ kth-arrival mass for both models. That mass was not quantitatively bounded in the frozen experiment, $\mathbb{E}[T_k] > a_{\max}$ alone is insufficient, and 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. These are mathematical results about the model, not causal statements about the physical source of X411's rates or activations.

3.6 Completion repetitions and finite-repetition uncertainty

Cell means are summaries of 20 accepted repetitions and are not uncertainty-free. For each cell we report the arithmetic mean, the sample standard deviation with denominator $n-1$, the coefficient of variation, the standard error of the mean, and the relative standard error of the mean. Cell means are formed by summing authoritative integer completion nanoseconds and dividing once, so the reported values differ from earlier serializations only at floating-point precision.

Interval overlays are nonparametric percentile intervals from 50,000 resampling draws with linear quantile interpolation at the 2.5th and 97.5th percentiles, under a frozen resampling contract fixed before the intervals were inspected. Resampling is block aware. For P3 and P4-A a draw samples the 20 accepted repetitions within each cell independently with replacement. For UCV a draw samples the ordered set of 20 complete protocol blocks with replacement once and applies that identical sampled block-index vector to every UCV cell, so completion runs are never treated as independent units across cells. The same paired UCV block draws support the U6-A, U6-B, and U6-C endpoint-slope overlays. Draw counts and the deterministic generator seed are recorded in the archived analysis contract.

These are *descriptive 95% finite-sample uncertainty intervals* for the quantity resampled. They are not confirmatory tests of cross-phase differences, model superiority, model equivalence, or causal placement effects; no significance is asserted from them. The interpretation classification for this evidence is qualitative with an uncertainty overlay.

3.7 Uncertainty and paired comparisons for the primary phases

The UCV intervals use the prospectively specified 50,000-draw percentile bootstrap with linear quantiles. Each draw resamples, independently and with replacement, the 10 complete input-characterization rounds and the 20 complete completion blocks. Each sampled input-characterization round retains its complete startup vectors; the startup empirical mixture is reconstructed within every bootstrap replicate before capacity- and activation-based predictions and the 20-run observed mean for every target–k cell are recomputed, followed by the unweighted MAPE across the 18 cells. Thus activation uncertainty is included, and neither individual workers nor the 360 completion runs are treated as independent model-comparison units. Reproducibility is bound to NumPy `Generator(PCG64)` with fixed seed `279745038652071891812215928358593692476` and deterministic fixed-index chunks. These intervals describe finite-sample uncertainty under

the frozen resampling scheme; UCV remains descriptive-only, and the intervals are not confirmatory tests of cross-phase differences, model superiority, model equivalence, or causal placement effects.

For P4-A model comparisons, the paired unit is one of the 36 configuration–k cell means. Each contrast uses the paired difference in cell APE. The preserved analysis used 50,000 paired-cell nonparametric bootstrap draws of the mean APE difference and two-sided paired Wilcoxon signed-rank tests. Holm adjustment was applied across exactly three prespecified comparisons: M1–M2a, M1–M2b, and M2a–M2b. The M2a–M2b adjusted value of 0.1327 therefore does not reject equal paired-error location at the 0.05 family-wise level; it neither establishes equivalence nor reverses the observed MAPE ordering.

The later M1–M2b UCV comparison is explicitly post-hoc and descriptive; its paired unit and resampling details are reported in the Supplement. No inferential p-value or prospective superiority claim is assigned.

3.8 Post-hoc timing-level interpretation diagnostics

Two outcome-informed diagnostics localize, but do not causally decompose, completion-level discrepancies. First, for every accepted completion run we evaluate the frozen model both at the run's observed activation vector and at an all-active counterfactual obtained by setting every activation time to zero. The difference is a model-implied all-active activation effect, not an additive component of the observed completion time. Second, for each placement and adjacent target pair relative to the smallest target, we compare the observed and frozen-model increments $T_k - T_{k_0}$. The same construction is applied to all P3, P4-A, and UCV placements. These diagnostics use completion outcomes, are labelled post-hoc interpretation diagnostics, and neither replace nor correct any primary prospective estimate. The UCV endpoint-slope summaries use the paired block resampling described above.

3.9 Post-hoc residual calibration and leave-one-k-out

Additive-offset and leave-one-k-out analyses ask how much residual structure can be absorbed by outcome-informed scalar calibration. They are post-hoc diagnostics, not prospective validation or corrected prospective performance; their estimators, tie rule, and cross-validation mechanics are reported in the Supplement. No offset is applied to any prospective result in Section 4.

3.10 Characterization burden accounting

Burden is scientific characterization measurement burden expressed in worker-hours; it is not cloud billing, energy, or end-to-end deployment time. For each representation we form the union of the unique acquisition records that representation requires, and sum their measured worker-time. Because the accounting is a union over unique records, shared records are counted once and burden is not additive across representations.

Burden is reported by explicit *scope*. Total required characterization is the union over all acquisition records the representation needs; capacity-only characterization is the strictly narrower union restricted to capacity acquisition; activation characterization is the corresponding activation-side union. A reduction computed on one scope is not a reduction on another, and the two are reported separately throughout. The accounting rule is unchanged across phases, but the frozen model definitions require different record sets: P4 M1 and M2a use the same placement-total capacity record union, whereas UCV U-M1 uses its phase-specific reference-placement capacity set. UCV burden is therefore computed from UCV acquisition records, and P4 burden figures are never reused as UCV burden. Node-local composition separately compares independently characterized node-local capacity sums with independently measured placement-total capacity. Relative SEM is a measurement-precision diagnostic, not a confidence interval for prediction error.

Zero-fit means no target completion outcome fits the analytic predictor. It still requires independently measured capacity and activation. The

retrospective fitted comparator and the outcome-conditioned diagnostics lie outside that prospective information boundary.

3.11 Evaluation design and phase roles

P3 and corrected P4-A each contain measured-regime prospective cells. The source-data configuration identities E6, B6, and P6 are displayed as P3-E6, P3-B6, and P3-P6 to provide phase context without redefining them. Supplementary Table S8 records the execution environment and acquisition design of each phase, and the accompanying Supplementary Methods give the capacity, activation, completion, freeze, and record-acceptance protocols. Exact environment-level reconstruction of the historical P3 campaign is nonetheless limited by metadata its frozen record does not preserve; Table S8 reports those fields as *not recorded* rather than inferring them from the later phases, which ran on separate clusters. This is a boundary on the historical environment record, not on the P3 measurements themselves, which are preserved in full.

UCV uses U6-A=(3,2,1), U6-B=(1,3,2), and U6-C=(2,1,3) over $k \in \{1, 3, 5, 10, 20, 40\}$. N=6 itself was already present in P3. The UCV novelty is the occupancy composition $\{3, 2, 1\}$ and its cyclic assignments. A predefined historical-data exclusion check found no prior target aggregate-capacity, startup-only, or completion observation for these configurations. UCV used fresh capacity characterization; historical P4 capacity coefficients were not prospectively reused.

UCV freezes 54 model–cell predictions before 360 accepted completion outcomes are opened. The UCV claim scope was fixed prospectively as descriptive-only before those target outcomes were opened rather than chosen afterwards. Before the freeze, a numerical-feasibility amendment – a uniform tightening of the numerical integration tolerances used to evaluate the predictions, applied identically to every prediction row and using no outcome information – was applied.

3.12 Retrospective fitted comparison

The fitted analysis asks a bounded substitution question: could direct fitting to preserved completion outcomes match the structured analytic reference within the evaluated measurement budgets? It is retrospective and cannot support a prospective performance claim. Evaluation uses leave-one-configuration-out folds and a fixed budget grid; H1 additionally consumes the frozen M2b prediction and is therefore not a standalone completion-only comparator. The comparison is deliberately information-asymmetric: C1 and C1-I extrapolate to each held-out configuration from completion outcomes observed in the other represented configurations without using the held-out configuration's completion outcomes, whereas M2b receives independently characterized state required for the target configuration but no target completion outcome. It should therefore be read as an information-source substitution comparison rather than a same-information model contest. Full model equations, fold mechanics, budget definitions, burden scopes, and all evaluated rows appear in Supplementary Methods and Table S5.

