REVISABILITY BUYS GEOMETRY-INDEPENDENCE: COMPETITIVE ANALYSIS OF BATCHED MEMORY CONSOLIDATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Moving memory items into network weights, known as consolidation, incurs costs that depend heavily on batching: items consolidated together share a fixed setup price and influence each other's utility based on batch composition. This paper establishes a pricing framework for this consolidation process, yielding a central negative result. The most straightforward consolidation strategy, where each batch fixes its internal alignment, lacks a bounded competitive ratio because an adversary can strategically flood any fixed configuration. Escaping this impossibility requires revisability: the ability to adjust a prior batch's alignment. We analyze two mechanisms for revisability (a lightweight approach of replaying past data during new consolidation steps and a costly joint refitting procedure) and derive their respective prices. Pre-registered experiments on real feature geometry show that while our core theoretical guarantees (the ledger, dominance arithmetic, scheduling reductions, and worst-case bounds) and the separation construction hold on natural data, the exact closed-form quantitative levels do not transfer directly. All pre-registered evaluation criteria are fully detailed in the appendix.

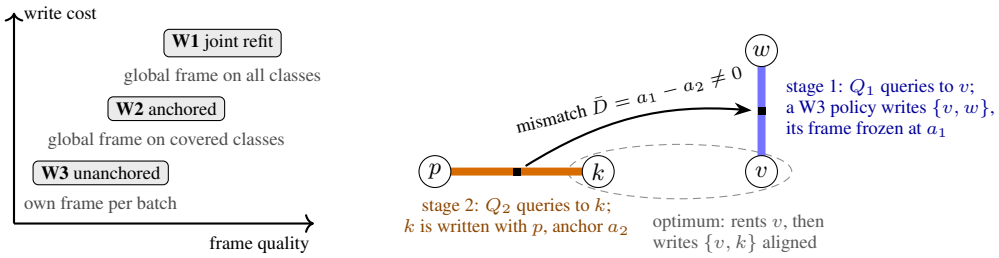
1 INTRODUCTION

Continual learning agents with a hybrid memory keep fast knowledge in an in-context store and slow knowledge in the weights, and move items from the first into the second by *consolidation*: each such move is one *write* of the item into the weights. A recent multi-author position paper names this store-to-weights move and leaves it unpriced (Dorovatas et al., 2026): no decision rule says when a write should occur, what it should contain, or what it costs in quality. Practitioner pipelines trigger consolidation by capacity thresholds, periodic flushes, or event triggers. None of these three rules keeps its cost within any fixed factor of what a rule that knew the future would pay, not even the capacity threshold when costs are known; and the one rule the economics dictates, writing whenever the accumulated cost of waiting reaches the setup price, stays within twice what that rule would pay (Appendix F.6).

This paper prices consolidation along the three axes it exposes. *When* to consolidate is a scheduling problem: an item not yet written accrues *rent*, a cost paid every time a query touches it, while each write event pays its setup price once; choosing the moment is a classical online problem, one in which each decision must be made before the future queries are known. *What to write together* is a geometry problem: items written in one batch share an internal reference frame, items in different batches generally do not, and the mismatch costs accuracy. *How* to write, through which fitting procedure (which stored data the write is allowed to see), is the question the field does not ask, and it decides whether worst-case guarantees exist at all.

Our answer is that *revisability buys geometry-independence*: guarantees that hold no matter what the classes' geometry is. The most straightforward consolidation strategy, the unanchored write, fits each batch in its own frame, reusing no stored examples from earlier batches; an exact per-pair cost accounting, the *ledger*, prices what that costs on every pair of classes written in different batches. Under the cost convention we adopt, charges count only error above the baseline every rule pays anyway, averaged over each pair's two classes; no policy built from unanchored writes alone keeps its cost within any fixed multiple of what a rule that knew the future would pay. The adversary that

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) The three write operators.

(b) Why pure unanchored writing has no finite ratio.

Figure 1: (a) The operators of Definition 1 on cost and frame quality (positions, not scale). (b) The two-phase adversary of Theorem 4 (Appendix G.4): a W3 batch needs two classes and only w is pending with v in stage 1, so a pure-W3 policy that writes v then must write $\{v, w\}$, freezing its frame at an anchor a_1 between v and w; the flooded class k can then only be written with p, at an anchor a_2 between k and p. The frozen rule's direction for (v, k) has no positive component along their displacement, so every flood query costs at least $\min(r, \ell\,\Xi)$, the rent or the priced excess floor of (5) (Lemma 25), while the optimum serves the flood without excess.

proves this works in two stages: it first queries one class for a fixed stretch, long enough that any policy must either lock in that class's frame or keep paying rent; it then floods a new class so heavily that no later unanchored write can align with the locked-in frame, while a policy that had kept paying rent writes the queried and the new class together and answers the flood at no error above the baseline (Figure 1b). A consolidation strategy that can rewrite a written frame escapes. When replay includes an example of every class written so far, at a setup price that does not grow with the store, the combined problem reduces to a classical networking rule, dynamic TCP acknowledgment, and its guarantees hold no matter what the geometry is; a full joint refit also escapes, but its cost grows with the whole store. We then tested, in a program registered before any run, which parts of this theory carry over to the geometry of features from real data; every registered criterion's outcome is reported in the appendix, negative results included.

The model. A frozen encoder maps inputs into $\mathbb{R}^d$; classes have Gaussian conditionals with shared covariance, whitened to the identity, so class geometry reduces to pairwise margins $\gamma_{rs} = \|\mu_r - \mu_s\|^2$ (squared distances). A consolidation write fits the cross-entropy head on a batch of classes through one of three operators (Definition 1, Figure 1a): W1, the full joint refit; W2, an anchored write that fits the incoming batch with a replay anchor drawn from the store; W3, an unanchored write that fits the batch alone while existing rows are held. W1 keeps the head's global frame, W2 keeps it on the classes its anchor covers, and each W3 batch is fit in its own. All error quantities are closed form in the margins and batch compositions. Costs follow Assumption A2: a write event pays a setup K (K_2 for an anchored, K_3 for an unanchored write) plus c per item, a query to an unwritten item pays rent r, and a query to a written class pays ℓ times the misalignment excess of its pairs with the other written classes. We study the population ledger, in which every effect that survives is caused by the write protocol rather than by estimation noise, and scope finite-sample effects in Appendix J; the frozen-feature, shared-covariance regime is the standard approximation for a fixed backbone with a trained head, and our program tests it at three testbeds, where the ledger's level does not transport and its direction test proves weak. The cross-entropy head and the prototype (nearest-class-mean) head coincide in the population limit on every pair the head carries in one frame (Proposition 10, App. B) and differ across frames by exactly the mismatch the ledger prices.

Contributions. (1) *An exact price for frame mismatch* (Sections 4 and 5): a closed-form misalignment ledger, an exact per-pair accounting of the error consolidation adds, for every cross-batch pair under W3, and a split of that error into a scalar part that per-class constant corrections remove and a vector part that no per-class constant touches: the gauge split. (2) *Batching dominance, in guarantee form* (Section 6): merging batches never worsens the guaranteed worst case, while the per-instance version is false: a single merge can damage aligned bystanders, the classes outside the merged batches. (3) *Consolidation timing* (Section 7): under covering anchors, scheduling is dynamic TCP acknowledgment at a query-driven rate; the rule that writes whenever accumulated rent reaches the setup price, break-even, pays at most $(2 + w^{\max}/K)$ times the optimum, where $w^{\max}$

is the largest rent one round can accrue; no rule that writes once a threshold is crossed matches a factor of 2 at fixed granularity; and when items arrive at a steady rate, the optimal period between writes obeys a square-root law. (4) *No finite ratio without revisability* (Section 8): under the differential cost convention and balanced-prior charge we adopt, no deterministic or randomized policy built from unanchored writes alone has a bounded cost ratio. (5) *Revisability restores the guarantees* (Section 8): anchored writes whose anchors cover every previously written class restore the classical guarantees independently of the geometry, at a setup price fixed independently of the store; priced against each other, a policy allowed to mix the two operators has optimal deterministic ratio $\Theta(P)$ in the premium $P = K_2/K_3$ of the anchored over the unanchored setup, exactly 2 at $P = 1$ in the fine-granularity limit (Appendix H.2). (6) *A pre-registered test* (Section 9): the ledger's exactness and the dominance arithmetic hold; the corrected separation theorem's bounds hold in every tested condition, after the program corrected an error in the theorem's first form; the separation construction is realizable on natural geometry; on real features the ledger's level and floor do not transport, and a trained model, given the same total training budget, orders the three operators opposite to the pre-run expectation.

2 Related work

Online algorithms. Section 7's scheduling layer reduces to classical dynamic TCP acknowledgment (Dooly et al., 2001; Seiden, 2000; Karlin et al., 2001) under its standard additive-delay objective (a sender decides when to acknowledge packets, each unanswered packet paying delay per round; other delay conventions, e.g. sum-monotone Bhore et al. 2026, are outside our claims), and it is the simplest member of the aggregation-with-delays family (Bienkowski et al., 2014; Buchbinder et al., 2008; Bienkowski et al., 2016; 2020; Azar & Touitou, 2019): the reductions are ours, the guarantees are not, and none of that line ties rent and error constants to statistical geometry. The rent-or-buy lineage (Karlin et al., 1988; El-Yaniv et al., 1999; Meyerson, 2005; Lotker et al., 2012) prices the item; we price the event, its operator and the premium. The two papers nearest to ours decide when to replay, by schedule search (Klasson et al., 2023) or by concurrent versus sequential rehearsal (Deng et al., 2025), and neither prices the write; Appendix K places both.

Memory systems for LLM agents. A newer wave builds memory for LLM agents directly: memory streams with recency and importance scoring (Park et al., 2023), operating-system-style paging between context and store (Packer et al., 2023), offline consolidation before queries (Lin et al., 2025), self-reorganizing memories (Xu et al., 2025), production memory layers (Chhikara et al., 2025), and a field survey (Hu et al., 2025). Each decides what to store, move, or overwrite by heuristic or by prompting the model itself. None of these systems writes into the weights, so none prices that write or chooses its operator.

3 The write operators

A consolidation event writes a batch B_t of classes into the head: a population cross-entropy fit on the batch's data under a count-weighted ℓ_2 ridge tending to zero. The ridge selects the batch-anchored member (a gauge choice, distinct from the replay anchor of Definition 1 below) of the fit's gauge family (the family of heads realizing the same population posterior), and every result below is stated under that selection (Appendix B). The operator is determined by what else the fit sees.

Definition 1 (Write operators)**. W1, full joint refit:** refit the head on the entire consolidated store, the batch included. The fit maintains the global frame; its cost scales with the whole store. **W2, anchored incremental write:** fit the incoming batch together with a replay anchor, a small set of stored exemplars of previously written classes, and refit the anchored classes' rows along with the batch's. The anchor's *coverage* is the set of previously written classes it contains. The write restores one frame across the batch and its covered classes, at the cost of the batch plus a few exemplars per covered class. **W3, unanchored incremental write:** fit the incoming batch alone, over the batch's own labels, with existing rows held fixed. The cheapest operator; the batch's rows are fit in the batch's own frame. A one-class batch is degenerate (no contrast, Remark 3, App. B), so W3 requires batch size at least two, and the objective normalizes within the batch alone, which is what makes the frame a property of the batch (Remark 4, App. D).

Remark 1 (Revisability separates the operators). The axis that matters below is whether an operator can *revisit* the frame of an already-written class. W1 revisits every written class and pays the whole store; W2 revisits exactly the classes its anchor covers and pays a few exemplars for each; W3 revisits nothing. The frontier of Section 8 studies policies built from W3 alone; the restriction is an assumption on the policy class (Assumption A1), and a policy that refits an already-written class does what W3 cannot and escapes the negative result by paying.

4 THE MISALIGNMENT LEDGER

Under W3 each batch B_t is fit in its own frame, and that frame is a population object we can write down. Let $\bar{\mu}_t = \sum_{r \in B_t} \pi_r \mu_r$ be the *batch anchor* and $\bar{S}_t = \sum_{r \in B_t} \pi_r \|\mu_r\|^2$ the *batch scalar*, the count-weighted means of the batch's class means and of their squared norms, with π_r class r's count share within B_t. The population optimum of the W3 fit is realized on one explicit family, directions $\mu_r - \bar{\mu}_t$ centred at the batch anchor with a bias convention fixed within the batch (Lemma 7, Appendix B), and the ridge's zero limit selects its minimum-norm member: the frame a W3 write leaves behind is its batch anchor.

Cross-batch pairs therefore pay a misalignment tax. For classes $r \in B_t$ and $s \in B_u$, write the displacement $\delta = \mu_r - \mu_s$, the margin $\gamma = \|\delta\|^2$, the anchor mismatch $\bar{D} = \bar{\mu}_t - \bar{\mu}_u$ and the scalar mismatch $\bar{\Delta} = \bar{S}_t - \bar{S}_u$. The two batches' heads decide

$$r \text{ over } s \quad \Longleftrightarrow \quad (\delta - \bar{D})^\top x \;\geq\; \tfrac{1}{2}\big(\|\mu_r\|^2 - \|\mu_s\|^2\big) - \tfrac{1}{2}\bar{\Delta}, \tag{1}$$

and both one-sided errors are exact (Lemma 11, Appendix C):

$$e_{s \to r} = \Phi\left(-\frac{\gamma/2 - \bar{\Delta}/2 + \bar{D}^\top \mu_s}{\|\delta - \bar{D}\|}\right), \qquad e_{r \to s} = \Phi\left(-\frac{\gamma/2 + \bar{\Delta}/2 - \bar{D}^\top \mu_r}{\|\delta - \bar{D}\|}\right). \tag{2}$$

The mismatch rotates the rule's direction to $\delta - \bar{D}$ and shifts its threshold by $\bar{\Delta}$; the formulas are exactly translation-invariant, the degenerate direction $\bar{D} = \delta$ gives balanced error exactly $\frac{1}{2}$, and at zero mismatch both errors equal the aligned Bayes value $e_{\mathrm{al}} = \Phi(-\sqrt{\gamma}/2)$, the paper's unit of account. We write e_{bal} for the realized balanced error, the mean of the two. Every quantity in Sections 5, 6 and 8 is assembled from (2). The aligned baseline falls fast in the margin, $e_{\mathrm{al}} = 0.309, 0.159, 0.0668, 0.0228, 0.0062$ at $\gamma = 1, 4, 9, 16, 25$, which is why the excess a flooded pair pays in Section 8, at least $\frac{1}{2} - e_{\mathrm{al}}$, is close to $\frac{1}{2}$ once classes are a few units apart.

5 THE SCALAR AND VECTOR GAUGE SPLIT

The two mismatch components have categorically different statuses (Proposition 13, proved in Appendix D). The scalar mismatch is removable for free: one additive constant per class, computed from the batch's own composition, cancels every $\bar{\Delta}$ term in every cross-batch pair rule at once. The vector mismatch is not: class-wise constants move a pair rule's threshold only, while $\bar{D}$ rotates its direction. The part of unanchored writing's cost that no choice of batch composition removes is therefore exactly the vector field $\{\bar{D}_{tu}\}$; class-wise *linear* corrections would move directions, but that is a refit, hence another operator. In short, *free class-wise corrections, including corrections built from the logarithm of the class counts, buy only the scalar gauge, and short of paying for anchors or a refit, batching is the only lever on the vector gauge.*

6 BATCHING DOMINANCE

Partition the incoming classes into W3 batches. Merging two batches replaces their anchors and scalars by count-weighted means. Write $\rho(P) = \max_{t \neq u} \|\bar{D}_{tu}\|$ for a partition's mismatch budget and $\Psi(\rho)$ for the worst balanced cross-batch error certified over all admissible geometries and partitions with budget at most ρ, over a fixed margin envelope (Definition 2, Appendix E).

Theorem 1 (Batching dominance, in guarantee form)**.** ***(a)*** *Merging t and u classifies every pair internal to the merged batch by the aligned Bayes pair rule, the unique minimizer of the balanced pairwise error; every such pair improves weakly, and strictly unless the pre-merge pair rule already coincided with the Bayes rule, which happens exactly on the exceptional set characterized in Appendix E.3 (the point $\bar{D} = 0$ for pairs with unequal norms, a ray for norm-equal pairs).* ***(b)*** *Every*

third-party mismatch vector becomes a convex combination, $\bar{D}_{(tu)v} = \lambda\,\bar{D}_{tv} + (1-\lambda)\,\bar{D}_{uv}$ with λ the count weight of t, so the mismatch budget never increases under a merge. ***(c)*** *Since Ψ is non-decreasing in ρ by definition, (b) implies that no merge ever worsens the certified bound $\Psi(\rho(P))$, and the one-batch write carries the best W3 guarantee.*

In words: merging batches into one bigger write never makes the guaranteed worst case worse, and strictly improves the merged pairs that were suffering from frame mismatch. Dominance holds in guarantee form only; per-instance dominance is false:

Example 2 (Bystanders)**.** *Take three batches T, U, V with $\bar{\mu}_T = \bar{\mu}_V$, so every T-versus-V pair is perfectly aligned once the offsets of Proposition 13 remove the scalar mismatch. Merging T with U moves the merged anchor toward $\bar{\mu}_U$, and $\bar{D}_{(TU)V} = (1-\lambda)(\bar{\mu}_U - \bar{\mu}_V)$ is nonzero as soon as $\bar{\mu}_U \neq \bar{\mu}_V$: every T-versus-V pair strictly worsens for generic placements (Appendix E.4), while the merged pair improves and the certified budget never rises; pairs not aligned before the merge may move either way, since the sign is decided by the mismatch's bearing and not by its length alone.*

The example motivates the *batch-partition problem*: given incoming classes' means and a batch-size cap $\bar{m} \geq 2$, partition them into batches of sizes in $[2, \bar{m}]$ minimizing $\rho(P)$ or $\Psi(\rho(P))$; without the cap, merging everything is optimal (Theorem 1(c)). It is a constrained geometric clustering problem, and it is open (Appendix E.5).

7 Scheduling: acknowledgment and the sleep law

Under W2 with covering anchors (or W1) the gauge cost is zero and what remains is pure scheduling, stated for $c = 0$ (Appendix F.6 carries a positive per-item price). It is the dynamic TCP acknowledgment problem of Dooly et al. (2001) at a query-driven accrual rate (Proposition 18, Appendix F.1): a query to an unwritten item accrues rent r, and a write event is the acknowledgment that clears all pending accruals at the shared setup K. Break-even (write when accrued rent reaches K) is $(2 + w^{\max}/K)$-competitive, with $w^{\max} = r \max_t |Q_t \cap \text{unwritten}|$ the largest single-round accrual: the constant is 2 in the fine-granularity limit $w^{\max}/K \to 0$, at most 3 once $w^{\max} \leq K$, and unbounded above that. No threshold rule is 2-competitive at fixed granularity, and Dooly et al. (2001)'s matching lower bound is asymptotic in the same sense, so there break-even achieves the best ratio any policy can achieve: its constant 2 matches the lower bound. The lower-bound instances of Seiden (2000), in which each pending item has unit weight, carry over unchanged, so $e/(e-1)$ is a randomized lower bound here too; the scope throughout is the additive-delay objective, and for sum-monotone, a batch-aware delay variant, break-even is $\Omega(n)$-competitive (Bhore et al., 2026), where we claim nothing. Appendix F.1 embeds the classical instances, which carries the lower bounds, and Appendix F.4 re-proves the upper bound at a time-varying rate. *In words:* once the operator restores the frame, the only question left is when to write, and it is classical: waiting costs rent, writing costs setup, and break-even pays at most twice the optimum plus a granularity term that vanishes at fine granularity. The stationary flow regime yields the sleep law:

Corollary 3 (Square-root sleep law)**.** *Items not yet written (entry-eligible) arrive at rate α with mean rent $\bar{\rho}$ per round. A periodic batch schedule with period T pays per-round cost $K/T + \alpha\bar{\rho}\,T/2$, minimized at*

$$T^* \;=\; \sqrt{\frac{2K}{\alpha\bar{\rho}}}, \tag{3}$$

the economic-order-quantity law in time, stated for $2K \geq \alpha\bar{\rho}$ so that $T^* \geq 1$. The period shortens as rents rise and lengthens as the setup grows, sign predictions the registered program confirms.

The law is a guarantee and not only an optimum: on discrete rounds, where the mean wait is $(T-1)/2$ (Appendix F.5), a period set to x times T^* pays exactly $(T^*(x + x^{-1}) - 1)/(2T^* - 1)$ times the rate at T^*, so a decade of mis-setting costs a factor above five; stochastic arrivals are classical through Proposition 18 (Appendix F.7). The capacity threshold and the write-on-first-query trigger have no competitive constant at all (Appendix F.6).

8 THE COMPETITIVE FRONTIER

Under W3 the quality axis reopens, and when-to-write couples with what-to-co-batch through the geometry of Section 6. Two hypotheses fix the regime.

Assumption A1 (Pure unanchored writes)**.** Each class is written at most once, in an unanchored batch of size at least two (Remark 3, App. B), and existing rows are held thereafter. Refitting a written class jointly with new arrivals is W2 or W1 by Definition 1 and belongs to the positive side below.

Assumption A2 (Costs)**.** A write event costs K plus c per item written. A query to an unwritten item pays rent r, the differential price of Appendix A.1, which nets out the aligned Bayes error that both stores pay on covered classes (Proposition 10, App. B). A query to a written class pays, for each other written class it is discriminated against, ℓ times the *excess* of the pair's balanced error over its aligned value (the *balanced-prior charge*; the one-sided charge of Remark 9 bills only the queried class's own error), with the scalar mismatch removed by the Proposition 13 offsets: internal pairs and anchored pairs pay nothing beyond the common baseline, and a cross-batch unanchored pair pays $\ell(e_{\mathrm{bal}} - e_{\mathrm{al}}) \geq 0$, the misalignment excess of Lemma 11 (nonnegative by Lemma 14, App. E). Appendix G.1 fixes this convention and explains why the gross convention (charging the baseline to written pairs only), under which the registered E-L2-4 ran, is not a competitive-analysis normalization.

The adversary controls the geometry of achievable frames, but only through convex hulls: a W3 batch's anchor is a convex combination of its members' means, so the anchors achievable from a pending pool are dense in the convex hull of the pool's means, a compact set fixed by the placements and independent of the policy (Lemma 22, Appendix G.2). On that domain the mismatch has a floor. Whenever the two classes' achievable-anchor sets are disjoint (for the norm-unequal pairs the construction uses, this is whenever their set difference misses the closed exceptional set $E^{\sharp}$ of Lemma 24), the flooded pair's balanced error stays strictly above e_{al}, and the worst case over the two anchor sets is attained (Lemma 24, Appendix G.3). In the norm-equal case, no condition phrased purely in distances between means, however large, implies this disjointness (Remark 8).

Theorem 4 (Absolute separation for pure W3)**.** *Let $K' = K + 2c$ be the cost of writing a pair, and consider the two-phase family of Appendix G.4 at margin γ and separation $s > 0$, whose stage-1 class is v and whose flooded class is k. The family is instantiated by two query streams: the* pure stream *is stage 1 only, Q_1 queries to v; the* composed stream *adds stage 2, Q_2 queries to the new class k. Every deterministic policy satisfying Assumptions A1 and A2 pays, against the offline optimum, at least*

$$\begin{cases} \dfrac{K' + \min\big(r,\ \ell\,\Xi\big)\,Q_2}{r\,Q_1 + K'} & \text{on the composed stream, if it writes } v \text{ in stage 1,} \\[2ex] \dfrac{r\,Q_1}{K'} & \text{on the pure stream, if it does not,} \end{cases} \tag{4}$$

where

$$\Xi \;=\; \tfrac{1}{2} - e_{\mathrm{al}} \;=\; \tfrac{1}{2} - \Phi\big(-\sqrt{\gamma}/2\big) \;>\; 0 \tag{5}$$

is the excess floor*: a lower bound on the misalignment excess the flooded pair carries under every admissible batch composition and every class-wise constant correction (Lemma 25, App. G), and the infimum of that excess as the composition is steered by multiplicity, an infimum over compositions that, unlike the anchor-set minimum of Lemma 24, is approached but never attained. Every randomized policy pays at least half of the corresponding bound in expectation against an oblivious stream. Consequently, for every C there are stage lengths, fixed in advance, on which every pure-W3 policy pays more than C times the optimum on one of the two streams: pure unanchored writing has no finite competitive ratio on any instance class containing one separated two-phase geometry, and no additive-constant relaxation rescues it.*

Proof idea (Appendix G.4, Figure 1b). Stage 1 queries v Q_1 times. A policy that does not write v in stage 1 pays rQ_1 on the pure stream, where the optimum writes $\{v, w\}$ once for K'. A policy that writes v in stage 1 must write $\{v, w\}$, the only admissible batch (w is the only class pending with v), which freezes v's frame at an anchor a_1 between v and w. In stage 2 the flooded class k can only be written with p, at an anchor a_2 between k and p, and at this placement the frozen head's

discriminant for (v, k) has no positive component along the pair's own displacement, so its balanced error is at least $\frac{1}{2}$ under every composition and every class-wise constant correction (Lemma 25). Each flood query then costs at least $\min(r, \ell\,\Xi)$, against an optimum that rents v and writes $\{v, k\}$ at the start of stage 2, at a cost of at most $rQ_1 + K'$. Taking $Q_1 > CK'/r$ and then Q_2 large makes one of the two streams cost more than C times its optimum. The bound is stated and priced under the differential accounting of Assumption A2 and the balanced-prior charge, and both choices carry weight (Remarks 12 and 9).

Proposition 5 (Anchored escape)**.** *Suppose every W2 write's anchor covers all previously written classes. Then the gauge cost is identically zero at population level, the coupled problem collapses to the scheduling problem of Proposition 18, and its guarantees hold uniformly over geometry: break-even is $(2 + w^{\max}/K)$-competitive, hence 2 in the fine-granularity limit, and $e/(e-1)$ is a randomized lower bound, each under Proposition 18's delay scope. With coverage A only, the same holds for all pairs inside A together with the batch, and pairs against uncovered classes keep the mismatch the ledger assigns them. Both statements are at a setup price fixed independently of the store, the instance class of Proposition 18; a setup that grows with the covered classes is outside every theorem here, and E-L2-11 measures that regime (Appendix I). Combined with Theorem 4: geometry-independent guarantees require an operator that can revisit written frames, and W2 is the cheapest, its cost scaling with the anchor's coverage in exemplars rather than with the store.*

The pair of results is the paper's thesis: *revisability buys geometry-independence, and anchoring is its cheap instrument under the gauge the batch-local fit selects.* W1 escapes too, at the price of the store, so anchoring is sufficient and not necessary. The gauge qualifier is a hypothesis, not information-theoretically forced: a member writing rows in absolute coordinates would share one frame across batches for free, and the ridge selects the batch-anchored member (Appendix D). The divergence is in the query count, not the per-query price ($\Xi < \frac{1}{2}$); under a one-sided charge the flood is free and the ratio bounded (Remark 9), and the gross convention gives the construction a per-instance ratio that is finite but unbounded across the class, which the program measured. The classical impossibility pattern transfers with its flood installed geometrically (App. G.7); the mixed problem is App. H; Figure 12 in the appendix draws the excess.

9 Experiments

Registered before any run (Appendix I; the frozen registration and its 93-amendment log ship in the supplement) and executed on smoke seeds, then on canonical seeds 11, 13, 17, the program asks how much of the population statement — the ledger's exact arithmetic, the dominance comparison, the separation bounds — carries over to real corpora and a trained model. Every measured number below enters through a gate that refuses any CSV off its pinned digest or hash and recomputes the manifest's verdicts from the verified rows. Seven synthetic lanes ran on CPU; the ledger's reach was tested on a small real corpus with non-Gaussian, unequal-covariance conditionals (the *reduced testbed*), then across eight frozen backbone-corpus cells at a hundred classes (Table 11) and on Pythia-160M; four experiments at measured constants ran last (E-L2-11 to E-L2-14). Appendix I lists every registered criterion with its outcome.

No registered test breached the ledger or the dominance arithmetic. E-L2-1's Monte Carlo covered 115 conditions, five margins by 8 mismatch families per seed, and put no cell-seed beyond the family-wise multiple-testing threshold; all 75 exceptional-ray tests returned exactly the aligned value, as Proposition 15's equality predicts. In E-L2-2, across 3780 randomly drawn geometries, no merged pair worsened by more than the stored values' resolution of 5.0×10^{-7}, while every bystander pair did worsen, as the theory predicts. In E-L2-3, the exact discrete-round minimizer fell inside the measured band around T^* in all 255 cell-seeds.

The frontier arithmetic survived in every measured cell. Under the differential cost convention the rent term already uses (Assumption A2), the corrected separation bounds hold in all 150 cell-seeds: Remark 11's fixed-composition bound, which fixes which classes share a batch, is exactly attained in the 72 cell-seeds where the rent term, not the write cost, is the active part of the bound, and the anchored policy's measured ratio sits at 2.000. At $Q_2 = 10^7$ (E-L2-9b) the best policy's measured ratio reaches the ceiling $1000.0 = rQ_1/K'$, with that policy writing *nothing* in stage one. The theorem is not contradicted: its adversary scales Q_1 with the flood Q_2, so the ratio diverges only in a joint limit that no finite run displays. The registration's first prediction, a plateau

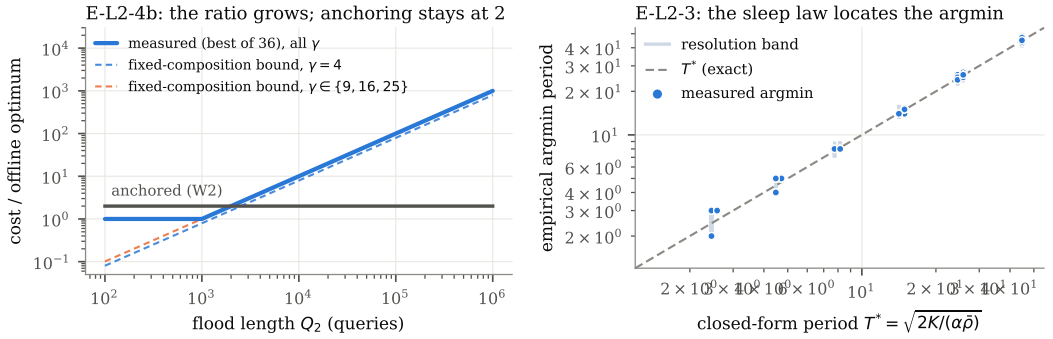


Figure 2: Left, E-L2-4b: the best ratio achieved by any pure-W3 policy against the exact optimum as the flood lengthens (γ sweep at $s = 12.9$; the four measured curves coincide to the printed precision, since the best policy rents k forever, a geometry-free branch, so one curve is drawn). Overlaid: the theorem's composition-free rate from (4); the fixed-composition bound of Remark 11 (dashed, everywhere above it); and the anchored break-even policy, the flat line at 2. Right, E-L2-3: the empirical argmin period against T^* with its resolution bands; both panels are drawn from the canonical CSVs. Takeaways: the unanchored measured ratio grows with the flood along the theorem's rate while the anchored policy holds at 2.000 in every cell (left); the exact discrete minimizer lands in the resolution band around T^* in all 255 cell-seeds (right).

under the gross convention, was withdrawn by amendment when the optimum was found to escape that baseline by writing the flooded class together with a new class; Appendix I records the full sequence; Figure 2 (left) draws the confirmed bounds.

On real geometry the gauge transports as a centring identity, and coverage buys the frame. At the reduced testbed (14.0 fit rows per dimension) the fitted rule's mean direction has cosine similarity 0.913 with the centred direction $\delta - \bar{D}$ and 0.642 with the raw δ; the raw cosine divided by the centred one is 0.7031, matching the centring prediction $\sqrt{1 - 1/m} = 0.7071$ at $m = 2$, so the fit keeps Lemma 7's gauge. Coverage, not size, buys the frame there: a covering anchor of size 1 or 3 left the pair's misalignment excess at 0.0079 and 0.0062, where an equal-sized non-covering anchor left it at 0.0440 and 0.0475 (paired difference -0.0382 at -7.7 standard errors). Across the eight backbone-corpus cells at a hundred classes (Table 12) the centring identity holds on all 742500 pairs, and coverage has the lower mean error on 7 of 8 cells. What does not carry over is quantitative: the level is over-predicted at the reduced testbed (0.1055 predicted against 0.0473 measured) and two orders of magnitude too low at a hundred classes, and the Gaussian floor's magnitude cannot be estimated there (Section 10); the appendix carries the per-cell detail.

On a trained model at a fixed step budget the pre-run expectation reverses (Figure 3). E-L2-8 wrote 100 classes (topics) into Pythia-160M through the three operators, each write implemented as a low-rank adapter (LoRA) update; the draft expected W3 worst and W1 best. Measured probe loss, lower is better: W3 1.7246, W2 1.8185, W1 1.9300, in every seed (seed-level $t \geq 4.1$). Two asymmetries favour the unanchored arm: every write is 200 steps (L2-A17(a)), so the joint refit gives each item a $10\times$ smaller share by the last round; and at answering time W3 draws on 10 resident adapters against W1's 1. The result measures the budget and the routing through the operators, not the operators alone.

At measured constants no proved bound was breached, the premium is measured, and the construction is real (Figure 4). E-L2-11: at full coverage the measured cost premium tracks the theory's prediction, within six percent on every seed (102.88/106.04/99.90 by seed against 100.0000), which checks the mixed-operator theory's predicted premium. E-L2-12: no proved bound was breached on the constant-setup arm the theorems cover, and the sleep law is in band on every stationary stream; break-even reached 2.2049, inside the 3.7731 that its row's parameters allow. A_{K_3}, Proposition 29's covering-anchored policy, beat break-even on none of 15 streams at the reference volume, at measured premiums of 147.1/155.8/152.2: at premiums of that order its promised improvement has no practical bite, a scope statement Section 10 carries. E-L2-14: the class means of the CIFAR-100 image benchmark contain a two-phase geometry of the kind the construction requires, and its unanchored ratio grows, as the flood lengthens, to 775.2295 against

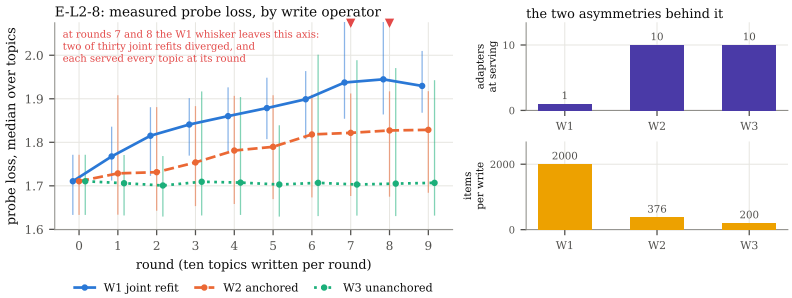


Figure 3: Measured probe loss by round for the three write operators, from `l2_e8_accel.csv`; probe loss is the held-out loss of reading a written class back out of the updated model, lower is better. Left: median over topics, whiskers at the 5th and 95th centiles, medians because W1 is bimodal at the two marked rounds (joint refits diverged). Right: resident adapters at answering time and training items per write at the last round (budget and routing, priced alongside the operator). The measured ordering runs opposite to the pre-run expectation.

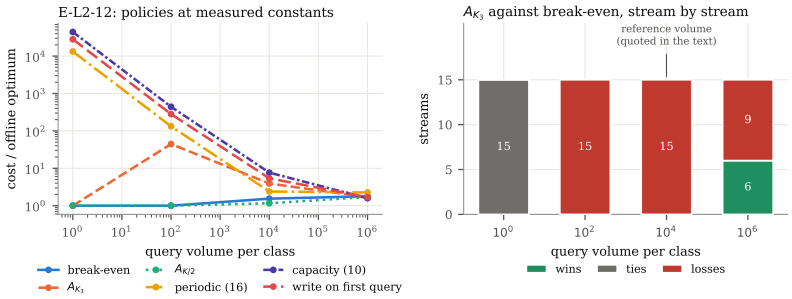


Figure 4: Measured, from `l2_e12_accel.csv`. Left: each policy's cost as a multiple of the offline optimum against query volume per class; at high volume writing is cheap, every policy nears the optimum, and the separations the theory is about live at the left of the axis. Right: per stream and volume, whether A_{K_3}'s cost lies above or below break-even's on the constant setup, counts printed; the marked column is the reference volume. All fifteen streams tie at unit volume.

the anchored 2.0000. E-L2-13's selection comparison on this testbed is read in Section 10; the appendix reports it under the registered filter and the unfiltered cells.

10 Limitations and Outlook

The ledger prices a population ideal: a batch-local fit over Gaussian-conditioned, whitened features, and it is pairwise. On real geometry its level and floor do not transport, its direction test certifies only a centring identity, and under the registered filter its ranking is empty; the closed forms are population objects in whitened coordinates, and at a hundred classes their levels do not transfer into operations. What proved itself is everything the paper proves: the exact ledger, the gauge split, the dominance arithmetic, the scheduling reduction with its inherited bounds, none breached on the constant-setup arm; the corrected separation bounds held in every tested cell; and the separation construction instantiated on natural geometry grows with the flood as the closed form says, though at $4.0/4.9/4.4$ times its level. The impossibility is convention-dependent (Remarks 12 and 9) and assumes the batch-anchored frame the ridge selects (Section 8); the operator ordering reverses on our one trained model, and we do not read that reversal as the operators alone. What the paper settles is where worst-case guarantees can exist under this cost convention: not for policies built from unanchored writes alone, and, for covering replay priced independently of the store, at the classical deterministic constant in the fine-granularity limit. Open: the batch-partition problem of Section 6; the optimal randomized constant at a time-varying rate, at least $e/(e-1)$ with no algorithm known to attain it (Appendix H.3); and the deterministic mixed-operator constant between $\max(2, 1+P/2)$ and $1+P$ (Appendix H.2). Geometry-independent guarantees are a theorem about the one operator whose anchor covers what it protects, and a measurement about everything else.

AI USE STATEMENT

The authors conceived, designed, and directed this research; every scientific decision in it — the pricing theory, the impossibility results, the pre-registered experimental programme, and the reading of its outcomes — was theirs. A generative AI assistant, working under their direction, assisted with drafting and editing the manuscript, exploring proof approaches that the authors then re-derived and checked, literature search and the verification of citation metadata against primary records, and implementing the experiment-runner, provenance-gate, figure, and build code that the authors specified and reviewed. No measured number was produced or altered by AI: every result comes from the registered runners on the canonical seeds through the fail-closed verification gate. The authors reviewed all AI-assisted content and take full responsibility for the paper. The submission form carries the same disclosure.

ETHICS STATEMENT

This is a theoretical paper with a pre-registered empirical program; no human subjects, personal data, or deployed systems are involved: the synthetic tiers are seeded draws, the frozen-feature tiers use standard public image benchmarks, and the model-scale tier writes synthetic statements. The consolidation policies studied are decision rules for memory systems; we see no direct path from the results to harm beyond the general dual-use character of more capable continual learning agents, which the position paper we build on discusses.

REPRODUCIBILITY STATEMENT

Every measured number traces to a canonical CSV with a manifest (code hashes, seeds, grid, tolerances, environment, per-cell falsifier verdicts) through the fail-closed gate of Appendix N; the frozen registration with its amendment log and SHA-256 escrow, the runners, the library with its self-test, the canonical and smoke CSVs, the figure scripts, and a one-command build that regenerates the manuscript's measured content ship in the supplement (Appendix J). Every closed-form table and figure is an evaluation of an equation stated in the text, recomputed by a companion build artifact that checks each tabulated constant through an independent implementation of the normal tail.

REFERENCES

Yossi Azar and Noam Touitou. General framework for metric optimization problems with delay or with deadlines. In *Proceedings of the 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 60–71, 2019.

Yossi Azar and Noam Touitou. Beyond tree embeddings – a deterministic framework for network design with deadlines or delay. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 1368–1379, 2020.

Sujoy Bhore, Michał Pawłowski, and Seeun William Umboh. Online TCP acknowledgment under general delays. *arXiv preprint arXiv:2604.13428*, 2026.

Marcin Bienkowski, Jarosław Byrka, Marek Chrobak, Łukasz Jeż, Dorian Nogneng, and Jiří Sgall. Better approximation bounds for the joint replenishment problem. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 42–54, 2014.

Marcin Bienkowski, Martin Böhm, Jarosław Byrka, Marek Chrobak, Christoph Dürr, Lukáš Folwarczný, Łukasz Jeż, Jiří Sgall, Nguyen Kim Thang, and Pavel Veselý. Online algorithms for multi-level aggregation. In *Proceedings of the 24th Annual European Symposium on Algorithms (ESA)*, pp. 12:1–12:17, 2016.

Marcin Bienkowski, Martin Böhm, Jarosław Byrka, Marek Chrobak, Christoph Dürr, Lukáš Folwarczný, Łukasz Jeż, Jiří Sgall, Nguyen Kim Thang, and Pavel Veselý. Online algorithms for multilevel aggregation. *Operations Research*, 68(1):214–232, 2020.

Marcin Bienkowski, Martin Böhm, Jarosław Byrka, Marek Chrobak, Christoph Dürr, Lukáš Folwarczný, Łukasz Jež, Jiří Sgall, Nguyen Kim Thang, and Pavel Veselý. New results on multi-level aggregation. *Theoretical Computer Science*, 861:133–143, 2021.

Niv Buchbinder, Tracy Kimbrel, Retsef Levi, Konstantin Makarychev, and Maxim Sviridenko. Online make-to-order joint replenishment model: primal dual competitive algorithms. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 952–961, 2008.

Niv Buchbinder, Moran Feldman, Joseph (Seffi) Naor, and Ohad Talmon. O(depth)-competitive algorithm for online multi-level aggregation. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1235–1244, 2017.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. Mem0: Building production-ready AI agents with scalable long-term memory. *arXiv preprint arXiv:2504.19413*, 2025.

Christian Coester and Alex Turoczy. Primal-dual online algorithms for the parking permit problem. In *34th Annual European Symposium on Algorithms (ESA)*, LIPIcs, pp. 81:1–81:16, 2026.

Junze Deng, Qinhang Wu, Peizhong Ju, Sen Lin, Yingbin Liang, and Ness B. Shroff. Unlocking the power of rehearsal in continual learning: A theoretical perspective. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *PMLR*, pp. 13277–13326, 2025.

Meng Ding, Kaiyi Ji, Di Wang, and Jinhui Xu. Understanding forgetting in continual learning with linear regression. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *PMLR*, pp. 10978–11001, 2024.

Daniel R. Dooly, Sally A. Goldman, and Stephen D. Scott. On-line analysis of the TCP acknowledgment delay problem. *Journal of the ACM*, 48:243–273, 2001.

Vaggelis Dorovatas, Malte Schwerin, Andrew D. Bagdanov, Lucas Caccia, Antonio Carta, Laurent Charlin, Barbara Hammer, Tyler L. Hayes, Timm Hess, Christopher Kanan, Dhireesha Kudithipudi, Xialei Liu, Vincenzo Lomonaco, Jorge Mendez-Mendez, Darshan Patil, Ameya Prabhu, Elisa Ricci, Tinne Tuytelaars, Gido M. van de Ven, Liyuan Wang, Joost van de Weijer, Jonghyun Choi, Martin Mundt, and Rahaf Aljundi. Position: Modular memory is the key to continual learning agents. *arXiv preprint arXiv:2603.01761*, 2026. Position paper at the International Conference on Machine Learning (ICML) 2026.

Ran El-Yaniv, Ron Kaniel, and Nathan Linial. Competitive optimal on-line leasing. *Algorithmica*, 25(1):116–140, 1999.

Yuyang Hu, Shichun Liu, Yanwei Yue, Guibin Zhang, et al. Memory in the age of AI agents. *arXiv preprint arXiv:2512.13564*, 2025.

Anna R. Karlin, Mark S. Manasse, Larry Rudolph, and Daniel D. Sleator. Competitive snoopy caching. *Algorithmica*, 3(1-4):79–119, 1988.

Anna R. Karlin, Claire Kenyon, and Dana Randall. Dynamic TCP acknowledgement and other stories about e/(e-1). In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC)*, pp. 502–509, 2001.

Marcus Klasson, Hedvig Kjellström, and Cheng Zhang. Learn the time to learn: Replay scheduling in continual learning. *Transactions on Machine Learning Research*, 2023.

Kevin Lin, Charlie Snell, Yu Wang, Charles Packer, Sarah Wooders, Ion Stoica, and Joseph E. Gonzalez. Sleep-time compute: Beyond inference scaling at test-time. *arXiv preprint arXiv:2504.13171*, 2025.

Zvi Lotker, Boaz Patt-Shamir, and Dror Rawitz. Ski rental with two general options. *Information Processing Letters*, 108(6):365–368, 2008.

Zvi Lotker, Boaz Patt-Shamir, and Dror Rawitz. Rent, lease, or buy: randomized algorithms for multislope ski rental. *SIAM Journal on Discrete Mathematics*, 26(2):718–736, 2012.

Mathieu Mari, Michał Pawłowski, Runtian Ren, and Piotr Sankowski. Online multi-level aggregation with delays and stochastic arrivals. In *35th International Symposium on Algorithms and Computation (ISAAC)*, LIPIcs, pp. 49:1–49:20, 2024.

Adam Meyerson. The parking permit problem. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 274–284, 2005.

Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G. Patil, Ion Stoica, and Joseph E. Gonzalez. MemGPT: Towards LLMs as operating systems. *arXiv preprint arXiv:2310.08560*, 2023.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST)*, pp. 1–22, 2023.