3.13 Provenance and evidence boundary

Three provenance facts bound the corpus. An earlier P4 campaign is excluded because its receive-order semantics violated the intended protocol, and it is named only as excluded provenance. The UCV numerical-feasibility amendment was verified as pre-outcome. After outcomes were opened, a transport/parser corrective applied the predefined decoder and deterministic earliest-valid chronology. Frozen predictions were unchanged. Of 362 physical attempts, 360 form the accepted analytical corpus and two later attempts at the affected planned position are permanently excluded as redundant; admission used neither outcome magnitude nor prediction-residual desirability.

The effect of that parser corrective is classified as a bounded scientific effect, and its evaluability boundary is stated explicitly rather than resolved. Although the preserved raw attempts can be decoded, re-decoding the first attempt does not

reconstruct the counterfactual campaign: immediate admission would have prevented two retries and changed the subsequent execution timing and cluster chronology. Whether the pre-corrective accepted-record set, the observed cell means, the primary UCV metric, the locked bootstrap interval, or the post-closure activation and Forward-10 diagnostics would have differed is therefore *not evaluable from the bound authority*. What the authority does establish is that the frozen predictions did not change. The current manuscript reports the locked post-corrective analytical corpus; the bound provenance identifies no further numerical correction to those locked values. We therefore do not claim that the corrective was shown to have zero effect on the primary UCV metric.

4 Results

4.1 Prospective prediction in measured placements

In P3, cell-mean MAPE was 1.82% for M1, 1.58% for M2a, and 1.56% for M2b. In corrected P4-A, MAPE was 4.69%, 2.54%, and 2.67%, respectively. M2a had the lowest observed P4-A MAPE, but the paired M2a–M2b comparison was not Holm-significant ($p_{\text{Holm}} = 0.1327$). The results support low prospective error in the measured regimes, not a statistically established M2b/M2a ordering and not equivalence.

For structural context only, the outcome-independent N=6 criterion selects P3-E6=(2,2,2) and P3-P6=(4,2,0). This **post-hoc descriptive subset** aligns with UCV on $N = 6$, three worker nodes, and $k \in \{1, 3, 5, 10, 20, 40\}$; M2b MAPE was 1.05% and signed MPE was −0.42%. Its by-k MAPE remained between 0.39% and 1.61% (Supplementary Table S10) and did not show the pronounced low-k concentration seen in UCV. Each by-k summary nevertheless averages only the two N=6 P3 cells, and at several k values the prediction error is comparable with the finite-repetition precision of the observed cell means; the subset is descriptive context rather than a precise cross-k shape estimate. P3 and UCV ran in separate environments, P3 preserves incomplete environment metadata, and no historical P3 startup vectors are preserved. The comparison is therefore matched structural context, not a controlled placement comparison, a causal effect size, or evidence that activation alone explains the cross-phase difference.

Table 3 Compact result-only view of prediction performance by phase. Evaluation roles, configuration scopes, and cell counts remain explicit in Table 1; estimator states remain explicit in Table 2. P3 and P4-A are prospective primary measured-regime evaluations. UCV predictions were frozen prospectively; after the corrective, their outcome evidence is descriptive historical evidence. Intervals are shown only where a registered phase-specific interval is available; the UCV intervals are descriptive 95% finite-sample uncertainty intervals. The final note preserves the predefined signed companion result. MAPE here is total-completion-time error, the phase primary metric; increment-based errors relative to each placement's smallest target are a distinct diagnostic estimand and are reported in Section 4.8.

Phase	Model	MAPE (%)	95% CI
P3	M1	1.82	
	M2a	1.58	
	M2b	1.56	
P4-A	M1	4.69	
	M2a	2.54	
	M2b	2.67	
UCV	U-M1	8.37	[6.90, 10.29]
	U-M2b	7.42	[5.99, 9.43]
	U-EQ	7.41	[5.99, 9.43]

Predefined companion: U-M2b signed MPE = +7.42%.

Restricting the two separate phase summaries to the shared target set $k \in \{1, 10, 40\}$ gave M2b MAPE of 2.51% over 27 corrected P4-A cells and 8.28% over nine UCV cells. The cellwise MAPE definition is unchanged and no repetitions or cells are pooled across phases. This deterministic descriptive check aligns the displayed k support only; it is not a controlled cross-phase comparison or evidence of a causal placement effect.

4.2 Capacity composition and empirical reuse in P4

P4 supplies the strongest direct evidence for a reusable capacity-state representation. Productive per-worker capacity was 15.197, 15.145, 15.061, and 12.926 Mhash/s at local occupancies one through four, respectively (Table 4). Occupancy four retained 85.05% of the occupancy-one rate, while fixed-occupancy cross-node CV remained

Table 4 Frozen P4 node-local productive capacity by local occupancy. Means are per worker across the nine node-level means; CV is dispersion across those homogeneous nodes at fixed occupancy. Relative capacity is normalized to occupancy one.

Occ.	Capacity (Mhash/s)	Rel. to $n = 1$ (%)	CV (%)
1	15.197	100.00	0.303
2	15.145	99.65	0.294
3	15.061	99.10	0.376
4	12.926	85.05	0.320

between 0.294% and 0.376%. Thus the composition result is not a consequence of a flat capacity curve: capacity changes materially at occupancy four even though occupancy-conditioned measurements remain stable across the homogeneous nodes. The measurements do not identify a hardware, scheduler, cache, or other physical mechanism.

Against that non-flat occupancy curve, the maximum absolute discrepancy between composed node-local capacity and independently measured placement-total capacity was 0.45% across the nine target configurations. The maximum occurred for the concentrated $N = 24$ placement $[4, 4, 4, 4, 4, 4, 0, 0, 0]$, which places four workers on each of its six active nodes, and was still only 0.45%. Composing node-local capacity therefore reproduced the independently measured placement total closely enough to serve as predictive state in this regime, which is what makes capacity-side reuse empirically available rather than assumed.

This supports comparatively reusable capacity characterization within the evaluated homogeneous P4 pool. It does not establish universal capacity composition, transfer to heterogeneous hardware, or prospective reuse of historical capacity coefficients in UCV.

The frozen P4 acquisition geometry also registers a deterministic capacity-characterization ratio. Let B_{NL} denote the measured worker-time of the shared node-local sweep and b_{PT} the mean worker-time for placement-total capacity characterization of one target. Their ratio, denoted here by $Q_{\mathrm{cap}} = B_{\mathrm{NL}}/b_{\mathrm{PT}}$ to avoid conflict with the fitted workstream's notation, was 4.9945, displayed as approximately 4.99 target placements. This is an acquisition-geometry ratio computed from measured characterization durations, not an outcome-performance measure. Below the measured P4 acquisition threshold, a shared node-local sweep need not be cheaper than a smaller target-specific reference set; UCV uses a different phase-specific required-record geometry and is not an independent validation of Q_{cap}. The ratio is not a universal break-even, billing threshold, minimum budget, or deployment rule.

4.3 Characterization burden by scope

Burden separates by scope (Table 5, Fig. 2). Over all nine P4-A target configurations, total required characterization was 25.50 worker-hours for both M1 and M2a and 16.44 worker-hours for M2b, a reduction of 35.52%. Restricted to the capacity-only scope, the same contrast was 20.35 versus 11.29 worker-hours, a reduction of 44.51%. P4 M1 and M2a have identical total and capacity burdens because their preserved definitions require the same placement-total capacity record union. The activation component was 5.15 worker-hours for all three representations, which is why the capacity-only reduction is the larger figure: the shared activation requirement dilutes the total-scope reduction. The two percentages describe different scopes and are not interchangeable.

Burden is also phase specific. In UCV, total required characterization was 2.31 worker-hours for U-M1 and 4.31 worker-hours for both U-M2b and U-EQ, with a common activation component of 0.98 worker-hours. U-M1 uses the preserved UCV reference-based acquisition definition and must not be read as having the same required-record geometry as P4 M1. These UCV figures come from UCV acquisition records; the P4 worker-hour figures are not UCV burden and are not transferred to it. That U-M2b and U-EQ share the same burden reflects a shared required record set, not model equivalence.