Manish Purohit, Zoya Svitkina, and Ravi Kumar. Improving online algorithms via ML predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

Steven S. Seiden. A guessing game and randomized online algorithms. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing (STOC)*, pp. 592–601, 2000.

Wujiang Xu, Zujie Liang, Kai Mei, Hang Gao, Juntao Tan, and Yongfeng Zhang. A-MEM: Agentic memory for LLM agents. In *Advances in Neural Information Processing Systems*, 2025. arXiv:2502.12110.

GUIDE TO THE APPENDICES

Appendices A and B build the world (economics, notation, substrate); C through F prove the economics of frames and schedules (ledger, gauge split, dominance, scheduling), each result the restructure moved is stated where it is proved, so the main text's pointers resolve to the statement and its proof together; G and H carry the competitive frontier and the mixed-operator problem, whose interior Appendix H.2 closes; I and J register the empirical program and the scope; K situates the related work; L through N are the companion material (worked instances, objections, code); O fixes the algorithms; P is the quantitative theory in figures and tables. Table 1 is the dependency map: every result, where it is proved, where it is evaluated, and what re-measures it.

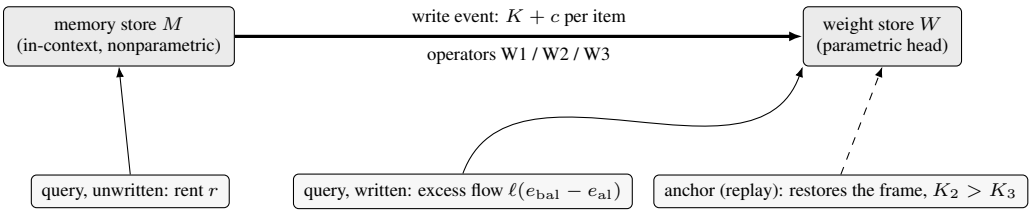


Figure 5: The cost flows of Assumption A2. Left: the in-context store, whose queries pay rent r for as long as their items stay unwritten. Right: the weight store, which a write event fills at a setup plus per-item cost through one of the three operators; a query to a written class pays the misalignment excess (middle) over the aligned baseline both stores share. The dashed arrow is the anchor: the priced revisit that zeroes the excess on the pairs it covers. The finding the diagram carries: an unanchored policy pays rent and excess for as long as it holds, and the one priced act that stops the excess is revisiting a frame — the paper's thesis that revisability buys geometry-independence, with anchoring as its cheap instrument. Every quantity in the diagram has a closed form in the margins and anchors.

A MODEL, CONVENTIONS, AND NOTATION

Rounds, items, stores. Time is discrete, $t = 0, 1, 2, \ldots$. Items (classes, or exemplars of classes, depending on the tier) live in an in-context store M before consolidation and in the weight store W

Table 1: Dependency map: result, proof site, quantitative site, and the registered experiment that re-measures it. Rows with no experiment are purely structural.

Result	Proof	Evaluation	Experiment
Remark 3 (singleton degeneracy)	App. B	n/a	n/a
Remark 4 (batch-local objective)	App. D	n/a	n/a
Definition 2 (guarantee functional)	App. E	n/a	n/a
Lemma 7 (per-batch gauge)	App. B	n/a	E-L2-5
Lemma 11 (misalignment ledger)	App. C	Figs. 10, 11; Tabs. 4, 5, 19, 20	E-L2-1
Prop. 12 (invariance)	App. C.3	n/a	n/a
Prop. 13 (gauge split)	App. D	Fig. 5	E-L2-5 (direction only)
Thm. 1 (dominance)	App. E	Fig. 18	E-L2-2
Prop. 15 (exceptional set)	App. E.3	Fig. 10 (left)	E-L2-1
Prop. 18 (reduction)	App. F.1	Tab. 6	E-L2-4b (anchored arm)
Cor. 3 (sleep law)	App. F.2, F.5	Fig. 15; Tab. 21	E-L2-3
Lemma 22 (confinement)	App. G.2	Fig. 17	n/a
Lemma 24 (mismatch floor)	App. G.3	Fig. 13; Tab. 17	E-L2-4
Prop. 33 (floor maximum)	App. P.2	Fig. 13	n/a
Thm. 4 (separation)	App. G.4	Figs. 12, 14; Tab. 18	E-L2-4, E-L2-4b
Lemma 25 (frozen-frame floor)	App. G.4	Tab. 18	E-L2-4b
Lemma 26 (free parking)	App. G.5	n/a	E-L2-4b
Prop. 5 (anchored escape)	App. G.8	n/a	E-L2-4b, E-L2-6
Prop. 27 (the mixed-operator bracket)	App. H.2	`l2_mixed.py`	E-L2-6
Prop. 28 (the comparison costs the premium)	App. H.2	`l2_mixed.py` (priced by enumeration)	n/a
Prop. 29 (the improved upper bound)	App. H.2	`l2_mixed.py` (against an enumerated optimum, 5420 streams)	n/a
Thm. 31 (no premium-independent constant)	App. H.2	`l2_mixed.py` (gadget minimax, on and off grid)	n/a
Prop. 30 (the classical constant, two classes)	App. H.2	`l2_mixed.py` (minimised over τ)	n/a
Cor. 32 (the bracket, exact at $P = 1$)	App. H.2	`l2_mixed.py` (strict-ratio witness)	n/a

after. A query addressed to an unconsolidated item pays *rent*: a per-query serving cost reflecting retrieval compute, retrieval failure risk, and context occupancy, netted against the consolidated state and denoted $r > 0$ per discriminating query. Rent accrues only while the item is unwritten; this is the delay-cost identification Proposition 18 makes exact. A write event pays a setup $K > 0$ plus $c \geq 0$ per item written (Assumption A2). The constants are symbolic throughout, their empirical calibration is E-L2-11, which measures the premium in the cost model's own units (E-L2-6 measures the error arms the premium buys), and nothing in the statements depends on their values, only on their signs and on $K > 0$, $r > 0$, $\ell > 0$.

The feature model. A frozen encoder maps raw inputs into $\mathbb{R}^d$. Classes r have Gaussian conditionals $N(\mu_r, \Sigma)$ with Σ positive definite and shared across classes; we whiten once, $\Sigma = I$, so all geometry is Euclidean in whitened coordinates and class structure reduces to the pairwise margins $\gamma_{rs} = \|\mu_r - \mu_s\|^2 > 0$ for distinct classes. The weight store is a linear cross-entropy head $h_r(x) = w_r^\top x + b_r$ over the label set it has been fit on; the in-context store is a prototype (nearest class mean) rule over stored exemplars. At population level, on a pair the head carries in a single frame and with matched label sets, these two rules coincide as classifiers, both Bayes at balanced priors; that is why the paper's error economics can be carried entirely by the head. Across frames they do not coincide, and that gap is exactly the misalignment excess this paper prices. The scope (matched coverage, balanced priors, one frame) is restated where it is used.

The population ledger. All error quantities in the main text are population (infinite-data) quantities: expectations over the class conditionals, with the head at its population optimum. Finite-

sample effects, estimation error of means, and optimization path effects are outside this ledger and are scoped in Appendix J. The population ledger is the right place to price protocols rather than implementations: every effect that survives here is intrinsic to the write operator and the schedule, and cannot be trained away.

Batches and the composed protocol. A batch B_t is a multiset of classes with multiplicities $n_r \geq 1$ (multiple exemplars of a class may be written together). Within a batch the training mixture weights are proportional to counts. The *composed protocol* subtracts the log-count offset from each class's logit at evaluation; composition is free (counts are known), and at population level it renders every within-batch pair rule Bayes at balanced priors. Class-wise constant corrections beyond log-counts are admitted in the same spirit: they cost nothing at population level and are estimated in practice; Proposition 13 is exactly the classification of what this entire family of free corrections can and cannot buy.

Error objects. All errors are pairwise. For a pair (r, s) with rule "decide r over s iff $g^\top x \geq c$," the one-sided errors are $e_{s\to r} = \Pr[g^\top X \geq c \mid X \sim N(\mu_s, I)]$ and $e_{r\to s} = \Pr[g^\top X < c \mid X \sim N(\mu_r, I)]$, and the balanced error is their mean. The multiclass error of the head is accessed through the standard envelope (worst pair, or the restricted union bound over pairs); the pairwise ledger is a deliberate modeling choice, made once here, and every theorem below is pairwise in its error objects.

Notation. μ_r: class mean (whitened). $\bar{\mu}_t, \bar{S}_t$: batch anchor and batch scalar. δ, γ: pair displacement and margin. $\bar{D}, \bar{\Delta}$: vector and scalar mismatch across batches. $e_{\mathrm{al}} = \Phi(-\sqrt{\gamma}/2)$: aligned (Bayes) balanced pair error. e_{bal}: realized balanced pair error. ϕ_{bal}: mismatch floor (Lemma 24). r, K, c, ℓ: rent, setup, per-item write cost, error currency. R: competitive ratio. $W1, W2, W3$: the operators of Definition 1. conv: convex hull; $\overline{S}$: closure. OPT: the offline optimum under the same operator class.

A.1 The rent, assembled

The rent r is the paper's price of not having written, and it is a composite object; stating its decomposition keeps the scheduling layer honest about what it prices. Per discriminating query, an unconsolidated item pays three components relative to the consolidated state: the serving differential (retrieval compute and context occupancy against amortized weight-store serving), the failure term $p(m)\,\ell\,(e_{\mathrm{miss}} - e_{\mathrm{W}})$, with $p(m) = \min(1,\, p_0 + \rho_{\mathrm{rot}}\, m)$ the retrieval failure probability at memory size m and ρ_{rot} the rot rate, and the coverage term when label sets differ. Two conventions matter. First, r is a *differential*: the consolidated state's per-query error on covered classes is the aligned Bayes error both stores share at population level, so the rent-side comparison nets it out; this is the store-neutrality fact, used here as a normalization and proved where it is load-bearing (Appendix B.3). Second, *every theorem here takes r constant*, with $p(m)$ at a fixed nominal store size for calibration only: m is occupancy, which a policy controls by writing, and Proposition 18's identification needs the rate exogenous. With that fixed, the identification uses that rent accrues *linearly* per round: each round of waiting costs what that round's queries cost, independent of history; nonlinear batch-delay objectives would break the inherited guarantees, which is exactly the scope sentence of Section 7. The rot rate enters the sleep law only through the mean rent $\bar{\rho}$, which is why Corollary 3's comparative static can say *the period shortens as rot raises rents* without a rot model of its own: rot is one of the addends of $\bar{\rho}$, and T^* is decreasing in $\bar{\rho}$.

The write-side constants are kept symbolic throughout: K (setup compute of a write event), c (per-item write cost), and ℓ (the error-to-cost currency that converts pairwise error flows into the same units as compute). Their empirical calibration is E-L2-6's measurement, not a modeling assumption; every theorem is stated for all positive values, and the closed-form regime of Theorem 4 is a condition on ℓ/r rather than a pinned constant.

B The Self-Contained Substrate

This appendix proves the population optimum of a cross-entropy head fit on a single batch, from the Gaussian model, with no imports. It is the substrate every later appendix stands on.

Table 2: Core symbols.

Symbol	Meaning
μ_r, Σ	class mean and shared covariance (whitened to identity)
δ_{rs}, γ_{rs}	pair displacement $\mu_r - \mu_s$ and margin $\|\delta_{rs}\|^2$
B_t, n_r	batch t and class multiplicities within it
$\bar{\mu}_t$, $\bar{S}_t$	batch anchor (count-weighted mean of means) and batch scalar (count-weighted mean of squared norms)
$\bar{D}_{tu}$, $\bar{\Delta}_{tu}$	cross-batch vector and scalar mismatch
$e_{s\to r}$, e_{bal}	one-sided and balanced pairwise errors
e_{al}	aligned (Bayes) balanced error $\Phi(-\sqrt{\gamma}/2)$
ϕ_{bal}	mismatch floor (Lemma 24)
$\Psi(\rho)$	guarantee functional at mismatch budget ρ (Definition 2)
$\rho(P)$	mismatch budget of a partition, $\max_{t\neq u}\|\bar{D}_{tu}\|$

Table 3: Economic and online symbols.

Symbol	Meaning
r	per-query rent of an unwritten item (differential, Appendix A.1)
$\bar{\rho}$	mean rent per item per round in the stationary flow
K, c	write event setup cost; per-item write cost
K_2, K_3	anchored and unanchored setups in the mixed-operator problem (Appendix H)
ℓ	error-to-cost currency
$K' = K + 2c$	cost of writing a pair
α	entry-eligible arrival rate per round
T^*	optimal period $\sqrt{2K/(\alpha\bar{\rho})}$
R	the registered (falsified) plateau ratio of the earlier separation theorem
$\Xi = \frac{1}{2} - e_{\mathrm{al}}$	excess floor: the per-query price, per unit currency, of Theorem 4
$\phi_{\mathrm{bal}} - e_{\mathrm{al}}$	the fixed-composition excess at a placement, which dominates Ξ
Q_1, Q_2	stage-1 and stage-2 flood lengths
p_W	stage-1 write probability of a randomized policy
OPT	offline optimum within the same operator class

Lemma 6 (Population CE optimum on a mixture)**.** *Let $C \geq 2$ classes with means $\{\mu_r\}_{r\leq C}$ and shared identity covariance be mixed with strictly positive weights π_r (summing to one). The population cross-entropy risk of a linear softmax head,*

$$\mathcal{R}(h) \;=\; \mathbb{E}\big[-\textstyle\sum_r \mathbf{1}\{Y=r\}\log q_h(r \mid X)\big], \tag{6}$$

is minimized, in function space, exactly on the Bayes posterior $q^*(r \mid x) = \Pr[Y = r \mid X = x]$, and the set of heads realizing q^* is the gauge family

$$w_r \;=\; \mu_r + a, \qquad b_r \;=\; -\tfrac{1}{2}\|\mu_r\|^2 + \log\pi_r + c_0, \qquad a \in \mathbb{R}^d,\ c_0 \in \mathbb{R}. \tag{7}$$

On this family the ridge penalty is strictly convex in a. Which centering it selects depends on how the penalty is weighted, and we fix the convention here because the rest of the paper rests on it: under the count-weighted *ridge $\sum_r \pi_r\|w_r\|^2$, each row penalised in proportion to the data it is fit from, the minimizer is $a^* = -\bar{\mu}$ with $\bar{\mu} = \sum_r \pi_r\mu_r$ the count-weighted mean; the unweighted penalty $\sum_r\|w_r\|^2$ would select the unweighted mean of the distinct class means instead, and the two coincide exactly at equal multiplicities. We use the count-weighted ridge throughout, so that the batch anchor is the count-weighted mean and the achievable-anchor set of Lemma 22 is the dense set of count-fraction combinations rather than the finite set of subset centroids; consequently the ridge-regularized risk with parameter λ has minimizers converging, as $\lambda \to 0^+$, to the minimum-norm member $w_r = \mu_r - \bar{\mu}$. The bias gauge c_0 is unconstrained by a ridge on the weights alone and is fixed by convention (sum-zero biases within the fitted label set).*

Proof. Strict properness of the log loss gives $\mathcal{R}(h) - \mathcal{R}(q^*) = \mathbb{E}\,\mathrm{KL}\big(q^*(\cdot \mid X)\,\|\,q_h(\cdot \mid X)\big) \geq 0$ with equality iff $q_h = q^*$ almost surely under the mixture's marginal on X. For the shared-

covariance Gaussian mixture, Bayes' rule gives

$$\log q^*(r \mid x) \;=\; \mu_r^\top x - \tfrac{1}{2}\|\mu_r\|^2 + \log \pi_r - \log Z(x), \tag{8}$$

so every head of the form (7) realizes q^* (the softmax cancels any common affine offset $a^\top x + c_0$ added to all logits), and conversely a head realizing q^* must satisfy $h_r(x) - h_s(x) = \log q^*(r \mid x) - \log q^*(s \mid x)$ for all x up to a x-null set, which forces the linear differences and hence membership in the family, the two Gaussians having analytic likelihood ratios. For the ridge selection: on the family, $\sum_r \|w_r\|^2 = \sum_r \|\mu_r + a\|^2$, a strictly convex quadratic in a with minimizer $a^* = -\big(\sum_r \mu_r\big)/C$ in the unweighted sum; under the count-weighted penalty fixed in the statement the weights carry through, giving $a^* = -\sum_r \pi_r \mu_r$; the unweighted penalty gives $-C^{-1}\sum_r \mu_r$, and the two agree at unit multiplicities. Standard epi-convergence of the ridge objectives (their minimizers are eventually confined to a bounded set, and their Γ-limit is the unregularized risk plus the indicator of the family) yields convergence of minimizers to the minimum-norm element. The bias convention is separate: differences $h_r - h_s$ never see c_0, and every pairwise statement below is gauge-invariant. □

Lemma 7 (Per-batch gauge)**.** *Fit W3 on batch B_t with $|B_t| \geq 2$, composed counts within the batch, whitened coordinates, and the sum-zero bias convention. The population optimum is realized on the family*

$$h_r(x) \;=\; (\mu_r - \bar{\mu}_t)^\top x \;-\; \tfrac{1}{2}\|\mu_r\|^2 \;+\; \tfrac{1}{2}\bar{S}_t, \qquad r \in B_t, \tag{9}$$

with the ridge's zero limit selecting the minimum-norm representative, whose directions are centered at the batch anchor. The bias constant $\frac{1}{2}\bar{S}_t$ is a within-batch convention; its arbitrariness across batches is the scalar gauge freedom Proposition 13 shows is harmless.

Proof of Lemma 7 (per-batch gauge). Apply Lemma 6 to the batch B_t as its own mixture: the label set is B_t, the weights are the batch's count fractions n_r/n_{B_t}, and the composed protocol evaluates at balanced priors, which is equivalent (for pairwise objects) to removing the $\log \pi_r$ from the bias differences; with the sum-zero bias convention the bias of class r is $-\frac{1}{2}\|\mu_r\|^2$ plus the common constant chosen so the biases sum to zero within the batch. That constant is $\frac{1}{2}\bar{S}_t$ up to the count weighting: with uniform within-batch counts the biases $b_r = -\frac{1}{2}\|\mu_r\|^2 + c_0$ sum to zero iff $c_0 = \frac{1}{2}\bar{S}_t$ with $\bar{S}_t = |B_t|^{-1}\sum_r \|\mu_r\|^2$, which is (9); with multiplicities, $c_0 = \frac{1}{2}\bar{S}_t$ holds verbatim for $\bar{S}_t$ the count-weighted mean of squared norms, and the pair-rule differences below are unchanged. The ridge selection centers the directions at the batch anchor $\bar{\mu}_t$ (the count-weighted mean of the batch's means), which is exactly the minimum-norm member of the batch's gauge family. □

Remark 2 (Why the anchor is the frame). The batch anchor $\bar{\mu}_t$ is not a summary statistic of convenience; it is the head's frame. The directions $\mu_r - \bar{\mu}_t$ and the bias constant $\frac{1}{2}\bar{S}_t$ together are a complete description of what the fit can express about absolute positions: two heads fit on the same batch agree on all decisions regardless of gauge representative, and two heads fit on different batches disagree on cross pairs precisely through $\bar{D}$ and $\bar{\Delta}$. Everything in the ledger follows from this.

Remark 3 (Singleton degeneracy). A singleton unanchored batch is degenerate: a one-class cross-entropy fit has no contrast, the population risk is minimized by a constant predictor, and no direction is identifiable. W3 therefore requires batch size at least two. We state this as a hypothesis of every W3 result; it is also the formal reason a singleton write has to be anchored.

Proposition 8 (Singleton degeneracy, formal)**.** *If B contains a single class r, the population cross-entropy risk over the single-class mixture is $\mathcal{R}(h) = 0$ for every head with $q_h(r \mid x) \equiv 1$, i.e., for every head whose logit for r exceeds its (empty set of) competitors; the risk is constant on this set and contains no information about μ_r, and no direction is identifiable. The population optimum of a singleton unanchored batch is therefore total gauge freedom, which is the formal content of Remark 3.*

Proof. With one class the label is constant, $\Pr[Y = r] = 1$, and $\mathcal{R}(h) = -\mathbb{E}[\log q_h(r \mid X)] = -\log 1 = 0$ whenever $q_h(r \mid \cdot) \equiv 1$; the softmax with a single logit is identically 1 for every value of the logit, so the risk is 0 on all of the head space and its argmin carries no information. □

The composed protocol at the batch level. Within a batch, composition (subtracting $\log$ counts) restores balanced-prior decisions on every within-batch pair; the within-batch pair rule is the Bayes bisector rule, with error e_{al}. This is why Theorem 1(a) can speak of merged pairs as aligned: after a merge, yesterday's cross pair is today's within pair. Across batches, composition removes the count-odds offsets, and Proposition 13(a) removes the scalar-gauge residue $\frac{1}{2}\bar{\Delta}$; what no free correction removes is $\bar{D}$, the difference of anchors.

B.1 Composition within a batch, in full

Within a batch with multiplicities n_r, the training weights are $\pi_r = n_r/n_{B_t}$, so the population optimum's bias carries $\log \pi_r$ (Lemma 6), and the within-batch pair rule acquires the count offset $\tau_{rs} = \log(n_r/n_s)$ in its threshold. Composition removes it: subtracting $\log n_r$ from each class's logit at evaluation restores the balanced-prior threshold, and the within-batch pair rule becomes the Bayes bisector with error e_{al}, at every count vector with $n_r, n_s \geq 1$. Three remarks. First, composition is free at population level and estimation-cheap in practice (counts are bookkeeping), which is the standing justification for adopting it as the default protocol. The uncomposed variant is a protocol choice, not an alternative theory, and its pairwise errors are the same closed forms with τ added to the thresholds. Second, the scalar offsets of Proposition 13(a) extend the same logic across batches: the batch scalar $\bar{S}_t$ plays the role that counts play within a batch, a constant per class computable from the class's own write event. Third, the degenerate multiplicity $n_r = 0$ never arises within a batch, a class being in the batch iff it was written there; the entry case $(0 \to 1)$ is priced by the miss convention, outside this paper's scheduling questions.

B.2 The multiclass envelope

Lemma 9 (Pairwise envelope for multiclass error)**.** *Fix a head over a label set S with pairwise rules as in (1). The multiclass error of the argmax rule on class r satisfies*

$$\Pr_{x\sim N(\mu_r,I)}\big[\arg\max_{j\in S} h_j(x) \neq r\big] \;\leq\; \sum_{j\neq r} e_{r\to j}, \qquad \textit{and} \qquad \max_{j\neq r} e_{r\to j} \;\leq\; \Pr[\arg\max \neq r], \tag{10}$$

so the multiclass error is pairwise-additive up to the standard union bound and is lower-bounded by the worst pair.

Proof. The event $\{\arg\max \neq r\}$ is the union over $j \neq r$ of $\{h_j \geq h_r\}$, whose probabilities are the one-sided pairwise errors; the union bound gives the upper inequality, and the lower inequality holds since each pairwise event is contained in the union event. □

Every use of "error" in this paper is a pairwise object; where a multiclass reading is desired, the envelope is the bridge, and the guarantee functional of Definition 2 inherits it unchanged (the worst pair lower-bounds multiclass error and the restricted union bound upper-bounds it, so a pair-level guarantee controls multiclass error only up to the factor $C - 1$).

B.3 The two stores agree at population level

Proposition 10 (Store neutrality in the instantiation)**.** *At population level, over a common label set with shared covariance and balanced evaluation priors, and for a pair the head carries in a single frame (one W3 batch, W1, or a W2 write whose anchor covers the pair), the prototype rule (nearest class mean in whitened coordinates) and the composed head of Lemma 7 define the same classifier on every pair, and both are the Bayes rule; consequently the two stores' per-query errors on covered classes coincide, and the rent's differential normalization (Appendix A.1) nets to the serving and failure components alone.*

Proof. The prototype rule decides r over s iff $\|x - \mu_r\| \leq \|x - \mu_s\|$, i.e., $\delta^\top x \geq \frac{1}{2}(\|\mu_r\|^2 - \|\mu_s\|^2)$: the Bayes bisector. The composed head's pair rule is the same halfspace (Appendix B.1). Pairwise agreement over all pairs implies the argmax classifiers agree up to tie sets, which have zero measure under the class-conditionals. □

This is the scope statement behind several narrative claims: consolidation buys serving cost, retrieval reliability, and coverage, and per-query accuracy on covered classes is not among the buys; the accuracy effects that remain are exactly the misalignment objects of Lemma 11, which is why they can carry an entire paper.

C THE MISALIGNMENT LEDGER

C.1 THE PAIR RULE

Let $r \in B_t$, $s \in B_u$, $t \neq u$. By (9),

$$h_r(x) - h_s(x) \;=\; (\mu_r - \bar{\mu}_t)^\top x - \tfrac{1}{2}\|\mu_r\|^2 + \tfrac{1}{2}\bar{S}_t \;-\; (\mu_s - \bar{\mu}_u)^\top x + \tfrac{1}{2}\|\mu_s\|^2 - \tfrac{1}{2}\bar{S}_u, \qquad (11)$$

which regroups to (1): the effective direction is $\delta - \bar{D}$ and the effective threshold is $\frac{1}{2}(\|\mu_r\|^2 - \|\mu_s\|^2) - \frac{1}{2}\bar{\Delta}$. All constants that survive into the pair rule are exactly the two mismatch objects; this is Lemma 11's claim that the ledger is complete.

C.2 THE ONE-SIDED ERRORS

Lemma 11 (Exact cross-batch errors)**.** *In the setting of Section 4, the one-sided pairwise errors of the composed W3 head are*

$$e_{s\to r} = \Phi\!\left(-\frac{\gamma/2 - \bar{\Delta}/2 + \bar{D}^\top \mu_s}{\|\delta - \bar{D}\|}\right), \qquad e_{r\to s} = \Phi\!\left(-\frac{\gamma/2 + \bar{\Delta}/2 - \bar{D}^\top \mu_r}{\|\delta - \bar{D}\|}\right), \qquad (12)$$

with the degenerate direction $\bar{D} = \delta$ handled separately: there the pair statistic is identically zero and the balanced error is exactly $1/2$. At zero mismatch ($\bar{D} = 0$, $\bar{\Delta} = 0$) both errors reduce to the aligned Bayes value $e_{\mathrm{al}} = \Phi(-\sqrt{\gamma}/2)$. Under a global translation τ, $\bar{D}$ is invariant while $\bar{\Delta}$ shifts by $2\tau^\top \bar{D}$ and $\bar{D}^\top \mu_s$ by $\tau^\top \bar{D}$, and the shifts cancel: the ledger is exactly translation-invariant.

Proof of Lemma 11. Under $X \sim N(\mu_s, I)$ the statistic $(\delta - \bar{D})^\top X$ is Gaussian with mean $(\delta - \bar{D})^\top \mu_s$ and variance $\|\delta - \bar{D}\|^2$. Hence

$$e_{s\to r} \;=\; \Pr\!\big[(\delta - \bar{D})^\top X \geq \tfrac{1}{2}(\|\mu_r\|^2 - \|\mu_s\|^2) - \tfrac{1}{2}\bar{\Delta}\big] \;=\; \Phi\!\left(\frac{(\delta - \bar{D})^\top \mu_s - c}{\|\delta - \bar{D}\|}\right), \qquad (13)$$

with c the threshold, and the numerator simplifies:

$$(\delta - \bar{D})^\top \mu_s - \tfrac{1}{2}\big(\|\mu_r\|^2 - \|\mu_s\|^2\big) \;=\; \delta^\top \mu_s - \tfrac{1}{2}\|\mu_r\|^2 + \tfrac{1}{2}\|\mu_s\|^2 - \bar{D}^\top \mu_s \;=\; -\tfrac{\gamma}{2} - \bar{D}^\top \mu_s, \qquad (14)$$

using $\delta^\top \mu_s = \mu_r^\top \mu_s - \|\mu_s\|^2$ and $\gamma = \|\mu_r\|^2 - 2\mu_r^\top \mu_s + \|\mu_s\|^2$. Substituting $\bar{\Delta}/2$ from the threshold gives the first formula of (12); exchanging $r \leftrightarrow s$ (which sends $\delta \mapsto -\delta$, $\bar{D} \mapsto -\bar{D}$, $\bar{\Delta} \mapsto -\bar{\Delta}$) gives the second. At $\bar{D} = 0$, $\bar{\Delta} = 0$: both arguments are $-\sqrt{\gamma}/2$. At $\bar{D} = \delta$: the direction $\delta - \bar{D}$ vanishes, the pair statistic is the constant 0, the rule is constant, one class receives all queries, and the balanced error is exactly $1/2$; Appendix G.3 treats the approach to this degenerate direction, where the two one-sided errors race to $\{0, 1\}$ at rates driven by the norm difference $\|\mu_r\|^2 - \|\mu_s\|^2$. □

C.3 THE TRANSLATION-INVARIANCE CERTIFICATE

Proposition 12 (Exact translation invariance)**.** *Fix a pair (r, s) in batches (t, u) and translate the entire world by τ: $\mu_j \mapsto \mu_j + \tau$ for every class j, inputs included. Then $\bar{D}$ is unchanged, $\bar{\Delta} \mapsto \bar{\Delta} + 2\,\tau^\top \bar{D}$, $\bar{D}^\top \mu_s \mapsto \bar{D}^\top \mu_s + \tau^\top \bar{D}$, and γ is unchanged. The numerators of both formulas in (12) are therefore unchanged, and so are the errors. No relative placement of the two batches' internals is identifiable by the ledger beyond $\bar{D}$ and $\bar{\Delta}$: the formulas are functions of the labeled geometry only.*

Proof. $\bar{\mu}_t$ and $\bar{\mu}_u$ each translate by τ, so their difference is invariant. $\bar{S}_t = \sum_{r\in B_t} \pi_r \|\mu_r\|^2$ maps to $\sum_{r\in B_t} \pi_r \|\mu_r + \tau\|^2 = \bar{S}_t + 2\tau^\top \bar{\mu}_t + \|\tau\|^2$, and similarly for $\bar{S}_u$, so $\bar{\Delta} \mapsto \bar{\Delta} + 2\tau^\top(\bar{\mu}_t - \bar{\mu}_u) = \bar{\Delta} + 2\tau^\top \bar{D}$. The first numerator of (12) shifts by $-\frac{1}{2}(2\tau^\top \bar{D}) + \tau^\top \bar{D} = 0$. The second numerator: $\bar{D}^\top \mu_r \mapsto \bar{D}^\top \mu_r + \tau^\top \bar{D}$ enters with a plus sign against $+\bar{\Delta}/2$'s minus, and the same cancellation holds. Margins are translation-invariant. □

Table 4: Closed-form balanced errors e_{bal} for a parallel mismatch $\bar{D} = \beta\delta$ at $\gamma = 9$ (aligned value 0.0668), from (15) and its norm-equal degeneration. Norm-unequal pairs degrade strictly and monotonically toward the degenerate value; norm-equal pairs sit exactly on the Bayes value for every $\beta < 1$, the exceptional-set coincidence.

β	0	0.2	0.5	0.8	0.95	1 (degenerate)
norm-unequal pair	0.0668	0.0803	0.2507	0.5000	0.5000	$1/2$
norm-equal pair	0.0668	0.0668	0.0668	0.0668	0.0668	$1/2$

The certificate is recorded because a ledger that failed it would be wrong: batch-internal coordinate choices must not leak into cross-batch errors, and this is the exact sense in which they do not. It also delimits what any batching policy can respond to: only the relative anchors and scalars, equivalently $\bar{D}$ and $\bar{\Delta}$, never absolute placements.

C.4 Worked values and the balanced ledger

The balanced error of a misaligned pair is $e_{\mathrm{bal}} = \frac{1}{2}(e_{s\to r} + e_{r\to s})$. Take the Proposition 13 offsets in force ($\bar{\Delta} = 0$) and a mismatch parallel to the displacement, $\bar{D} = \beta\,\delta$ with $\beta \in [0, 1)$, so the induced rule keeps the Bayes direction with a scaled threshold. Two regimes split exactly. If the pair is *norm-equal* ($\|\mu_r\| = \|\mu_s\|$, equivalently the bisector passes through the origin), then $\delta^\top \mu_s = -\gamma/2$ and $\delta^\top \mu_r = +\gamma/2$, both numerators in (12) equal $\gamma(1-\beta)/2$, and both one-sided errors equal $\Phi(-\sqrt{\gamma}/2)$: the induced rule *is* the Bayes rule, for every $\beta \in [0, 1)$, despite nonzero mismatch. This is the germ of the exceptional set of Appendix E.3. If the pair is norm-unequal, the coincidence breaks: at the representative geometry $\mu_s = 0$, δ along the first axis (so $\|\mu_r\|^2 = \gamma$),

$$e_{\mathrm{bal}}(\beta) \;=\; \tfrac{1}{2}\left[\Phi\!\left(-\frac{\sqrt{\gamma}}{2(1-\beta)}\right) + \Phi\!\left(-\frac{\sqrt{\gamma}\,(1-2\beta)}{2(1-\beta)}\right)\right], \tag{15}$$

strictly increasing from the aligned value toward $1/2$ as $\beta \uparrow 1$. Table 4 evaluates both regimes at $\gamma = 9$; all entries are closed-form evaluations of Φ at the stated arguments, and E-L2-1 re-measures each family through the registered pipeline.

C.5 The catalog: special families in closed form

The ledger's formulas are most useful pulled apart by mismatch direction. Throughout, offsets are applied ($\bar{\Delta} = 0$) unless stated, and we use the representative geometry $\mu_k = 0$, $\mu_v = \delta$. This is a choice of placement, not a normalization: Proposition 12 certifies the raw ledger, but once the Proposition 13 offsets set $\bar{\Delta} = 0$ the cancellation it relies on is gone, so absolute placement enters the corrected ledger and norm-equality is origin-relative. At this placement the pair is norm-unequal ($\|\mu_v\| = \sqrt{\gamma}$, $\|\mu_k\| = 0$) and the exceptional set is $\{0\}$.

Aligned. $\bar{D} = 0$: both one-sided errors $\Phi(-\sqrt{\gamma}/2) = e_{\mathrm{al}}$; the Bayes pair.

Scalar only. $\bar{D} = 0$, $\bar{\Delta} \neq 0$ (before offsets): the pair rule keeps the Bayes direction with threshold shifted by $\bar{\Delta}/2$: $e_{k\to v} = \Phi(-\sqrt{\gamma}/2 + \bar{\Delta}/(2\sqrt{\gamma}))$, $e_{v\to k} = \Phi(-\sqrt{\gamma}/2 - \bar{\Delta}/(2\sqrt{\gamma}))$: a pure calibration tilt, symmetric to first order, removable by the offsets, and the reason composition and Proposition 13(a) are the same kind of move.

Parallel vector. $\bar{D} = \beta\delta$: equation (15) and Table 4; at this geometry the balanced error climbs from e_{al} to $1/2$ as β runs from 0 to 1.

Perpendicular vector. $\bar{D} = t\,u$ with $u \perp \delta$: then $\bar{D}^\top \mu_k = \bar{D}^\top \mu_v = 0$, both numerators are $\gamma/2$, the denominator is $\sqrt{\gamma + t^2}$, and

$$e_{\mathrm{bal}}(t) \;=\; \Phi\!\left(-\frac{\sqrt{\gamma}}{2\sqrt{1 + t^2/\gamma}}\right), \tag{16}$$

Table 5: The perpendicular family (16) at $\gamma = 9$ (aligned value 0.0668): closed-form evaluations; E-L2-1 re-measures three of the five listed t values ($t = 3$ and 6 directly, $t = 0$ through the aligned parallel cell).

t	0	3	6	9	18	$\to \infty$
e_{bal}	0.0668	0.1444	0.2512	0.3176	0.4026	$\to 1/2$

a one-parameter family strictly increasing from e_{al} at $t = 0$ toward $1/2$ as $t \to \infty$, never touching the exceptional set (which is parallel by definition). Table 5 evaluates it.

General direction. Writing $\bar{D} = \beta\delta + t\,u$ with $u \perp \delta$ in the span of $\{\delta, \bar{D}\}$, the two arguments of (12) are $-(\gamma/2)/g$ and $-(\gamma(1/2-\beta))/g$ with $g = \sqrt{(1-\beta)^2\gamma + t^2}$: the parallel and perpendicular families are the two axes of the full two-parameter catalog, and every mismatch is a convex combination of behaviors along them.

Degenerate ray. $\bar{D} \to \delta$: the two one-sided errors race to $\{0, 1\}$ at rates set by the norm difference; the balanced error has direction-dependent limits tabulated in Lemma 23(iii), and its value at the point is exactly $1/2$.

Reversed. $\bar{D} = \beta\delta$, $\beta > 1$: the induced halfspace flips orientation; the balanced error is $\frac{1}{2}[\Phi(-\frac{\sqrt{\gamma}}{2(\beta-1)}) + \Phi(\frac{\sqrt{\gamma}(\beta-1/2)}{\beta-1})]$ at the norm-unequal geometry fixed above, rising toward $\frac{1}{2}[\Phi(0) + \Phi(\sqrt{\gamma})]$ and never reaching $1 - e_{\mathrm{al}}$; the reversed Bayes rule itself, worth exactly $1 - e_{\mathrm{al}}$, is attained only for *norm-equal* pairs, where the whole ray $\beta > 1$ realizes it (Proposition 15(ii)).

C.6 The ledger is everything

Rent aside, every cost object this paper prices is a functional of (12): the guarantee functional of Definition 2 evaluates e_{bal} at the worst admissible $\bar{D}$; the dominance theorem compares e_{bal} across partitions; the mismatch floor minimizes e_{bal} over achievable anchors; and the separation theorem's numerator is $\ell(e_{\mathrm{bal}} - e_{\mathrm{al}})$ at the floor. Each is a different query on the same formula. This completeness is what makes the program falsifiable at the level of one formula family: E-L2-1's grid attacks the ledger directly, and every downstream claim inherits its verdict.

D The Gauge Split

Proposition 13 (Scalar removable, vector not)**.** ***(a)*** *The single class-wise offset vector $o_r = -\frac{1}{2}\bar{S}_t$ for every $r \in B_t$, applied simultaneously across all batches, removes every $\bar{\Delta}$ term from every cross-batch pair rule at once; one constant per class, needing nothing of other batches beyond their compositions.* ***(b)*** *No family of class-wise constant corrections removes the $\bar{D}^\top x$ term: constants move the pair rule's threshold only, while the vector mismatch rotates its direction, $(\delta - \bar{D})$. The composition-proof cost of unanchored writing is exactly the vector field $\{\bar{D}_{tu}\}$.*

Part (a) is one substitution into (1); part (b) is one line. This appendix proves both and scopes class-wise *linear* corrections: a refit, hence another operator.

D.1 Part (a): one offset vector kills every scalar mismatch

Proof of Proposition 13(a). Apply the class-wise constant $o_r = -\frac{1}{2}\bar{S}_t$ to every class r written in batch t. The pair rule (1) between $r \in B_t$ and $s \in B_u$ shifts by $o_r - o_s = -\frac{1}{2}(\bar{S}_t - \bar{S}_u) = -\frac{1}{2}\bar{\Delta}$, which cancels the threshold's $-\frac{1}{2}\bar{\Delta}$ term, leaving the threshold $\frac{1}{2}(\|\mu_r\|^2 - \|\mu_s\|^2)$ exactly. The cancellation is simultaneous over all cross pairs: each class's offset depends only on its own batch's scalar, which the writer knows from its own batch composition, so the correction is well defined batch by batch with no cross-batch coordination. After the correction, the entire residual effect of the frame mismatch on every pair in the ledger is the direction change $(\delta - \bar{D})$. □

D.2 PART (B): CONSTANTS MOVE THRESHOLDS, NOT DIRECTIONS

Proof of Proposition 13(b). Let $\{o_r\}$ be any family of class-wise constants, arbitrary. The pair rule between $r \in B_t$ and $s \in B_u$ becomes $(\delta - \bar{D})^\top x \geq \frac{1}{2}(\|\mu_r\|^2 - \|\mu_s\|^2) - \frac{1}{2}\bar{\Delta} + o_r - o_s$: the constants enter the threshold only. The direction $\delta - \bar{D}$ is untouched by every choice of $\{o_r\}$, and it differs from the Bayes direction δ whenever $\bar{D}_{tu} \neq 0$. Since the group of class-wise constant corrections acts on thresholds alone, and the misalignment's effect on the pair rule is exactly a rotation of the direction (plus a threshold shift, which part (a) already neutralizes), the vector field $\{\bar{D}_{tu}\}$ is invariant under the full correction group: no member of the group removes it, on any pair, ever. □

Remark 4 (The unanchored fit is batch-local, and that is a modeling choice). Under W3 the objective is the cross-entropy of the batch's data over the batch's own labels: the softmax normalizes within the batch and frozen rows do not enter its partition function, which is what makes the frame a property of the batch. The alternative reading, a softmax over all labels written so far with old rows frozen, is a different operator, suppressing old classes and leaving rows outside Lemma 7's family; we do not price it here (Appendix J).

D.3 WHAT A LINEAR CORRECTION WOULD BE

One can imagine class-wise *linear* corrections, $h_r \mapsto h_r + a_r^\top x + o_r$. This is not a correction but a refit: the corrected head's directions have changed, which is the defining property of writing new geometry through an operator (Definition 1).

It is worth being exact about why, because a natural objection lands here and the honest answer is a modeling assumption rather than an impossibility. The population optimum of a batch-local fit is the whole family $w_r = \mu_r + a$, $a \in \mathbb{R}^d$ (a common vector cancels in the batch's own softmax). The member $a = 0$ writes rows in the encoder's absolute coordinates, so *every* batch that selected it would share one frame and no cross-batch mismatch would exist at all. Equivalently, setting a_r to the writer's *own* batch anchor gives $a_r - a_s = \bar{D}_{tu}$ and realigns every cross pair at once, using no foreign information. So the mismatch is not information-theoretically forced; what forces it is which member of the gauge family the fit returns.

Lemma 7's vanishing-ridge selection returns $a = -\bar{\mu}_t$, the count-weighted batch anchor, and that is the convention every result here is stated under. Gradient descent from zero initialization is a *different* selection and we do not claim they agree: what $\sum_r \nabla_{w_r} \mathcal{R} = \mathbb{E}[(\sum_r q_r - 1)X] = 0$ conserves is the *unweighted* sum $\sum_r w_r$, so descent from zero lands on the unweighted centroid, which Lemma 6 identifies with the unweighted penalty's selection and which coincides with $\bar{\mu}_t$ only at equal multiplicities.

The difference is not cosmetic. Lemma 22's achievable-anchor set is the dense set of count-fraction combinations under the count-weighted ridge; under the unweighted gauge it would collapse to the finite set of subset centroids, so a policy could not steer its anchor by multiplicity at all. Every statement here about achievable anchors, the excess floor's infimum included, is made under the count-weighted ridge. Under the unweighted gauge Theorem 4 survives with the *sharper* fixed-composition floor of Remark 11 in place of Ξ, and Lemma 22 as stated does not.

We therefore state the batch-anchored gauge as a hypothesis on the operator (Remark 4 already scopes the objective the same way) rather than as a fact about information, and everything downstream is a theorem about heads produced that way. A writer that deliberately stores absolute rows has left W3 as we price it, and has done so by carrying its own coordinate frame between writes, which is what anchoring buys and what the impossibility is about.

D.4 THE CORRECTION GROUP, FORMALLY

The family of class-wise constant corrections acts on heads by $h_r \mapsto h_r + o_r$ with arbitrary $(o_r) \in \mathbb{R}^C$; this is an additive group $G \cong \mathbb{R}^C$, and its action on pair rules is $c_{rs} \mapsto c_{rs} + o_r - o_s$ with the direction fixed. Three structural facts. (i) The orbit of any pair rule under G is the set of all thresholds for its direction: G acts transitively on thresholds and trivially on directions. (ii) The mismatch vector field $\{\bar{D}_{tu}\}$ is G-invariant, in the strong sense that the map from batch compositions to pair directions factors through the quotient by G; this is the formal content of Proposition 13(b). (iii) The subgroup computable at zero information cost is the one spanned by composition (log-count offsets)

and the scalar offsets of Proposition 13(a), both functions of bookkeeping alone; enlarging the group to anything that moves directions leaves the constant-correction family altogether and changes the head's directions, which is a refit rather than a correction. The taxonomy of Definition 1 is therefore visible, from the correction side, as exactly the boundary between what bookkeeping can fix and what only an anchor can.

D.5 What a full affine correction would be, formally

Allowing $h_r \mapsto h_r + a_r^\top x + o_r$ with per-class vectors a_r is the group $\mathbb{R}^{dC} \times \mathbb{R}^C$; its action on pair rules moves directions, $\delta - \bar{D} \mapsto \delta - \bar{D} + a_r - a_s$. This group can trivially realign any single pair (take $a_r - a_s = \bar{D}$), and equally trivially it can implement any classifier whatsoever, which is the point: its action does not preserve the head's property of being a population fit to data the writer actually has. The vectors a_r that would undo $\bar{D}$ need not depend on any foreign object at all: $a_r = \bar{\mu}_t$, the writer's own anchor, already gives $a_r - a_s = \bar{D}_{tu}$. What the boundary rests on is therefore the gauge selection of Lemma 7, stated as a hypothesis in Appendix D.3, and not on an information argument; the G-invariance of the vector gauge is algebraic, and Proposition 13(b) is a statement about the constant-correction family only.

E Batching Dominance

E.1 Bayes optimality and uniqueness for the balanced pair

Lemma 14 (Unique Bayes rule for a balanced Gaussian pair)**.** *Let $\mu_r \neq \mu_s$ and X drawn from the balanced mixture $\frac{1}{2}N(\mu_r, I) + \frac{1}{2}N(\mu_s, I)$. Among all measurable decision rules, the balanced error $\frac{1}{2}\Pr[\text{say } s \mid r] + \frac{1}{2}\Pr[\text{say } r \mid s]$ is minimized exactly by the likelihood-ratio rule, which for equal covariances and equal priors is the bisector rule*

$$\delta^\top x \;\geq\; \tfrac{1}{2}\big(\|\mu_r\|^2 - \|\mu_s\|^2\big) \quad \textit{decide } r, \tag{17}$$

with minimum value $e_{\mathrm{al}} = \Phi(-\sqrt{\gamma}/2)$. The minimizer is unique up to sets of zero mixture measure, and among halfspace rules with normal parallel to δ the optimal threshold $\frac{1}{2}(\|\mu_r\|^2 - \|\mu_s\|^2)$ is unique.

Proof. Pointwise over x, the balanced error of a (possibly randomized) rule equals $\frac{1}{2}\int \big[q_r(x)\,\pi(\text{say } s \mid x) + q_s(x)\,\pi(\text{say } r \mid x)\big]\,dx$ with q_r, q_s the class densities, minimized pointwise by putting all mass on the argmin of $\min(q_r, q_s)$, i.e., the likelihood-ratio rule; for equal covariance Gaussians $\log(q_r/q_s) = \delta^\top x - \frac{1}{2}(\|\mu_r\|^2 - \|\mu_s\|^2)$ is affine, so the rule is the stated halfspace. The value: under $X \sim N(\mu_s, I)$ the statistic $\delta^\top X$ is $N(\gamma/2, \gamma)$, so $\Pr[\text{say } r \mid s] = \Phi(-\sqrt{\gamma}/2)$, and symmetrically. Uniqueness: any rule differing from the likelihood-ratio rule on a set of positive mixture measure pays strictly more there; the two Gaussians are mutually absolutely continuous, so the boundary $\{q_r = q_s\}$, a hyperplane, is the only tie set and has zero measure under both. Threshold uniqueness along δ: the two one-sided errors are $\Phi(\cdot)$ evaluated at strictly monotone functions of the threshold, and their mean is strictly unimodal (the derivative of the balanced error in the threshold is a nonzero scale mixture of density values). □

Definition 2 (Guarantee functional)**.** Fix a margin envelope $[\gamma_-, \gamma_+]$ with $\gamma_- > 0$. For a batch partition P of a label set with means in the envelope, let $\rho(P) = \max_{t \neq u} \|\bar{D}_{tu}\|$ be its mismatch budget. The guarantee functional is

$$\Psi(\rho) \;=\; \sup\Big\{ \max_{\substack{r \in B_t,\, s \in B_u \\ t \neq u}} e_{\mathrm{bal}}(r, s; \bar{D}_{tu}) \;:\; \text{admissible geometries and partitions with } \rho(P) \leq \rho \Big\}, \tag{18}$$

the worst balanced cross-batch error over configurations whose budget is below ρ.

E.2 The merge arithmetic and the proof of Theorem 1

Proof of Theorem 1. **(a)** After merging t and u into one batch, every pair between them is a within-batch pair, decided by the composed protocol's Bayes bisector rule (the composed protocol of Appendix B; Lemma 14). Before the merge, the same pair was decided by (1), a halfspace rule; by

Lemma 14 its balanced error is weakly larger, strictly larger unless the two halfspaces agree up to a mixture-null set, i.e., unless the pre-merge rule already was the Bayes rule. Appendix E.3 characterizes exactly when a nonzero mismatch still induces the Bayes rule: only for norm-equal pairs with $\bar{D}$ on the ray $\{\beta\delta : \beta < 1\}$ (with $\bar{\Delta}$ corrections applied), a set we call the exceptional set. **(b)** The merged batch's anchor is the count-weighted mean $\bar{\mu}_{(tu)} = \lambda\bar{\mu}_t + (1-\lambda)\bar{\mu}_u$ with $\lambda = n_t/(n_t+n_u)$, so for any third batch v,

$$\bar{D}_{(tu)v} \;=\; \bar{\mu}_{(tu)} - \bar{\mu}_v \;=\; \lambda(\bar{\mu}_t - \bar{\mu}_v) + (1-\lambda)(\bar{\mu}_u - \bar{\mu}_v) \;=\; \lambda\,\bar{D}_{tv} + (1-\lambda)\,\bar{D}_{uv}, \tag{19}$$

a convex combination, whence $\|\bar{D}_{(tu)v}\| \le \max(\|\bar{D}_{tv}\|, \|\bar{D}_{uv}\|)$: the budget $\rho(P)$ never increases under a merge. The merged pair's own mismatch vector disappears (internal pair). **(c)** Ψ is nondecreasing in ρ because the feasible set grows by inclusion. Under a merge, (b) gives $\rho(P') \le \rho(P)$, so the certified bound $\Psi(\rho(P')) \le \Psi(\rho(P))$: no merge ever worsens the certified guarantee; (a) additionally lowers the realized errors of internal pairs. Iterating merges to the one-batch partition gives budget 0 and $\Psi(0) = \Phi(-\sqrt{\gamma_-}/2)$, the aligned envelope over the margin floor: the one-batch write carries the best W3 guarantee. □

E.3 The exceptional set, characterized

Proposition 15 (Exceptional set)**.** *Fix a pair (r, s) across batches (t, u) with the scalar offsets of Proposition 13(a) applied, so the induced rule is the halfspace $(\delta - \bar{D})^\top x \ge \frac{1}{2}(\|\mu_r\|^2 - \|\mu_s\|^2)$ and the Bayes rule is $\delta^\top x \ge \frac{1}{2}(\|\mu_r\|^2 - \|\mu_s\|^2)$. The two rules define the same classifier if and only if either $\bar{D} = 0$, or the pair is norm-equal ($\|\mu_r\| = \|\mu_s\|$, so the common threshold is 0) and $\bar{D} = \beta\,\delta$ for some $\beta < 1$. Consequently:*

1. *for norm-unequal pairs, the balanced error is strictly above e_{al} at every $\bar{D} \ne 0$ except the degenerate point $\bar{D} = \delta$, where the rule is constant and the balanced error is exactly $1/2$;*
2. *for norm-equal pairs, the balanced error equals e_{al} exactly on the ray $E = \{\beta\delta : \beta < 1\}$, is exactly $1/2$ at $\bar{D} = \delta$, and is strictly above e_{al} everywhere else, with values running up to the reversed-rule bound $1 - e_{\mathrm{al}}$ (on the reversed ray $\{\beta\delta : \beta > 1\}$ the balanced error is exactly $1 - e_{\mathrm{al}}$, and generic misorientations land in between);*
3. *in every case $e_{\mathrm{bal}}(\bar{D}) \ge e_{\mathrm{al}}$ for all $\bar{D} \in \mathbb{R}^d$, with equality exactly on E (which collapses to $\{0\}$ for norm-unequal pairs).*

Proof. Two halfspaces with normals g and δ and thresholds c and c^* define the same classifier iff $g = \alpha\delta$ with $\alpha > 0$ and $c = \alpha c^*$ (same oriented scaling). Here $g = \delta - \bar{D}$ and $c = c^* = \frac{1}{2}(\|\mu_r\|^2 - \|\mu_s\|^2)$, so coincidence requires $\bar{D} = (1-\alpha)\delta$ with $\alpha \in (0, 1]$ and $(1-\alpha)\,c^* = 0$: either $\alpha = 1$, giving $\bar{D} = 0$, or $c^* = 0$, the norm-equal case, in which *no* upper bound on α survives and the whole ray $\beta = 1 - \alpha < 1$ coincides. The backward half $\beta < 0$ is easy to miss and is real: both numerators of (12) equal $\gamma(1-\beta)/2$ and $\|\delta - \bar{D}\| = |1-\beta|\sqrt{\gamma}$, so both arguments are $-\sqrt{\gamma}/2$ at every $\beta < 1$, however large $\|\bar{D}\|$ is. This proves the classification; claims (i) through (iii) then follow from Lemma 14 (uniqueness gives $e_{\mathrm{bal}} \ge e_{\mathrm{al}}$ with equality iff same classifier), the constant-rule computation at $\bar{D} = \delta$ (balanced error $1/2$), and the reversed-halfspace observation ($\alpha < 0$ orients the rule against Bayes, and the balanced error of the reversed rule is $1 - e_{\mathrm{bal}}(\bar{D}) > 1/2$). □

Remark 5 (Reading the correction). The source statement of the dominance theorem said merged pairs improve *strictly whenever the mismatch was nonzero*. Proposition 15 shows this is false on E: a norm-equal pair whose batch mismatch happens to be parallel to its own displacement was already at Bayes before the merge, and the merge changes nothing for it. The correction is small (a ray of mismatch vectors, measure zero among mismatches for $d \ge 2$) but it is load-bearing twice: the bystander strictness claim and the mismatch-floor lemma both route through it, and both are stated with it here. Theorem 1's guarantee-form content is untouched: weak improvement holds everywhere.

E.4 The bystander example, computed

Let T, U, V be batches with representative classes t_1, u_1, v_1, each batch having at least two classes placed symmetrically about its anchor (as in Appendix L.4, so Remark 3 is respected and only the

anchors matter below). Take $\bar{\mu}_T = \bar{\mu}_V$, so $\bar{D}_{TV} = 0$, $\bar{\Delta}_{TV} = 0$, every T-versus-V pair at aligned Bayes, and $\bar{\mu}_U$ distant. Merge T and U: the merged anchor is $\bar{\mu}_{(TU)} = \lambda\bar{\mu}_T + (1-\lambda)\bar{\mu}_U$, and

$$\bar{D}_{(TU)V} \;=\; \bar{\mu}_{(TU)} - \bar{\mu}_V \;=\; (1-\lambda)(\bar{\mu}_U - \bar{\mu}_V) \;\neq\; 0 \quad \text{whenever } \bar{\mu}_U \neq \bar{\mu}_V, \tag{20}$$

while the scalar mismatch after the Proposition 13 offsets is removed. Every T-versus-V pair moves from the Bayes rule to a misaligned rule with $e_{\mathrm{bal}} \geq e_{\mathrm{al}}$, strictly unless $(1-\lambda)(\bar{\mu}_U - \bar{\mu}_V)$ lands on that pair's exceptional ray E, a codimension-$(d-1)$ condition on the placement for $d \geq 2$ that fails generically.

Meanwhile the merged T-versus-U pair becomes internal (aligned), and the third-party mismatch vectors convexify: $\|\bar{D}_{(TU)V}\| = (1-\lambda)\|\bar{D}_{UV}\| < \|\bar{D}_{UV}\|$ in this configuration, so the budget falls. But the third parties' realized errors are not thereby guaranteed to improve, because a pair's error depends on its mismatch's direction relative to its own displacement, not on the norm alone, and the convex combination rotates as well as shrinks the vector.

The merge is globally beneficial or not depending on the demand weights of the affected pairs: the bystander damage scales with the T-versus-V query flow, the benefit with the T-versus-U flow. Per-instance dominance is false; batch composition is a first-class decision.

E.5 The batch-partition problem

Without constraints the guarantee-form dominance of Theorem 1 makes the partition problem trivial: merge everything (the budget is then zero). The problem becomes real under the constraints that make batching a decision at all, and we register it in that form. *Offline, capacitated:* given incoming classes with means and a batch-size cap $\bar{m} \geq 2$ (write compute, latency, or staleness budgets), partition into parts of sizes in $[2, \bar{m}]$ minimizing the budget $\rho(P)$, or the guarantee $\Psi(\rho(P))$, or a demand-weighted realized-error functional. Theorem 1(b) makes every merge monotone for the budget, so the cap is what stops the monotone improvement, and the combinatorial content begins: this is a constrained geometric clustering problem, to our knowledge open in this form, both in complexity and in approximation. *Online, coupled:* with arrivals over time, rent accruing on the unbatched, and anchors frozen at write time (Assumption A1), the partition choice couples to the schedule, and the coupled problem is exactly the frontier of Section 8 and Appendix H. The offline capacitated version is a clean open problem; the online coupled version is priced and stated as the mixed-operator problem in Appendix H. Neither is attacked by force in this paper, by design.

E.6 Merge sequences and the budget induction

Corollary 16 (Budget monotonicity along merge sequences)**.** *Let $P_0, P_1, \ldots, P_m$ be any sequence of partitions in which each P_{i+1} is obtained from P_i by merging two batches. Then $\rho(P_0) \geq \rho(P_1) \geq \cdots \geq \rho(P_m)$, and the set of mismatch vectors present at step $i+1$ is contained, up to convex combination with coefficients summing to one over pairs present at step i, in the convex hull of those present at step i.*

Proof. Induction over the sequence using Theorem 1(b) step by step: each surviving mismatch vector after a merge is a convex combination of two vectors present before it, and norms of convex combinations are bounded by the max of the norms combined. □

The corollary is the formal backbone of the claim that *coarser is never worse in guarantee form*: any coarsening reached by merges inherits a nonincreasing budget. It also sharpens the reading of the bystander example: the damage there is realized damage on specific pairs, against a certified bound that did not move upward; the guarantee form and the realized form of dominance simply measure different things, and both are needed.

E.7 The bystander trade-off, demand-weighted

The merge decision is economic, and the example of Appendix E.4 becomes a decision rule once query flows are attached. Let f_{XY} denote the per-round query flow discriminating pairs between batches X and Y. Merging T and U changes per-round flow cost by

$$\Delta \;=\; f_{TU}\big(e_{\mathrm{al}} - e_{\mathrm{bal}}^{TU}\big) \;+\; f_{UV}\big(e_{\mathrm{bal}}^{(TU)V} - e_{\mathrm{bal}}^{UV}\big) \;+\; f_{TV}\big(e_{\mathrm{bal}}^{(TU)V} - e_{\mathrm{al}}\big), \tag{21}$$

with e_{bal}^{TU} the pre-merge cross error on T-versus-U pairs and $e_{\mathrm{bal}}^{(TU)V} = e_{\mathrm{bal}}$ at mismatch $(1-\lambda)(\bar{\mu}_U - \bar{\mu}_V)$. The first term is negative (merged pairs improve, Theorem 1(a)) and the third is positive off the exceptional set (the bystander damage); the second is of either sign, because the convex combination rotates the third-party mismatch as well as shrinking it, and the pair's error depends on bearing. The merge is beneficial iff $\Delta < 0$: batch composition decisions are closed-form calculations on this ledger once flows are estimated, which is the operational content of Example 2 and the reason E-L2-2 registers the sign structure as a falsifiable prediction rather than a qualitative story.

E.8 The capacitated batch-partition problem, formalized

Problem 17 (Capacitated anchor-partition)**.** *Given means $\mu_1, \ldots, \mu_C \in \mathbb{R}^d$ and a capacity $\bar{m} \geq 2$, partition $\{1, \ldots, C\}$ into parts $P_1, \ldots, P_J$ with $2 \leq |P_j| \leq \bar{m}$ minimizing (a) the budget $\rho(P) = \max_{j \neq j'} \|\bar{\mu}_t(P_j) - \bar{\mu}(P_{j'})\|$, or (b) the demand-weighted realized functional $\sum_{j \neq j'} f_{P_j P_{j'}} \max_{r \in P_j, s \in P_{j'}} e_{\mathrm{bal}}(r, s)$.*

Without the capacity the problem degenerates by Corollary 16 (merge everything; budget zero; every pair aligned). With it, the problem is a constrained centroid clustering: the centers are batch anchors (means of part members), the part sizes are constrained to $[2, \bar{m}]$, and the objective couples parts through center differences rather than within-part spread, an unusual structure that standard k-means-style analyses do not cover. We register its complexity and approximation status as open, note that greedy agglomerative merging in order of anchor proximity is a natural heuristic whose budget behavior is monotone by Corollary 16, and observe that the demand-weighted objective (b) is the one an operator should actually optimize, since it is the realized ledger cost. The online coupled version, where arrivals constrain what can be co-batched and anchors freeze at write time, is Appendix H's mixed-operator problem.

F Scheduling: the Reduction and the Sleep Law

F.1 The weighted acknowledgment problem and the reduction

Proposition 18 (Reduction to dynamic TCP acknowledgment)**.** *Identify a round in which a pending item is queried with one round of that item's delay accruing at rate r, its accrued foregone rent with its accumulated weighted waiting, and the write event with the acknowledgment that clears all pending requests at setup K. Under W2 the schedule problem is the dynamic TCP acknowledgment problem of Dooly et al. (2001) at a query-driven accrual rate, into which the classical unit-weight instances embed verbatim and whose amortization Appendix F.4 re-proves: break-even, flushing when accrued rent reaches K, is $(2 + w^{\max}/K)$-competitive for $w^{\max} = r \max_t |Q_t \cap \text{unwritten}|$. Call $w^{\max}/K \to 0$ the* fine-granularity limit*: the constant is 2 there, at most 3 once $w^{\max} \leq K$, and unbounded above that.* No threshold rule is 2-competitive at fixed granularity*: Appendix F.4 shows the threshold-θ rule's worst ratio on the two-phase adversary is at least $\max\{1 + K/\theta,\ 1 + (\theta + w^{\max} - r)/K\}$, with equality when r divides θ, above 2 for every θ once a round carries two queries and $9/4$ for break-even at $w^{\max} = K/2$; Dooly et al. (2001)'s matching lower bound is asymptotic in the same sense, so break-even is optimal exactly where it is 2-competitive. The classical instances embed, so Seiden (2000)'s $e/(e-1)$ transfers as a randomized* lower *bound; Karlin et al. (2001)'s matching upper bound does not, and its natural transfer is refuted by a two-item instance (Appendix H.3). The scope is narrow: for sum-monotone (batch-aware) delay costs break-even is $\Omega(n)$-competitive and the deterministic ratio $\Theta(\log n)$ (Bhore et al., 2026), where we claim nothing.*

Proof of Proposition 18. The dynamic TCP acknowledgment problem with weights: requests arrive over discrete rounds, request j carrying weight w_j; the algorithm may acknowledge at any round, clearing all pending requests at cost K; the objective is $\sum_j w_j \cdot (\text{waiting rounds of } j) + K \cdot (\text{number of acknowledgments})$. Map an instance of our scheduling problem to an acknowledgment instance through the *accrual rate*, not through request arrivals: item i accrues delay at rate $\rho_i(t) = r$ on rounds at which it is queried and 0 elsewhere, so its accrued delay at round T is $\int_{a_i}^{T} \rho_i(t)\,\mathrm{d}t$, and a write event that includes i is the acknowledgment that stops its accrual. This is a nonnegative,

item-additive rate that may vary in time. The alternative gloss, that each query is a request of weight r arriving at its query round, is *not* the right mapping and we do not use it: under the acknowledgment objective such a request would pay r for every round it then waits, whereas Assumption A2 charges a query once, at its own round. The classical unit-weight instances are the constant-rate special case $\rho_i \equiv r$, which is what makes them embed verbatim. This is the linear (additive-in-time) delay objective with time-varying weights, and it is the class the classical analysis covers, since that analysis uses only additivity of delay over items and over rounds and nonnegativity of the rate; it is not the class of batch-aware delay costs, where the caveat below bites. A write event covering several items at the shared setup K is an acknowledgment clearing the corresponding request classes at once. Two boundary observations make the reduction airtight in both directions. First, every unit-weight acknowledgment instance is an instance of our problem restricted to flush-all policies, which is the direction that carries the lower bounds (write every pending item at once), so the classical lower-bound instances embed verbatim: no policy for our problem beats the deterministic bound 2 or the randomized bound $e/(e-1)$ on those instances. Second, the classical break-even policy (acknowledge when accumulated weighted waiting since the last acknowledgment reaches K) is a legal policy for our problem, and its classical analysis, which uses only the linearity of the delay objective and nonnegativity of weights, applies verbatim; hence the guarantees carry. The scope sentence: the classical guarantees are proved for the additive delay objective, and the rent identification reproduces it with a time-varying nonnegative rate, which the same amortization covers; for sum-monotone (batch-aware) delay costs, where the cost of a batch is not the sum of its items' accrued delays, the greedy guarantee fails (Bhore et al., 2026), and our claims are silent there. Randomization: the randomized *lower* bound transfers by the first observation alone, oblivious adversaries in both problems, since the classical instances embed and no policy for our problem can beat on them what no policy for theirs can beat. The matching *upper* bound does not transfer through the second observation: that observation is about break-even, which is deterministic, and Karlin et al. (2001)'s randomized threshold family is analysed at a constant accrual rate. We claim only the lower bound, and Appendix H.3 now shows the caution was warranted: read in the accumulated-delay coordinate, in continuous time so that the granularity term vanishes, that family is not $e/(e-1)$-competitive once the rate varies, on a two-item instance with a three-line closed form. □

Remark 6 (What batching adds over per-item break-even). Without a shared setup, the problem decomposes into independent per-item ski rental, where per-item break-even carries the classical constant 2 in the same fine-granularity limit. The shared setup K is the batching discount: one event covers many items. The acknowledgment framing is exactly the shared-setup problem, and the classical analysis already contains the optimal exploitation of it. What the classical problem does not contain is the quality axis (the geometry of Sections 3 through 5) that reopens under the unanchored operator; that coupling is Section 8's subject.

F.2 The square-root sleep law

Proof of Corollary 3. Run a periodic policy: write all pending items every T rounds. Event cost: K per period, K/T per round. Flow cost: items eligible for entry arrive at exogenous rate α per round (stationary flow approximation); an item arriving at a uniform phase inside the period waits on average $T/2$ rounds before its write, accruing mean rent $\bar{\rho}$ per round; Little-style accounting gives per-round rent flow $\alpha\bar{\rho}T/2$. The total per-round rate $f(T) = K/T + \alpha\bar{\rho}T/2$ is strictly convex in $T > 0$ with $f'(T) = -K/T^2 + \alpha\bar{\rho}/2 = 0$ at $T^* = \sqrt{2K/(\alpha\bar{\rho})}$, the stated law, and $f(T^*) = \sqrt{2K\alpha\bar{\rho}}$. Elasticities: T^* scales by $\sqrt{1/2}$ when K is halved and by $\sqrt{1/2}$ when $\bar{\rho}$ is doubled: each elasticity is exactly $1/2$ in absolute value, with the signs the corollary states. Discrete rounds: rounding T^* to a positive integer changes f by $O(1/T^{*2})$, below measurement resolution for the registered grids. □

F.3 The quoted classical results, stated with their scope

For reference, the two inherited statements exactly as used, credited and scoped. **Deterministic (Dooly et al., 2001):** for the dynamic acknowledgment problem with additive delay and unit request weights, the break-even policy is 2-competitive, and no deterministic policy achieves a ratio below 2 asymptotically in the setup-to-delay ratio. We quote it in that form. The extension our reduction needs, to a nonnegative accrual rate that varies in time, is *ours* and is proved in Appendix F.4; it uses only additivity across rounds and items and exogeneity of the rate, which is exactly where the

sum-monotone caveat bites. **Randomized (Seiden, 2000; Karlin et al., 2001):** against oblivious adversaries, the optimal randomized competitive ratio for the same unit-weight problem is $e/(e-1) \approx 1.582$, the lower bound Seiden (2000)'s and the matching algorithm a randomized threshold family of Karlin et al. (2001), not break-even, which is deterministic; optimality is against the same objective and adversary class, and we claim the transfer only for the classical instances that embed (Appendix F.1), with the same single-round-accrual caveat as the deterministic bound. **The caveat (Bhore et al., 2026):** the break-even greedy is not 2-competitive for arbitrary monotone batch delay costs. For sum-monotone (batch-aware) delays it is $\Omega(n)$-competitive and the deterministic ratio is $\Theta(\log n)$; deterministic 2-competitiveness does survive for max-monotone and for batch-oblivious continuous submodular delays. Our claims are made only for the linear objective, and the scope sentence appears at both use sites (Propositions 18 and 5). We quote these results; the reductions that transport them are ours and are proved in this appendix.

F.4 The classical amortized analysis, included for self-containment

The break-even guarantee is classical (Dooly et al., 2001); we include the two-sided amortization here because our reduction leans on it and because restating it makes the linear-delay scope visible line by line. Upper bound: the break-even policy acknowledges when the weighted waiting accumulated since the last acknowledgment reaches K; each acknowledgment's waiting is then less than K plus the largest single-round accrual $w^{\max} = r \max_t |Q_t \cap \text{unwritten}|$, so the total waiting is at most OPT's waiting plus K per acknowledgment, and comparing acknowledgment counts against OPT's (each OPT acknowledgment can be charged to the break-even acknowledgment that covers its requests) gives the factor $2 + w^{\max}/K$ with $w^{\max} = r \max_t |Q_t \cap \text{unwritten}|$ the largest single-round accrual, Q_t the multiset of round t's queries so that a class queried twice in a round counts twice, hence at most 3 in the fine-granularity regime $w^{\max} \le K$ and 2 in the limit $w^{\max}/K \to 0$ in which one round's accrual cannot by itself pay for an acknowledgment, and $o(1)$ above 2 as $w^{\max}/K \to 0$. The bound is not 2 at fixed granularity even inside the regime, and no threshold rule reaches it, by a two-line argument on the single-group two-phase adversary. Let the rule flush once accrued rent reaches θ, and let the adversary stop the stream at the flush. If it can land the accrual on θ exactly, the rule pays $\theta + K$ against $\min(\theta, K)$; if instead it lets the last round overshoot, the rule pays the accrual at the flush plus K against $\min(\text{accrual}, K)$. The worst ratio of the threshold-θ rule is therefore at least $\max\{1 + K/\theta,\ 1 + (\theta + w^{\max} - r)/K\}$, which exceeds 2 for every θ as soon as $w^{\max} > r$, that is, as soon as one round can carry two queries. *The second branch is an equality only when r divides θ, and we state the hypothesis because the paper carried this expression as exact and it is not.* The overshoot term $\theta + w^{\max} - r$ assumes the round before the flush sat one query short, at accrual $\theta - r$; accruals are sums of per-round amounts that are multiples of r, so $\theta - r$ is reachable exactly when $r \mid \theta$. Off that grid the largest reachable accrual below the threshold is $\lfloor \theta/r \rfloor r > \theta - r$, the flush overshoots further, and the true worst ratio is strictly larger: at $K = \theta = 6$, $r = 4$ the expression gives 2 and the enumeration gives $7/3$, and at $K = \theta = 12$, $r = 5$ it gives 2 against $9/4$. Every use we make of the expression is as a lower bound and survives; what does not survive is the claim, now withdrawn, that break-even attains 2 at $w^{\max} = r$, since the same two instances have $w^{\max} = r$ and exceed 2. The threshold minimizing the expression sits *below* break-even, at the crossing $\theta^* = \frac{1}{2}\big(\sqrt{(w^{\max} - r)^2 + 4K^2} - (w^{\max} - r)\big)$, so on the divisible grid, where the expression is the ratio, the classical rule is not even the best threshold rule; off it the comparison is between two lower bounds and we claim nothing. `verify_numbers.py` certifies the expression against a brute-force search over every stream of at most seven rounds, on a grid that now includes (K, θ, r) triples with $r \nmid \theta$ and checks the equality claim only where the hypothesis holds; the earlier grid fixed $r = 1$, under which every integer threshold is divisible and the check could not have failed. The break-even instance in the paper's units: At $K = 4$, $r = 1$, two classes pending from round 0 and both queried, in four rounds carrying two queries, then one, then two, then one, the accrual runs 2, 3, 5: it passes the threshold without stopping on it, so break-even fires at the fourth round and pays $5 + 4 = 9$ against an optimum of 4, a ratio of $9/4 > 2$ at $w^{\max} = 2 \le K$, inside the bound $2 + w^{\max}/K = 5/2$. Rounds that all carry two queries land on the threshold exactly and give 2 on the nose, which is why the pattern has to be uneven; `l2_mixed.py` simulates both. Outside the regime the failure is coarser still: a single round can carry a policy from below its threshold to more than a full write past it, and the one-item instance of Proposition 21 makes break-even pay $r + K$ against an optimum of K. Lower bound: the classical two-phase family (a request stream that stops just before break-even, followed either by silence or by a flood) forces every deterministic policy to choose between paying setup for nothing and paying rent forever; both branches cost $2 - \epsilon$

times OPT in the limit. Both arguments use the structure of the delay in exactly one place: waiting accumulates additively across rounds and across items, which is where a sum-monotone batch delay objective breaks the accounting (Bhore et al., 2026) and which a time-varying nonnegative accrual rate preserves.

F.5 THE EXACT DISCRETE PERIODIC COST

The fluid derivation of Corollary 3 can be made exact on discrete rounds, which sharpens what the registered test checks. Run the period-T policy on a stationary point process of arrivals with rate α over uniform phases: an item arriving at phase $\tau \in \{0, \dots, T-1\}$ within a period waits $T-1-\tau$ rounds until the flush, so the mean wait is $(T-1)/2$ exactly, the per-period flow is $\alpha T \bar{\rho}\,(T-1)/2$, and the per-round cost is

$$f_T \;=\; \frac{K}{T} \;+\; \frac{\alpha\bar{\rho}\,(T-1)}{2} \;=\; \left[\frac{K}{T} + \frac{\alpha\bar{\rho}\,T}{2}\right] - \frac{\alpha\bar{\rho}}{2}. \tag{22}$$

The bracket is the fluid objective of Corollary 3 and the correction is constant in T, so the two objectives have the *same* minimizer over real periods, $T^* = \sqrt{2K/(\alpha\bar{\rho})}$: the fluid law misstates the discrete cost level by $\alpha\bar{\rho}/2$ per round and the minimizing period not at all. Two hypotheses are carried from here on. First, $c = 0$: writing every arrival once adds the constant rate $c\alpha$ to f_T, which leaves the minimizer alone but does move ratios, and Appendix F.6 records by how much. Second, $T^* \geq 1$, equivalently $2K \geq \alpha\bar{\rho}$; below it f is increasing on $[1, \infty)$, so the constrained minimizer is the shortest legal period $T = 1$; the mean-wait term itself is zero there and nonnegative throughout the legal range. Finally, f_T is convex on $T > 0$ with minimum $f_{T^*} = \alpha\bar{\rho}(2T^* - 1)/2$, so f_{T^*} lower-bounds the cost of every legal integer period; the registered test's argmin target is the better of $\lfloor T^* \rfloor$ and $\lceil T^* \rceil$ on the integer grid (Algorithm 2), which is what E-L2-3 measures against.

Table 6: The reduction dictionary of Proposition 18: the classical unit-weight instances embed into ours verbatim, which carries the lower bounds and the scope sentences; the upper bound needs the time-varying-rate amortization of Appendix F.4.

Dynamic TCP acknowledgment	Consolidation scheduling (anchored)
one round of a pending request's weighted waiting	one query-round of a pending item, at rate r
acknowledgment clearing pending requests (cost K)	write event clearing pending items (cost K)
batch acknowledgment	batched write (shared setup)
additive delay objective	per-query rent accrual, additive in time and items
sum-monotone batch delays	no claim (the caveat)
oblivious request stream	oblivious query stream

F.6 WHAT THE DEPLOYED HEURISTICS COST

Production memory stacks do not run break-even. They run a periodic flush at a period someone chose, a capacity threshold that writes when the store fills, or an event trigger that writes an item the first time it is used. This appendix prices all three against the same optimum, in the anchored regime where the quality axis is closed and only scheduling remains, so that the comparison is with Proposition 18's constant and nothing else. The three propositions are elementary; we state them because the alternative is to assert in an introduction that deployed heuristics are unbounded and never show it.

Five standing conventions, each of which a reader is entitled to see stated. **(i) Batch size.** The size-at-least-two restriction of Assumption A1 is a hypothesis on W3 and is inactive here (Remark 3: singleton writes are exactly the ones that are always anchored), so singleton write events are legal in this appendix and the constructions below use them. This is a real difference from Appendix H's mixed problem and it is worth naming, since both appendices quote Remark 3. There a singleton write is anchored, an anchored singleton needs a frame to anchor to, and no frame exists before the first write (an anchored pair may open the store, a lone class may not), so the *first* write can never be a singleton and a one-class instance forces every policy to rent, which is exactly why the single-item ski rental embedding is unavailable there. Here the deployed heuristics are the object of study and they do write singletons, so we price them: the one-item instances below are read as instances of

the heuristic's own problem, in which a singleton first write costs $K + c$, and no claim about the mixed problem is made from them. **(ii) Within a round, writes precede queries** (Appendix G.1), which is the convention the runners implement, so an item written at round t is already written for round t's queries, and a policy that writes only in reaction to a query pays that query's rent. Under the opposite convention the trigger's ratio below is unchanged (the write it reacts to still follows the query) and the capacity threshold's becomes $r(\bar{m}-1)Q/(K+r(\bar{m}-1))$, the optimum's round-zero write now following that round's queries; neither conclusion changes. **(iii) The per-item price is** $c = 0$, as in Section 7 and as in every canonical run of the registered program. Carrying $c > 0$ moves the three ratios in known directions and kills none of them. In Proposition 19 the exact ratio becomes $(T^*(x+x^{-1})-1+\kappa)/(2T^*-1+\kappa)$ with $\kappa = 2c/\bar{\rho}$ the per-item price measured in rounds of rent, still unbounded in x at every κ (and the $T = 1$ corner becomes $(T^{*2}+\kappa)/(2T^*-1+\kappa)$, still unbounded in T^*), though damped toward 1 as $\kappa \to \infty$. In Proposition 20 the optimum's cost becomes $K + c(\bar{m}-1)$ and the ratio is still unbounded in the stream length. In Proposition 21 the policy pays $r + K + c$ against an optimum of $\min(r, K + c)$ and the ratio only grows. **(iv)** $T^* \geq 1$, per Corollary 3 and Appendix F.5. **(v) Fine granularity:** $w^{\max} \leq K$, where $w^{\max}$ is the largest rent a single round can accrue, so one round's accrual cannot by itself pay for a write. The condition is on the round's accrual and not on r alone: a round in which many pending classes are queried accrues $r|Q_t \cap \text{unwritten}|$, and $r \leq K$ does not bound that. Inside this regime the guarantee of Proposition 18 is at most 3; it is exactly 2 only in the fine-granularity *limit* $w^{\max}/K \to 0$, and the condition $w^{\max} \leq K$ does not by itself deliver the classical constant. What it delivers is a bound: *every* threshold rule overshoots its threshold by up to a full round's rent, break-even included, so the guarantee is $2 + w^{\max}/K$ (Appendix F.4), which convention (v) keeps below 3 and which grows without bound outside it. Convention (v) is therefore a hypothesis on the comparison, not a convenience: it is what makes the last row of Table 7 a genuine separation from the first three rather than a statement about round granularity.

Proposition 19 (A mis-set period, exactly)**.** *In the stationary regime of Corollary 3 under the conventions above, write* $T^* = \sqrt{2K/(\alpha\bar{\rho})}$*, let the periodic policy run at any period* $T \geq 1$*, and set* $x = T/T^*$*. Its per-round cost relative to the cost at* T^* *is exactly*

$$\frac{f_T}{f_{T^*}} \;=\; \frac{T^*\big(x + x^{-1}\big) - 1}{2\,T^* - 1}, \tag{23}$$

with no approximation. In particular the ratio grows without bound as $x \to \infty$*; at the shortest legal period* $T = 1$ *it equals* $T^{*2}/(2T^* - 1)$*, which grows without bound in* $K/(\alpha\bar{\rho})$*; and it equals* 1 *only at* $x = 1$*. The reference* f_{T^*} *is the minimum of* f *over real periods, so it lower-bounds the cost of every legal period; the best legal period costs at most* $2f_{T^*}$*, and against that reference the first two conclusions survive (the third becomes "equal to* 1 *only when* T^* *is itself an integer and* $T = T^*$*"), with a decade of mis-setting, taken over real periods* $T \geq 1$*, costing at least* 5.05 *times the best integer period at every* $(K, \alpha, \bar{\rho})$ *with* $T^* \geq 1$*, in either direction wherever the mis-set period is itself legal (upward always; downward once* $T^* \geq 10$*). A period chosen without knowledge of* $(K, \alpha, \bar{\rho})$ *therefore has no competitive constant uniform in them, in either direction of mis-setting.*

Proof. By Appendix F.5 the exact discrete per-round cost is $f_T = K/T + \alpha\bar{\rho}(T-1)/2$. Put $A = \sqrt{K\alpha\bar{\rho}/2}$ and $B = \alpha\bar{\rho}/2$, so that $T^* = A/B$ and, substituting $T = xT^*$,

$$f_T \;=\; \frac{K}{xT^*} + \frac{\alpha\bar{\rho}\,xT^*}{2} - B \;=\; A\big(x^{-1} + x\big) - B, \qquad f_{T^*} \;=\; 2A - B. \tag{24}$$

Dividing and using $A = B\,T^*$ gives (23). The three consequences read off it: the numerator is unbounded as $x \to \infty$; at $x = 1/T^*$, that is $T = 1$, it is $B(1 + T^{*2}) - B = B\,T^{*2}$ against $B(2T^* - 1)$, and $T^{*2}/(2T^* - 1) \geq T^*/2$; and $x + x^{-1} \geq 2$ with equality only at $x = 1$. For the legal reference, f is convex on $T > 0$ with real minimizer T^*, so $f_{T^*} \leq \min_{T \in \mathbb{Z}_{\geq 1}} f_T \leq f_{\lceil T^* \rceil}$, and (23) at $x = \lceil T^* \rceil / T^* \leq 1 + 1/T^*$ gives $f_{\lceil T^* \rceil}/f_{T^*} \leq 1 + 1/\big(T^*(2T^* - 1)\big) \leq 2$, so substituting the integer reference costs at most that factor. The 5.05 is (23) at $x = 10$ (equivalently at $x = 1/10$, since $x + x^{-1}$ is symmetric under $x \mapsto x^{-1}$), which decreases in T^* to the limit $10.1/2$; the integer reference is at least f_{T^*}, so the printed ratio overstates the ratio against it; what restores the bound is the rounding penalty, $\min_n f_n \leq f_{\lceil T^* \rceil} = f_{T^*}(1 + \eta)$ with $\eta \leq 1/(T^*(2T^* - 1))$, against a decade ratio of $5.05 + 4.05/(2T^* - 1)$. The requirement is $5.05\,\eta \leq 4.05/(2T^* - 1)$, that is $\eta \leq 4.05/\big(5.05(2T^* - 1)\big)$, which the available bound delivers for $T^* \geq 1.247$; below that the

algebra does not close and the claim rests on the record rather than on the chain. The numeric record of account (`verify_numbers.py`, Appendix J) confirms the minimum 5.0500 directly over a grid of 3504 cost tuples (the ratio's largest value on the sweep is 8.63, at the smallest legal T^*, where (23) is steepest), and the infimum 5.0500 is approached as $T^* \to \infty$. □

Proposition 20 (The capacity threshold has no competitive constant)**.** *Fix a capacity $\bar{m} \geq 2$ and let the policy write when the number of pending items reaches $\bar{m}$, and at no other time: no age trigger, no timeout, no end-of-stream flush. For every C there is an instance on which it pays more than C times the optimum, with all cost constants held fixed. (An age or timeout trigger degrades the policy to the periodic one of Proposition 19, whose ratio is finite at fixed constants and unbounded over them; an end-of-stream flush does not rescue it, since the rent is already sunk.)*

Proof. Let $\bar{m}-1 \geq 1$ items arrive at round 0 and let every one of them be queried in every round for Q rounds. The pending count never reaches $\bar{m}$, so the policy never writes and pays $r(\bar{m}-1)Q$. The optimum pays at most K, by writing all $\bar{m}-1$ as one event at round 0 before that round's queries are served (a singleton event when $\bar{m} = 2$, legal by convention (i)), with no rent and, under anchoring, no misalignment excess thereafter. The ratio is therefore at least $r(\bar{m}-1)Q/K$, which exceeds any C once Q is large, at fixed $r, K, \bar{m}$. The failure is not a badly chosen capacity: any fixed $\bar{m}$ has its own starving instance, one item short of the threshold. □

Proposition 21 (The write-on-first-query trigger has no competitive constant)**.** *Let the policy write an item the first time it is queried. On the one-item instance whose item arrives at round 0 and is queried once, at round 1, it pays $r + K$ against an optimum of $\min(r, K)$, so when $r < K$ its competitive ratio is at least $1 + K/r$, unbounded in the setup-to-rent ratio. The failure is the mirror of the capacity threshold's, which starves; the trigger buys. Without convention (v) the same instance gives $1+r/K$ and break-even pays exactly what the trigger pays, which is the separation-destroying regime convention (v) excludes and not a second failure of the trigger.*

Proof. One item arrives at round 0, is queried at round 1, and is never queried again. The trigger cannot act before the query it reacts to, so by convention (ii) that query pays r and the write costs K: $r + K$ in total. The optimum either writes at round 1 before the query is served, for K, or never writes, for r; it pays $\min(r, K)$, and under anchoring the rented round carries its rent and no misalignment excess. The ratio is $(r + K)/\min(r, K) = 1 + \max(K/r, r/K)$, which under convention (v) is $1 + K/r$ and is unbounded as $K/r \to \infty$. The break-even rule of Proposition 18 rents on this instance whenever $r < K$ (on a one-query round $w^{\max} = r$, so convention (v) holds here), its accrued rent never reaching the threshold, and so pays the optimum exactly; at the boundary $r = K$ it writes and pays $2K$ against an optimum of K, a ratio of exactly 2. What is uniform in K and r throughout the regime of convention (v) is its *bound* $2 + w^{\max}/K$, at most 3 there and exactly 2 in the limit; that is the whole content of waiting until the accrued rent has paid for the write, and it is what the trigger rule, whose ratio grows without bound, does not have. At $r > K$ break-even also writes here and also pays $r + K$: no threshold rule is 2-competitive when one round of rent outruns a write, which is why convention (v) is stated. □

Remark 7 (What the three failures have in common). Each heuristic fixes, in advance, a quantity the problem's own constants determine: the period fixes the setup-to-rent trade, the capacity fixes it in units of items rather than of accrued rent, and the trigger fixes it at zero. The break-even rule fixes nothing in advance; it measures accrued rent against the setup it would pay and acts when they meet, which is why its constant is $2 + w^{\max}/K$, uniform in K and r, for the additive delay objective of Proposition 18 at $c = 0$; it is the single-round accrual $w^{\max}$, not the arrival process as such, that the constant still sees. The three failures are not the same kind of failure, and Table 7 keeps them apart: only the capacity threshold is unbounded at *fixed* cost constants, while the period and the trigger are finite at every fixed tuple and unbounded only as the tuple varies. Corollary 3 is the stationary special case in which the meeting point can be computed once and scheduled, and Proposition 19 is the price of computing it wrong.

F.7 Stochastic refinements

The flow approximation treats arrivals as a stationary fluid. Under Poisson arrivals the exact cost of a periodic policy is the same expression plus an $O(1)$ rounding and discretization term, and the

Table 7: The deployed heuristics, priced against the same optimum in the anchored regime at $c = 0$ and fine granularity $w^{\max} \le K$ (conventions (iii) and (v)); the condition is on the round's accrual, not on r alone. Only the last row is competitive. The middle column is the ratio to the best legal period (first row) or to OPT (the rest); the last column separates failure at fixed costs from failure only over the cost class.

Policy	Ratio	Failure mode
Periodic flush at fixed T	$\big(T^*(x + x^{-1}) - 1\big)/(2T^* - 1)$ exactly, $x = T/T^*$	unbounded as $x \to \infty$, and $T^{*2}/(2T^* - 1)$ at $T = 1$: finite at fixed costs, no constant uniform in $K, \alpha, \bar{\rho}$
Capacity threshold $\bar{m}$	unbounded at fixed costs	starves one item below the threshold and never writes
Write on first query	$\ge 1 + K/r$	buys before the rent has paid for the purchase; no constant uniform in K/r
Break-even (Prop. 18)	$2 + w^{\max}/K$, hence 2 as $w^{\max}/K \to 0$; $e/(e-1)$ a randomized lower bound; the matching upper bound fails at time-varying rate (App. H.3)	none for the additive delay objective at fine granularity; $\Omega(n)$ for sum-monotone batch delays (Bhore et al., 2026)

exact optimal policy for the stochastic shared-setup problem is the acknowledgment policy under the reduction of Proposition 18; the classical literature owns that refinement, and we claim only the fluid law and its elasticities. The registered E-L2-3 tests the law directly: locate the empirical cost-minimizing period by grid search against (3) across a $(K, \alpha, \bar{\rho})$ grid, with displacement of the argmin beyond the grid resolution as the falsifier.

G The Competitive Frontier

G.1 The online problem, fixed precisely

An *instance* is a finite class set with means $\{\mu_j\} \subset \mathbb{R}^d$ chosen by an oblivious adversary, and a stream: a sequence of queries, each labeled by the class queried. A *policy* observes the stream round by round; at any round it may perform a write event on any set B of pending classes (arrived and not yet written) of size at least two (Assumption A1, a restriction on W3 that is inactive in the anchored regime by Remark 3), paying $K + c|B|$; written rows are held thereafter. An anchored write of two or more classes may be the first event, in which case it is a joint fit priced at K_2 and establishes the frame; an anchored singleton needs a frame to anchor to, so it is legal only once some class has been written (Remark 3). Within a round, arrivals precede writes and writes are performed before queries are served, so a class written at round t is already written for round t's queries; this is the convention the runners implement. A class's mean is revealed when the class arrives, so a policy's behaviour up to round t is a function of the prefix through round t alone.