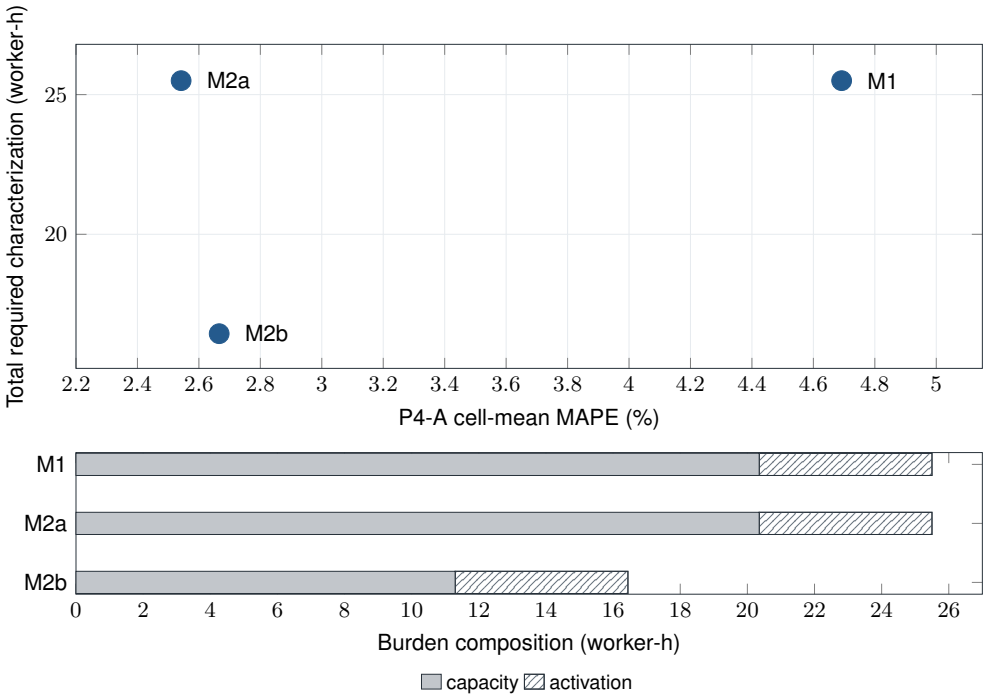


Fig. 2 P4-A prediction error versus characterization burden by representation. The trade-off plane shows cell-mean MAPE against total required characterization, and the aligned strip separates capacity and activation acquisition burden. The displayed M2a–M2b accuracy ordering was not statistically established under the prespecified Holm-adjusted comparison ($p_{\mathrm{Holm}} = 0.1327$); the figure does not establish equivalence. The reported M2b-versus-M2a reductions remain 35.52% for total required characterization and 44.51% for capacity-only characterization. These are scope-specific measurement-burden reductions, not performance effects.

Table 5 Characterization burden by **scope** and phase, in worker-hours. Worker-hours quantify scientific characterization measurement burden, not billed cloud cost, energy, or end-to-end deployment time. TOTAL REQUIRED CHARACTERIZATION is the union of unique acquisition records a representation requires; capacity-only is narrower. Required sets are model- and phase-specific. The short scope notes name only each row's representation; H1 completion-outcome acquisition presupposes the indicated M2b state. UCV burden is never inherited from P4.

Representation	Total required	Capacity only	Activation	Scope note
P4-A, all nine target configurations				
M1	25.50	20.35	5.15	Pooled
M2a	25.50	20.35	5.15	Placement total
M2b	16.44	11.29	5.15	Node-local composition
M2b vs M2a reduction: **35.52%** total required; **44.51%** capacity only.				
UCV, accepted characterization acquisition set (phase-specific; not P4 burden)				
U-M1	2.31	1.33	0.98	Pooled
U-M2b	4.31	3.33	0.98	Node-local composition
U-EQ	4.31	3.33	0.98	Equalized comparator
Retrospective fitted H1 information boundary; nine LOO folds at $r = 10$				
H1 incremental, conditional on M2b	12.98	–	–	Outcome acquisition; requires M2b
H1 standalone from scratch	29.43	–	–	Includes M2b characterization
M2b standalone reference	16.44	11.29	5.15	Frozen analytic reference

4.4 The structurally unseen $\{3, 2, 1\}$ placement family

Frozen U-M2b achieved 7.42% MAPE on the structurally unseen family, with the predefined bootstrap interval 5.99–9.43%; this prospective result is stated before any outcome-conditioned diagnostic. Its prospectively predefined signed companion MPE was +7.42%. Every U-M2b cell-level signed error was positive, so MPE equals MAPE in magnitude for this target set. U-EQ achieved 7.41% MAPE with interval 5.99–9.43%, and U-M1 achieved 8.37% with interval 6.90–10.29%. These observations are reported under the descriptive-only scope fixed prospectively before the target outcomes were opened.

U-M2b MAPE was 11.10% on U6-A, 8.00% on U6-B, and 3.14% on U6-C. The frozen result therefore varied across cyclic placements within the prespecified structurally unseen family. Because P3/P4 and UCV are separate experimental phases rather than a controlled between-phase treatment comparison, their differences in error are interpreted as an empirical transfer-scope observation, not as a causal effect of occupancy composition alone. This is descriptive evidence about one occupancy-composition family, not a proof of broad generalization.

The design also limits what UCV can adjudicate. U-M2b and U-EQ frozen predictions differed by only 0.000246 s on average and at most 0.000597 s, and discriminability is classified as low. The prespecified U6 family therefore offered little ex-ante leverage between these capacity representations, and their similar errors are evidence of low discriminability rather than of equivalence. No equivalence test was performed and none is claimed.

The capacity authorities explain this weak design leverage without resolving the observed error causally. Every $\{3, 2, 1\}$ target uses only local occupancies one through three. In the P4 occupancy curve, per-worker capacity at occupancies two and three retained 99.65% and 99.10% of the occupancy-one rate, compared with 85.05% at occupancy four; the fresh UCV capacity estimates likewise compose the target family only from occupancies one through three. The prospective UCV design therefore had little capacity-side leverage to distinguish composition from equalization. This explains the observed low ex-ante U-M2b/U-EQ discriminability as a design property, without implying equivalence or attributing the larger UCV residual to placement novelty alone.

U-M1 used the preserved equal-weight pooled reference constructed from UCV reference placements $[4, 2, 0]$ and $[2, 2, 2]$, the former of which exercises local occupancy four. That pooled reference capacity is lower than the composed-capacity value for each $\{3, 2, 1\}$ target, so the observed U-M1 disadvantage is partly reference-composition dependent and must not be read as a universal failure of pooled state.

Registered by-k, cyclic-spread, and post-hoc M1–M2b paired diagnostics are reported in the Supplement. They preserve the distinction among pooled signed mean-ratio error, cellwise MAPE, and post-hoc paired comparison and do not establish a physical placement mechanism or prospective model superiority.

4.5 X411: relative-precision heuristic passed and continuation gate failed

X411 ended at its pre-outcome gate with zero target-completion outcomes, and the characterization behind that decision was complete and sound. The campaign executed the prescribed characterization and prediction-freeze sequence without target-outcome access. The model-defined equal-share rates were 15.174 Mhash/s at occupancy one and 12.935 Mhash/s at occupancy four, a retention of 85.25%. Across the three measured nodes, node-total population CV was 8.60×10^{-4} at occupancy one and 8.40×10^{-4} at occupancy four. Composed and independently measured direct placement capacities differed by at most 0.103% across X411-A, X411-B, and X411-C. This supports close node-total composition in this environment; no equivalence margin or procedure was prespecified, and the direct capacities did not enter primary prediction.

A post-hoc outcome-free diagnostic of the frozen occupancy-four records retained all four simultaneous worker rates for 120 accepted node-rounds (480 worker observations). Median within-node CV was 0.1557 and median max/min rate ratio was 1.409. Thus $C(\text{node}, 4)/4$ is not empirically established as the physical worker-rate vector. The node total remains the exact sum of

Table 6 UCV evidence for the one structurally unseen $\{3, 2, 1\}$ placement family, separated by evidence class. The frozen-before-outcome primary block is retained as descriptive historical outcome evidence after the post-outcome corrective. The design-diagnostic block is outcome-independent and establishes low U-M2b/U-EQ discriminability, not equivalence. The final block is post-hoc and non-deployable.