Two adversary classes appear in this paper and every result names the one it uses, because against randomized policies they are not the same. The *oblivious* adversary fixes the whole instance, geometry and stream, in advance; Theorem 4 is stated against it, and its randomized half is a statement about expectations over an instance chosen before the policy's coins. The *adaptive-online* adversary may in addition choose a later arrival's placement as a function of the realized prefix; Appendix H.2's lower bound uses that freedom and is stated for deterministic policies only, where the two classes coincide because the prefix is a function of the instance. No randomized bound in this paper is claimed against the adaptive adversary.

Query charges: a query to an unwritten class pays rent r; a query to a written class pays the pairwise excess flow, ℓ times $(e_{\mathrm{bal}} - e_{\mathrm{al}})$ on each pair the query discriminates (the query's class against every other written class), with offsets applied so $\bar{\Delta} = 0$; internal pairs and anchored pairs have $e_{\mathrm{bal}} = e_{\mathrm{al}}$ and pay nothing.

Three remarks on the convention. First, what the netting removes is the aligned Bayes error, which both stores pay on a covered pair at population level (Proposition 10) and which no operator can avoid: it is the price of the classification problem, not of the protocol. The rent of Appendix A.1

is already defined net of it, and the flow charge is netted the same way so that the two halves of the objective are in the same units. We state precisely what this does and does not do. It is not an additive constant across policies: the number of baselines removed from a query is the number of classes the policy has written, which is a policy decision, so a policy that has written more classes has more removed. In the construction below the asymmetry runs in the policy's favour, since at the flood the policy has written four classes and the optimum two, so the lower bound we prove is conservative; and in the anchored regime the netting removes exactly the terms that Proposition 18's objective does not contain, which is what makes the reduction exact rather than approximate. A reader who prefers the gross objective may read Theorem 4 through the weaker convention-free statement recorded in Remark 12. Second, the *gross* convention, which charges the full balanced error to written pairs and nothing to queries served from memory, is not a competitive-analysis normalization: it lets an optimum avoid a baseline that every policy pays by co-batching a flooded class with a far, never-queried partner at negligible internal error, and it charges the anchored operator a geometry-dependent flow that the scheduling reduction of Proposition 18 does not contain. The registered experiment E-L2-4 was specified under the gross convention and its falsifier fired for exactly this reason (Section 9); the theorem below is stated and proved under the net convention, and the post-hoc program E-L2-4b measures it. Third, charging the balanced error per discriminating query is the balanced-prior evaluation convention of the pairwise ledger, and it is load-bearing: the one-sided convention recorded in Remark 9 does *not* yield a floor of the same shape at the construction below, and we withdraw any suggestion that it does. The infimum Ξ is approached through frames that are anti-aligned in the balanced sense but answer every query to the flooded class correctly: steering the batch composition toward $\alpha = \beta \to 0$ drives e_{bal} to $1/2$ while the flooded class's own one-sided error goes to zero, so under a one-sided charge the flood is free and the separation collapses to a bounded ratio. What diverges is the balanced-prior bill, which is the accounting this paper prices; Remark 11's fixed-composition rate does not have this property (at unit multiplicities the one-sided excess is a genuine 0.18 at the registered placement), which is a further reason to read the measured program against it.

The objective is total cost. OPT is the offline optimum over all schedules and batch compositions within the same operator class and cost model; the *competitive ratio* of a policy is the supremum over instances of $\mathrm{cost}/\mathrm{OPT}$, expectations for randomized policies against oblivious streams. A *pure W3 policy* satisfies Assumption A1.

Two cost-model consequences are used constantly. First, costs are nonnegative, so conditional lower bounds survive conditioning. Second, the excess flow on a query is at least the excess of any single one of its pairs, so single-pair floors suffice for lower bounds; conversely, OPT's cost is bounded above by exhibiting one strategy, and the strategy that matters writes the flooded class together with the class it is flooded against, an internal pair with zero excess.

G.2 Anchor confinement

Lemma 22 (Anchor confinement)**.** *A W3 batch's anchor is a convex combination of its members' means; over all batches drawable from a pending pool, the achievable anchors are the count-fraction convex combinations, dense in the convex hull of the pool's means; up to closure the achievable-anchor set is a compact convex set fixed by the placements and independent of the policy.*

Proof of Lemma 22. A batch's anchor is its count-weighted mean $\bar{\mu}_t = \sum_{r \in B_t} n_r \mu_r / \sum_r n_r$, a convex combination of its members' means with weights the count fractions; that convex combinations lie in the convex hull is the definition. Conversely, every convex combination with rational weights is realized by integer counts, and the rational-weighted means are dense in the hull; the achievable-anchor set of a pool is therefore dense in $\mathrm{conv}\{\mu_j : j \text{ pending}\}$, and its closure is that hull, a compact convex set determined by the placements alone. The denseness matters for adversarial constructions (the adversary may reason with the closed hull), and the exactness matters for policies: whatever batch a policy assembles from the pending pool, and however it splits or delays arrivals to steer anchors, the realized anchor is one of these convex combinations, never anything else. □

G.3 THE MISMATCH FLOOR, IN FULL

Throughout, fix the pair (v, k), write $\delta = \mu_v - \mu_k$, $\gamma = \|\delta\|^2 > 0$, and let $e_{\mathrm{bal}}(\bar{D})$ denote the balanced error of the induced pair rule at mismatch $\bar{D}$ with $\bar{\Delta} = 0$, a function $\mathbb{R}^d \to [0, 1]$ given by (12), with $e_{\mathrm{bal}}(\delta) := 1/2$ (the constant rule). Recall $e_{\mathrm{al}} = \Phi(-\sqrt{\gamma}/2)$.

Lemma 23 (Regularity of the floor function)**.** ***(i)*** *e_{bal} is continuous on $\mathbb{R}^d \setminus \{\delta\}$.* ***(ii)*** *$e_{\mathrm{bal}}(\bar{D}) \geq e_{\mathrm{al}}$ for every $\bar{D} \in \mathbb{R}^d$, with equality exactly on the exceptional set E of Proposition 15 ($E = \{0\}$ for norm-unequal pairs, $E = \{\beta\delta : \beta < 1\}$ for norm-equal pairs, the ray including its backward half $\beta < 0$, not the segment); in particular $e_{\mathrm{bal}}(\delta) = 1/2 > e_{\mathrm{al}}$.* ***(iii)*** *Every directional limit of e_{bal} at δ is at least e_{al}; for norm-equal pairs and u on the sphere it is $\lim_{t\downarrow 0} e_{\mathrm{bal}}(\delta + tu) = \frac{1}{2}\big[\Phi(-u^\top\mu_k) + \Phi(u^\top\mu_v)\big] \geq e_{\mathrm{al}}$, and for norm-unequal pairs it is $\frac{1}{2}$ in every direction, uniformly in the direction, so that e_{bal} extends continuously to δ there. Under equal norms equality with e_{al} is possible in the limit, at $u = -\delta/\|\delta\|$, which is why case (ii) of Lemma 24 needs the ray removed and case (i) does not.*

Proof. (i) On $\mathbb{R}^d \setminus \{\delta\}$ the formulas (12) are quotients of continuous functions with nonzero denominator $\|\delta - \bar{D}\|$. (ii) The induced rule at $\bar{D}$ is a halfspace rule; by Lemma 14 its balanced error is at least the Bayes value e_{al}, with equality iff it defines the same classifier as the Bayes rule, which Proposition 15 characterizes as exactly E; at $\bar{D} = \delta$ the rule is constant and the balanced error is $1/2$. (iii) Substitute $\bar{D} = \delta + tu$ into (12): with $w = tu$, the numerators are exactly $w^\top\mu_k$ and $-w^\top\mu_v$ when the pair is norm-equal, and in general differ from those by the constant $\frac{1}{2}(\|\mu_v\|^2 - \|\mu_k\|^2)$; dividing by $\|w\| = t$ and letting $t \downarrow 0$ gives $+\infty$ or $-\infty$ (one-sided errors racing to $\{0, 1\}$, balanced limit $1/2$) when the norms are unequal, and the stated finite limit when they are equal. In the finite case the limit equals $\frac{1}{2}[\Phi(-u^\top\mu_k) + \Phi(u^\top\mu_v)]$, and since $e_{\mathrm{bal}}(\delta + tu) \geq e_{\mathrm{al}}$ for every $t > 0$ by (ii) and the function of t is continuous on $(0, \infty)$, the limit is at least e_{al}. (The same continuity-along-rays argument proves the claim in the unequal-norm case, where the limit is $1/2$.) □

Lemma 24 (Mismatch floor)**.** *Fix a target pair (v, k) with $\delta = \mu_v - \mu_k$ and compact achievable-anchor sets A_1 (containing v's batch anchor) and A_2 (containing k's), and write $\phi_{\mathrm{bal}} = \min_{(a_1,a_2)\in A_1\times A_2} e_{\mathrm{bal}}(v, k;\, a_1 - a_2)$. Let $E^\sharp$ be the* closed *exceptional set: the single point $\bar{D} = 0$ when the pair is norm-unequal, and the closed ray $\{\beta\delta : \beta \leq 1\}$, which carries 0, the exceptional ray of Appendix E.3 and the degenerate point δ, when it is norm-equal. If $A_1 - A_2$ is disjoint from $E^\sharp$, then $\phi_{\mathrm{bal}} > e_{\mathrm{al}} = \Phi(-\sqrt{\gamma_{vk}}/2)$ and the minimum is attained. The two cases are not interchangeable: in the first, disjoint anchor sets are the whole hypothesis, e_{bal} extending continuously through δ with value $\frac{1}{2}$; in the second it does not extend, $E^\sharp$ is unbounded, and no separation stated in norms implies the hypothesis, however large (Remark 8).*

Proof of Lemma 24 (mismatch floor). Let $K = A_1 - A_2 = \{a_1 - a_2\}$, compact as the image of a compact set under a continuous map. Both cases are the same compactness argument on a domain where e_{bal} is continuous and strictly above e_{al}; they differ in how much has to be removed to get there.

Case (i), norm-unequal. By Lemma 23(iii) the limit of e_{bal} at δ is $\frac{1}{2}$ along every direction, and the convergence is uniform in the direction. Writing $w = \delta - \bar{D}$, the two numerators of (12) are $c - w^\top\mu_k$ and $-c + w^\top\mu_v$ with $c = \frac{1}{2}(\|\mu_v\|^2 - \|\mu_k\|^2) \neq 0$, while the denominator is exactly $\|w\|$; each numerator differs from $\pm c$ by at most $\|w\| \max(\|\mu_v\|, \|\mu_k\|)$, so the two arguments race to $\pm\infty$ together and the balanced value tends to $\frac{1}{2}$ regardless of direction. Setting $e_{\mathrm{bal}}(\delta) := \frac{1}{2}$ therefore extends e_{bal} continuously to all of $\mathbb{R}^d$. By Lemma 23(ii) the extension is pointwise $\geq e_{\mathrm{al}}$ with equality exactly on $E = \{0\}$, and $0 \notin K$ because $A_1 \cap A_2 = \emptyset$. A continuous function on a compact set attains its minimum, and here that minimum is strictly above e_{al}. Note what this does *not* require: K may contain δ, and in the applications of Appendix G.6 it always does.

Case (ii), norm-equal. Now $E = \{\beta\delta : \beta < 1\}$ is a ray, and e_{bal} genuinely fails to extend continuously to δ: Lemma 23(iii) gives the finite direction-dependent limit $\frac{1}{2}[\Phi(-u^\top\mu_k) + \Phi(u^\top\mu_v)]$, which equals e_{al} at $u = -\delta/\|\delta\|$. The hypothesis removes the whole closed ray, so K has positive distance from it, e_{bal} is continuous on an open neighbourhood of K, and it is pointwise $\geq e_{\mathrm{al}}$ there with equality only on E, which K avoids. Compactness again gives an attained minimum strictly above e_{al}. □

Remark 8 (Why the strengthened hypothesis, and why it cannot be dropped). The source sketch assumed only $0 \notin A_1 - A_2$. That is not enough once finding A-11 is on the table: if K meets the exceptional ray E, the floor is e_{al} *exactly* (attained, not merely approached), because on E the induced rule is the Bayes rule despite nonzero mismatch, and the separation mechanism dies for that pair. And if K approaches δ arbitrarily closely without containing it, e_{bal} is discontinuous at δ, the minimum over K need not be attained, and the infimum can sit at e_{al} (Lemma 23(iii): the direction $-\delta/\|\delta\|$ realizes the equality limit under equal norms). Norm-equality is read in the gauge the Proposition 13 offsets fix, as in Proposition 15; it is not a translation-invariant property of the pair. Disjointness from the whole closed ray is the clean hypothesis that makes compactness bite *for norm-equal pairs*; for norm-unequal pairs both objections evaporate, E collapses to $\{0\}$ and e_{bal} extends continuously through δ, which is case (i) of Lemma 24. A separation stated in norms cannot substitute for the ray hypothesis in the norm-equal case: the exceptional set is unbounded, so at $\mu_v = (1, 1.7)$, $\mu_k = (-1, 1.7)$ (norm-equal, $\delta = (2, 0)$) the singleton sets $A_1 = \{(-2B, 0)\}$, $A_2 = \{0\}$ are separated by $2B$ for every B and still give $\phi_{\mathrm{bal}} = e_{\mathrm{al}}$ exactly. An earlier form of this lemma asserted that a separation margin exceeding $\|\delta\|$ satisfies the hypothesis automatically; this instance is its counterexample, and that clause is withdrawn. In the norm-equal case compactness plus ray-disjointness is what bites; in the norm-unequal case compactness plus disjointness of the anchor sets is already enough.

Remark 9 (The one-sided convention). Under the one-sided convention (a query to k pays $\ell\, e_{k\to v}$ only), the balanced floor does not directly apply, because a one-sided error can sit below e_{al} even when the balanced error is above it. The translation normalization repairs this: by Proposition 12 the adversary may translate the world so $\mu_k = 0$, and then $e_{k\to v}(\bar{D}) = \Phi\big(-(\gamma/2)/\|\delta - \bar{D}\|\big)$ depends on $\bar{D}$ only through the distance to δ. It equals e_{al} exactly on the sphere $\|\delta - \bar{D}\| = \sqrt{\gamma}$ (which contains the aligned anchor $\bar{D} = 0$), exceeds e_{al} outside the closed ball of radius $\sqrt{\gamma}$ around δ, is below e_{al} inside that ball, and climbs to $1/2$ as $\|\delta - \bar{D}\|$ grows; in particular it is at least $\Phi(-\gamma/(2s))$ whenever $\|\delta - \bar{D}\| \geq s$. Separated constructions with $s > \sqrt{\gamma}$ therefore enjoy a one-sided floor $\Phi(-\gamma/(2s)) > e_{\mathrm{al}}$ with the same saturation to $1/2$ *whenever the hypothesis* $\|\delta - \bar{D}\| \geq s$ *holds*. It does not hold at the two-phase construction: at unit multiplicities $\|\delta - \bar{D}\| = \frac{1}{2}\sqrt{\gamma + s^2}$, which is below s once $s > \sqrt{\gamma/3}$, and at the registered placement $(\gamma, s) = (9, 12.9)$ it is 6.62 against 12.9. Worse, the policy can steer: driving the batch composition to $\alpha = \beta \to 0$ sends e_{bal} to $1/2$, which is where Ξ is approached, while sending $e_{k\to v}$ to *zero*. Under the one-sided charge the flood would then be free and Theorem 4 would collapse to a bounded ratio. We therefore claim nothing for the one-sided convention at the construction, and record that the divergence is a property of the balanced-prior accounting the paper prices, not of the service the flooded class receives: at the infimum the head answers every query to k correctly, and it is $e_{v\to k}$, on a class never queried in stage two, that carries the $1/2$. Remark 11's fixed-composition rate is free of this: at unit multiplicities the one-sided excess there is a genuine 0.18.

G.4 The two-phase construction and the proof of Theorem 4

Placement. The adversary fixes four classes. The flooding pair (v, k) sits at margin γ; normalize by translation so $\mu_k = 0$ and $\mu_v = \delta$ with $\|\delta\| = \sqrt{\gamma}$. Choose a unit vector $u \perp \delta$ and any separation $s > 0$, and place the partners

$$\mu_w \;=\; \mu_v + su \qquad \text{(far partner)}, \qquad \mu_p \;=\; 2\mu_k - \mu_v \qquad \text{(mirror partner: } \|\mu_k - \mu_p\| = \sqrt{\gamma}\text{)}. \tag{25}$$

The partners' role is to fix what a pure-W3 policy's frames can be, and nothing else. The far partner w is the only class pending with v in stage 1, so a policy that writes v there writes exactly the pair $\{v, w\}$, at an anchor $a_1 = \mu_v + \alpha s u$ with $\alpha \in [0, 1]$ the count fraction of w (Lemma 22); with unit multiplicities $\alpha = \frac{1}{2}$. The mirror partner p is the only class pending with k in stage 2 once v and w are frozen, so the policy's stage-2 write is exactly $\{k, p\}$, at an anchor $a_2 = -\beta\delta$ with $\beta \in [0, 1]$ the count fraction of p. Under the net convention the placement of the partners costs the optimum nothing wherever they sit, since the optimum never writes them; what the placement buys the adversary is that both of the policy's frames are pinned to one-parameter families that cannot reach each other. The resulting difference $\bar{D} = a_1 - a_2 = (1 + \beta)\delta + \alpha s u$ always has a δ-component of length at least $\|\delta\|$ and, whenever $\alpha > 0$, a component off the line of δ; at unit multiplicities $\|\bar{D}\| \geq \frac{3}{2}\|\delta\|$ with margin to spare. Two remarks on which case of Lemma 24 this is. The construction normalizes μ_k to the origin at distance $\sqrt{\gamma} > 0$ from μ_v, so the target pair is

norm-unequal and case (i) applies: what is needed is that the two anchor families be disjoint, which the δ-component of length at least $\|\delta\|$ gives. And α is strictly positive, w carrying at least one exemplar, so the composition never reaches the closure point $\alpha = \beta = 0$ at which $\bar{D}$ would equal δ; under case (i) even that point is harmless, since e_{bal} extends there with value $\frac{1}{2}$. The floor at the placement is exact, and it does not depend on the composition:

Lemma 25 (Frozen-frame floor at the placement)**.** *At the placement above, let the policy write v in a batch drawn from $\{v, w\}$ and k in a batch drawn from $\{k, p\}$, with any multiplicities, and let it apply any family of class-wise constant corrections to the head. Then the balanced error of the flooded pair (v, k) under the composed rule is at least $\frac{1}{2}$, and its misalignment excess is at least $\Xi = \frac{1}{2} - e_{\mathrm{al}}$.*

Proof. By Lemma 22 the realized anchors are count-fraction convex combinations, so $a_1 = \mu_v + \alpha s u$ and $a_2 = -\beta\delta$ for some $\alpha, \beta \in [0, 1]$, and $\bar{D} = a_1 - a_2 = (1 + \beta)\delta + \alpha s u$. Take first the corrections of Proposition 13(a), so $\Delta = 0$. With $\mu_k = 0$ and $u \perp \delta$ the two numerators of (12) are $\gamma/2$ and $-\gamma(\beta + \frac{1}{2})$, and the common denominator is $\|\delta - \bar{D}\| = \sqrt{\beta^2\gamma + \alpha^2 s^2}$. Writing $y = \gamma/(2\|\delta - \bar{D}\|) \geq 0$,

$$e_{\mathrm{bal}} \;=\; \tfrac{1}{2}\big[\Phi(-y) + \Phi\big((1 + 2\beta)y\big)\big] \;\geq\; \tfrac{1}{2}\big[\Phi(-y) + \Phi(y)\big] \;=\; \tfrac{1}{2}, \tag{26}$$

using $\beta \geq 0$ and monotonicity of Φ. For an arbitrary family of class-wise constants, Proposition 13(b) says the induced rule keeps the direction $g = \delta - \bar{D} = -\beta\delta - \alpha s u$ and acquires an arbitrary threshold. Set $a = g^\top \mu_k/\|g\|$ and $b = g^\top \mu_v/\|g\|$, so $b - a = g^\top \delta/\|g\| = -\beta\gamma/\|g\| \leq 0$. For any threshold t the balanced error is $\frac{1}{2}[\Phi(a - t) + \Phi(t - b)] \geq \frac{1}{2}[\Phi(b - t) + \Phi(t - b)] = \frac{1}{2}$, since $a \geq b$. (If $\|g\| = 0$ the rule is constant and the balanced error is exactly $\frac{1}{2}$.) Subtracting e_{al} gives the excess claim. □

Remark 10 (The floor is the frozen frame pointing the wrong way). The reason the bound is $\frac{1}{2}$ and not merely "above e_{al}" is worth naming: at this placement $g^\top \delta \leq 0$, so the frozen head's discriminant for the flooded pair is anti-aligned with the pair's own displacement. The head is not noisy about (v, k), it is oriented against it, and no threshold repairs a direction. Flipping the direction would be a rewrite of v's row, which is exactly what Assumption A1 forbids and what W2 buys back.

Remark 11 (Fixed unit composition, and the shape of the floor). At unit multiplicities ($\alpha = \beta = \frac{1}{2}$, the forced composition when each class is written once with weight one) the same computation gives $y = x = \gamma/\sqrt{\gamma + s^2}$ and

$$\phi_{\mathrm{bal}} \;=\; \tfrac{1}{2}\big[\Phi(-x) + \Phi(2x)\big] \;>\; \tfrac{1}{2}, \qquad \phi_{\mathrm{bal}} \downarrow \tfrac{1}{2} \text{ as } s \to \infty, \tag{27}$$

the fixed-composition floor, which is what Lemma 24 delivers for the singleton anchor sets $A_1 = \{a_1\}$, $A_2 = \{a_2\}$: it exceeds $\frac{1}{2}$ for every finite s, peaks at the universal height of Proposition 33, and is the sharper constant whenever the environment fixes unit compositions. Theorem 4 quotes the weaker $\Xi = \frac{1}{2} - e_{\mathrm{al}}$ because it holds against every composition the policy could choose.

The streams. Classes v, w arrive at round 0 and k, p at the start of stage 2. The *pure stream* is Q_1 queries to v (stage 1). The *composed stream* is the pure stream followed by Q_2 queries to k (stage 2). Both are oblivious; both phases are common prefixes in the sense that stage-1 behavior of any policy depends only on the stage-1 segment, whose distribution is identical under both continuations.

Proof of Theorem 4, deterministic case. Let π be deterministic and write $K' = K + 2c$. *The optimum is cheap on both streams.* On the pure stream, writing $\{v, w\}$ at round 0 costs K' and nothing afterwards (the pair is internal), so $\mathrm{OPT}_{\mathrm{pure}} \leq K'$. On the composed stream, renting v through stage 1 and writing $\{v, k\}$ at the first round of stage 2 costs $rQ_1 + K'$ and nothing afterwards: the flooded class sits in the aligned frame of its own flooding partner, and w, p are never written. Hence $\mathrm{OPT}_{\mathrm{comp}} \leq rQ_1 + K'$. *Stage 1.* If π does not write v during stage 1, its cost on the pure stream is rQ_1, at least rQ_1/K' times the optimum, which is the second bound of (4). Otherwise π writes v during stage 1, necessarily as the batch $\{v, w\}$ (the only admissible batch of the pending pool), paying K' and freezing the anchor a_1; by prefix-determinism it does the same on the composed stream. *Stage 2 floor.* On the composed stream, after stage 1 the pending pool is $\{k, p\}$ and v, w are frozen. Each stage-2 query to k costs π at least $\min(r, \ell\,\Xi)$: while k is unwritten it pays rent r; once written, its batch is $\{k, p\}$ at some anchor a_2 of the family above, and the query's excess

flow against v is at least $\ell\,\Xi$ by Lemma 25, whatever compositions and class-wise corrections π chose (the cross-batch excess against w only adds, and the internal (k, p) pair adds nothing). So $\mathrm{cost}(\pi) \geq K' + \min(r, \ell\,\Xi)\, Q_2$ on the composed stream, which against $\mathrm{OPT}_{\mathrm{comp}} \leq rQ_1 + K'$ is the first bound of (4). *Unboundedness.* Given any target C, take $Q_1 > CK'/r$, so that a policy of ratio below C on the pure stream must write v in stage 1, and then $Q_2 \geq C(rQ_1 + K')/\min(r, \ell\,\Xi)$; both are fixed before the policy runs, and one of the two streams then costs π at least C times its optimum. Since Q_2 may grow at fixed Q_1 against a fixed optimum, no additive constant rescues the ratio either. □

Proof of Theorem 4, randomized case. Let π be randomized and let $p_W = \Pr_\pi[v \text{ is written during stage 1}]$, the same under the pure and composed streams because stage-1 behavior depends only on the stage-1 segment, identically distributed under both. Two cases. *Case $p_W > 1/2$:* on the composed stream, condition on the event W that v is written in stage 1 (anchor a_1 forced). On W, every stage-2 query to k again costs at least $\min(r, \ell\,\Xi)$ pathwise, by Lemma 25 and regardless of the policy's stage-2 randomness; off W, costs are nonnegative. On W the policy has also paid K' for its stage-1 write, so $\mathbb{E}[\mathrm{cost}] \geq p_W\big[K' + \min(r, \ell\,\Xi)Q_2\big] > \frac{1}{2}\big[K' + \min(r, \ell\,\Xi)Q_2\big]$, which against $\mathrm{OPT}_{\mathrm{comp}} \leq rQ_1 + K'$ is exactly half the first bound of (4). *Case $p_W \leq 1/2$:* on the pure stream, $\mathbb{E}[\mathrm{cost}] \geq (1 - p_W)rQ_1 \geq \frac{1}{2}rQ_1$ against $\mathrm{OPT}_{\mathrm{pure}} \leq K'$: half the second bound. The adversary, knowing π's distribution but not its coins, selects the pure stream in the second case and the composed stream in the first; with (Q_1, Q_2) chosen as in the deterministic proof (with $2C$ in place of C), every randomized pure-W3 policy pays at least C times the optimum in expectation on one of the two streams of the family. □

Remark 12 (What the bound is made of, and what it is not). The numerator is the policy's best per-query option under the freeze: rent, or the floored excess. The denominator is bounded: the optimum rents through stage 1 and then writes the flooded class in the aligned frame of its flooding partner, after which every flood query is free of excess. The frozen frame is the whole mechanism. Co-batching is free alignment (Theorem 1(a): a merged pair is at Bayes), and the adversary sells it only to the prophet: the policy had to choose v's frame before k existed, and no unanchored write can ever put k into that frame (Lemma 22 confines k's anchor to the stage-2 hull, Lemma 25 prices the separation uniformly over the composition, that being the lemma the theorem uses, Lemma 24 reaching hulls only through its norm-unequal case (Appendix G.6), and Proposition 13 says no free correction repairs it). An earlier form of this theorem, registered before the experiments, charged the aligned baseline to written pairs and concluded a finite per-instance ratio $\min(r, \ell\phi_{\mathrm{bal}})/(\ell e_{\mathrm{al}})$; its optimum line assumed the optimum writes only the flooding pair, which is false under that convention (the optimum co-batches k with the far partner at negligible internal error), and the registered falsifier caught it (Section 9). Under the net convention the optimum's free state is not an artifact to be closed by placement; it is the content of the theorem. Two boundaries of the claim, stated here rather than left for a referee. The per-query price is bounded: $\Xi < \frac{1}{2}$ always, so the theorem is about an unbounded number of queries each paying a bounded positive excess, not about an unbounded price per query; in the language of the classical objective, the unbounded quantity is total normalized regret. And the absolute form is specific to the differential objective: under the gross convention the same construction still admits no envelope-uniform constant, because the optimum's escape route (co-batching the flooded class with the far partner, whose internal error is $\Phi(-\sqrt{\gamma + s^2}/2)$) drives its flood cost to zero as s grows, but at any fixed placement the ratio is then finite. The registered experiment measured the gross form and the post-hoc program the net form (Section 9).

G.5 Free parking for the optimum, as a lemma

The absolute form of Theorem 4 (no finite ratio, on a fixed instance) rests on the optimum having a state in which the flood is free. That state is not special to the two-phase family; it is a lemma about the ledger.

Lemma 26 (Zero-excess absorbing state)**.** *Fix an instance under Assumptions A1 and A2, a class k, and a nonempty set F of other classes, and suppose there is a round t at which every member of $F \cup \{k\}$ has arrived and none has yet been written. Then the schedule that writes $F \cup \{k\}$ as one unanchored batch at round t and writes nothing else serves every query after t, to k or to any member of F, at zero excess flow, since every pair inside the batch is internal and hence at its aligned Bayes value. Provided no class outside $F \cup \{k\}$ is queried after t, which the two-phase*

family satisfies, its total cost is the rent accrued before t plus the single write event $K + c(1 + |F|)$, independent of the length of the stream after t.

Proof. A within-batch pair is classified by the composed protocol's Bayes bisector (Lemma 7 with Appendix B.1), whose balanced error is e_{al}, so its excess is zero by definition; the batch has size at least two, so it is admissible under Assumption A1, and it is feasible at round t by hypothesis. Since nothing outside the batch is ever written, a query to a class in the batch is charged only against classes in the batch, and every such pair is internal. □

The three hypotheses are exactly what the lemma needs and exactly what the two-phase family supplies to the optimum: the classes are co-pending at a common round, and the optimum writes nothing else. They are also what a policy cannot in general arrange, because it must write before it learns which pairs will be queried, and the write-once boundary makes that choice final. That asymmetry, not the arithmetic, is the content: the optimum has somewhere free to park, and the policy, whose frames are frozen, does not. The lemma also shows what the anchored operator buys, and why it does not need the co-pending hypothesis: under W2 a policy can reach the zero-excess state at any later round by replaying the classes whose frames must agree, at the price of the anchor, which is why the coupled problem collapses to scheduling (Proposition 5).

G.6 Richer pending pools

The two-phase construction used forced batches (only one admissible partner per stage), which makes its anchors singletons. The general lemmas do not need that convenience, and it is worth recording what changes when pools are richer. If the stage-1 pool contains additional items $w_1, \ldots, w_m$, a policy may steer v's batch anchor anywhere in the convex hull of pool means (Lemma 22); the floor lemma then applies with $A_1 = \mathrm{conv}\{\mu_v, \mu_{w_1}, \ldots\}$ and A_2 the stage-2 analogue, but only through case (i), and it is worth being exact about why. Because $\mu_v \in A_1$ and $\mu_k \in A_2$, the difference set $A_1 - A_2$ *always* contains δ, so case (ii)'s ray hypothesis fails at every pool, however separated: no placement can prevent it, and any reading on which richer hulls avoid the closed ray is wrong. For a norm-unequal target pair this costs nothing, since case (i) asks only that the hulls be disjoint and lets $A_1 - A_2$ contain δ. For a norm-equal target pair Lemma 24 does not apply to hulls at all, and the floor must come from Lemma 25, which is what Theorem 4 in fact uses and which is stated for the batch composition directly rather than for an anchor set. Two consequences. First, giving the policy more co-batching freedom cannot restore competitiveness on separated geometries: the minimum in the floor is over a larger set, but for a norm-unequal pair the enlarged set still misses the origin as long as the hulls stay disjoint, and the floor survives. Second, the adversary's job is monotone in pool richness: hulls grow with the pool, so separation margins, not cleverness, carry the construction. The forced-batch instance is simply the minimal specimen of the species.

G.7 Which classical ingredients transfer

The classical impossibility pattern for one-way decisions under adversarial flow rests on three ingredients: an irrevocable act, adversary-controlled post-commitment flow, and a flood-immune zero-cost absorbing state for the optimum. All three transfer, and the comparison localizes how.

The table is the paper's answer to the re-skinning objection in the affirmative direction: what is inherited is the scheduling machinery, credited in Section 2; what is new is the geometry that prices the frames and installs the flood, and the operator boundary that decides whether the policy, too, can reach the free state: under anchoring it can, and the pattern dissolves into scheduling.

G.8 The anchored escape

Proof of Proposition 5. Under W2, the incoming batch is fit jointly with a replay anchor drawn from the store. Applying Lemma 6 to the union (batch plus anchor) as one mixture: the joint fit's gauge is centered at the union's mean, and the anchor's presence pins the batch's rows into the store's frame; at population level, with the anchor drawn from the same distribution as the store's composition, the cross-batch anchor mismatch is exactly $\bar{D} = 0$ and the scalar mismatch is removed by the Proposition 13(a) offsets, so the gauge cost is identically zero (finite-sample anchor noise

Table 8: Ingredient transfer into the pure-W3 frontier. All three ingredients of the classical pattern are present, the first two in geometrically installed form, and the impossibility is absolute.

Ingredient	Classical role	Status here
Irrevocable act	The promotion cannot be undone	Transfers, upgraded: irrevocability is the frozen row of Assumption A1, a definitional property of the operator, not a policy restriction
Adversarial post-commitment flow	A declared price process floods the committed state	Transfers, geometrically installed: anchor confinement (Lemma 22) fixes the flood's target through placements alone, so the oblivious stream needs no knowledge of the policy's realized batch choices
Zero-cost absorbing state	The optimum parks costlessly and the ratio diverges per instance	Transfers (Lemma 26): the optimum co-batches the flooded class with its competitors in one aligned frame and pays no excess afterwards, while the policy's frozen frame pays the excess floor on every query

is estimation-ledger and out of scope here). The objective then contains no geometry-dependent term: it is the rent-plus-setup objective of Section 7, and Proposition 18 applies with its guarantees, uniformly over the instance geometry, under its stated linear-delay scope; for sum-monotone batch delay costs the greedy guarantee fails (Bhore et al., 2026) and we claim nothing. Necessity is Theorem 4: the unanchored class admits no uniform constant. Sufficiency is this proposition. Within the taxonomy, revisability is competitively necessary and sufficient for geometry-independent guarantees, and anchoring is the cheapest revisable operator: W1 revisits every written class and escapes too, at the price of the store. □

H Priced Mixed-Operator Scheduling: the Order Settled, the Constant Open

The frontier resolution leaves a strictly smaller, priced question standing. **Mixed-operator scheduling.** At each write opportunity the scheduler holds three options per pending item: keep renting at r per query; write unanchored (W3) at setup K_3, accepting the excess floor of the achievable-anchor geometry on every later cross-batch query; write anchored (W2) at setup $K_2 > K_3$ with zero excess. Characterize the optimal competitive ratio as a function of the premium K_2/K_3 and the excess floor $\phi_{\mathrm{bal}} - e_{\mathrm{al}}$. Two boundary facts bracket the answer.

H.1 The two boundary facts, proved

Proposition 27 (The bracket)**.** *Fix the anchored setup K_2, understood as the full price of a write whose anchor covers the classes whose frames must agree (Proposition 5), with the per-item price absorbed into it as in Section 7 ($c = 0$). In the mixed-operator problem the break-even policy that writes anchored only is $(2 + w^{\max}/K_2)\, K_2/K_3$-competitive against the mixed-operator optimum, hence $2K_2/K_3$ in the limit $w^{\max}/K_2 \to 0$ and not at any fixed granularity; and the all-unanchored policy class has no finite ratio. Appendix H.2 closes the interior: the optimal deterministic ratio is $\Theta(K_2/K_3)$, so no premium-independent constant exists, and the bracket is $\big[\max\big(2 - r/K_3,\ 1 + (K_2 - r)/(2K_3)\big),\ 1 + (K_2 + w^{\max})/K_3\big]$, both ends strict (Corollary 32); its two ends differ by strictly less than a factor of two at every premium in the fine-granularity limit $w^{\max}/K_3, r/K_3 \to 0$, and approach that factor only as the premium grows, rather than differing by the premium. At $K_2 = K_3$ the two ends meet at 2 in that limit, recovering the classical deterministic constant.*

Proof. Upper: under W2 the excess is identically zero (Proposition 5), so the all-anchored break-even policy is A_θ of Proposition 29 at $\theta = K_2$, and that proposition gives the bound directly against $\mathrm{OPT}_{\mathrm{mixed}}$, at multiplier $\max(2, (2K_2 + w^{\max})/K_3) = (2 + w^{\max}/K_2)K_2/K_3$ since $K_2 \geq K_3$. *We route through it rather than through the acknowledgment optimum, and the reason is a hypothesis this proof used to skip.* Chaining $\mathrm{cost} \leq (2 + w^{\max}/K_2)\mathrm{OPT}_{\mathrm{ack}}(K_2)$ imports the classical amortization for the *unconstrained* break-even policy, and in the mixed problem no policy may open the store with a singleton (Remark 3), so the anchored break-even is suppressed until two classes are

pending and is not that policy. The gap is real and not small: at $K_2 = 4$, $K_3 = 2$, $r = 1$, one class queried alone for twenty rounds before a second arrives, the policy rents twenty and then writes for four, paying 24 against $\mathrm{OPT}_{\mathrm{ack}}(K_2) = 4$, a ratio of six. The proposition survives because the same rule binds the optimum, which also cannot write that singleton and pays 21 on that instance, for a ratio of 1.14; it is the route through $\mathrm{OPT}_{\mathrm{ack}}(K_2)$ that fails, and Proposition 29, whose policy carries the first-write rule in its definition and whose proof charges the rent before the first write separately, does not use it. The constant is 2 only in the limit $w^{\max}/K_2 \to 0$. Proposition 29 is proved in Appendix H.2 from the instance and the optimum alone and does not cite this proposition, so the appeal is not circular. The weaker reading, that $w^{\max} \le K_2$ suffices for the factor 2, is false, and the instance refuting it needs an *uneven* round pattern, so that the accrual steps over the threshold instead of landing on it. Take $K_2 = 4$, $K_3 = 2$, $r = 1$ and two classes, *both* queried, over four rounds carrying two queries, then one, then two, then one. Break-even's accrual runs 2, 3, 5 across the first three rounds and so passes K_2 without stopping on it; the trigger is tested at the start of a round, so the policy fires at the fourth, paying 4 on top of the 5 already rented, for 9 in all, against an optimum of 2 that writes the pair unanchored at round zero. The ratio is 4.5, above $2K_2/K_3 = 4$ and within the $(2 + w^{\max}/K_2)K_2/K_3 = 5$ this proof gives at $w^{\max} = 2$. Rounds that all carry two queries do *not* refute it: the accrual then lands on 4 exactly, and the ratio is $2K_2/K_3$ on the nose. `l2_mixed.py` simulates both patterns and prices each optimum by exhaustive enumeration over every legal write plan. The two comparisons the discarded route needed remain true and are worth recording, since they are what makes the premium the natural unit: any mixed schedule pays its rent, at least K_3 per write event, and a nonnegative excess, so $\mathrm{OPT}_{\mathrm{mixed}} \ge \mathrm{OPT}_{\mathrm{ack}}(K_3)$; and $\mathrm{OPT}_{\mathrm{ack}}(K_2) \le (K_2/K_3)\,\mathrm{OPT}_{\mathrm{ack}}(K_3)$, by running the latter's schedule at the higher setup. What they cannot supply on their own is the first link, for the reason above. Lower: two bounds combine. Appendix H.2 proves the premium-carrying bound $1 + (K_2 - r)/(2K_3)$ directly, and that is what makes the ratio grow with the premium; it is weak at small premium, where the classical constant $2 - r/K_3$ takes over, available here from a *two-class* instance (Proposition 30). The better of the two is $\max(2 - r/K_3,\ 1 + (K_2 - r)/(2K_3))$, and that is the lower end we state. (The single-item ski-rental embedding one would reach for first is indeed unavailable: under Remark 3 a lone class can be written by neither operator, W3 needing two members and a W2 singleton needing a frame that does not yet exist, so a single-item instance forces both the policy and the optimum to rent and its ratio is 1. Two classes suffice, and are legal.) The unanchored endpoint is Theorem 4. □

Before Appendix H.2 closes it, the bracket's width is the premium, and the premium is exactly where the coverage question enters: if a bounded anchor suffices for the frames a schedule actually needs to agree, K_2/K_3 is a constant and the bracket is a constant; if coverage must grow with the store, the upper end grows with it, and the interesting regime is the one in which a scheduler chooses coverage as well as timing. That choice is not in the problem as posed here, and stating it is one of the contributions of posing the mixed-operator problem. The interior question is how the achievable ratio moves between the classical constant and $2K_2/K_3$ as the premium and the excess floor vary: when the premium is small the anchored endpoint is essentially optimal, and when the excess floor is small relative to r/ℓ a scheduler may profitably take unanchored writes on pairs it expects never to be flooded. A natural candidate policy maintains accrued rent per pending item and the anchor-proximity graph of the pending pool; when an item's accrued rent reaches its write cost it writes the batch of that item and its nearest-anchor pending partner, anchored if the premium $K_2 - K_3$ is below the expected excess $\ell(\phi_{\mathrm{bal}} - e_{\mathrm{al}})$ per future discriminating query, unanchored otherwise. Appendix H.2 settles which of these readings is right: the premium is not an artifact of a lossy comparison but the price of an option the offline optimum holds and no online policy does, so the ratio grows with K_2/K_3 whatever the coverage regime, and the coverage question changes how fast rather than whether. E-L2-6 measures the one input, the premium, that the problem takes from the world.

H.2 The interior of the bracket, closed

Proposition 27 leaves the mixed-operator ratio below $(2 + w^{\max}/K_2)K_2/K_3$, and the question it poses is whether the premium is real or an artifact of comparing through the anchored optimum. It is real. This appendix proves that the optimal deterministic ratio is $\Theta(K_2/K_3)$: no premium-independent constant exists, and the surviving uncertainty is below a factor of two at every premium rather than a factor of the premium. It also supplies the lower end where the premium-carrying bound is weak, from the classical two-class instance, which pins the ratio to exactly 2 at unit premium in the fine-granularity limit. Throughout, $K_2 \ge K_3$ are the anchored and unanchored write prices

with the per-item price absorbed as in Proposition 27, $P = K_2/K_3$, and $w^{\max}$ is the largest rent a single round accrues under the policy in question (Appendix F.4). The results below are evaluated by `l2_mixed.py`. Its self-test prices the witness of Proposition 28 and the counterexample in Proposition 27's proof by exhaustive enumeration over every write plan (batch structure, timing and operator, at unit multiplicities; the optimum of each carries no cross-batch pair, so multiplicities cannot lower it), *simulates* the break-even and A_θ policies on the mirror family rather than asserting their costs (the family's two optima are the closed forms of Remark 13, checked by enumeration at $w = 2$ on the shortest legal stage), checks the two-class minimum with its argmin and every constant of Table 9, and, since amendment L2-A38, does two things it did not do before: it compares the gadget's minimax value with Theorem 31's closed form on and off its original four-point grid, where the form printed through L2-A37 was loose (Remark 15), and it prices A_{K_3} and break-even against an *enumerated* optimum on every stream of a family of 5420 short instances with arrival times, rounds carrying up to two queries and up to four classes, on which Proposition 29's bound holds strictly on every stream, at $\theta = K_3$ and at four other thresholds, while the strict ratio of A_{K_3} reaches 3.5 against the limit value $1 + P = 3$ (Corollary 32). The flood-length threshold of Theorem 31 is a derivation, and the self-test confirms only that a flood too short to reach it drops the gadget below the bound.

Proposition 28 (The comparison itself costs the premium)**.** $\sup \mathrm{OPT}_{\mathrm{anch}}/\mathrm{OPT}_{\mathrm{mixed}} = P$ *over instances, and the supremum is attained. No sharper comparison between the two optima exists, so the factor P in Proposition 27's chain cannot be removed by a better argument.*

Proof. Upper: recolour every unanchored event of an optimal mixed schedule as an anchored event on the same batch, at the full anchored price K_2, which by hypothesis is coverage-independent and so does not grow with the recoloured batch. Anchoring admits every batch W3 admits (Remark 3), the rent is unchanged, and the excess drops to zero (Proposition 5), so $\mathrm{OPT}_{\mathrm{anch}} \le R + K_2 E \le P\,(R + K_3 E) \le P\,\mathrm{OPT}_{\mathrm{mixed}}$, where R and E are the schedule's rent and event count and every event costs at least K_3. Attainment: let a at μ_a and its mirror b at $-\mu_a$ both arrive at round 0, with a queried in every round of a stage of length $T \ge K_2/r$ and b never queried. Writing $\{a, b\}$ unanchored at round 0 is legal (two pending classes) and free of rent (writes precede queries); it is the only batch, so no cross-batch pair exists and the excess is zero: $\mathrm{OPT}_{\mathrm{mixed}} = K_3$. An anchored schedule either pays K_2 or rents for T rounds at $rT \ge K_2$, so $\mathrm{OPT}_{\mathrm{anch}} = K_2$. □

Remark 13 (What the mirror pair does, and what it needs). Repeating the witness w times with pair i at $\pm z_i$ arriving at round iT gives $\mathrm{OPT}_{\mathrm{mixed}} = wK_3$ against $\mathrm{OPT}_{\mathrm{anch}} = wK_2$, and it is worth seeing why the excess stays at zero across w batches rather than only within one: every mirror pair's anchor is its own midpoint, which is the origin, so all w anchors coincide and every cross-batch mismatch $\bar{D}$ vanishes. The scalar mismatch does *not* vanish, since the batch scalars $\|z_i\|^2$ differs across batches; it is removed by the Proposition 13 offsets that Appendix G.1 puts in force, and without them the same family would carry an excess of order one. The stages are disjoint, so no round of the family carries more than one query and $w^{\max} = r$ on it, and $rT \ge K_2$ is tight: at $rT = K_2 - r$ an anchored schedule beats wK_2.