Evidence item	Metric	Display value	Evidence class / role
Frozen-before-outcome UCV primary; post-corrective outcome evidence is descriptive/historical			
Frozen U-M2b primary	MAPE	7.42%	Primary metric
Frozen U-M2b companion	Signed MPE	+7.42%	Predefined companion
U6-A frozen U-M2b	MAPE	11.10%	Configuration summary
U6-B frozen U-M2b	MAPE	8.00%	Configuration summary
U6-C frozen U-M2b	MAPE	3.14%	Configuration summary
Post-hoc design diagnostic (outcome-independent)			
U-M2b versus U-EQ displacement	Mean absolute prediction difference	0.000246 s	No equivalence claim
U-M2b versus U-EQ discriminability	Classification	LOW	Little design leverage
Post-hoc activation and temporal diagnostics			
Placement-pooled LOO realized activation	MAPE	1.46%	Excludes own vector; non-deployable
Same-run realized activation	MAPE	1.14%	Leakage-assisted; non-deployable
Forward-10 frozen later block	MAPE	9.82%	Temporal transfer
Forward-10 updated later block	MAPE	5.05%	Temporal transfer

the four rates, so this finding narrows worker-level interpretation without changing composition totals, primary predictions, gate arithmetic, or the historical stop.

The 10,000-replicate characterization-only gate gave global relative-RMS separation median 0.001637 (0.164%) with a 95% interval of 0.001041–0.002376. Its resolvability ratio was 2.453, above the frozen executed 2.0 characterization-uncertainty relative-precision heuristic. The same separation was far below the frozen 0.05 experimental-design continuation threshold. The predicted separation therefore passed the prespecified heuristic but remained far below the prespecified continuation threshold, and the frozen protocol did not authorize target-completion collection: target run count and target-outcome access count both remained zero (Table 7). This is a completed prospective negative design decision, not a failed execution, an outcome-based stopping rule, a hypothesis, significance, equivalence, or calibrated power analysis. Outcome-free sensitivity of these frozen statistics to alternative display cutoffs is reported in Supplementary Table S15.

4.6 Structural interpretation and terminal family closure

Representation discrimination lives in the full rate–activation interaction path. The exact identities in Section 3.5 show that the two model representations can differ through that path, not through rate heterogeneity alone and not through the post-full centroid term alone; for X411 the predicted separation remained small relative to the prespecified criterion. The theorem accepts arbitrary r_i and is unaffected by the observed occupancy-four heterogeneity. The historical gate correctly compared its two prespecified representations, X-M2b and X-EQ; it did not compare X-EQ with an unknown true physical worker-rate vector. Calling the identity an architectural explanation means that the predictor's equations account for the prediction displacement; it does not identify a physical mechanism.

Table 7 Prospective X411 $\{4, 1, 1\}$ characterization-only separation gate. The relative-RMS separation is computed across the 18 placement–k rows, each containing frozen X-M2b and X-EQ predictions. The interval and resolvability statistic use characterization uncertainty only. Passing the prespecified 2.0 rule means the estimated separation passed the characterization-uncertainty relative-precision heuristic; failing the separate prespecified 5% design threshold is not an equivalence result. Neither threshold is a formal inferential test. No target completion outcomes were collected.

Quantity	Frozen criterion	Result
Prediction cells	–	18 X-M2b; 18 X-EQ
Bootstrap replicates	–	10,000
Global relative-RMS median	≥ 0.050	0.001637 (0.164%)
95% characterization-only interval	–	[0.001041, 0.002376]
Resolvability ratio	≥ 2.0	2.453 (criterion met)
Design-separation gate	≥ 0.050	FAIL
Overall gate	both criteria pass	FAIL
Target completion runs	authorized only after PASS	0

The corrected structural evidence bounds stronger statements. Within the historical observed-capacity and activation region, the maximum screened relative separation was 0.003766 (0.377%); an earlier 0.049995 value came from an activation-dispersion extrapolation by a factor of 200 and is not historical evidence. For the X411-A-derived normalized certificate input $\rho = 0.853516$, $\mathrm{CV}(r)/2 = 0.0383$. This is an exclusion envelope, not a tight predictor: it is 23.4 times the frozen global separation median 0.00164 and 11.8 times the largest frozen cell separation 0.00325. The exact contention boundary at which this surrogate post-full envelope ceases to exclude 5% is $\rho_* = (101 - 15\sqrt{2})/98 = 0.8142$, not 0.75; it is neither a physical threshold nor an exact bound on the actual NHPP mean difference. Permitting 5% under the envelope does not show that completion separation attains 5%. The minimally constrained formal model class can attain 5%, but attainability within homogeneous GKE is not established because universal activation, rate-heterogeneity, and capacity-heterogeneity bounds are absent.

Corrected successor screening then exhausted the canonical six-worker, three-node family universe (Table 8). Five of seven families had been used scientifically, but historical use is not the same as current numerical evaluability or rankability. The active predictor requests $C(\mathrm{node}, n)$ for all three nodes without a zero-occupancy guard, so every zero-containing family requires undefined $C(\mathrm{node}, 0)$. In addition, the two unused families, $\{6, 0, 0\}$ and $\{5, 1, 0\}$, require unmeasured $C(\mathrm{node}, 6)$ or $C(\mathrm{node}, 5)$, exceed the four-worker-per-node measured design capacity, and break the equal-aggregate-rate premise; no capacity was invented. Node hardware homogeneity was not substituted for exact measured-capacity equality, and no family was proved exactly invariant from machine type alone. Consequently no unused family was eligible for ranking or execution authorization, and no successor target campaign ran under the stated screening constraints.

4.7 Completion finite-repetition uncertainty

Every cell mean summarizes 20 accepted repetitions and therefore has finite-sample uncertainty. The full cellwise dispersion and all-placement UCV endpoint-slope overlays are reported in Supplementary Table S6. Observed-minus-predicted endpoint-slope mismatches were -0.0147, -0.00260, and -0.00205 s per unit k for U6-A, U6-B, and U6-C, respectively. Every descriptive interval included zero; these overlays do not establish a capacity-side anomaly or statistical significance.

4.8 Post-hoc completion-level interpretation diagnostic

Across all accepted runs and target levels, the model-implied all-active activation effect accounted for a median 86.0–87.8% of predicted completion time across phases and placements. At the minimum target the range was 98.0–99.1%,

Table 8 Corrected terminal screening of the canonical six-worker, three-node occupancy-family universe. Historical use, current numerical support, measured-design capacity, and successor eligibility are separate questions. The active predictor evaluates all three node terms without a zero-occupancy guard; no missing capacities were invented.

Family	Historical use	Current predictor	X411 support	Successor status
$\{4,1,1\}$	X411	Defined	C_1, C_4 measured; within four-worker limit	Already used; not an unused successor
$\{3,2,1\}$	UCV1	Defined only if C_2, C_3 supplied	C_2, C_3 absent from X411; within limit	Already used; not rankable from X411 alone
$\{4,2,0\}$	P3-P6	Undefined: C_0 requested	C_0, C_2 absent; within limit	Already used; not currently rankable
$\{3,3,0\}$	P3-B6	Undefined: C_0 requested	C_0, C_3 absent; within limit	Already used; not currently rankable
$\{2,2,2\}$	P3-E6; X411 sentinel	Defined only if C_2 supplied	C_2 absent; within limit	Already used; not rankable from X411 alone
$\{6,0,0\}$	None	Undefined: C_0, C_6 requested	C_0, C_6 absent; exceeds limit	Unused and unsupported; not rankable
$\{5,1,0\}$	None	Undefined: C_0, C_5 requested	C_0, C_5 absent; exceeds limit	Unused and unsupported; not rankable
Successor-eligible rankable families				0
Eligible unused family selected				None
Successor target campaign performed				No

whereas at the maximum target it was 55.6–62.0%. These fractions describe a frozen-model counterfactual, not an additive decomposition of observed time. Total-time MAPE remains mathematically valid for the actual prediction target because operational activation was a prospectively characterized predictor input; it should not be read as search-only or offset-insensitive accuracy.