Proposition 29 (An improved upper bound: break even against the optimum's price, not your own)**.** *Let A_θ rent until the rent accrued since its last write reaches θ, the count starting, before any write, at the first round at which two classes are pending, then write every pending class as one covering anchored event at the same coverage-independent price K_2, and repeat. For every instance and every $\theta > 0$,*

$$\mathrm{cost}(A_\theta) \;\le\; \max\!\Big(1 + \tfrac{K_2}{\theta},\; \tfrac{\theta + w^{\max} + K_2}{K_3}\Big) \cdot \mathrm{OPT}_{\mathrm{mixed}}, \tag{28}$$

with no additive term. At $\theta = K_3$ the two branches are $1 + P$ and $1 + P + w^{\max}/K_3$, giving

$$\mathrm{cost}(A_{K_3}) \;\le\; \Big(1 + \tfrac{K_2 + w^{\max}}{K_3}\Big)\,\mathrm{OPT}_{\mathrm{mixed}}, \tag{29}$$

which is $1 + P$ in the fine-granularity limit: a factor $2P/(1 + P)$ better than the $2P$ Proposition 27 reaches in that same limit, approaching two as the premium grows. In that limit $\theta = K_3$ minimises the multiplier, the two branches crossing there; at fixed granularity the crossing sits below K_3, within $w^{\max}$ of it when $w^{\max} < K_3$, and we keep $\theta = K_3$ for its meaning, the optimum's own price. The multiplier is tight in the limit: on the family of Remark 13, with r dividing K_3 and $rT > K_3$, A_{K_3} pays exactly $(1 + P)$ times the optimum, against a multiplier of $1 + P + r/K_3$ there. The strict ratio does exceed the limit value $1 + P$ on short instances (Corollary 32).

Proof. A_θ holds exactly one frame at all times, so by Proposition 5 it pays no excess, and its cost is rent plus K_2 per event. Let t_0 be the first round at which two classes are pending. Before t_0 no policy, the optimum included, can write anything: fewer than two classes are pending, so neither operator has a pair, and an anchored singleton needs a frame that does not exist (Appendix G.1). The rent R_0 accrued before t_0 is therefore common to A_θ and to the optimum, and A_θ's counter starts at t_0, so its first epoch, like every later one, accrues at least θ before its event. This step is what the naive argument misses: with the counter running from the start, the first epoch's accrual after t_0 has no lower bound and the charging below fails on it.

Decompose the rounds from t_0 on into epochs: the first $[t_0, t_1]$ closed at t_0 (A_θ performs no event at t_0, so an optimum event there falls inside the first epoch), the later ones $(t_{i-1}, t_i]$, each ending with an A_θ event, and a final open epoch after the last event. Charge the optimum's cost to the epoch in which it is incurred; the epochs are disjoint, so the charges sum to at most $\mathrm{OPT} - R_0$. *An epoch ending with an event.* A_θ pays $K_2 + A_i$ there, with $\theta \le A_i < \theta + w^{\max}$: the trigger is tested at the start of each round before queries are served, so the accrual at the start of the epoch's last round is below θ, and one round adds at most $w^{\max}$. If the optimum performs an event inside the epoch it pays at least K_3 there, and A_θ's bill is at most $(K_2 + \theta + w^{\max})/K_3$ times it. If it performs none, call the epoch clean. Because A_θ writes *every* pending class at each event, a class it rents during epoch $i \ge 2$ arrived after the event at t_{i-1} (arrivals precede writes, so a class present at t_{i-1} was written there), and in the first epoch nothing was written by anyone through t_0; a clean optimum has written none of these classes before they arrived nor during the epoch, so it pays rent on at least the queries A_θ pays rent on, at least $A_i \ge \theta$, and A_θ's bill $K_2 + A_i$ is at most $(1 + K_2/\theta)$ times it. *The final epoch.* A_θ pays only rent there, below $\theta + w^{\max}$; if the optimum performs an event in it the bill is at most $(\theta + w^{\max})/K_3$ times the optimum's, and otherwise the optimum rents everything A_θ rents. Summing, $\mathrm{cost}(A_\theta) - R_0 \le \max(1 + K_2/\theta,\ (K_2 + \theta + w^{\max})/K_3)\,(\mathrm{OPT}_{\mathrm{mixed}} - R_0)$, and the multiplier is at least one, so the common R_0 is absorbed. The crossing of the two branches solves $\theta^2 + (K_2 + w^{\max} - K_3)\theta = K_2 K_3$, which is $\theta = K_3$ at $w^{\max} = 0$; at $w^{\max} > 0$ the first branch is below the second at K_3 and above it at $K_3 - w^{\max}$, which places the crossing as stated. For tightness, on the family of Remark 13 with r dividing K_3 and $rT > K_3$, A_{K_3} accrues exactly K_3 of rent on each arriving class, before the next pair arrives, and then pays K_2, for $w(K_3 + K_2)$ against wK_3. □

Remark 14 (Where the published policy loses the factor). Both links of Proposition 27's chain are simultaneously tight on the mirror family at $rT > K_2$ and in the limit $w^{\max}/K_2 \to 0$, so its $2P$ is not a slack estimate. Both qualifications are needed and both are easy to lose. At the family's boundary member $rT = K_2$ each pair's trigger point falls exactly on the next pair's arrival: at $w = 1$ the threshold never fires and the policy rents throughout, and at $w \ge 2$ it fires, but one covering event absorbs everything then pending. Either way the ratio is P, not $2P$. And at fixed granularity the constant is $(2 + w^{\max}/K_2)P$, which this family cannot exhibit: under break-even each a_i is written K_2/r rounds after it arrives and $T \ge K_2/r$, so at most one unwritten class is queried in any round and $w^{\max} = r$; the measured ratio is then exactly $2P$ at $rT > K_2$ when r divides K_2 (an overshoot adds at most r/K_3) and exactly P at the boundary, never more. Exceeding $2P$ takes the uneven-round instance of Proposition 27's proof, where two classes are queried in one round. `l2_mixed.py` simulates the family at both thresholds and at both readings of T. The defect is the threshold. Break-even waits until the accrued rent reaches its *own* buy price K_2, when the quantity governing the comparison is the *optimum's* buy price K_3. Lowering the threshold to K_3 buys back almost the whole factor of two, and the policy is otherwise unchanged.

Proposition 30 (The classical constant, from two classes)**.** *Under the arrival-revelation convention of Appendix G.1, every deterministic mixed-operator policy has competitive ratio at least* $2 - r/K_3$ *whenever* $r \le K_3$*, and when* r *divides* K_3 *this instance gives no more: a policy whose first write falls at round* $K_3/r - 1$ *achieves exactly that on it. Otherwise the integer minimum is the slightly larger value at the nearest admissible round, and we state the bound in the form that holds for every* $r \le K_3$*.*

Proof. Two classes are pending from round 0: a, queried once in every round, and b, never queried. The pair $\{a, b\}$ has size two, so it is a legal unanchored batch (Remark 3); the single-item obstruction recorded in Proposition 27 therefore does not apply, and the adversary's only freedom is when to stop. Every write, by either operator, costs at least K_3 (an anchored pair may be the first event,

Appendix G.1, but it costs $K_2 \geq K_3$), and the optimum's single write is the unanchored $\{a, b\}$ at K_3 exactly.

A deterministic policy either never writes, in which case the adversary extends the stream and the ratio grows without bound against an optimum of K_3, or it first writes at a round τ fixed by the prefix. It has then paid $r\tau$ in rent, round τ's own query being free because writes precede queries, plus at least K_3 for the event. The adversary stops there. The offline optimum, seeing $\tau + 1$ rounds of queries to a, pays $\min(K_3,\, r(\tau + 1))$: it writes $\{a, b\}$ at round 0, or it rents throughout. So the ratio is at least $(r\tau + K_3)/\min(K_3,\, r(\tau + 1))$. On the branch $r(\tau + 1) \leq K_3$ this equals $1 + (K_3 - r)/(r\tau + r)$, decreasing in τ and least at $r\tau = K_3 - r$, where it is $2 - r/K_3$; on the branch $r(\tau + 1) > K_3$ it equals $1 + r\tau/K_3$, increasing in τ and above $2 - r/K_3$ at every such τ. The minimum over τ is $2 - r/K_3$, attained at $r\tau = K_3 - r$. □

Theorem 31 (No premium-independent constant)**.** *Under the arrival-revelation convention of Appendix G.1, every deterministic mixed-operator policy has competitive ratio at least*

$$1 + \frac{K_2 - r}{2\,K_3} \;\geq\; \tfrac{1}{2}\big(1 + P\big) \quad \textit{when } r \leq K_3 \leq K_2. \tag{30}$$

Together with Proposition 29 the optimal deterministic ratio is $\Theta(P)$: it is not bounded by any constant independent of the premium. (Through amendment L2-A37 the bound was printed as the minimum of this quantity and P; the minimum came from pricing the anchored branch of the proof against an optimum of K_3 where the optimum is $\min(K_3, r(\tau + 1))$, and that branch never binds. Both forms are valid; this one is the gadget's value.)

Proof. Three classes are pending from round 0: H at the origin, N_1 at $(-1, -2)$ and N_2 at $(1, -2)$. Only H is queried in stage one, and N_1, N_2 are never queried; they exist to supply the batch-size partner and to be the reserve the optimum holds back. Let τ be the first round at which the policy writes, a function of the stage-one prefix alone. The adversary continues in one of two ways that share that prefix exactly, so the policy's conduct through round τ, including its round-τ write, is the same in both.

If the policy never writes, the adversary extends stage one and the ratio grows without bound against an optimum of K_3, which writes $\{H, N_1\}$ unanchored at round 0 for no rent and no excess. If the round-τ write is anchored, the adversary stops: the policy has paid $r\tau + K_2$ against an optimum of $\min(K_3, r(\tau + 1))$, which writes $\{H, N_1\}$ at round 0 or rents throughout. Writing $a = r\tau$, the ratio $(a + K_2)/\min(a + r, K_3)$ is least at $a = K_3 - r$, where it is $1 + (K_2 - r)/K_3$, above $1 + (K_2 - r)/(2K_3)$.

Otherwise the round-τ write is unanchored on some B with $|B| \geq 2$; only one such event can occur at round τ, since two disjoint unanchored batches would need four pending classes. The adversary picks the victim $v \in B$ given by Table 9, places a new class k at

$$\mu_k \;=\; 2\,\bar{\mu}\big(\{H, N_1, N_2\} \setminus \{v\}\big) \;-\; \mu_v, \tag{31}$$

which makes $\bar{\mu}(\{v, k\})$ coincide with $\bar{\mu}(\{H, N_1, N_2\} \setminus \{v\})$ exactly, and floods: k arrives at round $\tau + 1$ and is queried Q_2 times, and H is never queried again.

The optimum pays at most $\min(r(\tau+1)+K_3,\; 2K_3)$, by either of two plans. It may rent H through round τ and write $\{H, k\}$ at round $\tau+1$, one frame and no excess. Or it may write $\{H, N_1, N_2\}\setminus\{v\}$ at round 0, which is legal in every row of the table and contains H, so stage one is rent-free, and then write $\{v, k\}$ at round $\tau + 1$ before that round's queries: the two anchors coincide by the placement, so the mismatch is zero and so is the excess. This is the reserve: the optimum spends a never-queried class on the flood.

The policy pays at least $r\tau + K_3 + K_2$. The rent and the round-τ event are already spent. In stage two it either performs an anchored write, which costs K_2, or it does not, in which case v's frame is B forever, since an unanchored event cannot revisit a written class (Definition 1, Remark 1). Each of the Q_2 flood queries then costs r while k is unwritten, or, once k sits in an unanchored batch, at least ℓ times the excess of the single pair (k, v), single-pair floors sufficing for lower bounds (Appendix G.1). Table 9 gives that excess as an infimum over every admissible flood batch *and* every achievable anchor, so with $Q_2 \geq K_2/\min(r, \ell\,\Xi_{\mathrm{gad}})$ and $\Xi_{\mathrm{gad}} = 0.02305071$ the stage-two bill reaches K_2 either way.

Table 9: The lower bound's victim rule. $\Xi_{\mathrm{gad}} = 0.02305071$ is the smallest of the three infima and is the constant the proof uses. It is a different object from the excess floor $\Xi = \frac{1}{2} - e_{\mathrm{al}}$ of Theorem 4 and is named apart from it deliberately. Each infimum is taken over every admissible flood batch and, by Lemma 22, over every achievable anchor rather than over unit multiplicities: the unit-multiplicity excess of row one is 0.04121923, which the policy cuts to 0.559 of that value by steering its counts, so it is not a floor. These are case (i) of Lemma 24, and only case (i): the achievable difference set here contains the degenerate point $\bar{D} = \delta$, reached at the segment endpoints, so case (ii)'s ray-disjointness hypothesis fails at this gadget exactly as Remark 8 warns. What case (i) needs instead is that the pair be norm-*unequal* ($\|\mu_v\|^2 = 5$ against $\|\mu_k\|^2 = 4$ in each row) and the two anchor segments disjoint, and both hold: e_{bal} then extends continuously through δ with value $\frac{1}{2}$ by Lemma 23(iii), and the infimum over the compact segments is attained away from it. Attainment is not yet strictness: the minimum exceeds e_{al} because the achievable difference set also avoids $\bar{D} = 0$, the only point of the exceptional set E for a norm-unequal pair, and it does so with room ($\min\|\bar{D}\| = 1.789$, 1.789, 2.000 in rows one to three). The entries are that infimum, computed directly. Row three's infimum is approached only in the multiplicity limit. In row four the flooded class has no unanchored partner left, so the policy must anchor or rent.

First write B	victim v	μ_k	infimum of the (k, v) excess
$\{H, N_1\}$	N_1	$(2, 0)$	0.02305071
$\{H, N_2\}$	N_2	$(-2, 0)$	0.02305071
$\{N_1, N_2\}$	N_2	$(-2, 0)$	0.06807840
$\{H, N_1, N_2\}$	N_1	$(2, 0)$	no unanchored partner exists

With $a = r\tau$ as before: if $a + r \le K_3$ the ratio is at least $(a + K_3 + K_2)/(a + r + K_3) = 1 + (K_2 - r)/(a + r + K_3) \ge 1 + (K_2 - r)/(2K_3)$; otherwise it is at least $(a + K_3 + K_2)/(2K_3) \ge 1 + (K_2 - r)/(2K_3)$. Every branch therefore gives at least $1 + (K_2 - r)/(2K_3)$, which is at least $\frac{1}{2}(1 + P)$ whenever $r \le K_3 \le K_2$. The value is attained when r divides K_3: against this adversary a policy that writes unanchored at $r\tau = K_3 - r$ and anchors k on its arrival, before its first query is served, pays exactly this ratio, so the bound is the gadget's value, and `l2_mixed.py` confirms the identity on 420 grid points with K_3/r integral (and the minimax strictly above the bound off that grid). (An adversary free to stop after an unanchored write as well gets the two-class bound of Proposition 30 from the same gadget, which is why the corollary takes the larger of the two.) □

Corollary 32 (The bracket, and its unit-premium endpoint)**.** *For $r \le K_3 \le K_2$, every deterministic mixed-operator policy has ratio at least*

$$\max\Big(2 - \tfrac{r}{K_3},\ 1 + \tfrac{K_2 - r}{2K_3}\Big), \tag{32}$$

the two-class bound the larger when $K_2 < 2K_3 - r$ and the two equal at $K_2 = 2K_3 - r$, and A_{K_3} gives $1 + (K_2 + w^{\max})/K_3$ above; both ends are strict. The strict ratio of A_{K_3} does exceed the limit value $1 + P$ on short instances: 3.5 against $1 + P = 3$ at $(K_2, K_3, r) = (4, 2, 1)$ on two classes, a queried once, then twice, and b never, where the multiplier is $1 + (4 + 2)/2 = 4$. In the fine-granularity limit $w^{\max}/K_3, r/K_3 \to 0$ the two ends are $\max(2,\ 1 + P/2)$ and $1 + P$, whose ratio is $(1 + P)/\max(2,\ 1 + P/2) < 2$ at every $P \ge 1$ and tends to 2 as $P \to \infty$. At $P = 1$ the ends meet: the optimal deterministic ratio at $K_2 = K_3$ is exactly 2 in that limit.

Proof. The maximum of two valid lower bounds is a valid lower bound, and Proposition 30 and Theorem 31 are proved on separate instances; $2 - r/K_3 > 1 + (K_2 - r)/(2K_3)$ is the inequality $K_2 < 2K_3 - r$. The upper end is Proposition 29 at $\theta = K_3$, which carries no additive term. The witness is priced by enumeration in `l2_mixed.py`: rent $1 + 2$, then the anchored write at 4, against the optimum's unanchored pair at round 0 for 2; the same enumeration finds no instance of its 5420-stream family above the upper end. At $K_2 = K_3$ the lower end is $2 - r/K_3$ and the upper is $2 + w^{\max}/K_3$, which meet at 2 in the limit. That the two prices degenerate there to the acknowledgment problem, whose deterministic ratio is 2 asymptotically in the setup-to-delay ratio (Appendix F.4, quoting Dooly et al. 2001), is an independent check on the pair; the argument above does not use it. □

Remark 15 (The mechanism, and what the bound does not cover). The policy is caught between two facts it cannot reconcile. Rent pressure forces it to commit a frame once its accrued rent approaches

a write price. Committing a frame means choosing a partner, and the partner that would have been right has not arrived. The optimum, offline, holds a never-queried class in reserve at no rent and spends it on the flooded class, buying alignment at the unanchored price; the policy's only route back to alignment is the anchored write, at the anchored price. The batch-size hypothesis of Remark 3 is what makes a reserve necessary and therefore what makes it exploitable, and in the fourth row it bites hardest, leaving the flooded class no unanchored partner at all. Two limits are worth stating. The bound is for deterministic policies: Theorem 4's randomized argument does not transfer, because there the geometry is fixed in advance while here the adversary's placement answers the policy's batch, so a randomized policy would have to be handled by fixing one branch ahead of time and losing a further factor. And the gap between $\max(2,\, 1 + P/2)$ and $1 + P$ is open; the natural candidate for the lower end is a policy that takes its first unanchored write when its accrued rent reaches $K_3 - r$ and thereafter anchors every later arrival on its first query, before serving it. On this gadget, against an adversary free to stop or to flood, it attains exactly $\max(2 - r/K_3,\, 1 + (K_2 - r)/(2K_3))$, the corollary's lower end, which is $\max(2, 1 + P/2)$ as $r/K_3 \to 0$; the timing is essential, since the same repair after an immediate first write pays $(K_3 + K_2)/(K_3 + r)$ on the flood, which is $1 + P$ in the same limit. We could not bound the candidate in general. The four-point self-test grid on which the theorem's earlier form was checked is a caution of its own: the two forms coincide exactly when $K_2 \geq 2K_3 - r$, and every grid point satisfied it.

H.3 Directions

Four directions, priced by what this paper leaves open. *The mixed problem*: the mixed-operator problem itself, with the bracket of Proposition 27 as the target and the multislope and aggregation parents as the toolkit. *The capacitated partition*: the offline combinatorics of Appendix E.8, where the budget's convexification structure should buy approximation guarantees we have not pursued. *The estimation ledger*: the finite-sample boundary, anchor noise under W2 and mean estimation under W3, where the population prices here become the leading terms and the correction terms are the natural next derivation; E-L2-6's premium measurement is its first data point.

The randomized constant: $e/(e-1)$ is the optimal randomized constant for the constant-rate acknowledgment problem, the lower bound Seiden (2000)'s and the algorithm Karlin et al. (2001)'s, and the classical instances embed here (Appendix F.1), so the lower bound transfers. The upper bound does not, and we can now say why rather than only that we had not proved it. Our deterministic constant carries an additive $w^{\max}/K$ from the overshoot, and the natural conjecture was that the randomized one carries the same correction. That conjecture is *false*, and the witness is small enough to check by hand.

Work in continuous time, where $w^{\max}/K = 0$, so the overshoot term is not what is being measured, and read the classical randomized algorithm in the coordinate in which it is a legal policy here: after each acknowledgment at s, draw θ with density $e^x/(e-1)$ on $[0,1]$ and acknowledge at the first t with $A(s,t) = \theta K$, where $A(s,t)$ is the delay a phase beginning at s has accrued by t. In the accumulated-delay coordinate this *is* the classical algorithm, which is the reason to expect the transfer. Take $K = 2$ and two items: one arriving at time 1 and accruing at unit rate on $[1,2]$ only, one arriving at time 3 and accruing at unit rate on $[3,4]$ only. Each stops on its own; the offline optimum never acknowledges and pays the total accrual, $\mathrm{OPT} = 2$. For the algorithm, two identities do all the work: on any phase whose total remaining accrual is $C \leq K$ its expected cost is exactly $\frac{e}{e-1}C$, and the second item is invisible to the first acknowledgment when that acknowledgment falls before time 3, which happens with probability $(\sqrt{e} - 1)/(e-1)$. Hence

$$\frac{\mathrm{E[ALG]}}{\mathrm{OPT}} \;=\; \frac{e}{e-1}\Big(1 + \frac{\sqrt{e}-1}{2(e-1)}\Big) \;=\; 1.8806\ldots \;>\; \frac{e}{e-1} = 1.5820\ldots \tag{33}$$

The mechanism is specific to a rate that varies: an acknowledgment here is *worthless*, because the accrual it would have stopped was going to stop by itself, and the algorithm cannot tell that prefix from one in which the same item accrues forever.

At a constant rate an item's accrual stops only by being acknowledged, which is exactly the step the classical analysis uses. Nor is the gap a small constant. On instances built from many short pulses, each stopping on its own, the same family's expected cost approaches K times a renewal count with mean increment $K/(e-1)$, so its ratio rises toward e; a direct computation of the renewal equation gives 2.70 at the largest total accrual we evaluated, against $e = 2.7183$.

What stays open, and is now the sharper question, is the optimal randomized constant at time-varying rate: it is at least $e/(e-1)$ because the classical instances embed, we know no algorithm attaining it, and the pulse family does not lower-bound every algorithm, since on the pulse family never acknowledging is optimal. The exploratory script that produced the two-item figure, its exact backward induction, its independent Monte Carlo check and its reproduction of $e/(e-1)$ on the classical class, ships in the supplement, and the renewal-equation figure of 2.70 is computed by `verify_numbers.py`; none of it is a registered experiment or a measurement of anything, and the closed form above is the part that matters.

Learning-augmented schedules: with demand predictions available, the robustness-consistency trade-offs of the predictions line (Purohit et al., 2018) apply to the period choice and the anchor switch, and the closed forms here supply the robust side exactly. None of these is begun in this paper; each has its first stake in the ground here.

Parent literatures, pinned. The per-item three-option structure (rent, a cheap purchase with post-purchase quality loss, an expensive clean purchase) is a three-slope instance of multislope ski rental (Lotker et al., 2012), and its first two options, rent against a purchase that keeps paying a reduced rate, are the two general options problem of Lotker et al. (2008); the discounted-flow variant is competitive leasing (El-Yaniv et al., 1999), and interval-purchase versions are the parking permit problem (Meyerson, 2005). The batching side lives in the aggregation-with-delays hierarchy: multilevel aggregation is equivalent to TCP acknowledgment at depth one and to joint replenishment at depth two (Bienkowski et al., 2016; 2020; 2021), with online joint replenishment anchors (Buchbinder et al., 2008; Bienkowski et al., 2014), deadline variants (Buchbinder et al., 2017), and general delay frameworks (Azar & Touitou, 2019; 2020). In that vocabulary, the mixed problem is a priced depth-one aggregation problem whose batch quality is set by statistical geometry, a combination the hierarchy does not currently contain. The learning-augmented direction (Purohit et al., 2018) is the natural refinement when demand predictions are available. Appendix H.2 settles the problem's order; the constant between $\max(2, 1 + P/2)$ and $1 + P$ is what remains open.

I THE PRE-REGISTERED EMPIRICAL PROGRAM

This appendix reproduces the registered designs, statistics, and falsifiers as frozen before any run, followed by the amendment log and the outcomes. The registration document itself, with its freeze header, SHA-256 escrow, and timestamped amendments, ships in the supplement. Table 10 lists every registered criterion with its measured outcome; fired verdicts are in bold.

I.1 THE VENUES, NAMED

Three venues carried every measured number, and this subsection names each one so that a reader can tell what was and was not run before reading a single result. Nothing here is large. The synthetic tiers (E-L2-1 to E-L2-4, E-L2-9, and the four-experiment program E-L2-11 to E-L2-14) run on CPU from seeded generators and touch no pretrained weights at all. The vision venue is seven *frozen* backbones over two corpora: frozen means the trunk is loaded at its published checkpoint, its classifier head replaced by the identity, and every reported quantity computed on the extracted features, so no backbone parameter was ever updated by this paper. The language venue is a single model. We make no claim about behaviour at larger scale, and the negative results below are read at the scale in Table 11 and nowhere else.

The language venue, in full. E-L2-8 writes into `EleutherAI/pythia-160m`, a 160M-parameter decoder, through LoRA adapters. 100 topics, 20 statements per topic, 200 optimizer steps per write at learning rate 0.001, 600 adapter fits in total across the three operators, three seeds and the registered conditions. The base model is frozen and only the adapters are fit, which is what makes the serving-side asymmetry of amendment L2-A86 possible: W3 ends holding 10 resident adapters against the joint refit's 1. This is one model at one scale, and the reversal it shows is reported as a measurement at that scale rather than as a property of the operators.

What is not here. No experiment in this paper trains a backbone, a transformer trunk or any model above 160M parameters; no corpus exceeds 100 classes or 20,000 staged feature rows per cell; and

Table 10: Registered experiments, targets, measured outcomes, and falsifier verdicts (canonical seeds $11, 13, 17$). K-L2-n is experiment E-L2-n's falsifier, and fired verdicts are in bold: six fired, K-L2-4 and K-L2-9 for one cause, K-L2-11a, K-L2-12b, and K-L2-13a and K-L2-13b for one cause. The corrected theorem's post-hoc program is the E-L2-4b row.

ID	Target (closed form)	Measured; verdict
E-L2-1 ledger exactness	no cell beyond 4.5 s.e., at most 5 beyond 3 (L2-A7); ray equality	max 3.45 s.e. over 345 cell-seeds, 1 beyond 3 s.e. (0.93 expected), 0 family-wise; K-L2-1 not fired
E-L2-2 dominance	merged pairs improve, bystanders worsen, budget convexifies (forced limbs marked in App. I)	0 of 75915 merged pairs worsened; 2160 of 2160 bystanders worsened, 0 improved; convexity 4.1×10^{-15}; K-L2-2 not fired
E-L2-3 sleep law	$T^* = \sqrt{2K/(\alpha\bar{\rho})}$; elasticities $\pm\frac{1}{2}$	T_d in band in 255 of 255; elasticities 0.480, -0.480 vs 0.497, -0.499; K-L2-3 not fired
E-L2-4; E-L2-9 over the full lattice	plateau at $\min(r, \ell\phi_{\mathrm{bal}})/(\ell e_{\mathrm{al}})$, with ϕ_{bal} the mismatch floor of Lemma 24, the earlier theorem's bound	no plateau: ratio 999.0 against target 7.48 at $\gamma = 9$, anchored arm 317.0; the lattice minimum grew with the flood in 10 of 12 cells and plateaued at $2.4/5.3$ times the registered level in the other 2 (K-L2-4 **FIRED**, K-L2-9 **FIRED**)
E-L2-4b; E-L2-9b lattice (post hoc)	growth at least (4), anchored at most 2; the corrected rate at asymptotic flood	both bounds respected in all 150 cells (yes), tight against the fixed-composition rate in 72, anchored 2.000, brute force matched on $9000/9000$; the lattice minimum saturates at $1000.0 = rQ_1/K'$ in all 36 cell-seeds, the best policy writing nothing in stage one, so divergence needs Q_1 to grow too
E-L2-5/6/7 reduced venue	ledger reach: gauge direction, coverage effect, floor, level	margin falsifier 0 of 2400, 138 eligible (K-L2-5 not fired); any-margin 51.4%: no floor; gauge cosine 0.913 vs 0.642, ratio 0.7031 vs centring 0.7071 (weak); covered 0.0072 vs uncovered 0.0454, better/tied/worse 64/8/3 of 75, -7.7 s.e.; E-L2-7 on real means: 0/60 worsened, forced (App. I; K-L2-7 not fired)
E-L2-5/6/7 accel.; E-L2-10 map (8 cells)	the same at 100 classes; K-L2-10a/b: gauge and coverage on a majority of cells; K-L2-10c: no merged pair worsens	K-L2-5 not read (no eligible pair); gauge nearer on 8/8 cells, ratio 0.9496 to 0.9509 vs centring 0.9487 (K-L2-10a not fired, weak); level 0.0001 vs 0.0183; coverage 7/8 by mean, 2 beyond 2 s.e. (K-L2-10b not fired); 0 of 12000 merged worsened, 11612 unchanged at tolerance (K-L2-10c not fired)
E-L2-8 LoRA writes	measurement; the draft expected W3 worst, W1 best	reversed: 1.7246, 1.8185, 1.9300 in every seed, seed-level $t \geq 4.1$; time premium nil by design
E-L2-11, 13	coverage removes half the excess; ledger selection beats random and recency	depth 64 leaves $4.80/1.16/1.19\times$ the excess over 3 of 6 store sizes, against a baseline indistinguishable from zero; premium $102.88/106.04/99.90$ vs 100.0000; ledger rank correlation ≈ 0 (K-L2-11a **FIRED**, 11b not fired and read on 0 comparisons, 11c not fired; 13a **FIRED**, 13b **FIRED**)
E-L2-12, 14	no bound breached; A_{K_3} beats break-even; period in band; the construction exists and its ratio grows	break-even 2.2049 vs 3.7731; A_{K_3} wins on none of 15 streams at volume 10^4 (6 of 15 at 10^6) at premiums $147.1/155.8/152.2$; excess 0.2473 vs closed form 0.4943, ratio to 775.2295 at $4.0/4.9/4.4\times$ its predicted level, anchored 2.0000 (12a not fired, 12b **FIRED**, 12c not fired; 14a not fired, 14b not fired)

Table 11: The frozen-backbone venue (E-L2-5 to E-L2-7 at the anchor cell, E-L2-10 across the map), one row per cell. Every column but the checkpoint source is injected by the gate from that cell's own digest-verified manifest; the two source enumerations are not recorded in the manifests and are stated here from the pinned runner `l2_features.py`, which selects `IMAGENET1K_V1` for each torchvision trunk and the timm tag `vit_base_patch14_dinov2.lvd142m` for DINOv2. Samples are the feature rows staged per cell. The anchor cell carries twenty partitions and the sweep cells five, the axis cells existing for cross-cell comparison rather than within-cell precision.

backbone	source	weights	corpus	dim	samples	venue
ViT-B/16	torchvision	IMAGENET1K_V1	CIFAR-100	768	20,000	accelerator
ConvNeXt-T	torchvision	IMAGENET1K_V1	CIFAR-100	768	20,000	sweep
DINOv2-B/14	timm	LVD-142M	CIFAR-100	768	20,000	sweep
EfficientNetV2-S	torchvision	IMAGENET1K_V1	CIFAR-100	1280	20,000	sweep
ResNet-50	torchvision	IMAGENET1K_V1	CIFAR-100	2048	20,000	sweep
Swin-T	torchvision	IMAGENET1K_V1	CIFAR-100	768	20,000	sweep
ViT-B/16	torchvision	IMAGENET1K_V1	Food-101	768	20,000	sweep
ViT-L/16	torchvision	IMAGENET1K_V1	CIFAR-100	1024	20,000	sweep
bundled digits	scikit-learn	n/a	digits	64	1,797	reduced

no result is claimed to hold at a scale larger than the one it was measured at. The theory is scale free and the measurements are not, and where the two disagree the measurements are reported as what happened (Section 9).

I.2 PROVENANCE AND INFRASTRUCTURE

Numbers enter this manuscript through canonical CSVs produced by registered runners, carried by manifests, and injected by `fill_L2.py` into the named slots of Section 9 and the tables below. Two qualifications, stated here rather than left implicit. The gate signs the CSV, but the values it prints are read from the manifest's verdict block, and until amendment L2-A19 nothing covered that block: an adversarial audit of the executed program edited a manifest, left the CSV untouched, flipped K-L2-4 from FIRED to "not fired", and every instrument reported success. Since L2-A19 the gate hands the digest-verified rows to a registered analyzer, `l2_verdicts.py`, which recomputes the verdict block from those rows following each runner's own definition and compares it field by field; any disagreement aborts the build, naming the field, the manifest value and the recomputed value.

On the executed program this recomputes 1014 verdict fields across 42 canonical CSVs. 19 fields cannot be recomputed and the gate lists them by name on every run (l2_e10.cells_skipped, l2_e13.ledger_rank_correlation, l2_e13.ledger_rank_correlation_offsets, l2_e5.gauge_cos_ledger, l2_e5.gauge_cos_raw, l2_e5.max_grad_norm, l2_e8.NOT_CANONICAL, l2_e8.adapter_fits, l2_e8.batch_size_sign, l2_e8.by_condition, l2_e8.headline_arm, l2_e8_grid.NOT_CANONICAL, l2_e8_grid.by_operator, l2_e8_grid.by_operator_note, l2_e8_grid.headline_arm, l2_e8_grid.operator_separation, l2_e8_grid.operators_separated, l2_e9.max_gap_extrapolated, l2_e9b.max_gap_extrapolated): the reduced venue's gauge cosines, accumulated in the fitting loop of a runner that wrote them to no column (the accelerator runner writes them per pair, and there they are recomputed); every venue's maximum gradient norm, a property of the optimizer; the extrapolation maxima of E-L2-9 and E-L2-9b, built from live loop state and not printed; E-L2-8's condition summaries, batch-size sign and fit count, each of which one of its two CSVs carries for the other and is recomputed there, together with its prose fields; and E-L2-10's skipped-cell list, a skipped cell writing no row. Of the rest, one field, E-L2-5's kill margin, is a threshold frozen in the registration and is checked against it in each of nine files; two more, E-L2-1's family-wise threshold and E-L2-3's band width, are the registered constants of L2-A7(a) and (b), which the analyzer neither recomputes nor checks and this appendix prints from the registration rather than from the manifests. 28 further counts are pinned by the CSV only to an interval, the runner accumulating them at full precision while the CSV carries five decimals, and the gate reports the interval rather than asserting an exactness it does not have. The six brute-force statistics of E-L2-4 and E-L2-4b are recomputed from a sidecar CSV that carries no manifest digest, so for those the recomputation establishes arithmetic agreement with the runner and not tamper-evidence.

Second, what is recomputed is the summary, not the rows: the Monte Carlo draws, head fits and closed forms behind each row still rest on the runner and library pins together with the CSV digest.

Smoke seeds $\{0, 1, 2\}$; real seeds $\{11, 13, 17\}$; the sets are disjoint and smoke seeds are never canonical; audit instruments from the theory phase ran on container-local seeds excluded from canon, and their role is recorded as verification notes only. Runners: `l2_lib.py` (the Gaussian model, the exact closed forms of Lemma 11, the batch arithmetic of Theorem 1, the two-phase machinery, the exact schedule evaluator and the enumerated and brute-force optima), and `l2_runners.py` with one subcommand per CPU-tier experiment (`e1`, `e2`, `e3`, `e4`, `e4b`, `e9`, `e9b`; the anchor and program tiers run from `l2_features.py` and `l2_program.py`, whose manifests name them), each idempotent per cell, each writing its CSV with a manifest recording code hashes, seeds, grid, tolerances, library versions, timestamps, and the falsifier verdicts. Analysis only through the registered analyzers, and the gate prints no summary it could not re-derive without saying so.

What the manifest timestamps describe. Six canonical CPU-tier manifests carry an identical start and finish time with a resume count equal to their full cell count (E-L2-4b's, a seventh, carries a one-second window with the same property), and the three manifests of the accelerator venue's anchor cell carry a two-second window with a resume count equal to their full chunk count. The runners are idempotent per cell, so the final invocation of each found every row already present, computed nothing, and rewrote the manifest; the timestamps therefore describe that pass and not the computation, which the files' own modification times place across the preceding three quarters of an hour for the CPU tier and which the chunk record places for the anchor cell (its Stage A computed the sixty chunks; the reach map's Stage B then re-opened the cell and re-stamped it). This is what an idempotent resume does and no row is affected, but the field names imply otherwise, so we state

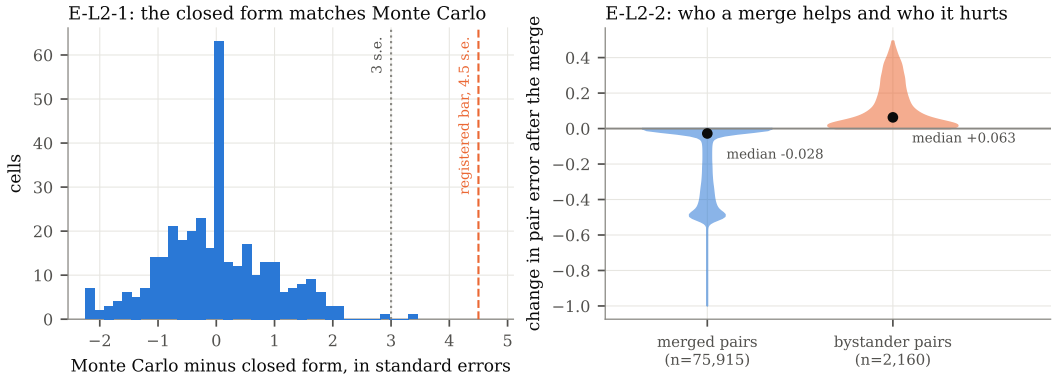


Figure 6: Measured, from `l2_e1.csv` and `l2_e2.csv`. Left, E-L2-1: the deviation of the Monte Carlo estimate from the closed form, in standard errors, one count per cell, against the registered bar of 4.5 s.e. and the library's 3 s.e. tolerance. The distribution is centred and no cell reaches the bar, which is K-L2-1 not firing. Right, E-L2-2: the change in pair error after a merge, for the pairs the merge covers and for the bystander pairs it does not, as densities with medians marked. The merged pairs improve and the bystanders degrade, which is the dominance sign with the cost the statement carries; the asymmetry in the counts is the design's, most pairs in a merge being merged pairs.

it rather than leave a reader to infer it. `check_submission.py` reports the condition on every build, and amendment L2-A21 registers the field-semantics repair, deferred for the same reason as the one below: the writer lives in the pinned library. A related gap belongs here: the surviving CPU-tier run logs are the discarded pre-L2-A12 run, not the reported one (the reduced anchor venue's post-L2-A12 log does survive, with verdicts identical to its shipped manifest), so the log directory is a record of a superseded execution and does not ship.

Instrument disclosure, and why it is not repaired. A final audit found that `l2_lib.py`'s `exceptional_hit` tests $\beta \in [0, 1)$ where Proposition 15 gives the ray $\beta < 1$: the shipped predicate implements the segment, and would classify Remark 8's own counterexample, which sits at $\beta = -B$, as non-exceptional. The mathematics in this paper is the ray; the library is what is wrong. It is not repaired, and the reason is the discipline rather than the effort: `l2_lib.py`'s hash is carried by every canonical manifest, so changing one character invalidates the whole executed program and would require re-running it under a new pin. No printed number moves, since the predicate sets a verdict *label* on cells whose exceptional status E-L2-1 determines from its own grid, and a norm-equal, exactly-parallel, backward-ray hit has measure zero under E-L2-2's random draws, which the analyzer's recomputation confirms cell by cell. So we record the defect, leave the pin intact, and repair it in the next round with the re-run that requires (amendment L2-A20). Compute class: CPU (two cores, under one hour for the full canonical program). Tolerances are fixed in `l2_lib.py` at the freeze and recorded in every manifest; they do not move afterward. Every falsifier outcome, fired or not, is reported in the manuscript, including fired falsifiers as stated negative results.

I.3 E-L2-1: EXACTNESS OF THE LEDGER

Design. Grid over $(\gamma, \|\bar{D}\|,$ direction of $\bar{D}$ relative to δ and $\mu_k, \bar{\Delta})$, with dedicated cells for: the degenerate direction $\bar{D} = \delta$; the exceptional segment (norm-equal pairs, $\beta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$); the reversed ray $\beta > 1$; norm-unequal pairs at the same β; and off-segment generic directions at several norms. The five degenerate cells are constant-rule cells: the sampler short-circuits to the closed form when the induced direction vanishes, so their fifteen cell-seeds agree by construction, are required to agree exactly, and are excluded from the deviation maximum (amendment L2-A31(f); the family-wise constants are computed on 345 rather than 330, the conservative direction). Monte Carlo conditional errors at fixed mismatch versus the closed forms (12), the balanced averages, and the exceptional-set predictions of Proposition 15 (equality on E, strictness off it). **Statistic.** Per-cell deviation of measured from closed form, in Monte Carlo standard errors. **Falsifier K-L2-1**, as amended by L2-A7(a) and relabelled by L2-A12(d), which is the form the verdict is read against.

The registered form was per-cell: any cell deviating beyond three standard errors kills Lemma 11 as stated. A grid of this size produces such cells under the null, so L2-A7(a) replaced it, before the run, with a family-wise criterion in two limbs: it fires if any cell exceeds 4.5 standard errors, *or* if more than five cells exceed three, five being the binomial 99.9% upper bound for 345 tests at $p = 0.0027$. Both limbs are read: 1 cells were flagged beyond three standard errors against 0.93 expected under the null, 0 beyond 4.5, and the measured maximum of 3.45 standard errors is above three and below the family-wise threshold, which is the situation the amendment anticipated. The per-cell form kills nothing; an exceptional-segment cell failing the equality prediction kills Proposition 15. **Output.** `l2_e1.csv`; slots `E-L2-1:MAXDEV`, `E-L2-1:EQCELLS`.

I.4 E-L2-2: DOMINANCE AND THE BYSTANDER

Design. Random geometries (class means drawn from separated mixtures), random admissible partitions (parts of size at least two), random merge sequences; plus the constructed bystander example of Appendix E.4 at random placements. **Predictions.** Merged-pair balanced errors weakly improve, strictly off the exceptional set; third-party mismatch vectors satisfy the convex-combination law exactly; bystander pairs strictly worsen off the codimension-$(d-1)$ coincidence set; the budget $\rho(P)$ is nonincreasing under every merge. **Falsifier K-L2-2**, as registered: any merged pair strictly worsening off the exceptional set, any bystander pair strictly improving, any budget increase under a merge, or a one-sided distribution of third-party effects beyond tolerance, kills Theorem 1 as stated; the tolerance is the sign tolerance of 10^{-3} on error differences, there being no Monte Carlo in this experiment (every quantity is closed-form arithmetic, amendment L2-A2). The third-party statistic is taken on a subsample, one representative pair per third-party batch per merged side, two per merge, rather than the full cross product the frozen text said, a deviation found by audit after the run and logged as amendment L2-A15; the merged-pair, bystander and budget statistics are over the full sets. **Output.** `l2_e2.csv`; slots `E-L2-2:SIGN`, `E-L2-2:CONV`.

What K-L2-2 can and cannot kill, and at what resolution (amendment L2-A31). An adversarial audit of the executed lane, after the fact, found that this falsifier is weaker than its registration reads, in three separate ways, and we state them rather than restate the verdict. First, the convexity limb is an identity, not a prediction: the merged anchor is by construction the size-weighted mean of the two anchors it replaces, so $\bar{D}_{(TU)V} = \lambda \bar{D}_{TV} + (1-\lambda)\bar{D}_{UV}$ holds in exact arithmetic for every draw, and the recorded residual 4.1×10^{-15} measures floating-point accumulation, not agreement with a theory. Second, the budget limb follows from that identity: a point on the segment between two anchors is no farther from any third anchor than the farther endpoint, so no pairwise maximum can rise and 0 increases was the only arithmetically possible outcome. Third, the sign limb is Bayes optimality: a merged pair's post-merge error is the Bayes balanced error of that pair and its pre-merge error is the balanced error of one particular linear rule for the same pair, so weak improvement cannot fail; in the constructed bystander geometry the two anchors coincide before the merge, so the pre-merge pair is already at its Bayes value and can only worsen. What this experiment does measure, and what no identity forces, is the *distribution*: that 58.6% of merged pairs improve strictly past the sign tolerance, that 94.0% of bystanders worsen past it, and that third parties move both ways (910 up, 1630 down, 3098 flat), which is the sign structure Appendix E.7's Δ calculus predicts and the only part of K-L2-2 a draw could have refuted. Alongside this: the runner records each difference at six decimals and then applies its strictness test to the rounded value, so the manifest fields named for a 10^{-12} threshold, and the count of worsened pairs printed in Table 10, resolve nothing finer than 5.0×10^{-7}; the manuscript now claims nothing below that, and reports both that 29.1% of merged pairs are separated far enough for both sides of the comparison to print as zero and that the smallest recorded bystander worsening, 3.0×10^{-6}, sits clear of the resolution. The full-precision differences are not in the archive and the pinned library hash forbids the re-run that would recover them, so this is a disclosure and a restatement of the claim at the resolution the evidence supports, not a repair.