For increments relative to each placement's smallest target, median absolute percentage errors 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 activation-time shift moves the model completion-time mean and distribution by the same amount while leaving relative activation staggering unchanged, which is why increments remove the common shift but not the stagger structure. These outcome-informed summaries are post-hoc interpretation diagnostics; they neither replace primary results nor authorize causal attribution.

4.9 Post-hoc residual calibration diagnostics

The frozen UCV residual is systematic in sign, which raises the question of how much a single additive correction could absorb. All results in this subsection consume held-out or in-sample cell outcome means and are post-hoc; none is a corrected prospective performance figure, and none is applied to the prospective results above.

The registered in-sample and leave-one-k-out results show that a scalar offset absorbs part, but not all, of the residual and performs least well at low k. Complete values and estimator sensitivity are reported in Supplementary Table S7 and Fig. S2.

4.10 Realized-activation and temporal diagnostics

Across the separate P4-A and UCV phases, the post-hoc same-run activation substitution reduced M2b MAPE from 2.67% to 0.90% and from 7.42% to 1.14%, respectively. It uses outcome-informed activation from the evaluated run and is non-deployable and explanatory only; these values are neither corrected prospective performance, causal attribution, nor a controlled cross-phase effect estimate. A separately frozen placement-pooled leave-one-run-out diagnostic excludes the evaluated run's vector, uses the 119 other accepted same-placement vectors, and reached 1.46% MAPE. This diagnostic is also post-hoc and non-deployable; the 1.46%–1.14%

difference only bounds the incremental association with evaluated-run information and is neither corrected prospective performance nor a causal decomposition.

The post-hoc two-by-two UCV sensitivity gave 7.42% MAPE with frozen activation and frozen capacity and 7.35% after substituting detector-window capacity; with same-run activation, the corresponding values were 1.14% and 1.15%. The capacity proxy is the trailing median of ten fixed 0.5-second detector buckets ending at steady declaration and covers that steady-declaration detector window, not the full productive-search interval. The two axes are not symmetric measurements, and the corpus cannot make them symmetric: no independently prespecified fixed-window throughput measurement spanning productive search was recorded for every required worker and run (Section 5.7), so no capacity axis comparable to the activation axis is available. This grid is therefore a sensitivity comparison rather than an error decomposition or causal attribution, and it is not corrected prospective performance or a deployable estimator. The detailed grid is retained in Supplementary Table S14 and does not exclude later within-run throughput changes.

A separate stopping-conditioned rate diagnostic is reported over all 18 placement–k strata in Supplementary Table S13. In the post-hoc late-k subset, defined after the primary diagnostic as $k \in \{20, 40\}$, the mean ratios were 1.001, 1.001, and 1.005 for U6-A, U6-B, and U6-C. This does not support a large persistent U6-A average-rate excess, provide an independent capacity validation, or establish causal attribution.

The post-hoc Forward-10 diagnostic reduced later-block UCV MAPE from 9.82% to 5.05%, but it establishes neither freshness nor an optimal refresh interval. Figure 4 summarizes the same-run activation and Forward-10 temporal-transfer diagnostics.

Frozen UCV mean-cell APE was concentrated at low k and was 16.91% at $k = 1$. This is a stratified descriptive summary, not a new primary metric. Detailed by-k values and last-activator distributions are reported in the Supplement. These post-hoc patterns are consistent with activation-state mismatch but do not identify a scheduler, hardware, or transport mechanism.

A direct state-distribution comparison sharpens that bounded interpretation (Fig. 3). For each U6 placement, sorting each preserved six-worker vector and averaging at each order statistic shows that the realized completion-run profile differs most from the frozen characterization at the earliest activation positions; the profiles are much closer at later positions. This post-hoc pattern is consistent with the concentration of frozen prediction error at low k and with the same-run sensitivity diagnostic, but it does not assign a causal fraction, establish a sole cause, exclude capacity or model mismatch, or provide a deployable correction.

At the earliest order statistic, the realized-minus-frozen shifts for U6-A, U6-B, and U6-C were -2.11, -1.59, and -0.66 s, respectively; the corresponding $k = 1$ frozen-prediction excesses over the observed cell means were $+2.08$, $+1.35$, and $+0.73$ s. The quantities are of similar scale across the three placements, consistent with small-k completion sensitivity to the earliest productive exposure. This is post-hoc state-distribution context, separate from the leakage-assisted same-run predictor replacement, and is not causal attribution.

The activation diagnostic remains an outcome-sensitive explanatory sensitivity, not relative attribution. The frozen occupancy-four capacity records additionally show substantial within-node worker-rate heterogeneity, while the primary representation assigns each node total equally across workers. Worker-rate allocation is therefore a plausible competing explanation for ramp-period discrepancies and precludes unique causal attribution. The available artifacts do not identify the physical mechanism underlying this heterogeneity; in particular, no CPU affinity or sibling mapping establishes a simultaneous-multithreading mechanism. UCV used fresh capacity characterization, and historical P4 capacity coefficients were not prospectively reused.

A separate post-hoc, outcome-free capacity sensitivity replaced the fixed final-occupancy rate during startup with the independently measured occupancy-conditioned capacities permitted by each frozen activation chronology. Among the evaluated low-k predictions, the maximum change magnitude was 10.221 microseconds (0.000103%). An independent implementation compared all 36

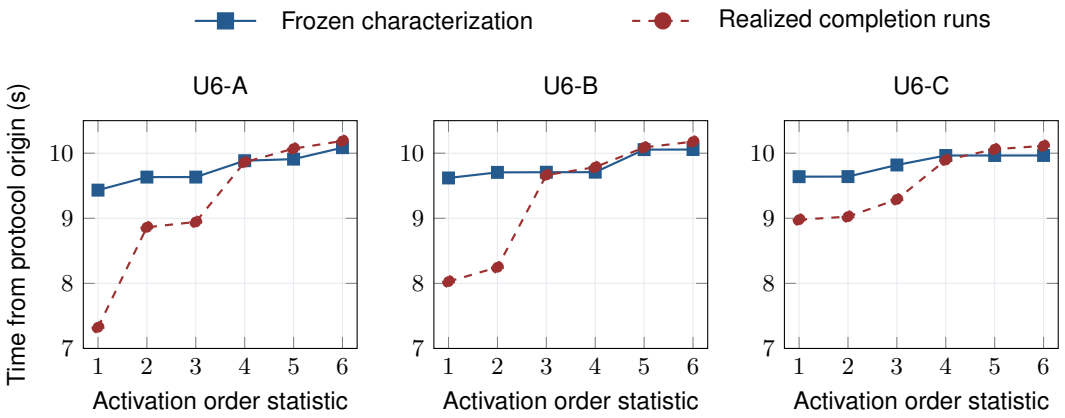


Fig. 3 Frozen versus realized UCV activation-order profiles. Each panel compares the mean within-vector activation order statistics from frozen characterization with those from accepted completion runs for one cyclic $\{3, 2, 1\}$ placement. Times use the existing protocol origin. The largest discrepancy appears in the early order statistics. This is post-hoc state-distribution evidence, not prospective validation, causal attribution, or a deployable refresh rule.

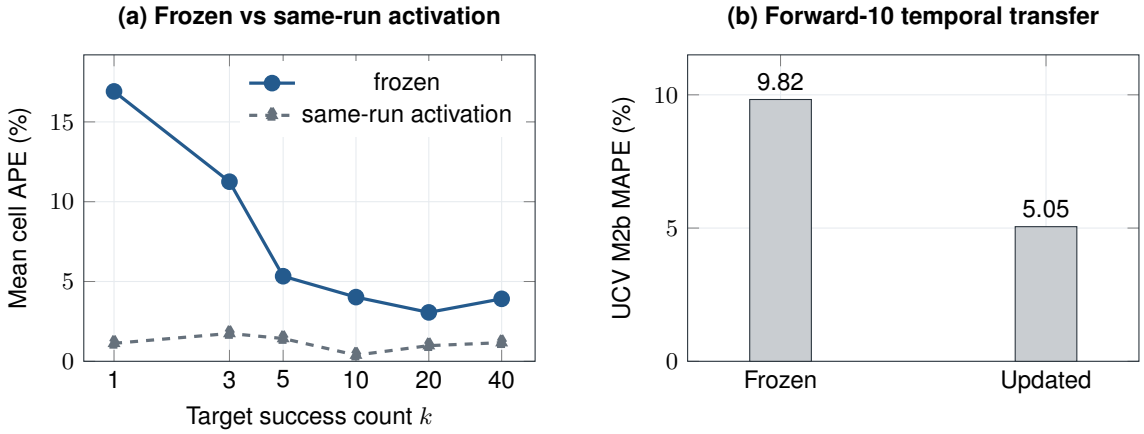


Fig. 4 UCV post-hoc activation and temporal diagnostics. (a) Frozen U-M2b mean cell APE by target success count versus the leakage-assisted, non-deployable same-run realized-activation diagnostic. (b) The separate Forward-10 later-block temporal-transfer diagnostic, shown as frozen and updated UCV M2b MAPE. Both panels are explanatory sensitivity analyses, not corrected prospective performance, a causal decomposition, or a deployable refresh policy.

fixed and alternative predictions with zero failures; the maximum disagreement was 6.03×10^{-11} s. This result narrows the specific measured-capacity substitution but does not establish negligible transient effects in general, exclude unmeasured dynamics, or identify activation as the physical mechanism for the UCV residual.

4.11 Retrospective fitted comparator

The fitted comparison is reported last because it is retrospective and is not prospective evidence. C0 is the k/N sanity model; C1 adds k, cluster-size, and placement terms; C1-I adds the prespecified limited two-way interactions; and H1 is a hybrid calibration that additionally consumes the frozen M2b prediction. At the best evaluated budget for each family, fold-mean MAPE was 11.16% for C0, 7.52% for C1, 7.40% for C1-I, and 2.61% for H1; the frozen M2b analytic reference was 2.67% (Supplementary Table S11). Fold distributions and the complete fixed-budget grid are reported in the Supplement and are not repetition-level confidence statements.

Neither standalone family C1 nor C1-I matched the reference at an evaluated budget, so the recorded status for both is no evaluated crossing on the grid $r \in \{1, 2, 5, 10, 20\}$. That status is bounded to the tested grid and locates no crossing outside it.

H1 reached approximately the same observed error regime as the analytic reference, but the comparison is not like-for-like because H1 consumes the frozen M2b structural prediction in addition to training-cell outcomes. At the standalone scope its information burden is larger than the M2b characterization burden it calibrates. Within the tested families and budget grid, standalone outcome fitting did not reach the analytic reference; this is not a universal result about regression or machine learning.

5 Discussion

5.1 What the evidence supports

The prospective, outcome-free evidence is the strongest layer, and it carries the paper's practical conclusion. Within the measured configurations, independently characterized state supported low-error prospective completion prediction without target-outcome fitting, with no established M2a–M2b accuracy ordering. P4 showed that occupancy-conditioned capacity is non-flat while node-local totals compose closely to placement totals in the studied homogeneous pools. That composition matters even when accuracy differences are small: for the registered P4 acquisition geometry, M2b reduced characterization burden, so a state representation can be useful through lower acquisition cost without claiming superior accuracy. The remaining phases bound rather than extend that conclusion. UCV provides descriptive historical evidence of a larger residual, but U-M2b and U-EQ were nearly indistinguishable before outcomes; their near-identical frozen predictions therefore indicated low pre-outcome design discrimination, while the subsequently similar errors do not establish equivalence. X411 then completed fresh occupancy-four characterization and prediction freeze before target outcomes; its separation passed the characterization-uncertainty heuristic but remained far below the continuation threshold, so the protocol stopped before target collection. Exact model algebra locates representation discrimination in the rate–activation interaction path and gives the stronger condition for complete-path invariance. Together, these results show that finer compositional state is not automatically valuable merely because it is more detailed: its value also depends on acquisition burden, reuse or freshness, pre-outcome discriminability, and structural interaction with activation timing.

5.2 What the evidence does not identify

The evidence does not establish equivalence between capacity representations, universal capacity transfer, arbitrary heterogeneous-cluster transfer, or universal superiority over fitted models. X411 supplies no target-outcome accuracy estimate because its prespecified gate stopped the campaign with zero target runs. The UCV outcome-conditioned diagnostics and the structural identity neither identify activation causality nor exclude capacity-side or model-side residual error. Complete-path invariance is exact under its stated model condition, whereas the post-full centroid displacement is not a universal shift of the actual expected completion time when pre-$a_{\max}$ arrival mass is non-negligible. A 5% separation is neither proved impossible nor proved attainable within homogeneous GKE. No scheduler, autoscaler, service-level, cost, energy, or deployment benefit was tested.

5.3 Which characterization rules are reusable, and where a bounded gap appears

Node-local capacity characterization showed close agreement with placement totals and can reduce acquisition burden in the measured environments. That property is distinct from whether redistributing a fixed aggregate capacity among workers materially changes predicted completion time. UCV offered little leverage on the latter question because its two representations were nearly indistinguishable by design. X411 addressed that question with fresh occupancy-four characterization and a pre-outcome gate: the model difference passed the prespecified characterization-uncertainty heuristic but remained far below the continuation threshold.

The completed X411 stopping decision sharpens the paper's message. Occupancy-aware node totals remain valid measured state, while X-M2b's worker vector is an equal-share construction and X-EQ is not declared equivalent. For the two prespecified representations and X411 activation distributions, the frozen X-M2b and X-EQ predictions differed too little to authorize outcome collection under the prespecified gate. Stopping avoided spending target-outcome budget on a campaign that had already failed its pre-outcome informativeness criterion; it was not an accuracy result or proof that M2b cannot help elsewhere. The exact interaction identity supplies an architectural account of this small displacement while preserving the distinction between a model explanation and a causal physical mechanism.

5.4 Why low k is the most sensitive regime

The concentration of error at low k is consistent with sensitivity to the earliest exposure path, offered as interpretation rather than mechanism. During the ramp, both activation timing and physical worker-rate allocation can change accumulated exposure. The realized-activation substitution shows outcome-sensitive explanatory leverage, while the occupancy-four records show that equal sharing can miss substantial within-node heterogeneity. These are competing explanations; the present data neither isolate their causal contributions nor identify a scheduler, hardware, transport, or simultaneous-multithreading mechanism. At larger k, early-path discrepancies are amortized over a longer productive interval.

The present instrumentation observes operational activation as a single protocol-defined time and cannot decompose it into provisioning, container startup, scheduling, warm-up, steady detection, and transport components. Those remain candidate mechanisms rather than identified causes, and no correlation reported here is treated as establishing one of them. Resolving which component dominates would require prospectively collected component-level timestamps or controlled interventions under diagnostic criteria frozen before outcome collection.

5.5 Characterization reuse and acquisition cost

P4 provides the direct prospective context for reusable compositional state characterization: node-local capacity measurements capture a non-flat occupancy curve, compose closely to placement totals, and lower the registered characterization burden at both reported scopes. Because the shared sweep has a fixed upfront cost, its advantage depends on target-set size; below the measured P4 threshold, it need not be cheaper than a smaller target-specific reference set. UCV has a different phase-specific required-record geometry and is not an independent validation of this ratio. The evidence supports composition-rule reuse within the measured homogeneous pools, not transfer of historical capacity coefficients across environments: each phase characterized its capacity state afresh. Within UCV's evaluated regime, fresh capacity characterization coexisted with the larger residual, while post-hoc activation substitutions had greater explanatory leverage; this is consistent with activation being more freshness-sensitive than capacity representation there, not with a general refresh policy. The fitted comparator reinforces the value of distinguishing structured characterization from outcome fitting: standalone fitted families did not reach the analytic reference over the evaluated grid, while the hybrid reached the same observed regime only by consuming the analytic prediction it calibrates, at a substantially larger standalone information burden.