I.5 E-L2-3: THE PERIOD LAW

Design. Stationary arrival streams at entry rate α, mean rent $\bar{\rho}$, setup K, across a grid spanning two orders of magnitude in each; periodic policies at a dense set of periods; total cost per round by simulation. **Prediction.** The simulated cost is minimized at $T^* = \sqrt{2K/(\alpha\bar{\rho})}$, with the elasticities of Corollary 3. **Falsifier K-L2-3**, as amended by L2-A3 and L2-A7(b), which is the form the verdict is read against. The registered form was per-cell: displacement of the empirical argmin

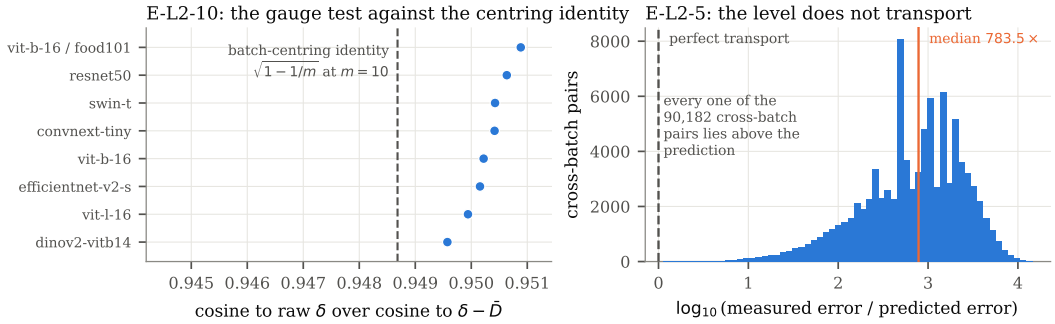


Figure 7: Measured, from `l2_e10.csv` and `vit_b_16__cifar100/l2_e5_accel.csv`. Left, E-L2-10: for each frozen backbone-corpus cell, the fitted rule's mean cosine to the raw δ divided by its mean cosine to the ledger's direction $\delta - \bar{D}$, against the batch-centring identity $\sqrt{1 - 1/m}$ at $m = 10$ classes per batch (dashed, 0.9487). The ledger's direction is the nearer of the two on every cell (8 of 8), and the ratio sits at the identity on every cell (0.9496 to 0.9509), which is the value a rule along $\delta - \bar{D}$ gives for isotropic means in batches of m: the test certifies that the fit keeps the batch-anchored gauge of Lemma 7, which is why K-L2-10a is reported as not fired and weak rather than as confirmed. The axis is truncated to the occupied range, which a dot plot may do and a bar chart may not. Right, E-L2-5: the distribution of $\log_{10}$ of the measured pair error over the ledger's predicted pair error, across every eligible cross-batch pair at the accelerator venue. Perfect transport is the dashed line at zero; every pair falls above it, at a median of 783.5×: the level does not transport.

beyond one grid cell plus tolerance anywhere on the grid kills Corollary 3 as a schedule-selection rule (the closed-form arithmetic itself is exact). "Plus tolerance" was undefined, so L2-A3 made it the statistical-resolution band and L2-A7(b) widened that band from two to five standard errors of the paired difference, family-wise over the hundred periods and 255 cell-seeds by a Bonferroni–Šidák argument, both before the run. That is a loosening and we name it as one, unlike L2-A12(d)'s relabelling; 255 of 255 cell-seeds place the exact discrete minimizer inside the band, 245 of them also within a single grid cell; the rest are reported as such rather than counted against the amended criterion, every one of them lying inside the band. **Output.** `l2_e3.csv`; slots `E-L2-3:ARGMIN`, `E-L2-3:ELAST`.

I.6 E-L2-4: THE TWO-PHASE DEMONSTRATION

Design. The placement of Appendix G.4: γ sweep at fixed separation s (with $u \perp \delta$), and s sweep at fixed γ; the pure-W3 policy family (write thresholds for both stages, including never-write, thirty-six policies); the anchored break-even policy; offline optimum by brute force over write sets on small instances (twelve rounds, three hundred streams per cell). **Predictions.** Evaluated pure-W3 ratios (exact closed-form costs, no Monte Carlo, amendment L2-A4(v)) plateau at the closed-form $R(\gamma) = r/(\ell\,\Phi(-\sqrt{\gamma}/2))$ in the closed-form regime ($\ell \geq 2r$) as the flood lengthens; the plateau climbs the γ sweep; the best flood-lengthening policy rents k forever exactly when the numerator's min selects rent; the anchored policy's ratio stays below 2 everywhere; the floor exceeds $1/2$ and tends to $1/2$ with the separation per (27). **Falsifier K-L2-4.** A missing plateau (ratios not flattening at R within tolerance), a plateau at the wrong level, or an anchored ratio exceeding 2 plus tolerance, kills the theorem's applicability as stated. **Output.** `l2_e4.csv`; slots `E-L2-4:PLATEAU`, `E-L2-4:ANCHORED`. **Registered as written; outcome: FIRED.** The registration above was frozen with the theorem's earlier form, whose bound $R = \min(r, \ell\phi_{\text{bal}})/(\ell e_{\text{al}})$ presupposed the gross cost convention and an optimum that writes only the flooding pair; the evaluated curves did not plateau and the anchored arm's ratio under the gross convention exceeded two as well (317.0 against the registered 2), both limbs firing; the cause is in the proof's optimum line, and the corrected theorem with its post-hoc program is recorded below (amendment L2-A8). The prediction table of Appendix I.22 is kept as registered and marked falsified.

I.7 E-L2-5: THE FROZEN-FEATURE ANCHOR

Design. One standard frozen-feature corpus; classes partitioned into real task splits as batches; heads fit per batch split; the vector gauge checked by cross-profile cosine methodology: the estimated cross-batch mismatch directions from fitted heads against the class-mean predictions, and the ordering of measured cross-batch errors against the ledger's e_{bal} predictions. **Prediction.** The ledger's ordering and invariance structure (translation invariance, scalar removability, vector non-removability) hold on real feature geometry up to estimation error. **Falsifier K-L2-5**, made precise by amendment L2-A9(a), which also fixes the fitting protocol the draft left open (each batch's head is a batch-local multinomial logistic fit over that batch's own labels on a random half of the corpus; pair errors are measured on the held-out half): the scope claim dies if more than five percent of cross-batch pairs measure *below* their aligned Bayes value by more than 0.05. (The margin makes this a test of gross violation; the direction itself is forbidden only under the instantiation's own model, and the venue diagnostic of amendment L2-A11(d), the any-margin count with its eligible-pair denominator, is reported alongside the falsifier precisely because on a real corpus the Gaussian aligned value need not be a floor at all.) A systematic level offset between predicted and measured is reported as an effect size rather than a kill, because the population ledger predicts the shape while a finite-sample head fit is an estimation-ledger object by this paper's own scope. The theorems are unaffected either way; the reach claim is what dies. **Venue.** Two are registered. The accelerator venue (a frozen vision backbone, one hundred classes in batches of ten) executed on 2026-09-12 under the authorization of amendment L2-A14, the session recorded as L2-A34, and is reported below, together with the reach map it anchors. The reduced venue of amendment L2-A10, added before any run of this experiment and changing no criterion, is a bundled handwritten-digit corpus of 1797 samples over ten classes in sixty-four raw pixel features, with batches of two classes; its class conditionals are neither Gaussian nor equal-covariance and several of its coordinates are constant, which is the point: it is real geometry that the instantiation does not describe. Results from the reduced venue are never reported as the accelerator tier. Amendment L2-A11, logged before the canonical reduced-venue run and after an adversarial audit of a first run whose CSVs were discarded, registered the optimizer budget (4000 full-batch steps, recorded terminal gradient norm), implemented the registered cosine gauge (slot `E-L2-5:COS`), added the any-margin diagnostic with its eligible-pair count, gave each experiment its own random stream, and hashed the corpus into every manifest. **Status.** executed on the reduced venue; executed on the accelerator venue, where the falsifier is not read (no eligible pair) (below). **Output.** `l2_e5.csv` and `vit_b_16__cifar100/l2_e5_accel.csv`; slots `E-L2-5:COS`, `E-L2-5:ORD`.

I.8 E-L2-6: THE ANCHOR PREMIUM

Design, as amended by L2-A9(b) to measure coverage and not only anchor size: for each cross-batch pair, three heads at the same batch composition, W3 (the batch alone), W2 with a replay anchor that covers the pair's foreign class, and W2 with an anchor of equal size that does not. This is the distinction Proposition 5 turns on, and separating the arms is what makes the premium interpretable. **Measurement.** Realized cross-batch error of each arm against the aligned value, and the extra data touched; the premium calibrates $K_2 - K_3$ for the mixed-operator problem. No falsifier. Amendment L2-A11(c): at the reduced venue the registered replay fractions both round to a one-class anchor, so size had a single level and duplicate configurations; the factor is now anchor size in classes, $k \in \{1, 3\}$, both arms carrying exactly k, and matched-pair counts are over distinct configurations. **Status.** executed on the reduced venue of L2-A10; executed on the accelerator venue (below). **Two numbers the gate emits that do not flatter the anchor (L2-A32).** The W3 arm, no anchor at all, measures a mean cross-batch error of 0.0451; the equal-sized non-covering W2 anchor measures 0.0454. On this corpus a replay anchor that does not cover the pair's foreign class buys nothing over no anchor, which is the coverage claim stated from the other side. And the covered arm's 0.0072 sits below the Gaussian aligned value 0.0358 in 75 of 75 matched rows: the floor does not transport to this geometry (Section 9), and a covered pair measuring under the aligned value is that failure, not the anchor outperforming Bayes. Both macros existed in the gate's output and were not printed before this amendment. **Output.** `l2_e6.csv`; slots `E-L2-6:PREMIUM`.

I.9 E-L2-7: DOMINANCE ON REAL FEATURE GEOMETRY

Design, registered by amendment L2-A9(c). The bystander of Example 2 needs a pre-merge *aligned* pair, which random partitions of a real corpus do not supply; on real features the pairs across two unmerged batches are third parties, whose realized errors may move either way by Theorem 1 itself, so a criterion that fired when they improved would fire on a correct prediction. E-L2-7 therefore tests what the theorem does predict on real geometry: merged pairs improve weakly, the mismatch budget never rises, the convex-combination law is exact, third parties move in both directions, and the bystander effect is reported conditionally on the lowest decile of pre-merge mismatch (of the third-party rows; the denominator is in the manifest), the nearest real analogue of an aligned pair. Amendment L2-A11(g), binding on the report: every E-L2-7 quantity is the ledger's closed form evaluated at the corpus's own class means, no held-out sample enters it, and the convexity and budget checks are identities of the merge construction that test the code path rather than the corpus, as is the sign check, since at the corpus's means the post-merge error is the Bayes value of the Gaussian form and the pre-merge error that of one linear rule (L2-A31(b) applied here; the corpus enters through its means alone). **Falsifier K-L2-7.** Any merged pair strictly worsening beyond the sign tolerance, any budget increase, or a convexity residual above 10^{-6}. **Status.** executed on the reduced venue of L2-A10; executed on the accelerator venue (below). **Output.** `l2_e7.csv` and its `_accel` and `_sweep` counterparts.

I.10 E-L2-8: THE WRITE-OPERATOR DEMONSTRATION ON A TRAINED MODEL

Design, registered by amendment L2-A9(d) and made executable by L2-A13(b). Topics arrive in batches and each write fits a LoRA adapter; the three operators differ in *which adapter serves a topic at probe time*. W3 trains a fresh adapter on its own batch and a topic is served forever by the adapter that wrote it, previous adapters frozen and routed per batch. W2 trains a fresh adapter on its batch plus a replay anchor drawn from previously written topics, and the topics the anchor *covers* are re-routed to the new adapter while the rest keep their old one, which is coverage in the sense Proposition 5 uses. W1 retrains one adapter on everything written so far and every topic is served by it. **Measurement.** Probe loss per topic per round on held-out statements of each topic, the revisit count, and the item and wall-clock premium; the routing is recorded per row so the arms can be checked to differ rather than assumed to. **No falsifier**: this tier measures. **What the routing costs, priced.** The three arms differ at probe time as well as at write time, and both differences favour the cheaper operator. At the last round W3 is served by 10 resident adapters and W1 by 1, so the unanchored arm holds the larger model at serving; and at a fixed per-write step budget W1's single adapter sees the whole store in the same number of steps, 10 times fewer gradient steps per item than W3's. Both are consequences of the operator definitions rather than instrument faults, and both are why this tier carries no falsifier: at matched steps the comparison prices the budget and the routing together with the operator, and separating them needs a capacity-matched and pass-matched arm this program did not register. **Instrument disclosure.** The first implementation of this runner stored one model object under every topic key, so all three operators probed whichever adapter had trained last and W3's routing, which is the whole content of the unanchored operator, was not implemented; amendment L2-A13(b) records the repair and the runner now ships a dry-run mode that exhibits the three arms' revisit counts before any accelerator time is spent. **Two joint refits diverged, and the routing spread each one.** Drawing the per-round distributions for Figure 3 showed what the round-level summaries had not: at rounds 7 and 8 the W1 probe loss is bimodal, a fraction of topics far above the rest (37 of 240 and 40 of 270 of the round's topic-probes above a loss of 3) while every other round and every other arm is tight. Each affected row traces to one serving adapter, and there are 2 of them over the three seeds, which is 2 of 30 refits; seed 11 has none. The cause is a diverged refit rather than a property of those topics, and the reason it is visible at all in a per-topic measurement is the asymmetry priced above: W1 serves every topic from one resident adapter, so a single bad refit reaches every topic at its round, where an unanchored write would have confined it to its own batch. Nothing here is tuned away, and nothing here moves a reported number: the headline losses and every separation are read at the last round, which carries no affected row (confirmed from the rows). Figure 3 plots medians with 5th-to-95th whiskers for this reason, a mean over a bimodal round describing no topic, and marks the two rounds rather than hiding them. Amendment L2-A88 logs the finding. **Status.** executed. **Output.** `l2_e8_accel.csv` for the headline condition and `l2_e8_grid_accel.csv` for the full grid, split as amendment L2-A17(e) requires so that the

gate's verdict analyzer recomputes one condition rather than an average over the grid file's three replay ratios and two batch sizes.

Result (executed 2026-09-12 under L2-A14's authorization, session L2-A34: 600 adapter fits over 8 conditions; the headline condition is batches of 10 at replay ratio 0.25, 3 seeds, 10 rounds). Final mean probe loss and worst-topic loss by operator, with the three-seed totals of re-routings and training items and the elapsed seconds of one seed's condition (probes included; the runner records no adapter-fit time): W3 1.7246 / 2.2945, 0 re-routings, 6000 items, 24 s; W2 1.8185 / 1.9991, 336, 8688, 24 s; W1 1.9300 / 2.0420, 1350, 33000, 23 s. The routing is as registered: W3 never re-routes a topic, W2 re-routes what its anchor covers, W1 everything. **The pre-run expectation reverses.** The registration makes this a measurement with no prediction (L2-A9(d)); the manuscript's pre-run draft stated an ordinal expectation, W3 worst and W1 best on loss with the item and time premium running the other way, and that is what reverses. Measured, W3 is best and W1 worst in every seed (per-seed means 1.767/1.710/1.697, 1.816/1.817/1.822, 1.929/1.934/1.927). Paired by (seed, topic) over 300 pairs the differences are -0.2055 ± 0.0062 for W3 against W1, -0.1116 ± 0.0050 for W2 against W1 and -0.0939 ± 0.0065 for W3 against W2, all 3 beyond two standard errors; but that pairing treats each fit as fixed, and one W1 adapter serves every topic of a seed, so the unit of replication for a fit-level effect is the seed. Per seed the differences are $-0.162/-0.224/-0.230$, $-0.112/-0.117/-0.105$ and $-0.050/-0.107/-0.125$, of one sign in every case, with t statistics of 9.4, 33.1 and 4.1 on two degrees of freedom: the reversal is robust, W3 against W2 the least secure of the three, and the per-topic standard errors above overstate the precision (the null comparison under Batch size below calibrates the fit-to-fit noise at about -0.0360). On the worst topic the order is not stable across seeds: W3's worst-topic loss by seed is 2.29/1.84/1.79 against W2's 1.95/1.99/2.00 and W1's 2.03/2.03/2.04, so the one seed in which a W3 adapter fit badly owns the headline worst-topic figure, and in the other two W3 has the best tail as well. **Why the sign, and what the instrument can show.** Amendment L2-A17(a) fixed every write at 200 optimizer steps whatever it trains on. Two consequences. The wall-clock premium is nil by construction, the three operators taking 24, 24 and 23 seconds per seed-condition, so the only premium this instrument exhibits is items touched (33000 for W1 against 6000 for W3). And the joint refit spreads a constant budget over a growing store: the loss of the batch written in a round, measured at that round and averaged over seeds, is 1.708 for W1 at the first round, where it coincides with W3's 1.708 because the first write is the same fit, and 1.926 at the last, peaking at 2.355 where single-seed fits went bad (the worst, 3.17 at round 9), while W3's stays between 1.693 and 1.821 (worst single-seed fit 2.07). The reversed order is a statement about the operators at matched optimizer steps, not at matched passes over the data; a matched-epoch protocol is a different registration and none was run. **Interference.** A topic's loss at the final round minus its loss at the round it was written, over the topics that were re-routed at least once: none for W3, which never re-serves a topic (drift 0); 0.1070 ± 0.0043 over the 225 topics W2's anchor re-routed, 95% of them worsened, so on this instrument re-routing costs the covered topic; -0.0059 ± 0.0230 over the 270 topics W1 re-served, zero on average with 77% worsened and most of the rest improved, the joint refit moving every topic every round. **Replay ratio** (W2 at batches of 10, ratios 0.10, 0.25, 0.50): mean loss 1.7749, 1.8185, 1.8668, worst-topic loss 2.3343, 1.9991, 2.0636, at 135, 336, 675 re-routings: more coverage, more dilution, higher mean loss at fixed steps. **Batch size**, the registered batching-dominance sign, full batches of 10 against batches of 2 at the headline ratio, the difference of final mean losses paired by (seed, topic) over 300 pairs: W3 -0.0082 ± 0.0166 (by seed $+0.063/-0.017/-0.071$), W2 0.0124 ± 0.0068 ($-0.001/+0.003/+0.036$), W1 -0.0360 ± 0.0037 ($-0.036/-0.040/-0.032$). The W1 entry is not a batching effect and cannot be read as one: the joint refit's final adapter is a fresh fit of every written topic in both arms, 2000 items against 2000, the same statements, so the two arms differ only in the fit's random stream, and the entry is this instrument's fit-to-fit noise on identical data, consistent in sign across seeds because the small-batch arm's last fit follows fifty rounds rather than ten. Read against that calibration, neither W3's nor W2's sign is beyond noise (0.5 and 1.8 per-topic standard errors, with per-seed signs that disagree), and the small-batch W3 arm's worst topic reaches 4.4402 against 2.2945 at full batches, one adapter per two topics being the fragile configuration. The batching-dominance sign is therefore not read on this instrument. **Instrument disclosure.** The runner's separation statistic, and the analyzer's copy of it, key the per-topic dictionary by topic alone, so with three seeds in one file the last seed's rows overwrite the others and the manifest's "topics: 100" is one seed's hundred; the paired figures above are the gate's, over every (seed, topic) pair, and the manifest's single-seed values fill no slot (L2-A36(e)). The tier cannot fire a falsifier and is reported as measured.

I.11 E-L2-5, 6, 7 AT THE ACCELERATOR VENUE

Status. executed (2026-09-12, authorized by L2-A14, session L2-A34). This is a *separate* venue from the reduced one reported in Section 9, and amendment L2-A10 binds us never to report one as the other: the two fill separate macro families from separate files, and the gate holds each manifest to the pin of the venue it records, refusing an accelerator manifest in the reduced venue's slot even under a rename. The protocol is E-L2-5, E-L2-6 and E-L2-7 exactly as registered, at the hundred-class scale of amendment L2-A9 with full cross-batch coverage (L2-A13(c)), on the frozen backbone and corpus of E-L2-10's anchor cell, which this *is* (L2-A14).

Result (270000 cross-batch pairs over twenty partitions of a hundred classes, seeds 11, 13, 17). *E-L2-5.* The registered falsifier is **not read (no eligible pair)**: it counts pairs measuring more than 0.05 below their aligned value, and 0 of 270000 pairs are eligible, because at this venue the Gaussian aligned value never exceeds 0.0083, below the 0.05 margin. The runner's string for that state is "not fired"; a falsifier that examined no pair has not passed, and we report it as not read (L2-A36(c)). The any-margin diagnostic, 38270 pairs (14.2%) below the aligned value, is exactly the count of pairs with zero held-out error, 38270, and that is arithmetic rather than coincidence: the aligned values have median $< 10^{-5}$ against a held-out resolution of 0.0063 (at least 79 held-out samples per class; one mistake moves a pair error by half the reciprocal), only 0.03% of them reaching that resolution, so a pair can measure below its aligned value only by measuring zero. At this venue the floor statistic measures separability and says nothing about the floor. *The gauge.* The fitted rule's cosine to $\delta - \bar{D}$ is 0.850 against 0.807 to δ; the centred direction is nearer on 270000 of 270000 pairs, the per-pair margin having standard deviation 0.0066 and minimum 0.0118, and the ratio of the two mean cosines is 0.9502, which the reach map below reads against what centring alone predicts. *The level.* Predicted 0.0001 against measured 0.0183 mean cross-batch error, correlation 0.38 against 0.40 for the aligned value alone, median absolute gap 0.0135, at 4000 head steps with terminal gradient norm 6.8×10^{-5}: the population ledger under-predicts the measured error by two orders of magnitude, the whitening estimated from 10000 rows in 768 dimensions (13.0 per dimension) leaving anisotropy the closed forms cannot see, the mechanism the E-L2-11 design paragraph below names. *E-L2-6.* W3 at 0.0134 against a covering anchor at 0.0127 and an equal-sized non-covering one at 0.0134 (aligned 0.0000); the covered arm is no worse than the uncovered one on 64 of 90 matched configurations, which is the runner's count and includes ties: strictly better on 42, tied on 22, worse on 26, a paired difference of -0.0006 at -1.1 standard errors, at 20.0% more data touched (0 of 90 covered rows below the aligned value at the CSV's five-decimal resolution); by anchor size $k \in \{1, 3\}$ the covering arm reads 0.0126 and 0.0129 against 0.0133 and 0.0134. As at the reduced venue the non-covering anchor buys nothing over no anchor; unlike the reduced venue, where the covering anchor's advantage was -0.0382 at -7.7 standard errors, here it is within noise, and Table 12 shows the same across the map. *E-L2-7.* 0 of 1500 merged pairs worsened, 15 improved and 1485 were unchanged at the sign tolerance (against 0 unchanged of 60 at the reduced venue: at these separations the closed forms sit near zero, so the verdict below is doubly empty, forced by Bayes optimality and numerically flat), 12 of 12000 third parties up and 59 down, 0 budget increases, convexity residual 3.6×10^{-16}, 6 of 1200 near-aligned-decile rows worsened (**K-L2-7** not fired; forced as at the reduced venue, closed forms at the corpus's means). **Timestamps.** The three anchor-cell manifests were rewritten by the reach map's Stage B, which re-opened the cell after Stage A had computed all sixty of its chunks; their start and finish times describe that two-second pass and the chunk record's own modification time places the computation (the timestamps paragraph above). The reduced venue's finding stands on its own and is not amended by this one.

I.12 E-L2-10: THE REACH MAP

Design, registered and frozen by amendment L2-A14. The reduced venue produced a three-part finding on one small corpus: the vector gauge transports, the aligned floor does not, and coverage rather than anchor size buys the frame. One corpus cannot separate a property of the ledger from a property of that corpus, so E-L2-10 repeats the three measurements of E-L2-5, E-L2-6 and E-L2-7 across a grid. **Grid**, fixed before any run: an anchor cell (a frozen vision transformer on a hundred-class benchmark at twenty partitions, which *is* E-L2-5/6/7's accelerator venue and is reported as both), a backbone axis of six further frozen backbones on the same corpus at five partitions each, and a corpus axis of one further corpus at five partitions. This is not the full cross product and the manuscript says so wherever the result appears; a cell that cannot be built is recorded as skipped

with its reason and is never counted as a pass. **Falsifiers.** *K-L2-10a*, a kill: the fitted discriminant's mean cosine to $\delta - \bar{D}$ must exceed its mean cosine to the raw displacement δ on a majority of cells, and a failure kills the claim that the vector gauge is a property of the ledger rather than of one corpus. *K-L2-10b*, a kill: a covering anchor must beat an equal-sized non-covering anchor on a majority of *eligible* cells, a cell counting as eligible only if some arm measures a nonzero error, since a corpus separable enough that every arm measures exactly zero shows nothing; the exclusion is counted and reported, and this eligibility rule is stated before the run rather than after, which is the lesson of amendment L2-A11(d). *K-L2-10c*, a kill: no merged pair worsens on any cell, tested on closed forms at each corpus's class means as amendment L2-A11(g) requires us to say. **Reported, not killed.** The below-aligned fraction, with the prediction that it is smaller on backbones closer to the shared-covariance Gaussian model; if it is instead near-constant across every cell, the honest reading is that the floor's failure is a property of this paper's idealization rather than of any corpus, which is a stronger negative result about our instantiation than one venue could support. The level is an effect size as at every venue. **Instrument.** The batch-local head fit dominates the tier's cost and runs on an accelerator when one is present; each cell records the difference between the accelerator and reference backends in both fitted coefficients and measured pair errors and is refused, not reported, if either exceeds 10^{-9}. The comparison is made on the whitened features the experiment fits: on the raw corpus, whose covariance is singular when coordinates are constant, the two backends diverge by an order-one amount after four thousand steps while every measured pair error stays bit-identical, so that form of the check raises a false alarm and gives a false all-clear at once. **Status.** executed. **Not executed, reported as such.** Amendment L2-A27 extended this experiment, before any run, with a modality block (two sentence encoders on CLINC150 and Banking77, and the CLIP ViT-B/16 image tower on CIFAR-100 and Food-101), a covariance-heterogeneity diagnostic H per cell, the prediction P10-5 that the ledger's level error correlates with H across cells while the gauge cosine does not, and the scope falsifier K-L2-10d (Spearman $|\rho| < 0.5$ over at least twelve cells). L2-A22(e) placed it first in the cut order, and it was cut: none of its cells ran, H was not measured, and K-L2-10d is not read. The level failure this paper reports is therefore unexplained, which is what L2-A27 said would be reported in that case. **Output.** `l2_e10.csv`.

Result (executed 2026-09-12 under L2-A14's authorization, session L2-A34; every row of the map was checked against the cell's own three manifests, 144 field comparisons over 24 files, each of whose verdict blocks the gate re-derived from that cell's rows; Table 12). 8 cells read, 0 skipped; the accelerator and reference head backends agreed on each cell's check fit (L2-A14(a)): recorded pair-error disagreement 0, with the coefficient disagreement recorded per cell in its `backend_check.json` at between 3 and 7×10^{-16}. *K-L2-10a* not fired: the centred direction $\delta - \bar{D}$ is nearer to the fitted rule than the raw δ on 8 of 8 cells, smallest cell margin 0.0421, and on every one of the 742500 pairs across the map. **What that certifies, audited for vacuity.** The margin is nearly the same number at every cell because the ratio of the two cosines is the same number, 0.9496 to 0.9509 across 7 backbones and 2 corpora, against $\sqrt{1 - 1/m} = 0.9487$, the cosine between $\delta - \bar{D}$ and δ for independent isotropic class means in batches of $m = 10$ ($\|\delta\|^2 = 2s$, $\delta^\top \bar{D} = \|\bar{D}\|^2 = 2s/m$), and therefore the ratio any rule lying along $\delta - \bar{D}$ produces. The same identity holds at the reduced venue, whose batches of 2 give $\sqrt{1 - 1/m} = 0.7071$ against a measured ratio of 0.7031, so the reduced venue's "the gauge transports" is downgraded with it (L2-A37(a)). By Lemma 7 the population rows of a batch-local head are $\mu_r - \bar{\mu}_t$, so a cross-batch difference of rows is $\delta - \bar{D}$ exactly, and a ridge-fitted softmax head started at zero keeps its rows summing to zero at every step; the fitted rule leaves that direction only through the fit's own noise, and for the two cosines to cross the noise would have to carry a component along $\bar{D}$ of $(1 - \sqrt{1 - 1/m})\sqrt{m}$, about a sixth at $m = 10$, of the rule's component along $\delta - \bar{D}$. The kill therefore tests that the heads keep the centred gauge on every cell, which a converged fit keeps by construction and could lose only by tilting a sixth of its length toward $\bar{D}$, and that the whitened class geometry is near enough to isotropic for the centring arithmetic to hold, which it does to two decimals; it does not measure how well either direction describes the rule, both cosines sitting near 0.85. We report it as not fired and as weak. *K-L2-10b* not fired: a covering anchor beat an equal-sized non-covering one by mean error on 7 of 8 eligible cells (0 saturated), DINOv2-B/14 the exception, but beyond two standard errors of the paired difference on only 2 cells (Table 12). The runner's configuration count, covered no worse than uncovered, is a majority on every cell, and it counts ties: strictly better on 11 to 53 of 90 configurations per cell and tied on 13 to 60, a strict majority on 2 cells and more wins than losses on 6. The reduced venue's coverage effect, -0.0382 at -7.7 standard errors, does not transport in size to a hundred classes, where the anchor cell reads -0.0006 at -1.1. *K-L2-10c* not fired:

Table 12: The reach map (E-L2-10), one row per cell, all at a hundred classes in batches of ten on seeds 11, 13, 17: the fitted rule's cosine to $\delta - \bar{D}$ and to δ and their ratio (the centring value for a rule along $\delta - \bar{D}$ is 0.9487); the fraction of pairs below the aligned value, which at these venues is the fraction with zero held-out error; the three E-L2-6 arms; the paired difference covered minus uncovered in standard errors (z); and the matched configurations on which the covering anchor was strictly better, tied, and worse. Every entry is injected by the gate from the cell's own digest-verified files.

backbone	corpus	parts	$\cos(\delta - \bar{D})$	$\cos \delta$	ratio	zero %	W3	cov.	unc.	z	better/tied/worse
ViT-B/16	CIFAR-100	20	0.850	0.807	0.9502	14.2	0.0134	0.0127	0.0134	-1.1	42/22/26
ViT-L/16	CIFAR-100	5	0.885	0.840	0.9499	14.8	0.0121	0.0113	0.0122	-1.6	32/36/22
ResNet-50	CIFAR-100	5	0.853	0.811	0.9506	5.8	0.0235	0.0213	0.0227	-1.7	41/19/30
ConvNeXt-T	CIFAR-100	5	0.851	0.809	0.9504	14.5	0.0127	0.0126	0.0126	-0.0	25/33/32
Swin-T	CIFAR-100	5	0.856	0.813	0.9504	9.8	0.0166	0.0128	0.0170	-5.4	53/21/16
EfficientNetV2-S	CIFAR-100	5	0.873	0.830	0.9502	5.3	0.0230	0.0209	0.0217	-0.9	37/24/29
DINOv2-B/14	CIFAR-100	5	0.889	0.844	0.9496	45.7	0.0047	0.0046	0.0041	+1.2	11/60/19
ViT-B/16	Food-101	5	0.862	0.820	0.9509	0.7	0.0484	0.0410	0.0478	-4.6	49/13/28

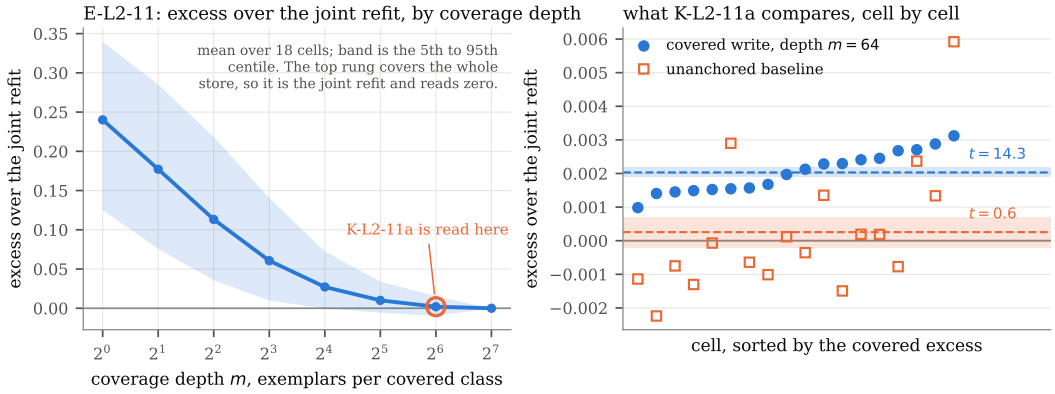


Figure 8: Measured, from `l2_e11_accel.csv`. Left: the anchored write's excess over the joint refit against coverage depth m, mean over the eighteen unbalanced cells with a 5th-to-95th centile band. Deeper coverage buys the excess down monotonically; the top rung covers the whole store, so it *is* the joint refit and reads zero by construction, and the criterion is therefore read one rung below it. Right, on an axis two orders of magnitude finer than the left panel's, where the left panel's curve has already flattened onto zero: the two quantities K-L2-11a actually compares, cell by cell, with each one's mean and standard-error band. The covered write at $m = 64$ sits at $t = 14.3$ above the joint refit while the unanchored baseline is indistinguishable from it at $t = 0.6$. The criterion asked whether some covering write shallower than the store beats not covering, and this is the picture of it firing.

0 of 12000 merged pairs worsened and 11612 were unchanged at the sign tolerance, closed forms at each corpus's means, which at these separations are near zero: forced by Bayes optimality and numerically flat. **The floor, reported and not killed.** The below-aligned fraction ran from 0.7% to 45.7%, but it is not the statistic the registration meant. At every cell the aligned values sit below the held-out resolution on all but at most 0.44% of pairs (their maximum over the map is 0.0399), and the 103470 below-aligned pairs are exactly the 103470 pairs with zero held-out error; the fraction is a separability statistic, largest on DINOv2-B/14 and smallest on Food-101, and the registered prediction about proximity to the Gaussian model is not testable here. The reduced venue's floor finding stands on its own. **The level.** The predicted mean error sits 35 to 1539 times below the measured one across the cells, and the predicted values correlate with the measured ones less than the aligned value alone does on 8 of 8 cells: the population ledger's level does not transport at a hundred classes on any backbone or corpus tried. **K-L2-5 across the map** is not read on any cell (0 of 742500 pairs eligible; 8 of 8 cells), for the reason given under the accelerator venue.

I.13 E-L2-11: THE PREMIUM, AND THE PRICE OF COVERAGE

Design, registered and frozen by amendments L2-A22 and L2-A23. Every constant in Appendices F through H is symbolic, and the one the mixed-operator bracket turns on, the premium $P = K_2/K_3$, has never been measured. E-L2-11 measures it. At store size n with incoming batch B, three arms are fit and every cross-batch pair evaluated: W3 on B alone with pairs served by the composed rule; W2 on B plus m exemplars of each written class, for m across four octaves and at full coverage; and W1, the joint refit, which is the reference. **Two conventions the tier turns on.** The excess is referenced to W1 and not to the Gaussian aligned value, because the reduced venue measured 51.4% of pairs below that value, so $e_{\mathrm{bal}} - e_{\mathrm{al}}$ is not a nonnegative quantity on real features and cannot be the price an operator pays; and the cost model is written down before the run, $\mathrm{steps} \times 3n_{\mathrm{fit}}dK$ for a head serving K classes, which makes full coverage cost $((|B| + n)/|B|)^2$, quadratic in the store-to-batch ratio where the analysis assumes a constant per event. **The scale every level statement should be held against.** Every closed form in the ledger is a population object stated in whitened coordinates, and the anchor tier estimates that whitening from the corpus it measures: 1797 samples in 64 dimensions, split in half, so the fit half is 898 rows, 14.0 per dimension. The accelerator venue keeps two hundred images per class and is not much better conditioned. Residual anisotropy that the estimate leaves behind enters the measured pair errors without appearing in the predicted ones, which is the mechanism by which a level statement can fail while a direction statement holds, and it is why E-L2-11's manifest records the estimate's quality and E-L2-13 reports the ledger's rank correlation two ways. A pilot at six rows per dimension found the ledger's ranking carrying no rank content at all; one at twenty was unaffected (amendments L2-A29(x), L2-A30(i)). **Where the kill is read, and why not at full coverage.** At m equal to everything stored, the covering write's fit set *is* the joint refit's, so its excess over W1 is zero by construction and no corpus could make it otherwise; a falsifier arithmetic forbids from firing is not one. That arm is therefore a null control, reported and required to read zero, a nonzero value being an instrument fault rather than a finding, and the kill is read at the deepest coverage depth in the grid that is strictly shallower than the store: the deepest write that is still a write and not a refit (amendment L2-A29(f)). The balanced control is likewise kept out of every primary statistic and reported beside it (L2-A29(g)). **Falsifiers.** *K-L2-11a*, a kill: if coverage at that depth leaves at least half the unanchored excess standing on a majority of store sizes and cells, coverage does not buy geometry-independence on real features and Proposition 5 is scoped to the population. *K-L2-11b*, a kill: non-monotone excess in m above break-even kills the coverage mechanism as the ledger describes it. *K-L2-11c*, a kill on the cost model: a measured premium deviating from the model by more than a quarter means the competitive analysis is denominated in units the hardware does not use. *K-L2-11d* is a measurement and not a kill: whether the break-even coverage as a fraction of what is stored falls with the store. **Registered prediction, from a disclosed pilot.** Below a break-even coverage a covering write leaves the pairs *worse* than no covering write at all, the mechanism being a bias-variance trade in which under-covered anchoring pays neither the frame bias nor the estimation variance well. **Status.** executed.

Result. Premium at full coverage 99.9001 against the model's 100.0000 (by seed 102.88/106.04/99.90), worst deviation 39.5162% and median 14.1104% along the sweep; break-even coverage 128.0000 exemplars per class (by store size 10 128; 20 128; 30 128), 1.2800 of what is stored, trending flat in the store size, and sitting at the deepest rung the sweep holds: True (where that reads True the sweep did not bracket the break-even and the figure is a lower bound on it); excess at coverage depth 64 was 0.0020 against the unanchored 2.6×10^{-4}, a ratio of 4.80/1.16/1.19 over the 3 of 6 store sizes whose baseline is positive, with the balanced control at 0.0718 and the full-coverage null control at 0.0000 (it is zero by construction; a nonzero value would mean two fits over the same rows disagreed); serving differential 2.3×10^{-5} per query (head 5.0×10^{-7}, nearest-neighbour store 2.4×10^{-5}), over 18 cells. Refitting the largest head at twice the step count moved the pair errors by 0.0056 against a held-out resolution of 0.0063, so the fits are converged at the resolution the measurement has: True (where that reads False the tier's numbers depend on the step budget and are reported as such). The registered FLOP cost model against the clock: a wall-clock floor of 1.0079 s, a median deviation of 88.6622% over the 96 fits above that floor, so the model's shape is what the analysis uses and its wall-clock is not what an accelerator charges. **K-L2-11a** FIRED, **K-L2-11b** not fired, **K-L2-11c** not fired; K-L2-11d is a measurement, not a kill, the break-even coverage as a fraction of the store reported above. **Reading**, with what each half of the ratio can carry. The two quantities K-L2-11a compares are not equally resolved. The covered excess at depth 64 is 0.0020 with a standard error over the 18 cells of 1.4×10^{-4}, a t of 14.3, and the

same sign on 18 of 18 cells: a partial covering write leaves a firmly measured positive excess. The unanchored baseline is 2.6×10^{-4} with a standard error of 4.6×10^{-4}, a t of 0.56, and a positive sign on only 8 of 18 cells: at a hundred classes there is no unanchored excess to remove, which is the reach map's finding in the tier's own units. So the *ratio* the kill is read on has a denominator indistinguishable from zero and its value is not interpretable, while the *direction* is: partial coverage does not reduce the excess, it adds one, which is what the disclosed pilot predicted below the break-even coverage. Two further scope statements the ratio needs. It is keyed by store size, not by seed, and enters only for the 3 of 6 sizes whose unanchored baseline is positive at all ($4.80/1.16/1.19$ at those sizes); and the registered "majority of cells" limb is vacuous here, this tier running on one backbone-corpus cell, so what was read is a majority of store sizes. The break-even coverage reported above is the top rung of the grid, 128.0000 exemplars per class against a largest class of 121: that rung is the joint refit, which is why it reads zero and why the kill is read one rung below it. K-L2-11b is printed as its manifest records it, and it examined nothing: with the break-even at the top rung, the depths at or above it number one per store size and the monotonicity test made 0 comparisons. The premium is the reading here with the most behind it and it is still three numbers, not one: the manifest's 99.9001 against the model's 100.0000 is one seed's full-coverage point, the three seeds giving $102.88/106.04/99.90$; and K-L2-11c is read over every paired grid point, not at full coverage, where 33 of 324 exceed its quarter bar (18 of them in the balanced control) at a median deviation of 14.1104% and a worst of 39.5162%. What the tier fixes is the premium's order at the largest store, not a constant: E-L2-12 runs at $147.1/155.8/152.2$ by seed, and the cost model makes it grow with the store by construction.

I.14 E-L2-12: THE SCHEDULING POLICIES, RUN AT MEASURED CONSTANTS

Design, registered and frozen by amendment L2-A24. Every scheduling result in this paper is a worst-case bound and not one has been run. E-L2-12 runs ten policies over three stream families at the constants E-L2-11 measures: break-even, the A_θ family, the sleep-law period, the three deployed heuristics of Appendix F.6, the natural mixed policy of Appendix H, and the two pure operators as references. The third stream family is the one to watch: each arriving class draws a heavy-tailed popularity and is queried in proportion to it, which is the *natural* analogue of Theorem 4's adversarial flood with nobody arranging it. The currency ℓ is swept over decades and every policy is reported as a Pareto frontier with the interval on which it wins, because choosing a value would make each comparison a statement about that choice. The optimum is exact by dynamic programming for the all-anchored restriction, a band between the proved lower bound $\mathrm{OPT}_{\mathrm{ack}}(K_3)$ and a hindsight upper bound for the mixed problem, and exact by enumeration on instances of at most twelve arrivals. **Two arms, because the theorem fixes the setup price and the measurement does not.** Proposition 18 fixes K and proves $2 + w^{\max}/K$ for that instance class. E-L2-11's own cost model makes the anchored setup grow with the store, which is a class no theorem here covers and is exactly what the Directions passage sets aside. Running only the second and comparing it to the first reports a theorem as refuted by an instance it never claimed; a dress rehearsal of the runner on real features did precisely that, at 2.99 against a bound of 2.00, before any canonical seed ran. The tier therefore runs both: a constant-setup arm, which holds the anchored setup at the reference state and is the arm every falsifier below is read on, and a growing-setup arm, which prices each write by the state it happens in and is reported as a *measurement* of what an endogenous setup price costs the classical guarantee (amendment L2-A29(a), (b)). The ratio compared against a bound is on writes plus rent, the cost the theorem is about, and not on the currency-weighted total, which the dynamic programme does not charge itself (L2-A29(d)). **Falsifiers.** *K-L2-12a*, a kill and the sharpest registered anywhere in this program: on the constant-setup arm, any policy exceeding its proved bound on any stream fires it, and since the theorems are worst-case a single violation means the cost model or the theorem is wrong, with K-L2-11c saying which. *K-L2-12b*: if A_{K_3} does not beat break-even where the measured premium exceeds two, Proposition 29's improvement has no operational content. *K-L2-12c*: if the square-root period falls outside the band on *stationary* streams at measured constants, Corollary 3 fails as a selection rule in the units that matter. **Status.** executed.

Result. Over 15 streams at query volumes 1.00/100.00/10000.00/1000000.00, the write crossover at 10000.0000, rent priced at 3.9×10^{-5} s per query and a measured premium reference K_2/K_3 of 147.1429: break-even reached 2.2049 against the 3.7731 its own row allows, and came within 0.3307 of a bound at its tightest point; A_{K_3} reached 47.1931 against 156.8394, tightest margin 108.7868, beating break-even on 0 of 15 (streams with measured premium > 2; 15 ran; 0 tied, the two policies

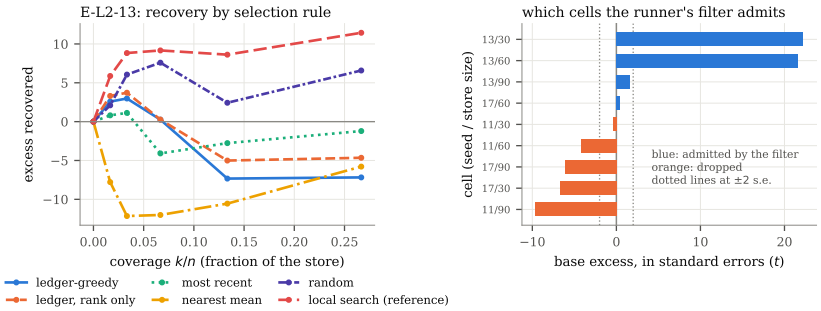


Figure 9: Measured, from `l2_e13_accel.csv`. E-L2-13 compares six rules for choosing which classes an anchor covers, by the misalignment excess each recovers against the covered fraction of the store, on the cells a pre-registered filter admits (those whose base excess is positive). Section 10 reads the result as a scope statement; Appendix I reports it with and without the filter.

having written identically); the square-root period against the empirical argmin, stream by stream: 3 of 3 stationary streams in band: S1:0.5 in band (T*=49.7, nearest 24, argmin 16); S1:1.0 in band (T*=49.7, nearest 24, argmin 16); S1:1.5 in band (T*=49.7, nearest 24, argmin 16) (5 families, [28] arrival window, 4 query volumes); the natural mixed policy won on the currency interval empty, over 840 bound comparisons, with 0 anchored rows below the anchored optimum (an arithmetic impossibility, reported because three separate pricing defects were caught by it before anything else showed them). The deployed heuristics, whose ratios the three propositions of Appendix F.6 say are unbounded over the cost class, ranged capacity 1.000 to 46888.066 over 2940 settings; periodic 1.000 to 46888.072 over 3360 settings; trigger 1.528 to 44157.727 over 420 settings as their one fixed parameter moved. With the setup price made endogenous instead of fixed, break-even's worst ratio rose to 3.8148 (A_{K_3} to 3.5623), a factor of 1.7302 on the guarantee it holds to at a fixed price: that factor is the measurement this arm exists for and no falsifier is attached to it. The L2-A24 economization check, which prices the tier at a reduced step count only if a refit at the full count agrees on every cross-batch pair: held: fit at 1000 steps (L2-A24). **K-L2-12a** not fired, **K-L2-12b** FIRED, **K-L2-12c** not fired; K-L2-12d is a measurement, not a kill, the currency interval reported above.

I.15 E-L2-13: LEDGER-GUIDED SELECTIVE COVERAGE

Design, registered and frozen by amendment L2-A25. If full coverage is quadratic in the store, the only escape is covering less, and the ledger claims to know which pairs a frozen frame will damage. Choosing the k classes to cover so as to remove the most excess maximises the induced excess inside the covered set together with the batch, which is *densest-k-subgraph plus a modular term* on the excess graph: the gain of adding a class rises with the set, so the objective is supermodular, greedy carries no constant guarantee, and this tier tests the ledger's *ranking* rather than an algorithm's optimality. Six selection rules are compared at matched exemplar budget: greedy on the ledger's closed form, greedy on its ordering alone (registered separately because the reduced venue found the gauge transports while the level does not), nearest mean, most recently written, random, and local search on the measured graph as an upper reference *for the registered objective*, which is not the same as an upper bound on realised recovery: covering a class changes the fit, so every pair's excess moves and the objective is a model of the outcome rather than the outcome. A rehearsal of the runner on real features found that arm scoring below ledger-greedy on realised recovery, which is a fact about the objective and not a fault in the search (amendment L2-A29(n)). The primary statistic is registered before the run as the area under the recovery curve *restricted to* covering at most half the store, with its reason: a disclosed pilot found selection stops mattering once coverage is dense, and the sparse regime is the only one in which selective coverage is cheaper than covering everything. **Falsifiers.** *K-L2-13a*, a kill on the instrument's operational content: losing to random selection means the ledger has nothing to say beyond its population statement. *K-L2-13b*, the bar that matters: losing to a recency replay buffer means the ledger adds nothing a practitioner does not already have. *K-L2-13c* measures whether the coverage needed falls with the store, which is what would make the premium sublinear. **Status.** executed.

Result. Restricted recovery area, ledger-greedy -3.5278 and its rank-only variant -2.6640, against most-recent -2.0323, nearest-mean -8.4870 and random 4.5180, with local search on the measured graph at 9.1649; the coverage recovering nine tenths of the excess was 0.0167 of the store, over 9 cells, of which 4 had an excess to recover and 4 produced a curve under both compared rules. Separately, and with no threshold attached: the ledger's predicted excess ranked the cross-batch pairs against the measured one at Spearman 0.0058 over 9 cells, and -0.0037 with the two batches' scalar gauges equalised instead, both indistinguishable from zero: the CSV stores six decimals and a perturbation below that resolution moves either median by more than its own value, so neither sign is evidence. Both are withdrawn from the gate's strict recomputation and named among the fields it does not verify (amendment L2-A40, corrected by L2-A74: the values do reproduce from the shipped rows, bit for bit, on the platform that produced them, which the withdrawal first denied; what they do not do is survive a sub-resolution perturbation or reproduce on a second host, where the same analyzer over the same rows returned a different sign on the first of them). The fired reads below are read against that. The selection rules rank by the first, which is the form the runner's own fit matches – gradient descent from zero keeps a batch's biases summing to zero, and the plug-in biases sum to zero exactly at that gauge – and the second is reported beside it so that which convention a fitted head follows is a measurement of this program rather than an assumption inside it (L2-A29(z)). That number is what a fired K-L2-13a has to be read against, because a selection rule built on a prediction with no rank content scores at chance, and the kill would then have fired on the coordinates rather than on the ledger: a disclosed pilot found the correlation strong on a small real corpus and absent on a synthetic one whose whitening was estimated from too few rows for its dimension, which is why E-L2-11's manifest carries the whitening's estimation quality (amendment L2-A29(x)). Ledger-greedy beat random on a majority of compared cells: False; it beat the recency buffer on a majority: False. **K-L2-13a** FIRED, **K-L2-13b** FIRED; K-L2-13c is a measurement, not a kill, the coverage fraction reported above. **Reading**, and what the kills were decided on. Both fire on a venue where the ledger's ranking is indistinguishable from zero. The attribution to the whitening is post hoc and is stated as such: amendment L2-A29(x) disclosed a pilot in which the ranking carried no rank content at fewer rows per dimension than this venue has, and registered *no* prediction for the venue, which came in at 13.0 rows per dimension. *The eligibility rule decides both verdicts, and we state that plainly.* A cell enters the comparison only if its uncovered base excess is positive, which is a filter inside the registered runner and not a choice made after the rows were seen; 4 of 9 positive, the other 5 negative, so both kills were decided on 4 cells of 9 (seed 13: 3; seed 17: 1). Removing the filter and recomputing the same restricted area over all 9 registered cells, ledger-greedy wins 6 of 9 against random and 6 of 9 against recency, at which bar neither fires. The verdicts reported here are therefore the frozen instrument's, and a reader who prefers the registration's unqualified "majority of cells" gets the opposite reading; both numbers are above, and the gate derives the second from the same rows by the runner's own definition. Nor are the dropped cells merely unresolved, which an earlier form of this paragraph asserted and the rows refute: 4 of 5 of them carry a base excess resolved at $|t|$ between 4.2 to 9.6, negative and firmly so, while 2 of 4 of the cells the filter keeps sit inside two standard errors of zero. What the filter selects for is the sign of a quantity that is well measured and mostly negative at this size, not the presence of signal. The paper reads both kills as a limit of the instrument's reach on real features at this size rather than as a refutation of the population statement, which no experiment at this resolution can test.

I.16 E-L2-14: THE SEPARATION CONSTRUCTION, INSTANTIATED

Design, registered and frozen by amendment L2-A26. Theorem 4's construction needs four classes in a particular arrangement, it is worst-case, and this paper never claims it occurs naturally. Whether a natural corpus contains it is a question about the corpus. E-L2-14 searches each corpus's measured class means for the quadruple maximising the construction's own separation margin subject to the roles the proof requires, reports the best margin per corpus, and then runs the two-phase stream on the best quadruple against the exact optimum, exactly as E-L2-4b does on synthetic geometry: the same registered stage-threshold family, the same segment-cost evaluator, and the same optimum enumerated over batch structures and stage-boundary write times, with only the four class means replaced by the ones the corpus supplies. The anchored policy is run on the same instantiated construction, because the theorem's content is the contrast between the two and not the level of either. **Falsifier.** *K-L2-14a*, scope, and it can fire: if no quadruple in any corpus achieves a margin sufficient for a measurable divergence, the construction is not realizable on natural geometry and the theorem's reach is limited to adversarially placed classes, which is a narrowing the manuscript

would then state where the theorem is stated. *K-L2-14b*: if the ratio does not grow with flood length on the best quadruple, the mechanism does not operate at natural margins. **Status.** executed.

Result. The search returns one quadruple per seed, its own argmax of predicted excess, 2 distinct across the three; the best (classes 59/75/52/47 in the roles v, w, k, p, margin $\gamma = 30.1878$) carried a measured excess of 0.2473 (cifar100; by seed 11 0.2473; 13 0.2081; 17 0.1783, spread 0.0690) against the closed form's 0.4943 at the same means, the registered synthetic placement's 0.5139, and a measurability threshold of 0.0063, the corpus's own held-out resolution (L2-A29(j)); the best online unanchored policy's ratio ran 1.0000 to 775.2295 over flood lengths 1000 to 1000000 (maximum 775.2295), against 2.0000 for the anchored policy on the same instantiated construction and 166.3571 for the closed form at this placement; the ratio grew with flood length: yes. **K-L2-14a** not fired, **K-L2-14b** not fired. **Reading.** Neither kill fires: CIFAR-100's own class means contain quadruples in the roles the proof requires, the best carries an excess more than an order of magnitude above the held-out resolution, and on it the unanchored policy's ratio grows with the flood while the anchored policy holds 2.0000, which is the theorem's contrast on natural geometry. The level is the usual story: the closed form at the same means predicts 0.4943 against a measured 0.2473, twice what a fitted head shows. The ratio's level is off in the opposite direction and by more: the best policy's ratio at the longest flood is $4.0/4.9/4.4$ times the closed form's R at the same placement, by seed. The mechanism, not the level, is what the theorem claims, and the mechanism operates.

The fired falsifiers, read together. Four fired at measured constants, two registered as kills, one as a scope criterion, and one (K-L2-13b) registered without a class as the bar that matters, and one narrative holds all four. Coverage at depth did not remove the unanchored excess (K-L2-11a: at the deepest coverage the covering write's excess is firmly positive at $t = 14.3$ while the unanchored baseline it is measured against is indistinguishable from zero at $t = 0.56$, so coverage did not remove an excess, it added one where there was none to remove); the mixed operator's theoretical improvement did not beat break-even on a single one of the 15 qualifying streams at the reference volume 10^4, losing strictly on every one there and winning on only 6 of 15 at 10^6, at measured premiums of $147.1/155.8/152.2$ by seed (K-L2-12b); and the ledger's selection signal carried no rank content at this venue's resolution (K-L2-13a and K-L2-13b, read against the withdrawn correlations above and decided on the 4 cells the eligibility rule left: a rule built on a prediction with no rank content scores at chance, so the pair fired on the coordinates rather than on the rule). Together they say what the anchor tier's readings already said: the closed forms are population objects in whitened coordinates, and at a hundred classes with the whitening estimated from the corpus, the population content does not transport into operations. Untouched is everything the paper proves: the exact ledger, the gauge split, the dominance arithmetic, and the scheduling reduction with its inherited bounds, none of them contradicted by a measurement here (K-L2-12a never breached on the constant-setup arm; the square-root period in band on every stationary stream), and the separation construction instantiated on natural geometry, its ratio growing with the flood as the closed form says it will, though at $4.0/4.9/4.4$ times the level the closed form gives (K-L2-14a and K-L2-14b both not fired). The scope sentence the data supports: geometry-independent guarantees are a theorem about an operator whose anchor covers what it protects, and a measurement about everything else.

I.17 E-L2-9: THE FULL POLICY LATTICE

Design. The scale-up of E-L2-4 at asymptotic fidelity, registered separately: the complete stage-threshold lattice (201 thresholds per stage from 0 to $4K'$ plus never, both stages) on the constructed placement at $\gamma \in \{4, 9, 16, 25\}$ and three separations ($1.5\sqrt{\gamma}$, 1.47γ, 2.94γ), floods $Q_2 \in \{10^5, 10^6, 10^7\}$, exact closed-form costs. **Prediction (as registered).** The lattice minimum of the ratio plateaus at R to machine precision after extrapolation in $1/Q_2$. **Falsifier K-L2-9**, as amended by L2-A5: the extrapolated limit missing R by more than 10^{-6} relative, or the lattice minimum attained away from the rent-forever stage-two choice in the closed-form regime. **Output.** `l2_e9.csv`. **Outcome: FIRED**, for the same reason as K-L2-4; the post-hoc E-L2-9b under the net convention is recorded below.

I.18 POST-HOC PROGRAM E-L2-4B AND E-L2-9B (AMENDMENT L2-A8, LOGGED AFTER K-L2-4 FIRED)

The registered E-L2-4 and E-L2-9 were executed as frozen and their falsifiers fired. The corrected theorem (Theorem 4) changes the cost convention from gross to net (Assumption A2, Appendix G.1) and its prediction from a plateau to unbounded growth at a closed-form rate. The post-hoc program re-runs the identical grids, policy family, enumerated optimum, brute-force validation, and anchored arm under the net convention. **Predictions.** The best family ratio on the two-stream family at flood Q_2 is at least $\min\big((K' + \min(r, \ell(\phi_{\mathrm{bal}} - e_{\mathrm{al}}))\, Q_2)/(rQ_1 + K'),\ rQ_1/K'\big)$: the first branch is the cost of a policy that has written v in stage 1, and a policy that writes nothing in stage 1 caps its ratio at rQ_1/K' however long the flood, so at fixed Q_1 the family's ratio saturates rather than diverging. Divergence needs Q_1 to grow with Q_2, which the registered grids do not do. The first branch binds with equality whenever the full stage-2 excess makes renting the flooded class the cheaper instrument (then the best policy rents k forever); the anchored break-even policy's ratio is at most 2 against the W2 optimum (the classical constant, attained exactly on the two-phase family); the enumerated optimum equals the brute-force optimum on every twelve-round instance. **Status.** Post hoc: the predictions were written after the registered falsifier fired and after the corrected proof was written, and they are reported as a confirmation of the corrected theorem's arithmetic, not as a pre-registered test. The development run of the runner at smoke scale had already exhibited the growth before the canonical E-L2-4 executed; the canonical run was nevertheless executed as registered, and the post-hoc program was logged after it. **Outputs.** `l2_e4b.csv`, `l2_e4b_bruteforce.csv`, `l2_e9b.csv`.

Optimality checks. The enumerated optimum equalled brute force on all 9000 of 9000 two-phase instances; on the four-class arm amendment L2-A12(c) added, restricting write times to the arrival rounds cost nothing on all 9000 of 9000; and the anchored arm sat at 2.000.

I.19 RUNNER CONTRACTS AND THE FREEZE PROTOCOL

The registered analyzer recomputes a manifest's verdict block from the digest-verified rows (1014 field comparisons across the canonical CSVs of this build) and the gate fails closed on any field that does not follow. On the executed tiers 19 named fields are outside its reach and are printed as such (l2_e10.cells_skipped, l2_e13.ledger_rank_correlation, l2_e13.ledger_rank_correlation_offsets, l2_e5.gauge_cos_ledger, l2_e5.gauge_cos_raw, l2_e5.max_grad_norm, l2_e8.NOT_CANONICAL, l2_e8.adapter_fits, l2_e8.batch_size_sign, l2_e8.by_condition, l2_e8.headline_arm, l2_e8_grid.NOT_CANONICAL, l2_e8_grid.by_operator, l2_e8_grid.by_operator_note, l2_e8_grid.headline_arm, l2_e8_grid.operator_separation, l2_e8_grid.operators_separated, l2_e9.max_gap_extrapolated, l2_e9b.max_gap_extrapolated); on the accelerator program of amendments L2-A22 to L2-A29 none is, container fields included, and a tamper in any one of the ninety-eight is caught (L2-A29(y)). Each runner subcommand exposes `--seeds` (a comma list; real seeds require `--canonical`, which refuses smoke seeds) and `--out`, resumes idempotently from a partial CSV (rows are keyed by cell parameters and completed cells are skipped), refuses to run unless the library self-test passes, and writes the JSON manifest beside its CSV. The library `l2_lib.py` owns the model, the closed forms, the tolerance constants, and the policy implementations; runners contain no mathematics, so a tolerance change is a library change, hashed into every manifest. The freeze sequence as executed: tolerances locked in `l2_lib.py`; the design document gained its frozen header with timestamp and the SHA-256 of the draft body; smoke-seed validation ran (seeds 0, 1, 2) over every runner's full grid; the canonical seeds then ran under the same pinned code. The amendment log falls into four groups, and each amendment's own header states its kind.

The CPU tier, before its runs (L2-A1 to L2-A10). A1 completed E-L2-1's grid and replication count and fixed a degenerate-orientation defect in the library; A2 implemented E-L2-2 as registered, with per-pair records and the convexity residual; A3 set E-L2-3's grid, its random rents, and the statistical-resolution band that made "plus tolerance" precise; A4 fixed E-L2-4's two-stream ratio, its exact enumerated optimum validated by brute force, the W2 cost model for the anchored arm, and the grids; A5 set E-L2-9's venue and extrapolated statistic; A6 implemented the gate as a macro file; A7 replaced two per-cell thresholds with family-wise kill criteria, added E-L2-2's second mixture scale, and fixed the rounding reference for the elasticity fit; A9 registered the anchor tier's protocol and A10 its reduced venue, which added a venue and changed no criterion. The exception in this

group is A8, the post-hoc program of Appendix I.18, labelled as post hoc at every site because it was logged after K-L2-4 and K-L2-9 fired.

Supersessions and repairs, before the runs they govern (L2-A11 to L2-A19). A11 and A12 superseded the first reduced-venue and first CPU-tier runs respectively; each was logged before the re-run it governs and after the runs it discards, each discarding the CSVs it supersedes, a distinction the headers state rather than fold into a blanket claim. A13, A16, A17 and A18 repaired the accelerator package before any run of it; A14 registered E-L2-10; A19 closed a defect in the gate itself.

Disclosures that changed no executed statistic (L2-A20, L2-A21). A20 disclosed a library-versus-manuscript boundary discrepancy on the exceptional set; A21 disclosed manifest timestamps that record a re-stamp rather than the run. Neither touched a pinned instrument or a criterion.

The accelerator program, before its runs (L2-A22 to L2-A30). A22 fixed the program's conventions, its cut order, and the pilots it discloses; A23 to A26 registered E-L2-11 through E-L2-14; A27 extended E-L2-10 with a modality axis; A28 recorded the instrument conventions settled while the program's runner was built, including the unit in which its costs are denominated, two cost-model repairs the pilots found, and a check A24 had registered that the runner asserted rather than performed; A29 records what an adversarial read and a dress rehearsal of that runner on real features found, including three falsifiers that were firing or unable to fire on artefacts of the instrument rather than facts about the corpus; A30 carries the same pass over the anchor tier. Every one of these was logged before any run of the tier it governs, none of which had executed at the time (the anchor tier and the program tier have since run, on 2026-09-12, under L2-A14's authorization and the session record L2-A34).

After execution, different in kind (L2-A31 onward). A31 is labelled so in the log: it records an adversarial audit of the EXECUTED CPU tier and of the build chain, and every item in it is a disclosure about a published number or a repair in the presentation and build layers, never a change to a pinned instrument or a criterion. The amendments past it keep the same discipline. After it: L2-A32 reconciles an external audit of an earlier build, L2-A33 an area-chair read of the main text, L2-A34 is the authorization record for the program tier's run, L2-A35 an instrument note on the delivered notebook, L2-A36 the intake record of the accelerator outputs with the gate repairs it forced, L2-A37 a hostile pass on that intake's text; L2-A38 to L2-A72 are the audits, the program tier's intake and the text passes of 2026-09-12 and 13, each named by its header (the vacuity audit of the Appendix H.2 evaluator with one theorem tightened, an analyzer repair and a precision withdrawal at the program tier's intake, a bibliography pass, an appendix claims audit, and the clarity passes); and L2-A73 reconciles two copies of this log kept concurrently by two working sessions on 2026-09-12 and 13, whose entries it interleaves by timestamp and renumbers, both escrow records kept. Two amendments are *post hoc* and are labelled so wherever they appear: A8 above, and A15, which records an E-L2-2 sampling deviation that an adversarial audit of the executed program found rather than one planned. Of the rest, none was motivated by a result: each repaired a defect visible by inspection, each was logged before the experiments it affects executed, and each amendment's own header states which of the two it is.

I.20 Manifest schema

Every canonical CSV carries `name.csv.manifest.json` with, at minimum: `runner` and `runner_sha256`; `lib_sha256`; `seeds`; `grid` (the full parameter dictionary); `tolerances` (the library constants at run time); `python` and `numpy` versions (the SciPy version is not recorded, though E-L2-1's saturated-cell branch calls it); `started_utc` and `finished_utc`; `chunk_plan` and `resumed_from` (or `null`); and `falsifier_verdicts` (the kill criteria evaluated per cell, recorded whether or not fired). The fill script refuses any CSV whose manifest is missing, whose runner hash is not pinned, or whose seed list intersects the smoke set. The schema is deliberately boring: every field is a check someone can run without trusting anyone.

I.21 Tolerances and power, pre-registered

Monte Carlo replication counts are set so that three standard errors of every cell statistic are at most one-third of the smallest distinguished effect (2×10^7 draws per class per cell per seed for E-L2-1, amendment L2-A1); counts are grid parameters, recorded in manifests, never post-hoc. The

Table 13: Pre-registered tolerances (locked in `l2_lib.py` at freeze; descriptive until then). MC standard errors at the registered replication counts hold three-sigma bands below one-third of the smallest effect the predictions distinguish.

Experiment	Tolerance	Smallest distinguished effect
E-L2-1	3 MC s.e. per cell; $z = 4.5$, family-wise rate 0.0023 (L2-A7, relabelled by L2-A12(d))	exceptional-set equality vs strict excess ($\geq 10^{-3}$ in e_{bal})
E-L2-2	sign, 10^{-3}	weak vs strict improvement off the exceptional set
E-L2-3	1 grid cell, or the 5 s.e. resolution band (L2-A3, A7)	argmin displacement of T^* under two-decade grids
E-L2-4	5% of R	plateau level and its climb across the γ sweep (registered form)
E-L2-9	10^{-6} relative, extrapolated	plateau level at asymptotic fidelity (registered form)
E-L2-5	5% of pairs below aligned by > 0.05 (L2-A9)	measured error below the aligned Bayes value, which the ledger forbids
E-L2-6	none	measurement only
E-L2-7	sign, 10^{-3}; convexity residual 10^{-6}; any budget increase	a merged pair worsening or a batch budget rising under a merge on real features

Table 14: E-L2-4 plateau targets as registered, at $r = 1$, $\ell = 2$ (kept verbatim; the plateau row was falsified, the floor row stands). The evaluated best-family ratios at the longest flood were 999.0, 999.0, 999.0, 999.0 under the registered (gross) convention, growing without a plateau.

γ	4	9	16	25	note
target plateau R (falsified)	3.15	7.48	21.98	80.52	$\pm 5\%$ tolerance
anchored bound	< 2	< 2	< 2	< 2	$+$ tolerance
floor at $s = 12.9$	0.5548	0.5807	0.5546	0.5176	Prop. 33 scaling

two-phase family of E-L2-4 enumerates the thirty-six pure-W3 policies: stage-1 write thresholds $\{0, 0.25, 0.5, 0.75, 1\} \times K'$ plus never, stage-2 thresholds the same list, anchors forced by batch composition; a policy's ratio is the larger of its ratios on the pure and the composed stream, the theorem's adversary; the anchored arm runs break-even at both stages under the W2 cost model against the W2 optimum. Optima are exact: enumeration over batch structures and stage-boundary write times, validated by brute force over every write-set sequence on twelve-round instances (three hundred two-phase streams per cell and seed with random stage split), where the family's mean ratio over streams is also recorded.

I.22 Pre-registered prediction targets, tabulated

The closed-form evaluations of Appendix P double as the numeric targets the falsifiers test; fixing them here, before any run, is what makes the program a registration rather than a fit. E-L2-1's target cells are the catalog tables (Tables 19 and 20) plus the exceptional-set equality prediction; E-L2-3's is the closed form T^* evaluated on its registered grid (the `Tstar` column of `l2_e3.csv`), which Table 21 illustrates on a coarser grid chosen for exact square-root ratios and does not reproduce; E-L2-4's are Table 18 and the floor family of Table 17. One further target table pins the two-phase demonstration's plateau arithmetic and the anchored arm's bound, so that a fired falsifier is unambiguous.

The floor row at $s = 12.9$ (the $\gamma = 9$ peak) is read off Proposition 33's curve: the fixed abscissa sits after the peak for $\gamma = 4$ ($s^* \approx 5.5$, decayed to 0.5548), exactly at it for $\gamma = 9$ (0.5807), and before it for $\gamma \in \{16, 25\}$ ($s^* \approx 23.2, 36.4$, still rising at 0.5546 and 0.5176), so a single sweep at fixed s crosses the peak row-wise; the values are evaluated from (27) and machine-checked (Appendix P's build artifact). The corrected theorem's per-query rates at these cells are the excess floors 0.3962, 0.5139, 0.5318, 0.5114 (Table 16); at $\ell = 2r$ the rent-or-write switch selects renting for $\gamma \in \{9, 16, 25\}$ and writing for $\gamma = 4$ on the (v, k) term alone, and renting everywhere once the far partner's cross-batch excess is included, which is what E-L2-4b measured.

Table 15: Slot map: the registered slot identifiers, their canonical CSVs, and the content the gate injects (as macros, one per quantity; Section 9 and this appendix reference the macros only). Slots of unexecuted experiments render as "registered, not executed".

Slot	Source CSV	Content
E-L2-1:MAXDEV	l2_e1.csv	worst deviation (s.e.), flagged cells, family-wise verdict
E-L2-1:EQCELLS	l2_e1.csv	exceptional-set equality check (rows, failures, worst deviation)
E-L2-2:SIGN	l2_e2.csv	merged, bystander, third-party sign counts and violations
E-L2-2:CONV	l2_e2.csv	convex-combination law residual; budget increases
E-L2-3:ARGMIN	l2_e3.csv	argmin displacement, band membership, fired count
E-L2-3:ELAST	l2_e3.csv	measured elasticities with intervals and the rounding reference
E-L2-4:PLATEAU	l2_e4.csv	best-family ratio per cell and flood against the registered target (fired)
E-L2-4:ANCHORED	l2_e4.csv	anchored max ratio under the registered convention
E-L2-4b	l2_e4b.csv	bound check, anchored ratio, brute-force agreement (post hoc)
E-L2-9, 9b	l2_e9.csv, l2_e9b.csv	lattice minima and extrapolation
E-L2-5:COS, :ORD	l2_e5.csv	pairs below aligned, correlation, median gap (executed)
E-L2-8	l2_e8_accel.csv	probe loss, routing, items and seconds per operator; with l2_e8_grid_accel.csv, the paired separations, drift, replay-ratio series and batch-size sign (executed)
E-L2-10:MAP	l2_e10.csv	per-cell gauge margin and ratio, zero-error fraction, coverage arms, dominance counts, backend agreement, each checked against the cell's own _accel or _sweep files (executed)
E-L2-6:PREMIUM	l2_e6.csv	unanchored, covered and uncovered arms (executed)
E-L2-7	l2_e7.csv	merged, budget, convexity, third-party counts (executed)

J Reproducibility and Scope

Reproducibility. The code package comprises l2_lib.py (with its self-test, which asserts a fixed set of the manuscript's constants against hand-computed reference values and must pass before any runner executes; the table-complete check is verify_numbers.py), l2_runners.py with the registered subcommands, the frozen registration with its amendment log and hash escrow, the canonical CSVs and manifests (and the smoke-seed counterparts), fill_L2.py, the figure scripts, and a build script that regenerates every measured number and figure of this manuscript from the canonical CSVs in one invocation; environments are captured in every manifest.

Theory code from the verification phase (the instruments that checked the ledger, the dominance arithmetic, and the two-phase bounds before the registration) is seed-labeled and excluded from canon by protocol; every quantity it informed is re-measured by the registered program, and one quantity it had appeared to confirm, the plateau of the earlier separation theorem, was refuted by the registered program because the instrument's optimum was two hand-picked candidates rather than an enumeration, which is the reason the registered runners compute optima by enumeration and validate them by brute force.

Scope, stated where it binds. The population ledger excludes transient and optimization-path effects: the head is at its population optimum at every write, and finite-sample effects (anchor estimation under W2, mean estimation generally) are estimation-ledger objects, outside this paper's prices and inside its falsifier E-L2-5's measurement reach. The error ledger is pairwise by design, with multiclass error accessed through the standard envelope; every theorem's error objects are pairwise. The Gaussian shared-covariance instantiation is the frozen-feature working approximation; E-L2-5, executed at the reduced venue, at the accelerator venue and across the reach map's eight

cells, tests its reach on real feature geometry, and its finding narrows scope rather than the mathematics: the level does not transport, and the direction statistic is decided by batch centring. The write-once hypothesis of the frontier is a restriction on the policy class, stated as Assumption A1 and discussed in Remark 1, and the cost constants are symbolic by construction, with the anchor premium's empirical calibration deferred to E-L2-6. The mixed-operator problem of Appendix H is closed to within less than a factor of two by Appendix H.2, exactly at unit premium, and the residual constant is stated as open.

Threats to validity, enumerated. (i) The population ledger idealizes the write as an instantaneous refit to the population optimum; real optimization paths, learning rates, and replay budgets are transients this model prices at zero, which is the scope sentence of Appendix A, and the falsifier program measures the ledger's realized accuracy rather than assuming the idealization downward. (ii) The pairwise ledger is a surrogate for multiclass error; the envelope of Lemma 9 bounds the surrogate's optimism, and the demand-weighted objectives use the surrogate consistently on both sides of every comparison. (iii) The Gaussian shared-covariance assumption is the frozen-feature working approximation; E-L2-5 tests the invariance structure, not the distributional fit; at the reduced venue the aligned floor does not transport; across the reach map the level is under-predicted by a factor of 35 to 1539 and the floor sits below the measurement's resolution; and the direction test as instrumented is decided by batch centring at both venues (Appendix I), which narrows scope without touching any proof. (iv) The write-once hypothesis of Section 8 is a restriction on the policy class, not a definition: a system that refits an already-written class escapes the negative result, and it escapes by paying either the store (W1) or the anchor (W2), which is the positive side. We name it as a restriction because calling it definitional would hide what the theorem is about. (v) The constants K, c, ℓ are symbolic; conclusions that depend on their ratios (the rent-or-write switch) are stated as conditions, and E-L2-6, executed at three venues, and E-L2-11 measure the one ratio (the anchor premium) that the mixed-operator bracket is stated in. (vi) The W3 fit is batch-local (Remark 4); the global-softmax variant with frozen rows is a different operator and is not priced here. (vii) The excess convention of Assumption A2 nets a policy-dependent number of baseline terms, which in our construction favours the policy, so the lower bound is conservative; Appendix G.1 states this and Remark 12 gives the convention-free weaker statement. (viii) One registered falsifier fired; the theorem it tested was corrected and the correction is stronger, but the corrected theorem's own empirical check is post hoc and is labeled as such.

What would falsify what, and what did. K-L2-1 kills the ledger (not fired); K-L2-2 kills the dominance theorem's strictness content (not fired); K-L2-3 kills the sleep law as a selection rule (not fired); K-L2-4 and K-L2-9 killed the separation theorem in its registered form (fired; the corrected theorem and its post-hoc program are Section 8 and Appendix I.18); K-L2-5 kills the instantiation's scope claim (not fired at the reduced venue of amendments L2-A10 and L2-A11; not read at the accelerator venue and on every reach-map cell, no pair being eligible at the registered margin); K-L2-7 kills the dominance sign on real means (not fired at every venue, and forced); K-L2-10a, b and c kill the claim that the gauge, the coverage effect and the dominance sign are properties of the ledger rather than of one corpus (none fired; the first is reported as weak, the second as within noise on most cells, and the third as numerically flat, for the reasons given in its subsection); K-L2-10d, the scope falsifier of the L2-A27 modality extension, is not read, that extension having been cut under L2-A22(e); E-L2-6, E-L2-8 and the program's measurement tiers measure and cannot fire. Each is reported as an outcome in Table 10.

K Extended Related Work

K.1 The acknowledgment line: what is inherited, what is not

The dynamic TCP acknowledgment problem (Dooly et al., 2001; Seiden, 2000; Karlin et al., 2001) contributes the scheduling skeleton of Section 7: the objective's shape, the break-even policy, and the deterministic constant 2 and randomized lower bound $e/(e-1)$ in the unit-weight, fine-granularity case. It does not contribute what Section 7 actually proves, and an earlier form of this paragraph said it did. The classical result is quoted for additive delay with unit request weights; the extension to a query-driven, time-varying accrual rate is ours (Appendix F.4), the deterministic constant at fixed granularity is $2 + w^{\max}/K$ rather than 2, the threshold lower bound showing no threshold

rule reaches 2 there is ours, and Karlin et al. (2001)'s matching randomized upper bound does not transfer, its natural transfer being refuted by a two-item instance (Appendix H.3). What transfers verbatim is the embedding of the classical instances and hence Seiden (2000)'s lower bound. On the algorithms side the rent-or-buy frontier has moved since Meyerson (2005): Coester & Turoczy (2026) determine the parking-permit problem's deterministic ratio exactly, and Mari et al. (2024) study aggregation with delays under stochastic arrivals, the setting the sleep law's Poisson reading gestures at.

K.2 The continual-learning side: scheduling and pricing replay

Two lines are the nearest neighbours of what we do and neither prices the write operator the way we do, which is the sense in which the gap we claim is a gap. Klasson et al. (2023) pose replay *scheduling* as a decision problem and search over when to replay which task, which is our schedule without a price attached to the write event; the policies they compare are the objects our Section 7 costs. Deng et al. (2025) analyse rehearsal in overparameterized linear models and compare concurrent with sequential rehearsal, finding sequential revisiting better when tasks are less similar and proposing a hybrid; that is an ordering question adjacent to our schedule, asked without a price on the write, and it does not touch what a batch should contain, which our E-L2-6 measures and our E-L2-11 fails to convert into an operational rule. Ding et al. (2024) give closed-form forgetting for linear regression under task ordering, the nearest analogue to the ledger in a model where the fit is solvable; their object is the loss under a sequence, ours is the pairwise error under a frame. Our claim is not that the write event has gone unpriced, but that pricing it, scheduling it, and composing it have not been done in one object with a competitive guarantee attached. Our contribution on that line is the identification (rent as weighted waiting, write as acknowledgment), the batching-quality coupling that the classical problem cannot express, and the separation theorem, whose adversary installs the flood geometrically rather than by declaring prices. The line's own frontier, general delay costs (Bhore et al., 2026) and aggregation hierarchies (Bienkowski et al., 2016; 2020; 2021; Buchbinder et al., 2017; Azar & Touitou, 2019; 2020), is precisely where our scope sentences point: we claim nothing beyond linear delay, and our mixed problem is priced depth-one territory with a quality axis the hierarchy does not carry.

K.3 The rent-or-buy lineage

Ski rental and its relatives (Karlin et al., 1988; El-Yaniv et al., 1999; Meyerson, 2005; Lotker et al., 2012; 2008) price the per-item decision: when to convert a per-use payment into a one-time purchase. This paper's objects are one level up: the purchase is a batch event with a shared setup, and its quality depends on composition and operator. The multislope and two-option variants are the closest classical relatives of our open problem, and the learning-augmented line (Purohit et al., 2018) is the natural refinement once demand predictions exist; none of these contain a notion of write quality, because none of them have a second store to be misaligned with.

K.4 Consolidation, batching, and sleep in learning systems

The consortium position paper (Dorovatas et al., 2026) names the agenda: modular memory, consolidation for stability and specialization, and heuristic or learned policies for it. Empirical work on sleep-like consolidation trains consolidation mechanisms on schedules and reports accuracy and stability gains, which our Corollary 3 reads in reverse: the schedule is not a biological tribute but the solution of a priced batching problem, and its period follows the setup-to-rent ratio.

Replay-style methods that interleave past-class exemplars with new batches are, in the taxonomy of Definition 1, anchored incremental writes; the anchor is what lets the write revisit the frames of the classes it covers, which is the mechanism this paper prices and the reason W2 inherits the classical guarantees (Proposition 5).

Read the other way, Theorem 4 is an argument for why replay budgets are not merely a variance-reduction device: coverage is what buys a competitive guarantee, and a replay set that never revisits a class leaves that class's frame frozen forever. We are not aware of a treatment that prices the write operator, its schedule, and its composition jointly, which is the slot this paper occupies.

K.5 THE MODEL-REALISM DEFENSE, STATED ONCE IN FULL

A reviewer seeing Gaussian class-conditionals and linear heads will ask whether the theorems are about learning or about statistics homework. Three-part answer. First, the frozen-backbone regime (fixed encoder, trained or refit head) is the dominant deployed continual learning setting and the one where consolidation is a real production decision; the theory targets exactly it. Second, class-conditional Gaussian structure with shared covariance is the standard working approximation of feature geometry in that regime, and every falsifier anchors on real frozen features, so the approximation's reach is measured, not presumed. Third, the population cross-entropy head and the prototype rule agree as classifiers at balanced priors (Proposition 10), so the ledger's objects are store-properties rather than head idiosyncrasies; the price of the idealization is scoped in Appendix J and its estimation-side boundary is where finite-sample work belongs. What the model deliberately does not contain, transients, optimization paths, non-Gaussian feature tails, is the complement of what the theorems are about: the protocol-intrinsic economics of writing memory.

L WORKED MICRO-INSTANCES

This appendix walks every main theorem through one small instance each, with every number an exact evaluation of a formula from the text. It exists because the paper's objects are few and explicit, and a reader should be able to hold one world in mind and watch all of them act in it.

L.1 A BATCHING WORLD

Take $d = 2$ and six classes in three two-class batches:

$$T = \{(1,1),\,(-1,-1)\}, \qquad U = \{(2,3),\,(2,1)\}, \qquad V = \{(1,-1),\,(-1,1)\}, \tag{34}$$

with anchors $\bar{\mu}_T = (0,0)$, $\bar{\mu}_U = (2,2)$, $\bar{\mu}_V = (0,0)$: the T-versus-V pairs are perfectly aligned, the T-versus-U and U-versus-V pairs carry mismatch $\pm(2,2)$, and the scalar offsets of Proposition 13(a) are in force throughout.

The bystander event. Merge T and U: the merged anchor is $\bar{\mu}_{(TU)} = (1,1)$, and $\bar{D}_{(TU)V} = \frac{1}{2}\bar{D}_{TV} + \frac{1}{2}\bar{D}_{UV} = (1,1)$, the convex-combination law of Theorem 1(b) in action. The bystander pair $\big((1,1),(1,-1)\big)$ has $\delta = (0,2)$, $\gamma = 4$, aligned error $e_{\mathrm{al}} = \Phi(-1) = 0.1587$. After the merge its rule has direction $\delta - \bar{D} = (-1,1)$ and the errors evaluate to $\Phi(-\sqrt{2}) = 0.0786$ and $\Phi(0) = 0.5$: the balanced error moves from 0.1587 to 0.2893. This is Example 2 with coordinates: the pair was at Bayes and the merge nearly doubled its error, one of its two one-sided errors going to chance.

The merged pair improves. The pair $\big((-1,-1),(2,1)\big)$ across T and U has $\delta = (-3,-2)$, $\gamma = 13$, and pre-merge mismatch $\bar{D}_{TU} = (-2,-2)$, direction $\delta - \bar{D} = (-1,0)$: the one-sided errors evaluate to 0.3085 and 0.0062, balanced 0.1574. After the merge the pair is internal, aligned, at $e_{\mathrm{al}} = \Phi(-\sqrt{13}/2) = \Phi(-1.803) = 0.0357$: a $4.4\times$ improvement.

A third party improves too, by rotation. The pair $\big((2,1),(-1,1)\big)$ across U and V has $\delta = (3,0)$, $\gamma = 9$. Pre-merge mismatch $\bar{D}_{UV} = (2,2)$ gives direction $(1,-2)$ and one-sided errors 0.0221 and 0.7488, balanced 0.3855. Post-merge $\bar{D}_{(TU)V} = (1,1)$ gives direction $(2,-1)$ and one-sided errors 0.0221 and 0.2512, balanced 0.1366: a $2.8\times$ improvement. Note the mechanism: the budget only fell from 2.83 to 1.41 in norm, but the *direction* rotated off the pair's worst bearing, which is where the error actually lives; the guarantee of Theorem 1(c) tracked the budget, the realized error tracked the geometry.

L.2 A SECOND BATCHING WORLD: PERPENDICULAR MISMATCH

The first world's merge damaged bystanders through a mismatch bearing toward the pair's own displacement; this one sits at the opposite bearing. Take

$$T = \{(1,1),\,(-1,-1)\}, \qquad U = \{(1,3),\,(-1,1)\}, \qquad V = \{(2,0),\,(-2,0)\}, \tag{35}$$

so $\bar{\mu}_T = (0,0)$, $\bar{\mu}_U = (0,2)$, $\bar{\mu}_V = (0,0)$. The pair $\big((1,1),(-1,1)\big)$ across T and U has $\delta = (2,0)$, $\gamma = 4$, threshold 0 (the pair is norm-equal), and mismatch $\bar{D} = (0,-2)$ with induced direction $(2,2)$: one-sided errors 0.5 and 0.0786, balanced 0.2893, against the aligned 0.1587: damage, but graded rather than catastrophic, and note that a norm-equal pair still degrades here because the mismatch is *not* parallel to its displacement (the exceptional set is a ray, not a neighborhood). Merging T and U (anchor $(0,1)$) makes the pair internal: 0.1587, recovered.

The third party $\big((1,3),(2,0)\big)$ across U and V has $\gamma = 10$: pre-merge mismatch $(0,2)$ gives one-sided errors 2.0×10^{-4} and 0.7602, balanced 0.3802; post-merge mismatch $(0,1)$ gives 0.0127 and 0.1855, balanced 0.0991: a $3.8\times$ improvement under a merge that merely halved the mismatch norm, again through rotation of the bearing. Both constructed worlds improve their third parties; the worsening direction exists by the Δ calculus of Appendix E.7 (its second term carries either sign) but is not exhibited by hand here, and the random grids of E-L2-2 are registered to span it; we prefer this honesty to a decorative instance.

L.3 A SCALAR WORLD: OFFSETS RESTORING BAYES

Take $T = \{(2,0),(0,2)\}$ and $U = \{(1,-1),(1,3)\}$: both anchors are $(1,1)$, so $\bar{D} = 0$, while $\bar{S}_t = 4$ and $\bar{S}_u = 6$ give $\bar{\Delta} = -2$: a pure scalar mismatch, the case Proposition 13(a) owns. The pair $\big((2,0),(1,-1)\big)$ has $\delta = (1,1)$, $\gamma = 2$, aligned error $e_{\mathrm{al}} = \Phi(-1/\sqrt{2}) = 0.2398$. Before offsets, the ledger gives one-sided errors $\Phi(-2/\sqrt{2}) = 0.0786$ and $\Phi(0) = 0.5$: a calibration tilt masquerading as interference, balanced 0.2893. After the offsets $o_r = -\frac{1}{2}\bar{S}_t$ and $o_s = -\frac{1}{2}\bar{S}_u$, both one-sided errors return exactly to 0.2398: the pair is at Bayes, the scalar gauge is gone, and nothing about the vector geometry was ever touched. This is the whole content of the split in six numbers: what composition buys (this), and what it cannot (the worlds above).

L.4 A MERGE CASCADE: WATCHING THE BUDGET CONVEXIFY

Four singleton-class batches on a line, anchors at $0, 1, 3, 7$ (realized by pairs of classes placed symmetrically about those points): the initial budget is $\rho = 7$, carried by the $(0,7)$ pair. Merge the middle two: their anchor moves to 2, the $(0,7)$ pair persists, and ρ stays 7 (no merge can fix the pair it does not touch). Merge the result with the first batch: the new anchor is $4/3$, and the mismatch to the last batch is $\frac{1}{3}\cdot 7 + \frac{2}{3}\cdot 5 = 5.667$, the convex-combination law of Theorem 1(b) with visible arithmetic: the budget falls from 7 to 5.667, not to the 5 a norm-only reading would expect, because the combination is of mismatch vectors, here signed and collinear.

Merge everything: one batch, no cross pairs, $\rho = 0$. The path $7 \to 7 \to 5.667 \to 0$ is the guarantee-form theorem's trajectory in one line: monotone, convexifying, and indifferent to the realized errors along the way, which the earlier worlds showed can move either way.

L.5 A SLEEP INSTANCE

Take $K = 18$, $\alpha = 0.5$ arrivals per round, $\bar{\rho} = 1$: the law gives $T^* = \sqrt{2 \cdot 18/0.5} = 8.49$, and the exact discrete cost of Appendix F.5 evaluates to $f_8 = 4.00$, $f_9 = 4.00$, $f_4 = 5.25$, $f_{16} = 4.875$, with the minimum 3.99 at $T = 8.49$: the integer plateau straddles T^*, as the exact-discrete analysis of Appendix F.5 predicts with no approximation at all in the argmin. Quadrupling rot (so $\bar{\rho} = 4$) moves T^* to 4.24; quadrupling the setup moves it to 16.97: both elasticities one half in absolute value, both signs as Corollary 3 states.

L.6 A TWO-PHASE INSTANCE

Run the construction of Appendix G.4 at $\gamma = 9$, $s = 6$, $u = (0,1)$, and the translation normalization $\mu_k = (0,0)$, $\mu_v = (3,0)$: then $\mu_w = (3,6)$, $\mu_p = (-3,0)$ (mirror), $a_1 = (3,3)$, $a_2 = (-1.5,0)$, $\bar{D}^* = (4.5,3)$ with $\|\bar{D}^*\| = 5.41 > \|\delta\| = 3$ and a component off the line of δ: off the closed ray with margin. At this fixed unit composition the floor evaluates at $x = \gamma/\sqrt{\gamma + s^2} = 9/6.708 = 1.342$: $\phi_{\mathrm{bal}} = \frac{1}{2}[\Phi(-1.342) + \Phi(2.683)] = \frac{1}{2}[0.0899 + 0.9963] = 0.5431$, above one half as (27) promises, and the fixed-composition excess is $\phi_{\mathrm{bal}} - e_{\mathrm{al}} = 0.5431 - 0.0668 = 0.4763$. Theorem 4 quotes the composition-free $\Xi = \frac{1}{2} - e_{\mathrm{al}} = 0.4332$, which is what the policy pays even if it

Table 16: The theorem's excess floor $\Xi = \frac{1}{2} - e_{\mathrm{al}}$, and the sharper fixed-composition excess at the reference separation $s = 12.9$, as the flooding pair's margin grows (exact evaluations).