5.6 Systems implications

Finer compositional state should not be acquired merely because it is measurable. Its value depends on predictive fidelity, prospective availability, characterization burden, reuse or freshness, and whether rate–activation structure can create design-relevant separation. X411 demonstrates why pre-outcome discriminability belongs in that assessment: occupancy four was characterized, yet the two frozen representations remained too close to test the intended practical distinction under the frozen design. The protocol therefore stopped before target-outcome collection, preserving the outcome budget without converting the

gate into an accuracy, scheduling, refresh, or placement-policy result.

5.7 Limitations and threats to validity

The workload is controlled SHA-256 independent-trial search. Its identification advantage depends on the ideal-hash/random-output assumption that distinct counter-derived digests behave as independent uniform outputs; under that model $p = 2^{-24}$ is known independently of cluster measurements. Deterministic SHA-256 execution does not prove that probabilistic assumption. This design permits characterization fidelity and completion-model error to be examined without estimating an application-specific success process from target outcomes, but does not establish transfer to arbitrary workloads. The fixed-exposure Barbour–Hall probability bound does not bound clock-time MAPE. The experiments use homogeneous GKE environments and operational activation is protocol-defined rather than pure physical startup latency.

The primary model assigns each active worker $C(\text{node}, n)/n$ at its final local occupancy; the transient sensitivity uses the same model-defined equal-share representation at current occupancy. Occupancy-four records show material physical within-node worker-rate heterogeneity, so this construction is not a directly observed physical worker-rate vector and neither analysis tests arbitrary within-node allocation. The outcome-free transient sensitivity bounds only its explicit equal-share substitution. The artifacts preserve no CPU affinity, sibling mapping, or pod-to-core topology sufficient to identify the physical mechanism.

No independently prespecified fixed-window throughput measurement spanning productive search was recorded: one accepted worker has no productive trajectory bucket, and 668 of 2,160 worker records have no complete productive bucket. These counts describe post-hoc diagnostic-bucket availability, not missing primary completion records, and prevent construction of an independently prespecified, outcome-independent fixed productive-search throughput window for every required worker and run. A post-hoc stopping-conditioned mean rate can nevertheless be computed for all 2,160 workers from preserved productive-attempt counts and same-worker steady-to-stop timestamps; because each interval ends at the target-dependent stop, it is a sensitivity diagnostic rather than an independent capacity measurement. Low-k intervals are especially short and vulnerable to start–stop effects and stopping-condition selection. The stopping-conditioned rate is not inserted into the two-by-two state diagnostic, which therefore does not exclude within-run throughput changes after steady declaration.

UCV is descriptive historical evidence for one structurally unseen occupancy-composition family after the disclosed post-outcome transport-parser corrective. Its frozen predictions were unchanged, but no clean rerun was performed and the correct-decoder-from-start chronology cannot be reconstructed; the counterfactual accepted-record set and primary metric are therefore unknowable from the bound evidence. N=6 itself was represented in P3, and the P3 N=6 subset performance is post-hoc descriptive. U6 U-M2b–U-EQ prospective discriminability was low, so their observed near-equality cannot establish representation equivalence, and the physical origin of cyclic residual heterogeneity remains unresolved. Broader unseen-placement or workload claims require new prospectively selected campaigns spanning additional occupancy compositions, worker counts, and workload families.

X411 covers one fresh $\{4, 1, 1\}$ family on three n2-custom-6 worker nodes in one GKE 1.35.6 environment and one zone. It characterizes prediction separation, not completion accuracy: the prospective gate stopped the campaign with zero target outcomes. Its 95% interval quantifies characterization uncertainty under the frozen bootstrap, not equivalence, completion variability, or generalization across clouds, versions, topologies, schedulers, or heterogeneous hardware. Direct placement-capacity measurements support close node-total composition agreement but were excluded from primary prediction; no composition-equivalence margin or procedure was prespecified.

The structural result is conditional on the stated independent-trial non-homogeneous Poisson-process model and ideal-hash assumption. Complete-path invariance is exact under the per-activation-group balance condition; zero rate–activation covariance gives only equality after

full activation. The post-full displacement and CV(r)/2 envelope apply to the stated surrogate and approximate actual mean completion only when pre-$a_{\max}$ arrival mass is negligible; they are not tight predictors or clock-time error bounds. The ρ boundary is therefore a conservative screening boundary, not evidence that actual separation reaches 5%. The observed X411 displacement is explained by the model's architecture, not causally attributed to a physical mechanism, and neither universal impossibility nor homogeneous-GKE attainability follows. The successor closure is likewise model- and universe-bounded: zero-occupancy families are unsupported by the active predictor, not impossible under every model.

Completion repetitions are finite: every cell mean summarizes 20 accepted repetitions, and within-cell dispersion was largest in UCV, where relative standard errors of the cell mean reached 5.69%. The finite-repetition intervals are descriptive overlays, not tests, and all three UCV endpoint slope-mismatch intervals include zero.

The offset-calibration, leave-one-k-out, realized-activation, and Forward-10 diagnostics are post-hoc and do not provide corrected prospective or deployable performance. The placement-pooled leave-one-run-out diagnostic excludes the evaluated run's activation vector but uses other repetitions from the same placement–k cell, so it is not fully out-of-sample on the activation side; the same-run diagnostic additionally uses evaluated-run information. Freshness support is moderate and no optimal refresh interval was established. The later capacity checkpoint is non-target and limited to its observed interval, so it cannot establish absence of drift. The fitted comparator is retrospective, and the UCV M1–M2b paired comparison is post-hoc descriptive. No scheduler, autoscaler, service-level, downstream-policy, heterogeneous-cluster, or arbitrary-workload benefit was tested.

An earlier P4 campaign is excluded because it violated the predefined receive-order protocol. The locked post-corrective analytical corpus has no further identified numerical correction; implementation details are provided in the supplement and archived materials.

6 Conclusion

For independent-trial SHA-256 search under an ideal-hash model in homogeneous GKE, independently characterized state supported low-error prospective prediction in the prospectively evaluated P3 and corrected P4-A configurations. Node-local totals composed closely to placement totals and reduced total characterization burden by 35.52% for the nine-target P4 geometry. What did not follow is equally definite. UCV remains descriptive historical evidence after its corrective; X411 collected zero target outcomes after its clean pre-outcome gate found separation far below the 5% threshold despite passing the relative-precision heuristic. Occupancy-four heterogeneity limits the physical interpretation of equal-share worker rates, while the exact model analysis locates discriminative value in rate–activation interaction and its stronger complete-path invariance condition. Richer state is therefore useful not merely because it is measurable, but when it is prospectively available, characterization-efficient, reusable or sufficiently fresh, and structurally capable of changing prediction by a design-relevant amount.

Acknowledgements

Generative AI tools assisted with software implementation, workflow development, and manuscript drafting and editing. The authors independently designed the study, made all scientific decisions, verified the results, and take full responsibility for the work.