γ	1	4	9	16	25	36	49	64
e_{al}	0.3085	0.1587	0.0668	0.0228	0.0062	1.35e−3	2.33e−4	3.2e−5
$\Xi = \frac{1}{2} - e_{\mathrm{al}}$	0.1915	0.3413	0.4332	0.4772	0.4938	0.4987	0.4998	0.5000
$\phi_{\mathrm{bal}}(12.9) - e_{\mathrm{al}}$	0.2068	0.3962	0.5139	0.5318	0.5114	0.5015	0.5000	0.5000

steers its anchors by multiplicity. With $r = 1$ and $\ell = 2$ the policy's per-query price on the flood is $\min(1,\ 2 \times 0.4332) = 0.866$ once it has written $\{v, w\}$, while the optimum, renting v through the Q_1 stage-1 queries and then writing $\{v, k\}$ for $K' = 1$, pays $Q_1 + 1$ in total: at $Q_1 = 1000$ and a flood of 10^6 queries, every pure-W3 policy *that has written* $\{v, w\}$ in stage one pays at least $(1 + 0.8663856 \times 10^6)/1001 = 865.5$ times the optimum on this world's composed stream; a policy that writes nothing in stage one instead caps its ratio at $rQ_1/K' = 1000$, the branch E-L2-9b evaluates, and the ratio keeps growing with the flood, while the anchored break-even policy pays exactly twice the optimum on the same streams. The companion table shows how the excess grows with the flooding pair's margin: it approaches $1/2$, so the price per query approaches half the currency once the margin is a few units.

M Objections, Answered

Objection 1: these are LDA theorems. The frozen-backbone regime is the deployed setting where consolidation is a production decision; shared-covariance Gaussian structure is the standard working approximation of its feature geometry, measured on real features by E-L2-5 at three venues, where its level fails and its floor is unreadable at a hundred classes (Section 9); and the population head and the prototype rule agree as classifiers (Proposition 10), so the ledger prices stores, not head idiosyncrasies. Appendix K states the full defense once.

Objection 2: the experiments are synthetic. The registered program of Appendix I is designed falsifiers first, executed on canonical seeds through a fail-closed pipeline, and reported in full, including the falsifier that fired and changed a theorem; the frozen-feature tiers ran at two venues and across a reach map of eight backbone-corpus cells, the model-scale tier ran on a trained Pythia-160M language model, and each is reported with what it could and could not read; the four-experiment program of amendments L2-A22 to L2-A26 executed last and fired four of its registered falsifiers, two kills, one scope criterion and one the log classes only as the bar that matters, all reported as fired with their readings; across the whole program six falsifier verdicts fired, and the gate names them so the count cannot go stale (EelevenVerdictA, EfourVerdict, EnineVerdict, EthirteenVerdictA, EthirteenVerdictB, EtwelveVerdictB). The false-attestation principle applies to ourselves first.

Objection 3: the write-once hypothesis does the work. It does, and that is the finding rather than a concealment. The hypothesis says the policy never revisits a written frame, which is exactly what the cheap operator cannot do; a policy that revisits escapes the bound, and Proposition 5 prices the two ways to revisit, the store (W1) and the anchor (W2). We state it as an assumption on the policy class (Assumption A1) rather than fold it into a definition, because the content of Theorem 4 is precisely that revisability is competitively necessary.

Objection 4: the scheduling half is TCP acknowledgment, so what is new. The reduction is credited and the inherited constants are quoted, not re-proved; the new objects are the misalignment ledger that prices frames through statistical geometry, the bystander counterexample, and a separation theorem whose flood is installed geometrically through anchor confinement rather than declared as a price process, so the oblivious adversary needs no knowledge of the policy's realized batch choices (Appendix G.7).

Objection 5: the separation is only a cost-ratio argument, and the per-query price is bounded. Both halves are true and neither is hidden. The rate is closed form and bounded by $\frac{1}{2}$: the policy pays at least $\ell\,\Xi$ on every flood query (Table 16) while the optimum pays nothing after one write (Lemma

26), so what diverges is the query count times a bounded positive excess. The normalization is the one the rent already uses, net of the baseline both stores share, and Remark 12 records what changes under the gross convention: the same construction still admits no envelope-uniform constant, and the per-instance ratio becomes finite but grows without bound as the partners separate. The earlier, gross-normalized form of the claim predicted a finite per-instance plateau and was falsified by its own pre-registered test, which the paper reports (Section 9).

Objection 6: population idealization. Transients and optimization paths are real and outside the model; the model's job is to price protocols, and every price here is protocol-intrinsic. The estimation ledger (finite-sample anchor noise, mean estimation) is the natural next layer, and E-L2-5/E-L2-6 measure exactly its boundary on the anchor side.

N Code and the Provenance Gate

The provenance gate is `fill_L2.py`. It loads each canonical CSV with its manifest, refuses any CSV whose manifest is missing, whose runner or library hash differs from the values pinned at the freeze, whose seed list intersects the smoke set or is not the registered real set, or whose manifest is not marked canonical; it then writes one macro per registered quantity into a generated file that the manuscript inputs, and nothing else in the manuscript carries a measured numeral. Slots whose experiment is registered but not executed render as "registered, not executed"; a CSV for such an experiment is rejected until a fill rule is registered for it. For the reach map the gate also opens every cell's own files and refuses a map row that does not follow from them (L2-A36(b)), and a falsifier whose eligible set is empty is rendered as not read whatever string its runner wrote (L2-A36(c)). The registered runners follow the contract of Appendix I.19 and contain no mathematics beyond calls into `l2_lib.py`, whose self-test asserts a fixed set of tabulated constants (the aligned table, the $\gamma = 9$ rows of the catalog and floor tables, the worked worlds, the two-phase arithmetic, the agreement of the segment formula with the explicit schedule evaluator, and the agreement of the enumerated optimum with brute force on a short stream) to 2×10^{-4} against stored reference values computed by hand; it does not cover the other γ columns of the catalog and floor tables, the sleep-law grid or the divergence table, which only `verify_numbers.py` covers. That tolerance is four times the half-unit of a four-decimal print, so the self-test cannot see a last-digit error in a printed constant, and three of its stored reference values are themselves off in the fourth decimal (inside its tolerance, outside the paper's); the library is hash-pinned into every canonical manifest and is left as it ran, and the printed constants are instead held to 5×10^{-5} by `verify_numbers.py`, the resolution at which an audit found three last-digit errors in the worked worlds (L2-A32) and at which those constants are now checked on every build. The full sources ship in the supplement.

Glossary

Batch anchor $\bar{\mu}_t$: the count-weighted mean of a batch's class means; the frame a W3 write leaves behind. **Batch scalar** $\bar{S}_t$: the count-weighted mean of squared norms; the removable half of the frame's residue. **Vector mismatch** $\bar{D}$: the difference of two batches' anchors; the unremovable half. **Scalar mismatch** $\bar{\Delta}$: the difference of batch scalars; removed by one offset vector per batch (Proposition 13). **Aligned error** e_{al}: the Bayes balanced error of a pair at margin γ, equal to $\Phi(-\sqrt{\gamma}/2)$. **Mismatch floor** ϕ_{bal}: the best balanced error an adversarial placement can force onto a pair given compact achievable anchors (Lemma 24). **Rent** r: the per-query price of not having written (Appendix A.1). **Write event**: one consolidation operation through an operator, at setup K plus c per item. **Operators W1, W2, W3**: full joint refit; anchored incremental write; unanchored incremental write (Definition 1). **Guarantee functional** Ψ: the worst balanced cross-batch error at a mismatch budget (Definition 2). **Budget** $\rho(P)$: the largest mismatch norm across a partition's batch pairs. **Bystander**: a pair aligned before a merge and damaged by it (Example 2). **Exceptional set**: the mismatch ray on which a norm-equal pair is exactly at Bayes despite nonzero mismatch (Proposition 15). **Excess floor** $\Xi = \frac{1}{2} - e_{\mathrm{al}}$: the per-query price Theorem 4 forces a pure-W3 policy to pay on a flood, uniformly over its batch compositions and corrections, with no finite ratio. **Sleep law**: the square-root optimal period of periodic consolidation (Corollary 3). **Composed protocol**: evaluation-time subtraction of count (and scalar) offsets; free, and it buys exactly the scalar gauge. **Provenance gate**: the fail-closed pipeline (fill_L2.py) through which alone measured numbers may enter.

Algorithm 1 Break-even flush (the anchored scheduling policy of Propositions 18 and 5; $(2 + w^{\max}/K)$-competitive under the linear-delay scope, hence 2 in the fine-granularity limit).

$A \leftarrow 0$ ▷ accrued foregone rent since last write
for each round t with query set Q_t **do**
  $A \leftarrow A + r \cdot |Q_t \cap \text{unwritten}|$
  if $A \geq K$ **then**
    write all pending items as one batch (anchored: W2 with a replay anchor)
    $A \leftarrow 0$
  end if
end for

Algorithm 2 Periodic sleeper (the stationary regime of Corollary 3; period $T^* = \sqrt{2K/(\alpha\bar{\rho})}$, evaluated per Appendix F.5 and rounded to the integer grid).

estimate α (entry-eligible arrival rate) and $\bar{\rho}$ (mean rent) from bookkeeping
$T \leftarrow \arg\min_{T' \in \{\max(1,\lfloor T^* \rfloor), \max(1,\lceil T^* \rceil)\}} \left[K/T' + \alpha\bar{\rho}(T'-1)/2\right]$ ▷ best legal period
for each round t **do**
  if $t \bmod T = T - 1$ **then**
    write all pending items as one batch (anchored)
  end if
end for

O ALGORITHMS

This appendix fixes pseudocode for every policy and pipeline the paper names, so that each theorem's subject, each corollary's rule, and each registered experiment's machinery has one executable reading. The first three are the scheduling policies; the fourth is the batching decision rule the ledger prices; the remaining three are the registered program's machinery.

Reading the algorithms against the theorems. Algorithm 1 is the policy whose guarantee Proposition 18 inherits and whose anchored precondition Proposition 5 pins; Algorithm 2 is its stationary cousin and Corollary 3's subject; Algorithm 3 is the mixed problem's natural occupant, inside the bracket of Proposition 27; Algorithm 4 is the ledger's decision calculus, whose sign structure E-L2-2 tests; Algorithms 5 through 7 are the empirical program's spine, and the gate is the only path by which the measured numbers of Section 9 entered this manuscript.

P QUANTITATIVE RESULTS FROM THE CLOSED FORMS

This appendix is the paper's quantitative theory in one place: every theorem with a closed form gets its evaluation, its plot, and its table, so that the manuscript's quantitative content stands on its own beside the measured program. Where a figure overlays nothing measured, that is the point: these are the targets the registered program was set to reproduce, and Appendix I's falsifiers are sharp because the targets are explicit; one of them was missed, and Section 9 says which and why.

P.1 THE CATALOG, PLOTTED

P.2 THE FLOOR, ITS SHAPE, AND A UNIVERSAL MAXIMUM

Proposition 33 (Universal floor maximum)**.** *Write $\overline{\Phi}(x) = \Pr[Z > x]$ for the standard normal upper tail. At the constructed placement of Appendix G.4, the floor as a function of $x = \gamma/\sqrt{\gamma + s^2} \in (0, \sqrt{\gamma})$ is $\phi_{\text{bal}}(x) = \frac{1}{2}\left[\overline{\Phi}(x) + 1 - \overline{\Phi}(2x)\right]$, unimodal in x with maximizer*

$$x^* \;=\; \sqrt{\tfrac{2}{3}\ln 2} \;\approx\; 0.680, \qquad \phi_{\text{bal}}^{\max} \;=\; \phi_{\text{bal}}(x^*) \;=\; 0.5807, \tag{36}$$

both independent of γ, and attained inside the domain exactly when $\gamma > \frac{2}{3}\ln 2 \approx 0.462$; below that margin the domain stops short of x^*, ϕ_{bal} is increasing throughout it, and the supremum is the

Algorithm 3 Anchor-premium switch (the natural mixed policy of Appendix H, the natural occupant of Proposition 27's bracket, now known to be $\Theta(K_2/K_3)$ wide).

maintain accrued rent a_i per pending item i and the anchor-proximity graph of the pending pool
for each round t **do**
  for each query to unwritten i **do** $a_i \leftarrow a_i + r$
  end for
  $S \leftarrow \{\, i : a_i \geq K + c \,\}$ ▷ per-item break-even against the write price
  if $S \neq \emptyset$ **then**
    $j \leftarrow$ nearest-anchor pending partner of $\arg\max_i a_i$ ▷ batch size ≥ 2
    $\Delta \leftarrow (K_2 - K_3)$ versus $\ell\,(\phi^{\text{est}}_{\text{bal}} - e^{\text{est}}_{\text{al}})\cdot$ expected future discriminating queries
    write $\{i, j\}$ anchored (W2) if the premium is below the expected mismatch saving, else unanchored (W3)
    reset a_i, a_j
  end if
end for

Algorithm 4 Bystander-aware merge (the decision rule priced by equation in Appendix E.7; Theorem 1 guarantees the certified budget never rises, this rule optimizes the realized ledger).

Require: batches T, U candidate for merge; query flows f_{TU}, f_{UV}, f_{TV}; all pairwise e_{bal} from Lemma 11 at current and post-merge mismatches
$\Delta \leftarrow f_{TU}(e_{\text{al}} - e^{TU}_{\text{bal}}) + f_{UV}(e^{(TU)V}_{\text{bal}} - e^{UV}_{\text{bal}}) + f_{TV}(e^{(TU)V}_{\text{bal}} - e_{\text{al}})$
return merge iff $\Delta < 0$ ▷ first term negative, third positive off the exceptional set, second of either sign

boundary value $\phi_{\text{bal}}(\sqrt{\gamma}\,) < 0.5807$*, approached as* $s \to 0$*. In the separation coordinate the peak sits at* $s^* = \sqrt{\gamma^2/x^{*2} - \gamma}$*, asymptotic to* $\gamma/x^* \approx 1.47\,\gamma$*. Consequently the placement can push the floor as high as* 0.5807 *but never higher, at any margin, the fixed-composition excess is at most* $0.5807 - e_{\text{al}}$*, and the rent-or-write switch* $\ell(\phi_{\text{bal}} - e_{\text{al}}) \geq r$ *(beyond which a flooded policy prefers to rent forever) is reached at* $\ell \geq r/(0.5807 - e_{\text{al}})$ *when* s *is chosen near* s^**.*

Proof. Differentiate: $\mathrm{d}\phi_{\text{bal}}/\mathrm{d}x = \frac{1}{2}[2\,\mathrm{Z}(2x) - \mathrm{Z}(x)]$ with Z the standard normal density; the unique zero solves $\mathrm{Z}(x) = 2\,\mathrm{Z}(2x)$, i.e., $-x^2/2 = \ln 2 - 2x^2$, i.e., $x^2 = \frac{2}{3}\ln 2$; the sign pattern (negative below the crossing of $2Z(2x) - Z(x)$ from positive density growth, positive after) gives the maximum, which lies in the open domain $(0, \sqrt{\gamma})$ precisely when $\gamma > x^{*2} = \frac{2}{3}\ln 2$; otherwise the derivative has no zero in the domain, ϕ_{bal} increases to the boundary, and the "no higher" half of the statement still holds since $\phi_{\text{bal}} \leq \phi_{\text{bal}}(x^*)$ pointwise on $(0, \infty)$; and the value evaluates as stated since it depends on x^* alone. The separation coordinate inverts $x = \gamma/\sqrt{\gamma + s^2}$. □

P.3 The regime map

P.4 The sleep law, plotted

P.5 Geometry, drawn

P.6 Extended excess-floor table

P.7 The extended catalog, tabulated

The two tables carry the paper's practical message in compact form. Parallel mismatch is catastrophic at any margin (half the table jumps to chance before β reaches one), which is the exceptional-set geometry's evil twin: the same parallelism that hides damage at norm-equal pairs concentrates it at norm-unequal ones. Perpendicular mismatch is graded and margin-controlled, and its scaling in $\theta = t/\sqrt{\gamma}$ says the operative question is always mismatch relative to the pair's own scale, never absolute.

Algorithm 5 Two-phase world generator (the registered construction of Appendix G.4, used by E-L2-4; parameters (γ, s), translation normalization $\mu_k = 0$).

$\mu_k \leftarrow (0,0)$; $\mu_v \leftarrow (\sqrt{\gamma}, 0)$ ▷ $u \perp \delta$ along the second axis
$\mu_w \leftarrow \mu_v + s\,u$ ▷ far partner: fixes the stage-1 anchor a_1 and the floor's value
$\mu_p \leftarrow 2\mu_k - \mu_v$ ▷ mirror partner: fixes the stage-2 anchor a_2; closed-form floor
stream: Q_1 queries to v, then Q_2 queries to k ▷ pure and composed streams share the stage-1 prefix
return world ▷ forced batches $\{v, w\}$ then $\{k, p\}$; anchors a_1, a_2; floor by (27)

Algorithm 6 Registered runner contract (the subcommands `e1`, `e2`, `e3`, `e4`, `e4b`, `e9` and `e9b` of `l2_runners.py`; the freeze protocol of Appendix I.19).

Require: grid $\mathcal{G}$, seeds $\{11, 13, 17\}$ (canonical) or $\{0, 1, 2\}$ (smoke, never canonical)
partition $\mathcal{G}$ into chunks below the runtime bound
for each chunk (resumable, idempotent) **do**
  for each cell **do**
    if cell present in partial CSV with matching manifest **then** skip
    end if
    compute the cell statistic (Monte Carlo, simulation, or exact evaluation) against `l2_lib.py` closed forms
    append row keyed by cell parameters; evaluate the cell's falsifier clause
  end for
end for
emit CSV plus manifest (code hash, seeds, grid, tolerances, versions, verdicts)

P.8 The sleep grid, tabulated

Every row of Table 21 is a square root in K and every column a square root in $(\alpha\bar{\rho})^{-1}$: the one-half elasticities of Corollary 3 are the table's arithmetic, and the cost row shows the batching discount's quantitative size at these scales: a sixteen-fold rise in rent pressure moves the optimal cost rate only about three-and-a-half-fold, because the period contracts as the square root while the cost grows as the square root of the pressure.

P.9 Where misalignment is cheap: excess over aligned

P.10 Directional slices of the guarantee functional

P.11 Reading the quantitative theory

Three sentences assemble this appendix. The catalog (Figures 10 and 11) is what a misalignment costs, direction by direction. The floor (Figure 13, Proposition 33, Table 17) is the best a pure unanchored writer can hope to face, and its universal maximum says the placement's power is bounded too. The regime map (Figure 14) and the excess-floor table (Table 18) are the frontier itself: what a flooded unanchored policy pays per query, and through which instrument. Every number is a formula evaluated; the formulas are proved in the main appendices; and the measured column of this story is Section 9.

Index of Figures and Tables

List of Figures

1 The operator taxonomy and the frozen-frame mechanism 2
2 E-L2-4b's growing ratio and E-L2-3's located argmin 8
3 E-L2-8's measured probe loss and the two asymmetries behind it 9
4 E-L2-12's policy costs by query volume and the per-stream outcomes at constant setup 9

Algorithm 7 Provenance gate (`fill_L2.py`; the only path by which a measured number may enter this manuscript).

Require: canonical CSV directory, tex file, pinned runner hashes

1: inventory the named slots in the tex; fail on any unregistered slot
2: **for** each referenced CSV **do**
3: fail closed if the CSV, its manifest, or its pinned hash is missing, or if seeds intersect the smoke set
4: **end for**
5: **if** freeze not landed **then** refuse
6: **end if**
7: emit one macro per slot from its CSV and manifest; slots of unexecuted experiments render as registered, not executed; report the falsifier verdicts alongside

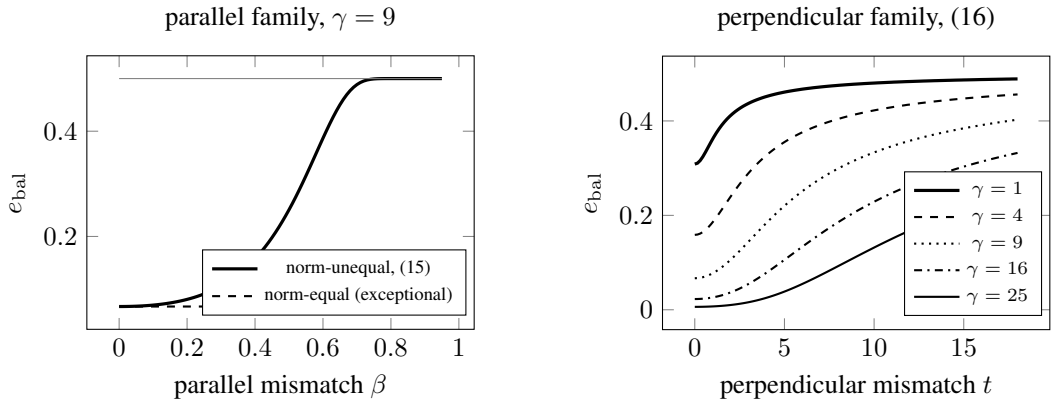


Figure 10: The misalignment catalog of Appendix C.5, evaluated from its equations. Left: the parallel family at $\gamma = 9$; the norm-equal pair sits on the exceptional ray at the Bayes value for every $\beta < 1$ (Proposition 15), while the norm-unequal pair climbs toward $1/2$ per (15). Right: the perpendicular family (16) is strictly increasing in t from the aligned value toward $1/2$, steeper for smaller margins. E-L2-1 re-measures both families cell by cell.

5 The cost flows: rent, write events, misalignment excess and the anchor 12
6 E-L2-1's Monte Carlo deviations and E-L2-2's merged and bystander pairs 48
7 E-L2-10's gauge-cosine ratio by cell and E-L2-5's transport ratio 50
8 E-L2-11's excess against coverage depth and the pair K-L2-11a compares 56
9 E-L2-13's recovery by selection rule and the runner's eligibility filter 59
10 The misalignment catalog, evaluated from its equations 75
11 Directional slices of the two-parameter catalog 76
12 The fixed-composition excess of the two-phase construction 76
13 The mismatch floor against the separation 77
14 The regime map of the frontier, in closed form 78
15 The sleep law's exact discrete cost and its argmin 78
16 The ledger's geometry: two unanchored batches and their anchors 79
17 The two-phase construction at the worked placement 79
18 The bystander world: a merge that damages aligned pairs 80
19 The capacitated batch-partition problem 80

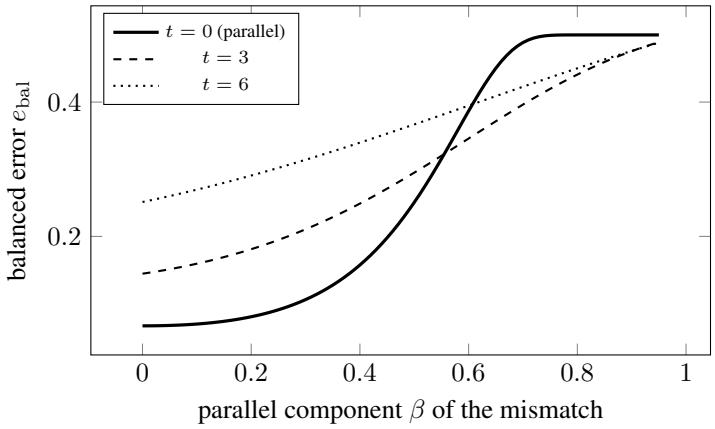


Figure 11: Directional slices of the full two-parameter catalog at $\gamma = 9$: the balanced error over the (β, t) plane of Appendix C.5's general direction, sliced at fixed perpendicular component t. Perpendicular spread flattens and lifts the surface: with t large the pair is far from degeneracy and from the exceptional ray simultaneously, and e_{bal} is lifted toward $1/2$ across the whole β range without concentrating there: at $t = 6$ it runs from 0.2512 at $\beta = 0$ to 0.4762 at $\beta = 0.9$, against 0.0668 to 0.5 at $t = 0$. The ridge at $\beta \uparrow 1$, $t = 0$ is the degenerate ray. E-L2-1 sweeps this surface on a grid including exceptional and degenerate cells.

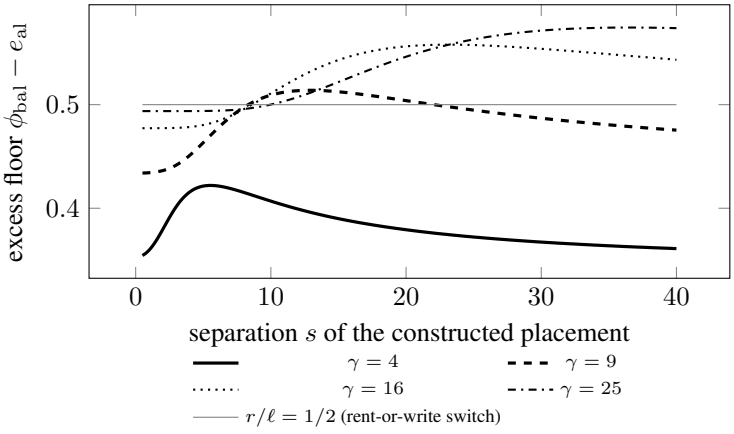


Figure 12: The fixed-composition excess $\phi_{\mathrm{bal}} - e_{\mathrm{al}}$ of the two-phase construction, evaluated from the formula as a function of the separation s for four margins, with the rent-or-write switch $r/\ell = 1/2$ drawn as the horizontal reference; the theorem's composition-free floor $\Xi = \frac{1}{2} - e_{\mathrm{al}}$ is tabulated beside it rather than drawn, and lies strictly below $1/2$ at every margin (Appendix P tabulates the values and their provenance). It is bounded away from zero everywhere and peaks at the universal height of Proposition 33 minus e_{al}. Multiplied by the currency ℓ it is the per-query price a pure unanchored policy pays on the flood once it has written the frame, against an optimum that pays nothing; the horizontal line is where renting the flooded class becomes the cheaper option at $\ell = 2r$. E-L2-4b evaluates the resulting cost growth exactly (Section 9).

20 The price of misalignment over the aligned baseline 81
21 Slices of the guarantee functional 82

List of Tables

1 Dependency map: each result, its proof site and its quantitative site 13

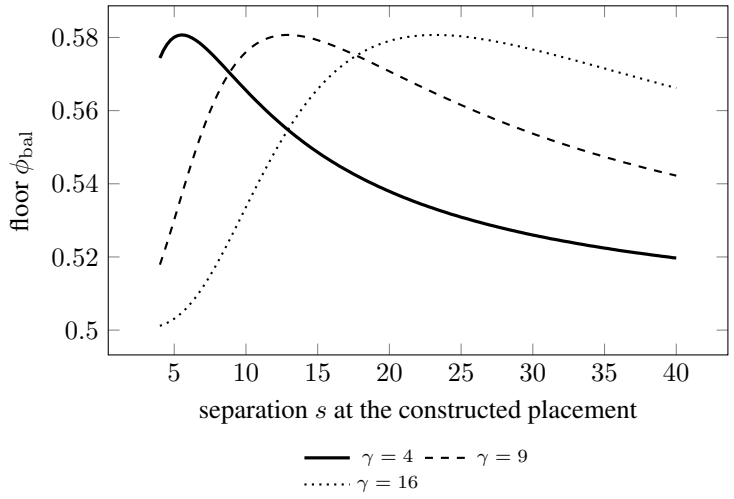


Figure 13: The mismatch floor (27) as a function of the separation s (every $s > 0$ is legal; the separation clause of Lemma 24 holds at every s here, since at unit composition $\|\bar{D}\| \geq \frac{3}{2}\sqrt{\gamma}$ with a component off the line of δ), evaluated from the formula. Each curve rises above $1/2$, peaks at the universal height of Proposition 33, and decays back to $1/2$; the peak walks outward linearly in γ. E-L2-4's γ and s sweeps re-evaluate this family exactly.

Table 17: The floor (27) at $\gamma = 9$ across the legal separation domain (closed-form evaluations; E-L2-4 evaluates six of the seven listed separations through the same closed form, $s = 12$ being outside its sweep; the form's independent checks are the self-test's identity at the anchors and `verify_w2.py`).

s	6	8	10	12	12.9	16	24	$\to \infty$
ϕ_{bal}	0.5431	0.5643	0.5760	0.5803	0.5807	0.5779	0.5633	$\to 1/2$

2 Core symbols. 15
3 Economic and online symbols. 15
4 Closed-form balanced errors for a parallel mismatch 19
5 The perpendicular family at the representative margin 20
6 The reduction dictionary: consolidation terms and their acknowledgment counterparts 28
7 The deployed heuristics, priced against the same optimum 31
8 Ingredient transfer into the pure-W3 frontier 38
9 The lower bound's victim rule, row by row 43
10 Registered experiments, targets, measured outcomes and falsifier verdicts 46
11 The frozen-backbone venue, one row per cell 46
12 The reach map, one row per backbone-corpus cell 56
13 Pre-registered tolerances, locked in at freeze 64
14 E-L2-4 plateau targets as registered 64
15 Slot map: registered slot identifiers and their sources 65
16 The theorem's excess floor across margins 70
17 The floor at the representative margin across the grid 77

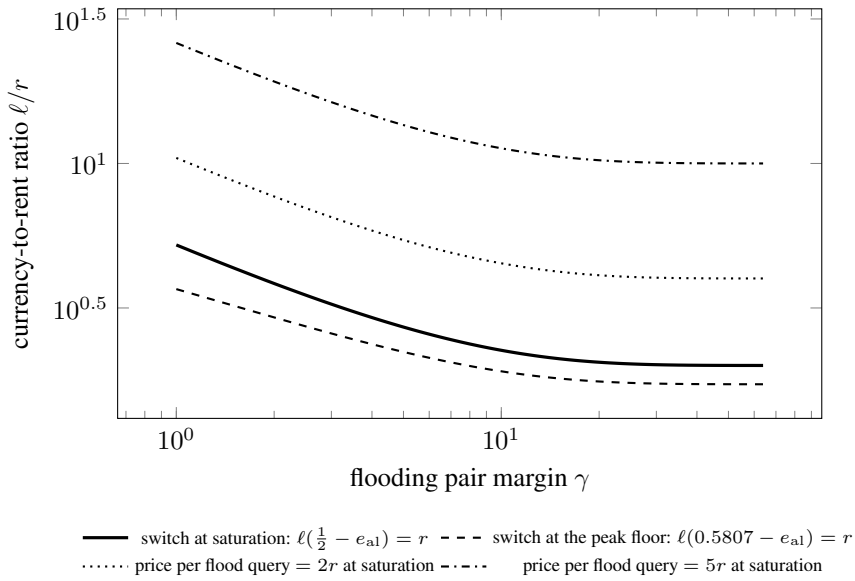


Figure 14: The regime map of the frontier, in closed form: above the solid curve, $\ell(\phi_{\mathrm{bal}} - e_{\mathrm{al}}) \geq r$ at saturated separation and a flooded pure-W3 policy's cheapest option is to rent the flooded class forever, paying r per query against an optimum that pays nothing; below it, writing and paying the excess floor is cheaper, and the per-query price is $\ell(\phi_{\mathrm{bal}} - e_{\mathrm{al}})$. The dotted and dash-dot curves are level sets of the per-query price in units of rent. In every regime the price is paid on every flood query and the ratio grows without bound; the map only says which of the two instruments the adversary's flood extracts it through. E-L2-4b evaluates both regimes exactly (Section 9).

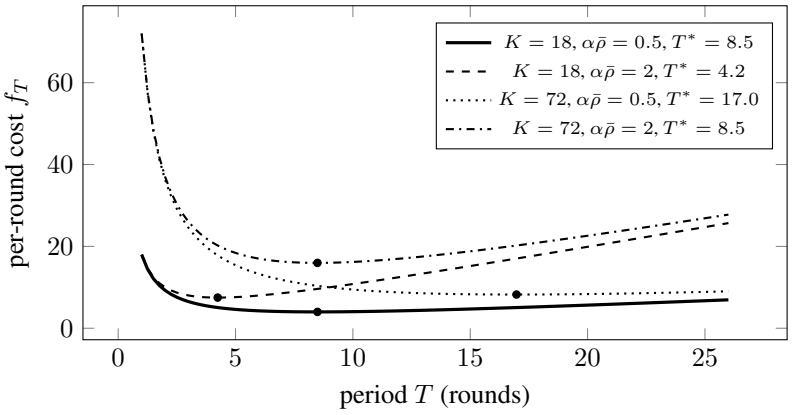


Figure 15: The sleep law's exact discrete cost $f_T = K/T + \alpha\bar{\rho}(T-1)/2$ (Appendix F.5), with the closed-form optima marked; the argmin markers sit exactly at $T^* = \sqrt{2K/(\alpha\bar{\rho})}$ with no approximation. Quadrupling rot halves the optimal period; quadrupling the setup doubles it: the one-half elasticities of Corollary 3 are the curvatures visible here. E-L2-3 locates these argmins by grid search as its falsifier.

18 The excess a flooded pure-W3 policy pays 79
19 The parallel family at the representative norm-unequal pair 80
20 The perpendicular family at scaled mismatch 81
21 The optimal period and its neighbours across the grid 81

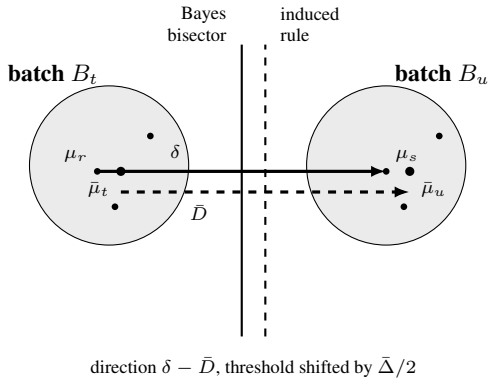


Figure 16: The ledger's geometry: two unanchored batches with anchors $\bar{\mu}_t$, $\bar{\mu}_u$; the pair displacement δ; the vector mismatch $\bar{D}$; and the induced pair rule of (1) against the Bayes bisector. Everything the cross-batch pair pays is a function of these three objects and $\bar{\Delta}$, which Proposition 13 removes.

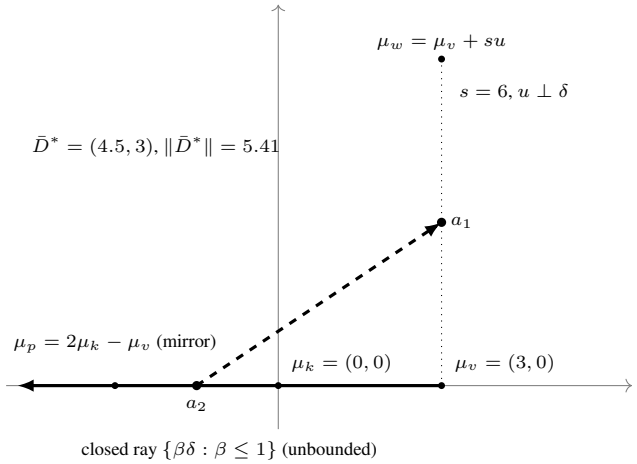


Figure 17: The two-phase construction of Appendix G.4 at the worked values $\gamma = 9$, $s = 6$: the far partner fixes the stage-1 anchor a_1, the mirror partner fixes the stage-2 anchor a_2, and the induced anchor difference clears the closed *ray* (the thick arrow, unbounded to the left; it is a ray and not the segment $[0, \delta]$, which is what Remark 8's counterexample turns on) with margin, giving the floor $\phi_{\text{bal}} = 0.5431$ by (27) and the excess floor 0.4763.

Table 18: The excess a flooded pure-W3 policy pays per query: the theorem's composition-free $\Xi = \frac{1}{2} - e_{\text{al}}$ of (5), and the sharper fixed-composition values $\phi_{\text{bal}} - e_{\text{al}}$ at three separations (closed-form evaluations, four decimals). The price per flood query is ℓ times the entry; it is never small, the fixed-composition value is largest near $s^* \approx 1.47\gamma$ (Proposition 33), and it is paid on every query of an unbounded flood. E-L2-4b evaluates 6 of the 12 cells in the $\gamma \in \{4, 9, 16, 25\}$ columns of the three ϕ_{bal} rows; $\gamma = 1$ is not in its grid.

γ	1	4	9	16	25	36	64
e_{al}	0.3085	0.1587	0.0668	0.0228	0.0062	0.0013	$3.2{\cdot}10^{-5}$
$\phi_{\text{bal}}(s{=}6) - e_{\text{al}}$	0.2232	0.4216	0.4763	0.4839	0.4941	0.4987	0.5000
$\phi_{\text{bal}}(s{=}12.9) - e_{\text{al}}$	0.2068	0.3962	0.5139	0.5318	0.5114	0.5015	0.5000
$\phi_{\text{bal}}(s{=}24) - e_{\text{al}}$	0.1997	0.3734	0.4965	0.5578	0.5604	0.5342	0.5028
$\Xi = \frac{1}{2} - e_{\text{al}}$	0.1915	0.3413	0.4332	0.4772	0.4938	0.4987	0.5000

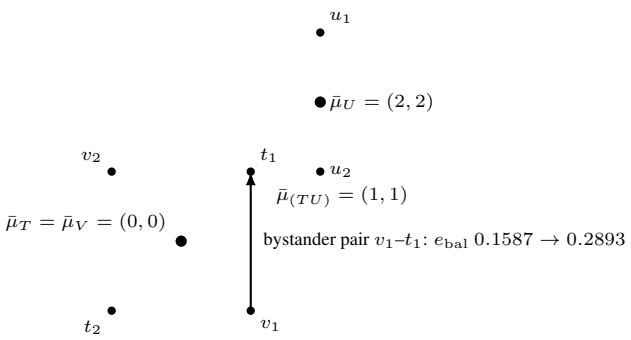


Figure 18: The bystander world of Appendix L.1: merging T (points t) with U (points u) moves the merged anchor to $(1, 1)$; the perfectly aligned T-versus-V bystander pair shown degrades from the Bayes value 0.1587 to 0.2893, one of its two one-sided errors going to chance, while the merged pair improves $0.1574 \to 0.0357$ and the drawn third party improves $0.3855 \to 0.1366$ by rotation (all values closed-form evaluations from the ledger).

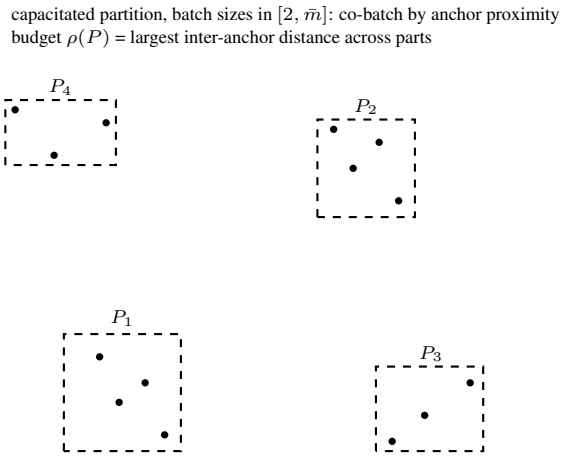


Figure 19: The capacitated batch-partition problem of Appendix E.8: incoming classes (points) with a size cap must be grouped so that batch anchors (part centroids) are mutually close; the realized objective weights each part pair by its query flow. Without the cap the problem degenerates (merge everything, Corollary 16); with it, the combinatorics begin, and the natural heuristic is greedy proximity merging, whose budget is monotone by the same corollary.

Table 19: The parallel family (15) at the representative norm-unequal geometry, across margins and mismatch magnitudes (closed-form evaluations; the norm-equal row is the exceptional-set constant). E-L2-1 re-measures every cell.

e_{bal}	$\beta = 0$	0.2	0.5	0.8	0.95	degenerate
$\gamma = 1$	0.3085	0.3099	0.3293	0.4697	0.500	$1/2$
$\gamma = 4$	0.1587	0.1661	0.2614	0.4993	0.500	$1/2$
$\gamma = 9$	0.0668	0.0803	0.2507	0.5000	0.500	$1/2$
$\gamma = 16$	0.0228	0.0365	0.2500	0.5000	0.500	$1/2$
$\gamma = 25$	0.0062	0.0156	0.2500	0.5000	0.500	$1/2$
norm-equal, any γ	e_{al} for every $\beta < 1$ (the exceptional set, Proposition 15)					

Table 20: The perpendicular family (16) at scaled mismatch $\theta = t/\sqrt{\gamma}$ (closed-form evaluations; the collapse across γ at fixed θ is the scaling of (16): only the aligned level changes). E-L2-1 remeasures twelve of the fifteen cells, three of them through the aligned parallel cell; the $\theta = 4$ column is measured nowhere.

e_{bal}	$\theta = 0$	0.5	1	2	4
$\gamma = 1$	0.3085	0.3274	0.3618	0.4115	0.4517
$\gamma = 9$	0.0668	0.0899	0.1444	0.2512	0.3580
$\gamma = 25$	0.0062	0.0127	0.0385	0.1318	0.2721

Table 21: The optimal period $T^* = \sqrt{2K/(\alpha\bar{\rho})}$ and the exact discrete minimum cost $f_{T^*} = \sqrt{2K\alpha\bar{\rho}} - \alpha\bar{\rho}/2$ on an illustrative grid chosen so the square-root structure is visible in exact ratios (closed-form evaluations, not measurements; E-L2-3's own registered grid is $K \in \{1, 3, 10, 30, 100\}$ with $\alpha, \bar{\rho} \in \{0.1, 0.3, 1, 3, 10\}$ and shares no K value with this table).

T^*	$\alpha\bar{\rho} = 0.25$	0.5	1	2	4
$K = 8$	8.00	5.66	4.00	2.83	2.00
$K = 18$	12.00	8.49	6.00	4.24	3.00
$K = 32$	16.00	11.31	8.00	5.66	4.00
$K = 72$	24.00	16.97	12.00	8.49	6.00
f_{T^*}	e.g., $K = 18$: 2.88, 3.99, 5.50, 7.49, 10.0 across the same columns				

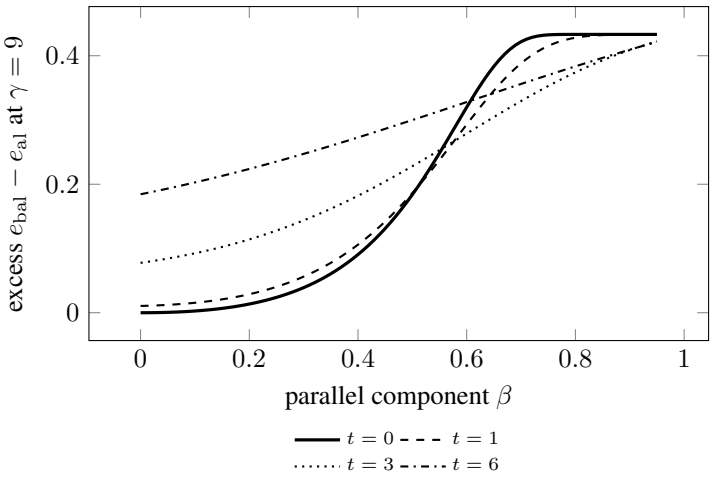


Figure 20: The price of misalignment over the aligned baseline, evaluated from the catalog: the excess is small only when the mismatch is nearly perpendicular *and* small relative to the pair's scale, and it is worst along the parallel direction at moderate magnitude, the same direction where the exceptional set hides all damage at norm-equal pairs. Read with Table 19: the $t = 0$ curve is the table's $\gamma = 9$ row minus its first entry.

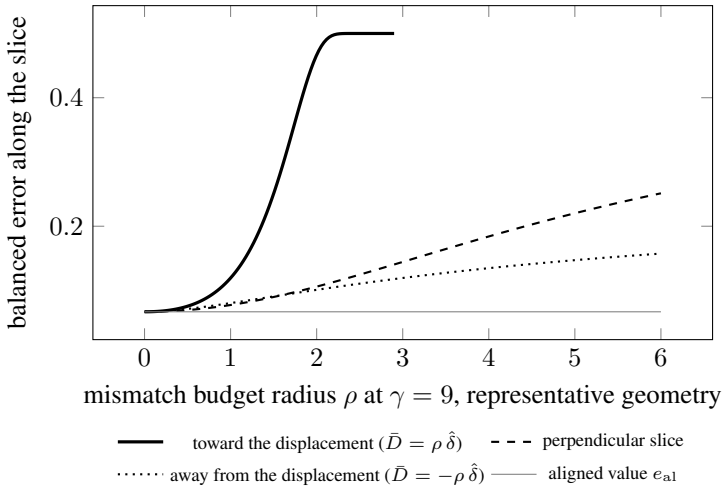


Figure 21: Slices of what the guarantee functional Ψ of Definition 2 supremizes over: the balanced error at fixed mismatch radius, along the three bearings that matter (toward the pair's own displacement, perpendicular, and away), evaluated from (12) at the representative geometry. The toward-slice dominates and terminates at the degenerate point (radius $\|\delta\| = 3$), which is why the budget alone certifies honestly while the realized error depends on bearing; the away-slice is the cheapest direction and the perpendicular family sits between, as Table 20 quantifies.