References

[1] Luby, M., Sinclair, A., Zuckerman, D.: Optimal speedup of Las Vegas algorithms. Information Processing Letters **47**(4), 173–180 (1993) https://doi.org/10.1016/0020-0190(93)90029-9

[2] Truchet, C., Richoux, F., Codognet, P.: Prediction of parallel speed-ups for Las Vegas algorithms. In: 42nd International Conference on Parallel Processing, pp. 160–169 (2013). https://doi.org/10.1109/ICPP.2013.25

[3] Truchet, C., Arbelaez, A., Richoux, F., Codognet, P.: Estimating parallel runtimes for randomized algorithms in constraint solving. Journal of Heuristics **22**(4), 613–648 (2016) https://doi.org/10.1007/s10732-015-9292-3

[4] Yoo, M., Kim, G., Kim, K.: Empirical analysis of waiting time and overhead in container-based hash search using Poisson and Gamma distributions. In: KIPS ASK 2026 (2026). Paper ID KIPS_C2026A0088; in Korean. https://ask.kips.or.kr/programBook

[5] Sodhi, S., Subhlok, J., Xu, Q.: Performance prediction with skeletons. Cluster Computing **11**(2), 151–165 (2008) https://doi.org/10.1007/s10586-007-0039-2

[6] Venkataraman, S., Yang, Z., Franklin, M., Recht, B., Stoica, I.: Ernest: Efficient performance prediction for large-scale advanced analytics. In: 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16), pp. 363–378 (2016)

[7] Alipourfard, O., Liu, H.H., Chen, J., Venkataraman, S., Yu, M., Zhang, M.: CherryPick: Adaptively unearthing the best cloud configurations for big data analytics. In: 14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17), pp. 469–482 (2017)

[8] Yadwadkar, N.J., Hariharan, B., Gonzalez, J.E., Smith, B., Katz, R.H.: Selecting the best VM across multiple public clouds: A data-driven performance modeling approach. In: Proceedings of the 2017 Symposium on Cloud Computing, pp. 452–465 (2017). https://doi.org/10.1145/3127479.3131614

[9] Klimovic, A., Litz, H., Kozyrakis, C.: Selecta: Heterogeneous cloud storage configuration for data analytics. In: 2018 USENIX Annual Technical Conference (USENIX ATC 18), pp. 759–773 (2018)

[10] Bader, J., Lehmann, F., Thamsen, L., Leser, U., Kao, O.: Lotaru: Locally predicting workflow task runtimes for resource management on heterogeneous infrastructures. Future Generation Computer Systems **150**, 171–185 (2024) https://doi.org/10.1016/j.future.2023.08.022

[11] Yildirim, E., Hussein, M., Titov, M., Kilic, O.O.: Predicting runtime and resource utilization of jobs on integrated cloud and HPC systems. Future Generation Computer Systems **176**, 108230 (2026) https://doi.org/10.1016/j.future.2025.108230

[12] Tian, H., Yu, M., Wang, W.: CrystalPerf: Learning to characterize the performance of dataflow computation through code analysis. In: 2021 USENIX Annual Technical Conference (USENIX ATC 21), pp. 253–267 (2021)

[13] Hsu, C.-J., Nair, V., Menzies, T., Freeh, V.W.: Micky: A cheaper alternative for selecting cloud instances. In: 2018 IEEE 11th International Conference on Cloud Computing (CLOUD), pp. 409–416. IEEE, Piscataway, NJ, USA (2018). https://doi.org/10.1109/CLOUD.2018.00058

[14] Fekry, A., Carata, L., Pasquier, T., Rice, A., Hopper, A.: To tune or not to tune? in search of optimal configurations for data analytics. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2494–2504. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3394486.3403299

[15] Will, J., Thamsen, L., Scheinert, D., Bader, J., Kao, O.: C3O: Collaborative cluster configuration optimization for distributed data processing in public clouds. In: 2021 IEEE International Conference on Cloud Engineering (IC2E), pp. 43–52. IEEE, Piscataway, NJ, USA (2021). https://doi.org/10.1109/IC2E52221.2021.00018

[16] Scheinert, D., Thamsen, L., Zhu, H., Will, J., Acker, A., Wittkopp, T., Kao, O.: Bellamy: Reusing performance models for distributed dataflow jobs across contexts. In: 2021 IEEE International Conference on Cluster Computing (CLUSTER), pp. 261–270. IEEE, Piscataway, NJ, USA (2021). https://doi.org/10.1109/Cluster48925.2021.00052

[17] Garralda-Barrio, M., Eiras-Franco, C., Bolón-Canedo, V.: Adaptive incremental transfer learning for efficient performance modeling of big data workloads. Future Generation Computer Systems **166**, 107730 (2025) https://doi.org/10.1016/j.future.2025.107730

[18] Mars, J., Tang, L., Hundt, R., Skadron, K., Soffa, M.L.: Bubble-up: Increasing utilization in modern warehouse scale computers via sensible co-locations. In: Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture, pp. 248–259 (2011). https://doi.org/10.1145/2155620.2155650

[19] Delimitrou, C., Kozyrakis, C.: Quasar: Resource-efficient and QoS-aware cluster management. In: Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems, pp. 127–144 (2014). https://doi.org/10.1145/2541940.2541941

[20] Zhang, Y., Hua, W., Zhou, Z., Suh, G.E., Delimitrou, C.: Sinan: ML-based and QoS-aware resource management for cloud microservices. In: Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, pp. 167–181 (2021). https://doi.org/10.1145/3445814.3446693

[21] Faisal, A.B., Martin, N., Bashir, H.M., Lamelas, S., Dogar, F.R.: When will my ML job finish? toward providing completion time estimates through Predictability-Centric scheduling. In: 18th USENIX Symposium on Operating Systems Design and Implementation (OSDI 24), pp. 487–505. USENIX Association, Santa Clara, CA (2024)

[22] Li, Z., Wang, Y.: An adaptive read/write optimized algorithm for Ceph heterogeneous systems via performance prediction and multi-attribute decision making. Cluster Computing **26**, 1125–1146 (2023) https://doi.org/10.1007/s10586-022-03764-3

[23] Zanussi, L., Tessera, D., Massari, L., Bermejo, B., Juiz, C., Calzarossa, M.: Load-aware predictive auto-scaling framework for cloud environments. Cluster Computing **29**, 197 (2026) https://doi.org/10.1007/s10586-026-05944-x

[24] Rossi, A., Visentin, A., Carraro, D., Prestwich, S.D., Brown, K.N.: Forecasting workload in cloud computing: towards uncertainty-aware predictions and transfer learning. Cluster Computing **28**, 258 (2025) https://doi.org/10.1007/s10586-024-04933-2

[25] Casalicchio, E.: A study on performance measures for auto-scaling CPU-intensive containerized applications. Cluster Computing **22**, 995–1006 (2019) https://doi.org/10.1007/s10586-018-02890-1

[26] Liu, P., Guitart, J.: Performance characterization of containerization for HPC workloads on InfiniBand clusters: An empirical study. Cluster Computing **25**, 847–868 (2022) https://doi.org/10.1007/s10586-021-03460-8

[27] Cardas, C., Aldana-Martín, J.F., Burgueño-Romero, A.M., Nebro, A.J., Mateos, J.M., Sánchez Sánchez, J.J.: On the performance of SQL scalable systems on Kubernetes: A comparative study. Cluster Computing **26**, 1935–1947 (2023) https://doi.org/10.1007/s10586-022-03718-9

[28] Heydari Beni, E., Truyen, E., Lagaisse, B., Joosen, W., Dieltjens, J.: Reducing cold starts during elastic scaling of containers in Kubernetes. In: Proceedings of the 36th Annual ACM Symposium on Applied Computing, pp. 60–68 (2021). https://doi.org/10.1145/3412841.3441887

[29] Dean, J., Barroso, L.A.: The tail at scale. Communications of the ACM **56**(2), 74–80 (2013) https://doi.org/10.1145/2408776.2408794

[30] Barbour, A.D., Hall, P.: On the rate of Poisson convergence. Mathematical Proceedings of the Cambridge Philosophical Society **95**(3), 473–480 (1984) https://doi.org/10.1017/S0305004100061806

Statements and Declarations

Funding

Min Su Yoo received scholarship support through the BK21 FOUR program at Chonnam National University while conducting this work. This study was financially supported by Chonnam National University (Grant number: 2026-0428-01).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Min Su Yoo: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, and Writing – original draft. Hyunglip Bae: Supervision and Writing – review & editing. Both authors reviewed and approved the final manuscript.

Data Availability Statement

The complete data, code, frozen authorities, and offline reproduction checkers supporting this article are publicly available in Zenodo at https://doi.org/10.5281/zenodo.22643283 (concept DOI: https://doi.org/10.5281/zenodo.22053883). X411 target-completion data do not exist because the prespecified pre-outcome continuation criterion was not met; the target-completion run count was zero.

Code Availability

The analysis code and offline reproduction checkers are included in the verified Zenodo archive identified in the Data Availability Statement.

Ethics Approval

Not applicable; this computational systems study involved no human participants or animals.

Consent to Participate and Publish

Not applicable.

Supplementary Information

Detailed methods, diagnostics, tables, and provenance are provided in the Supplementary Information (Online Resource 1; `ESM_1.pdf`).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf